67
1-25-80
98 BNL 51091

#1

Dr. 2044

MASTER

EEMIS DATA SECTOR CORRESPONDENCE WITH CONCEPTUAL DATABASE DESIGN

K. CROTEAU, A.S. KYDES, AND D. MAIER

July 1979

Prepared for the
ENERGY EMERGENCY MANAGEMENT INFORMATION
SYSTEMS PROGRAM OFFICE, ENERGY INFORMATION ADMINISTRATION

NATIONAL CENTER FOR ANALYSIS OF ENERGY SYSTEMS
DEPARTMENT OF ENERGY AND ENVIRONMENT

BROOKHAVEN NATIONAL LABORATORY
UPTON, NEW YORK 11973

Robb

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

EEMIS DATA SECTOR CORRESPONDENCE WITH CONCEPTUAL DATABASE DESIGN

K. CROTEAU, A.S. KYDES, AND D. MAIER

July 1979

DISCLAIMER

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

NATIONAL CENTER FOR ANALYSIS OF ENERGY SYSTEMS
DEPARTMENT OF ENERGY AND ENVIRONMENT
BROOKHAVEN NATIONAL LABORATORY
ASSOCIATED UNIVERSITIES, INC.

UNDER CONTRACT NO. DE-AC02-76CH00016 WITH THE
UNITED STATES DEPARTMENT OF ENERGY

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency, contractor or subcontractor thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency, contractor or subcontractor thereof.

Printed in the United States of America
Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161
Price: Printed Copy \$10.00; Microfiche \$3.50

ABSTRACT

The purpose of this report is fivefold: (1) It provides an introduction to database systems and criteria that are important in the selection of a commercial database management system. (2) It demonstrates that the mapping of the EEMIS data sector structure into the conceptual database model is complete and preserves the hierarchies implied by the EEMIS data structure. (3) It describes all fields and associated field lengths. (4) It provides accurate formulae for estimating the computer storage requirements for the conceptual models at the facility level. (5) It provides the details of the storage requirements for oil refineries, petroleum crude storage facilities, natural gas transmission and distribution facilities, and company description records.

ACKNOWLEDGMENT

The authors gratefully acknowledge the typing and editorial efforts of Betty Knittweis, Mary Lou Weis, Susan White-DePace, and Gwen Weaver in the preparation of this report. Of course, all errors in this document are the sole responsibility of the authors.

**THIS PAGE
WAS INTENTIONALLY
LEFT BLANK**

CONTENTS

I.	Introduction to Database Systems.....	1
	Database Models.....	1
	Conceptual Level Versus Physical Level.....	3
	Database Management Systems.....	4
	Judging a Database Management System.....	5
	System Backup.....	7
II.	Field Codes and Sizes in the EEMIS Database	7
	Field Sizes.....	7
	Shorthand Fields.....	12
	ID Codes.....	13
	Energy Codes.....	14
	Domains and Representation for Energy Code Subfields.....	14
	Geographic Codes.....	17
	Facility IDs.....	20
	Link IDs.....	22
	Caveat on Codes and IDs.....	24
III.	Mapping and Size Calculations.....	35
	The Mapping Algorithm.....	37
IV.	Cross-Reference Maps and Sizes.....	45
	Domestic Crude Oil Production.....	49
	Crude Oil Bulk Terminal.....	50
	Strategic Petroleum Reserve.....	58
	Oil Refinery.....	59
	Oil Product Bulk Terminal.....	67
	Crude Oil Storage.....	68
	Petroleum Product Storage.....	71
	Natural Gas Production.....	72
	Gas Processing Plant.....	76
	LNG Terminal.....	83
	Underground Storage.....	84
	Supplemental Fuel Storage.....	88
	Transmission and Distribution Facility.....	92
	Demand Facility.....	99
	Large End-Use Consumer.....	103
	Foreign Country.....	108
	Transportation: Links, Junctions, Nodes.....	112
	Company Description Records.....	119
	RES Functional Hierarchy.....	123
	Demand Sector Hierarchy.....	123
	Energy Forms.....	127
	Conceptual Record Geographic Hierarchy.....	129
V.	Case Studies - Storage Requirements.....	136
	Summary of Storage Considerations.....	150
	References.....	151

FIGURES

1	Example of a hierarchial database record.....	2
2	Natural gas energy hierarchy.....	25
3	Petroleum energy hierarchy.....	26

TABLES

1	Breakdown of Shorthand Field Lengths (Bytes).....	12
2	Field Name Lists and Definitions.....	27

I. INTRODUCTION TO DATABASE SYSTEMS

DATABASE MODELS

Database models are systematic methods and notations for specifying the conceptual organization of data. There are three main database models: relational, hierarchical, and network.

The basic building block in all three models is the field (or attribute). A field represents a slot to hold a single data item. The field name is, naturally, the name we use to refer to the field. Each field has a set of allowable values that it may hold, such as all dates in the twentieth century or any salary between 5,000 and 150,000 dollars. The set of allowable values is the domain of the field. Fields are grouped into units called records. The manner in which fields are grouped into records plus the connections between the records is mainly what characterizes the different database models. In the actual database, values are placed into each field of a record to form a record instance. There may be many instances of a single record. For example, we might have a record in the conceptual model that contains fields with information about oil refineries. In the database itself, we would have one instance of this record for each oil refinery, with appropriate values inserted in each field.

The relational database model⁵ is the easiest to understand. In the relational model the records are simple lists of fields called relation schemes; a record instance is called a tuple, and the set of all record instances for a specific relation scheme is organized together as a relation. For example, we might have a relation scheme with fields for refinery name, refinery capacity, unit cost, and refinery zip code. Each tuple would consist of three values, one for each field, such as Standard Oil No. 43, 400 bbl/day, \$20, 08132. The set of all such tuples forms a relation on a relation scheme. Thus another way to view a relation is as a table, with the field names in the relation scheme labeling the columns and each tuple as a row in the table. A relational database scheme is a collection of relations. The main benefit of the relational model is that a table is an easy concept for neophyte users to grasp.

The hierarchical database model⁷ has records with a more complex structure than in the relational model. Each record is divided into segments, which are also groups of fields. Within each record, there is a treelike structure

relating the segments. Every segment but one is a child of its parent segment, the segment without a parent is the root segment. For example, we might have a record with three segments. The root segment has fields for refinery name and refinery capacity. The root segment has a child segment with fields for product type and month. This child segment has its own child segment with fields for destination sent and quantity (see Figure 1).

In a record instance of a hierarchical database record, the segments behave in a special way. Associated with each segment instance of a parent segment there is a list of segment instances for each child segment.

Figure 1

Example of a hierarchical database record

In our example, for each segment instance of the root segment (that is, each refinery name/capacity) there is a list of product type/month segment instances. Further, for each product type/month segment instance, there is a list of destination sent/quantity instances. Thus, a record instance in a hierarchical database consists of treelike lists of data items.

The network database model⁷ has records similar to those in the relational model; however, the grouping of record instances is more complex than simple tables. Record instances are linked into sets on the basis of equal field values. Suppose we have a record with fields for refinery name, product type, month, and total output. We might have all instances of the record with the same refinery name value linked together into a set: the set of all monthly reports for a refinery. We might simultaneously link record instances together on equal values of product type, the set of all monthly reports for a given product type. Sets can also be defined on instances of different records.

There is a fourth database model under development, the functional model,⁹ which does not conform to the field-record methodology. In the functional model, data are represented through functions. For example, we might have a function CAP that returns the capacity of a refinery and a function OUTPUT that returns the total output of a refinery for a given month. For example, CAP (Standard Oil No. 43) would give the capacity of Standard Oil Refinery No. 43 and OUTPUT (Standard Oil No. 43, March 75) would give the total output for this refinery during March 75.

Comparison of database models is difficult with such a brief description of each database model. The relational model is quite easy for the non-computer-specialist to grasp. The hierarchical model supports structures where there are multiple levels of detail within each record. The network model is well-suited for embodying connections between various data objects, although it is the most difficult for the casual user to grasp. Very little has been studied about the functional model. The choice of a database model depends on the inherent structure of the information to be captured, the availability of supporting software, and the proclivities of the database designer.

CONCEPTUAL LEVEL VERSUS PHYSICAL LEVEL

When the data requirements for a database are first discussed, they are likely to be hazy and ill-defined. Before the database is implemented, it is necessary to have a clear and precise mental picture of the information to be represented. This cohesive mental image is called the conceptual model of the database. In some texts, the term logical is used in place of conceptual.

For actual implementation of the database, it is necessary to specify the details of how the data will be stored in core, on disks, on tapes, and so forth. These details are the physical level of representation of the database (sometimes termed the internal level).

There is sometimes a level above the conceptual level, called the external (or user) level. At this level, views of the conceptual model are defined for each user. A view is essentially the subset of the conceptual model that is of interest to a particular user, although a view can imply transformation and re-formatting of data from the conceptual level.

In some texts, the conceptual level is further divided. In these cases, the mental picture of the database is the conceptual model, and the expression of the conceptual model in the format of a given database system is the logical model.

DATABASE MANAGEMENT SYSTEMS

The primary purpose of a database management system (DBMS) is to facilitate the transition from conceptual model to physical implementation. The DBMS provides a data description language, which is used to describe the conceptual model in a form which the DBMS can understand. The data description language of a given DBMS is designed for one of the three major database models: relational, hierarchical, or network. Consequently, a specific DBMS is most suited for one class of conceptual models.

The DBMS usually has software to perform the physical storage organization of the database with a little human help. Decisions must be made as to what form the data will take in the storage devices, which data will occupy which devices, and how the data are organized on a particular device.

The DBMS usually performs many other functions besides physical storage organization. The DBMS can handle physical storage allocation as more storage space is needed for the database. The DBMS supports a data manipulation language, which handles insertion, update, and retrieval of data in the database. The DBMS generally has some software for data validation: checking whether values are in the proper format and in the proper range. A DBMS often provides data management facilities: programs that can compute the current size of the database, programs that keep track of the usage of different components of the database, and programs that can back up the database.

JUDGING A DATABASE MANAGEMENT SYSTEM

Most commercial DBMS provide all the functions outlined above. This section deals with other features, in addition to those already listed, that determine which DBMS is most suitable for a given application. Some of these considerations are addressed by many systems, others are not adequately addressed by any.

Support at the External Level. Does the DBMS provide an explicit mechanism for defining individual user views on top of the conceptual model? Is the user view limited to strict subsets of the conceptual model?

Data Independence. The DBMS should provide independence for the user from knowledge of the physical storage structure. He should not have to know that there are four record instances in each physical disk record and that the information he wants is on disk three. Also, the user should not need to know the access techniques of the system he uses. The data should be presented to him in the form his view indicates. He should not have to know that the refinery name field is the fourth through eighth word in a record instance. Furthermore, the user should be independent of changes to the physical structure and access techniques. If the database administrator decides to move a database file from tape to disk, the user should not have to rewrite his application programs to accommodate the changes. The data independence concept also applies at the external level. A user should retain his own view regardless of changes in the conceptual model, as long as the data he requires still remain in the model.

Extensibility and Alterability. Small changes in the conceptual model should not mean large changes at the physical level. In particular, a small change to the model should not mean the entire database must be recopied from scratch. Similarly, extensions to the conceptual model that do not imply changes in the existing portions of the model should require very little change in the existing structure at the physical level, although the change may bring about extensive additions to the physical structure. In general, unrelated portions of the conceptual model should not impinge heavily on each other in the physical representation.

Self Description. The DBMS should support a description of the current conceptual model. Furthermore, this description should be accessible by the same methods used for access to the actual data in the database. Many DBMS have a data dictionary with information about the various data items in the database, but few allow users to access this information as they would regular data. A user should be able to ask "What information do you have about each refinery?" as easily as "What was the production of unleaded gasoline in New Jersey in 1977?" In other words, a user should be able to find out about the database from the database.

Record Keeping. The data management routines of a DBMS usually contain some record keeping programs, but they never do all the record keeping they could or should. Unorganized databases become choked by the unautomated record keeping processes that contain information such as what jobs must be run, what data items are invalid, who gathers data for a particular record. All records necessary for the well-being and smooth management of the database should be part of the database.

Data Manipulation Language. What environment does the data manipulation language need? Can it be used interactively, or must it be used from within some programming language? How complex is the manipulation language? There should be some subset that can be quickly assimilated by casual users. How powerful is the manipulation language? Are there operations on the database it cannot perform?

Ease of Use. This consideration follows closely on the last. Can the DBMS handle users that are not computer specialists? It is unreasonable to expect that all features of the database be available to the neophyte user, but if there are to be casual users, there must be some set of useful features that are easy to use. Perhaps the interactive query system could help a user reformulate queries rather than simply rejecting incorrectly formatted queries. The interactive system could provide more than just flat answers to queries. For example, for the questions "How many refineries in Georgia produce both leaded and unleaded gasoline?", an answer of "none" may be true, but an answer of "none: there are no refineries in Georgia producing unleaded gasoline" is much more useful in general.

Range of Conceptual Models Supported. Although the data description language of a DBMS may be based on the relational, hierarchical, or network database models, the DBMS may support only a limited version of one of these database models. That is, not all the conceptual models based on the particular database model are supported. The data description language may not be able to directly express all aspects of the mental models that are desired.

Integrity - Data Screening. Many systems support data screening on a field by field level. A value for a field can be checked for format and for being in a given range. There are more complex integrity constraints that often cannot be enforced by the DBMS. For example, if we had a record that listed percentage of ownership of one company by other companies, the percentages for all record instances should not sum to more than 100. Perhaps we want to disallow two separate characterizations of the same oil product type. There could also be statistical data screening routines: 90% of the values for a field must fall within a given range. Of course a DBMS will not be able to handle all possible integrity constraints, but a good one should handle more than field by field screening.

Application Library. Certain application programs are of interest to so many users that they should be provided as part of the DBMS. Examples are report generation programs and archiving routines. The DBMS should provide some general purpose application programs so that users are not duplicating each other's efforts.

SYSTEM BACKUP

The DBMS usually provides routines that will copy the whole database onto external storage so it can be restored after a hardware failure. Some systems keep track of all transactions within the database since the last backup, so that users can be informed what data may have been lost. In some systems the backup is automatically performed at regular intervals, such as every night. Related to backup is archiving. Archiving is the removal of portions of the on-line database that are not of current interest; these are transferred onto other storage media in case they are needed in the future.

Data Pedigrees. Not all information is easily coded into a number. In some databases, much of the information is a matter of opinion. In these databases, the DBMS should be able to identify a datum with its reliability.

Different values for a field may come from different sources. The reliability of data should be carried over into reports generated from the database. A report could tell, for example, if the data were unreliable, incomplete, or out-of-date.

Security. Some databases contain privileged and sensitive information, and access to such information must be restricted. However, there may be other parts of the database at which anyone may be allowed to look. The computer operating system security protection may not be sufficient in such cases, if it can only grant or refuse access to the DBMS as a whole and not by parts. The DBMS must have the means to grant selective access to the database.

Representation of Domain Values. The form that the user wants to use to specify domain values may not be very efficient for storage. The user may have a field with only two possible values, "male" and "female." To save space, coded into "0" and "1." The question is "Who does the encoding and decoding?" Is it the user or the DBMS? For a system with casual users, the coding should not be placed on the user, as it is another hindrance to easy interaction with the system.

Data Compression. If the database is going to be large, the DBMS should have the means to automatically compress the data. Choosing shorter representations for domain values has already been mentioned. The DBMS can also delete space for fields that have no value. Textual information is usually quite redundant and can often be compressed by 40-50% with various schemes.

Data Transmission. The information in some databases is meant for transmission to remote locations. A national database may have users in every state. The DBMS for such a database should have its own telecommunication routines.

No DBMS has all the features described above, nor is it necessary that one should, since no particular database will require all the features. The importance of various features will depend on the uses and requirements of the database.

II. FIELD CODES AND SIZES IN THE EEMIS DATABASE

This section has three parts. The first part describes the significant fields and the decisions that were made regarding their lengths and contents, the second part details the "shorthand" fields that appear in the conceptual design, and the third part defines the fields that appear in the database together with their lengths as they are presently coded. The field sizes for the EEMIS database have been derived from information provided by the Stony Brook Laboratory for Software Systems Research⁶, which developed the conceptual design of the database, and from recommendations provided by the EEMIS design group at BNL. Further discussion and information from the U.S. DOE may lead to revisions in the field sizes; many of the fields that appear in this report may not be available in the interim EEMIS. Decisions regarding the ultimate sizes of fields will be made on a dynamic basis, and the sizes provided should be viewed as good first order estimates.

Many of the size estimates come from data that have been recorded in the commercial and industrial sectors; such items as quantities, percentages, and costs are readily available from trade journals and reports. Other data are based on expert judgement or common usage, such as dates, names, addresses. For fields that have derived representations that were developed as part of the design process, such as the energy or facility codes, information is provided that justifies their lengths.

A detailed list of fields together with their lengths appears later in this section; we first provide a guide to the design and decision-making process that led to the development of the field sizes within the system.

FIELD SIZES

Most of the names used within the database (e.g., facility names, individuals' names, addresses) are described as thirty-byte alphanumeric fields. Enough identification must be provided so that the essential information of the name is preserved while unnecessarily long names that can be abbreviated are shortened. Thirty bytes is long enough to record most names fully and to abbreviate long names meaningfully. Additionally, many mailing labels accommodate thirty-character names (they are usually one inch wide by thirty

characters long), so the storage field could be used to generate mailing lists without conversion. The lengths of city, state, and zip were chosen to allow all three fields to appear on one thirty-character line.

Special Fields representations, specifically Energy codes, Geographic codes, Facility ID, and Link ID, are described completely in a later portion of this section. Energy codes have been developed as eight-character codes, composed of four one-character subfields and one four-character subfield that describe the level, sector, category, group, and type of energy in question. Geographic codes are also eight characters with two subfields, a one-character geographic level, and a seven-character geographic descriptor. Facility codes are sixteen characters long and contain five subfields that describe the location, level, sector, function, and identifier of the particular facility. Link IDs are twenty-four characters long, composed of seven subfields detailing transportation mode, level, energy form, input facility, output facility, location, and the identifier of the particular link. Alphanumeric representation for IDs or codes were used wherever possible because the process of quickly and correctly encoding the conceptual models into ADABAS schemata is significantly facilitated whenever mnemonic alphanumerics are used. This encoding preference increases redundancy in storage but makes the debugging process much quicker. The choice of alphanumeric codes and IDs for the interim EEMIS was prompted by the perceived need to have the EEMIS system implemented and debugged at the earliest possible date. The use of numeric codes will reduce the storage requirements by nearly 50% but will require substantially larger amounts of time to debug, and delays in implementation could kill the development of the EEMIS system altogether. The recommended strategy is to implement the interim system quickly by using alphanumerics and later develop the more efficient storage approach using numeric IDs and codes.

Quantities and capacities were determined from industry data. Typical figures are, for example, in billions of cubic feet of natural gas and millions of barrels of oil. Standardization of the field size to handle both types of figures led to a use of eight bytes, which is sixteen digits using packed decimal representation. This field could represent a scaled value based on some agreed upon factor (e.g., the value could represent thousands of cubic feet, so a recorded value of 15 would actually be 15,000). Since the scale of the values is relative, any agreed upon factor could be used and the field would still record meaningful information. There should be agreement

for a given energy type as to the base units, such as barrels of oil or tons of coal. However, the actual unit can vary for different quantities of oil. A refinery at the facility level might record output in thousands of barrels. The national aggregate refinery might measure output in millions of barrels. The field held allows a choice of both base unit and multiplier (essentially a limited form of scientific notation).

Dates within the database will be recorded as the seven character YYYYMMDD format. This is a standard calendar approach and was adopted because of its extensive use in a wide range of applications. Note that any seven character arrangements of the data can be used (e.g., the international format of MMMDDYY) so that any anticipated data arrangement can be accommodated. Date fields are used in two places within the database: as a snapshot date for storage holdings and as a recording period in quantity records. Full dates are used within the system to allow the possibility of staggered or frequent reporting dates or changes in the period of reporting. If evaluation of the data recording process shows that all reported data will arrive on a monthly basis, then the date field could be modified to YYYYMM.

Percentages and costs provide much of the useful information for calculations within the database. Because of the potential importance of pricing considerations the cost field is a twelve-digit field (six bytes) with an implied decimal point between the third and fourth digits (i.e., when used for unit pricing) and no decimal for total price. Fractional cents are critical when dealing with large bulk quantities of oil and gas and are necessary to record. Likewise, percentages are critical in certain areas (sulfur content, for one) and are recorded as four-digit fields (two bytes) with an implied decimal point between the second and third digits (i.e., xx.xx).

States, zip codes, and phone numbers are based on their everyday usage. The unique two character postal code is used for states, five digits suffice for zip codes and ten digits for phone numbers (area code + seven digits). If legislation to increase the zip code to eight or ten digits is enacted by the federal government then this field will have to be changed.

For most of the remaining fields, the definitions naturally led to field size decisions. The reader is directed to the BNL issue paper² and the Stony Brook report⁶ for further discussion on the field size selection.

SHORTHAND FIELDS

To facilitate the development of the conceptual models, certain fields were designed to represent sets of fields. These are listed in Table 1 with the shorthand name appearing under the column "Field/Segment Name"; the fields that compose them are listed under "Specific Contents." The actual size of the individual fields is given in the column labeled "Length," with a total for the shorthand field.

Table 1: Breakdown of Shorthand Field Lengths (Bytes)

Field/Segment Name	Specific Contents	Length (Bytes)
<u>Location</u>	Street Address	30
	City	20
	State	2
	Zip	5
	Phone Number	10
	Contact Person	30
	Latitude (DDMMSSL)	9
	Longitude (DDMMSSL)	<u>9</u>
		115
<u>Q-Field</u>	From Date	7
	To Date	7
	Reported Quantity	9
	Computed Quantity	9
	Units	3
	Aggregate	<u>1</u>
		36
<u>QP-Field</u>	Q - Field	36
	Price	<u>6</u>
		42

Field/Segment Name	Specific Contents	Lengths (Bytes)
<u>S - Field</u>	On Date	7
	Reported Quantity	9
	Computed Quantity	9
	Units	<u>3</u>
		28
<u>RN - Field</u>	Route Name	30
	# of Links	<u>1</u>
		31
<u>RL - Field</u>	Link #	1
	Link ID	<u>26</u>
		27

ID CODES

The identification code conventions adopted for use within the EEMIS database are particularly important since they (1) provide the means for understanding the cross referencing that occurs between data elements of the database, and (2) provide information for determining the storage needs of data elements. Because these codes have no real-world equivalents, and because they are composed of subfields that themselves contain important information, they are described in more detail for future reference.

ENERGY CODES

We plan to give every energy form a unique code so that we do not have to repeat the description every place the energy type information is wanted. This code has certain properties.

1. The format is consistent for all types. For example, the energy code field in a record accommodates an oil product type or a coal type.

2. The format allows different levels of aggregation and shows the relationships between levels.

We propose five levels of aggregation:

<u>Level</u>	<u>Example Values</u>
1. Universal	universal energy form
2. Sector	petroleum, gas, coal, electricity
3. Category	crude oil, oil product
4. Group	domestic crude, foreign crude, SPR, gasoline
5. Type	East Texas, North Slope, Type I, leaded gas

Universal energy form is all energy from all sectors. The energy code field is composed of five subfields: energy level, energy sector, energy category, energy group and energy type. A one character representation is provided for the first four fields and four characters for the last. For example:

DOMAINS AND REPRESENTATION FOR ENERGY CODE SUBFIELDS

For each subfield of energy code, the domain and its proposed representation are given. Note that the representation for one subfield may depend on the value of another. Also, for each energy form at a given level the full representation for the energy code for that form is provided by concatenating

the values for the subfields. For example, an energy form at the category level will have no values for energy group and energy type. When there is no value we use blanks, denoted as b or _____ (for a string of trailing blanks).

1. Energy Level: This subfield determines the level at which the energy form is described. Universal is for all energy forms combined; note that it needs no description other than level, since there is only one energy form at the universal level.

<u>Domain Values</u>	<u>Representation</u>
Universal	U
Sector	S
Category	C
Group	G
Type	T
Full code for Universal energy form:	U _____.
(The line _____ stands for trailing blanks.)	

2. Energy Sector

<u>Domain Values</u>	<u>Representation</u>	<u>Full Codes for Sectors</u>
petroleum	P	SP _____
natural gas	N	SN _____
coal	C	SC _____
electricity	E	SE _____
b (if energy level is universal)		

3. Energy Category: This subfield specifies a grouping between energy sector and energy group.

<u>Energy Sector</u>	<u>Domain Values</u>	<u>Representation</u>	<u>Full Codes for Categories</u>
petroleum	crude oil	C	CPC _____
	oil product	P	CPP _____
natural gas	wet gas	W	CNW _____
	dry gas	D	CND _____
b (if energy level is universal)			

4. Energy Group: An energy group is a set of individual types that have the same form of description. That is, the information to be represented about each type in the group has the same format.

Energy Sector

<u>Category</u>	<u>Domain Values</u>	<u>Representation</u>	<u>Full Codes for Groups</u>
petroleum-crude	domestic crude	D	GPCD_____
	foreign crude	F	GPCF_____
	strategic petroluem	S	GPCS_____
petroleum-prod	gasoline	G	GPFG_____
	jet fuel	J	GPPJ_____
	distillate fuel oil	D	GPPD_____
	residual fuel oil	R	GPPR_____
	light products	L	GPPL_____
	(incl. natural gas liquids)		
	other	O	GPPO_____
natural gas-wet	wet gas	W	GNWW_____
natural gas-dry	methane	M	GNDM_____
	liquid natural gas	L	GNDL_____

b (if energy level is universal,
sector, category)

5. Energy Types: This subfield is the lowest level in the energy field hierarchy.

Energy Sector

Category-Group

<u>(Representation)</u>	<u>Domain Values</u>	<u>Representation</u>	<u>Full Codes for Types</u>
PCD	East Texas	ETEX	TPCDETEX
	North Slope	NSLP	TPCDNSLP
	...	...	...
PCF	Arabian Light	ARLT	TPCFARLT
	Iranian Light	IRLT	TPCFIRLT
	...	...	...
PCS	type 1	TYP1	TPCSTYP1
	type 2	TYP2	TPCSTYP2

<u>(Representation)</u>	<u>Domain Values</u>	<u>Representation</u>	<u>Full Codes for Types</u>
PPG	leaded	LEAD	TPPGLEAD
	unleaded	UNLD	TPPGUNLD
	aviation	AVIA	TPPGAVIA
PPG	naptha type	NAPT	TPPJNAPT
	kerosene type	KERO	TPPJKERO
PPD	heating	HEAT	TPPDHEAT
	diesel	DIES	TPPDIES
PPR	No. 5 and No. 6	5OR6	TPPR5OR6
	fuel oil		
	other	OTHR	TPPROTHR
PPL	ethane/ethylene	ETHN	TPPLETHN
	liquified gases	LIQG	TPPLLIQG
	propane	PROP	TPPLPROP
	petrochemical	PCHM	TPPLPCHM
PPO	other	OTHR	TPPOOTHR
NWW	wet gas	WETG	TNWWWETG
NDM	methane	METH	TNDMMETH
NPL	liquid natural gas	LIQN	TNPLLIQN

bbbb (if energy level is anything
but type.

GEOGRAPHIC CODES

We need to easily identify geographic areas at various levels, namely;

Countries

Regions

States

Counties

as well as areas that do not conform to political boundaries: oil fields, natural gas fields, demand sector areas. We call such areas non-political areas.

The geographic code field is divided into two subfields, geographic level and geographic descriptor. The proposed representation for the domains of these two subfields requires one character and seven characters respectively, for an eight-character representation of the geographic code.

1. Geographic Level: This subfield defines the type of area we wish to describe.

<u>Domain Values</u>	<u>Representation</u>	
foreign country	F	
domestic crude production region	U	
natural gas production region	G	Non-political identifiers
demand sector area	D	
national	N	
region	R	
state	S	
county	C	

2. Geographic Descriptor: This subfield describes the area being coded. The subfield contents depend on the contents of the geographic level subfield.

A. Foreign Country: The geographic descriptor is the first seven characters in the name of the country if the geographic level is F (foreign).

<u>Domain Values</u>	<u>Representation</u>	<u>Complete Code</u>
Mexico	MEXICOb	FMEXICOb
Saudi Arabia	SAUDIAR	FSAUDIAR
Canada	CANADAb	FCANADAb
Venezuela	VENEZUE	FVENEZUE

B. Domestic Crude Oil Production Region, Natural Gas Production Region, Demand Sector Area: The geographic descriptor is a unique identifier for the region or area. For example, Northwest Texas might be a crude oil production region with representation NWTEXAS and complete eight-digit code UNWTEXAS.

C. National, Regional, State, County: These areas form a strict hierarchy similar to the energy hierarchy. The geographic descriptor is broken down further into three subfields: region descriptor, state descriptor, and county descriptor. The following representation is proposed.

Geographic Descriptor

i. If the geographic level is national: region, state and county descriptors are blank.

Full code: N _____

ii. Region Descriptor: The proposed representation is one character.

<u>Domain Values</u>	<u>Representation</u>	<u>Full Code for Regions</u>
Pacific Northwest	P	RP _____
New England	N	RN _____
Mid Atlantic	M	RM _____
Northeast	E	RE _____ etc.

b (if geographic level is national)

iii. State Descriptor: The proposed representation is the two-character postal code.

<u>Domain Values</u>	<u>Representation</u>	<u>Full Code for States</u>
New Jersey	NJ	SENJ _____
New York	NY	SENY _____
Oregon	OR	SPOR _____ etc.

bb (if geographic level is region)

iv. County Descriptor: The proposed representation is four letters that uniquely identify a county within a state.

<u>Domain Values</u>	<u>Representation</u>	<u>Full Code for Counties</u>
Passaic	PASS	CENIPASS
Mercer	MERC	CENJMERC
Bergen	BERG	CENJBERG etc.

bbbb (if geographic level is region)

FACILITY IDS

The facility ID field is composed of five subfields: facility sector, facility function, facility level, geographic code, facility identifier. We describe the domain and proposed representation for each field separately.

1. Facility sector: This subfield describes the sector of the energy system with which this facility is associated. The proposed representation is one character.

<u>Domain Values</u>	<u>Representation</u>
all sectors combined	A
petroleum	P
natural gas	N
demand-facility	D
demand-large consumer	L
import-export	X
transportation junction	T

2. Facility function: This subfield describes the function of a facility within its sector. The proposed representation is two characters.

<u>Facility Sector</u>	<u>Domain Values</u>	<u>Representation</u>
all sectors combined	all functions	LL
petroleum	all functions	LL
	domestic crude oil production	DC
	crude storage	CS
	product storage	PS
	crude bulk terminal	BT
	strategic petroleum reserve	SR
	refinery	RE
	product bulk terminal	PT
natural gas	all functions	LL
	production	PR
	processing plant	PP
	liquid natural gas terminal	LG
	underground storage	US

<u>Facility Sector</u>	<u>Domain Values</u>	<u>Representation</u>
	supplemental fuel storage	SS
	transmission and distribution	TD
demand-facility	all functions	LL
	residential/commercial	RC
	utility	UT
	transportation	TR
	military	MI
	residential	RE
	commercial	CO
demand-large consumer	same as demand-facility except no residential	
import-export	all functions	LL
	foreign country	FC
transportation	all functions	LL
junction	junction	JN

3. Facility Level: This subfield gives the level of aggregation for the facility. The proposed representation is one character.

<u>Domain Values</u>	<u>Representation</u>
unspecified	U
country/national	N
region	R
state	S
individual facility	F

At this point, we can define the field facility type. It is the concatenation of the subfields facility sector, facility function and facility level. The proposed representation is therefore four characters.

Examples:

PBTS: petroleum/crude bulk terminal/state level

NPPF: natural gas/processing plant/individual facility

Note: Facility types whose representations begin with ALL, PLL, NLL, DLL, LLL, XLL, and TLL are dummy facilities; they are not defined in the

facilities portion of the conceptual model. Such facilities are used when the exact source or destination of a shipment is unknown. Of course, shipments to or from such facilities cannot be faithfully incorporated in any network analysis.

4. Geographic Code: This field was described in the last section. It contains the geographic location most appropriate for the facility.

5. Facility Identifier: This subfield is used to ensure uniqueness among facility ID values. The proposed representation is four characters.

The field "facility ID" is the concatenation of the subfields facility sector, facility function, facility level, geographic code and facility identifier. This representation is therefore sixteen characters.

Examples:

P B T S S E N J b b b b b b b b

Petroleum/crude bulk terminal/state level/State/East/New Jersey/no identifier

N P P F C E N J B E R C S T D O

Natural gas/processing plant/individual facility/county level/East/New Jersey/Bergen County/ STDO identifier.

LINK IDs

The link ID field is composed of seven subfields: link mode, link level, input facility type, output facility type, link energy form, geographic code and link identifier.

We describe the domain and proposed representation for each domain separately.

1. Link Mode: This subfield describes the means of transportation along the link. The proposed representation is one character.

<u>Domain Values</u>	<u>Representation</u>
multi-modal	M
rail	R
barge	B
truck	T
pipeline	P
tanker	K
oxcart	O
null (when two facilities are on the same site)	

2. Link Level: This subfield specifies whether the link has an actual physical counterpart or represents an aggregation of real-world links. The proposed representation is one character.

<u>Domain Values</u>	<u>Representation</u>
physical	P
aggregate (megalink)	A

3. Input Facility Type: This subfield gives the facility type at the link origin facility. The domain is the same as for the facility type field.

4. Output Facility Type: This subfield gives the facility type at the link destination facility. The domain is the same as the facility type field.

5. Link Energy Form: This subfield gives the form of energy the link handles, possibly down to the group level. This subfield is itself composed of the first four subfields of the energy type field: energy level, energy sector, energy category, and energy group.

At this point we can define the field link type. It is the concatenation of the subfields link mode, link level, input facility type, output facility type, and link energy form. The representation is therefore fourteen characters.

Examples:

a. R A P B T S P B T S C P C b

is an aggregate rail link between two state level crude bulk terminal facilities carrying crude oil.

b. P P N P R F N P P F G N W W

is a physical pipeline link between an individual natural gas production facility and an individual natural gas processing plant, carrying wet natural gas.

6. Geographic Code: This subfield was described in a previous section. It contains the geographic location most appropriate for the link.

7. Link Identifier: This subfield is used to ensure uniqueness among link ID values. The proposed representation is four characters.

The field "link ID" is the concatenation of the field's link mode, link level, input facility type, output facility type, link energy form, geographic code and link identifier. The representation is therefore twenty-six characters.

Examples:

a. R A P B T S P B T S C P C b R E b b b b b b P O S 3

might represent a link of the type in Example (a.) above connecting two states in the Northwest.

b. P P N P R F N P P F G N W W C E N J b b b b S T O 8

might represent a link of the type in Example (b.) above connecting two facilities in the state of New Jersey.

To further illustrate the energy hierarchy discussed in the previous pages. Figures 2 and 3 show the natural gas energy hierarchy and the petroleum energy hierarchy. Note that as one proceeds from left to right in the structure, the level of detail increases (from sector to category to group to type) and the divisions become finer; for example, in Figure 2, from natural gas to wet and dry gas to methane and LNG.

CAVEAT ON CODES AND IDs

We do not intend that codes and IDs defined in this section be visible to users normally, although certain subfields may be of interest. We assume such subfields would be presented in an interpreted version rather than in their internal representation to the user of the EEMIS database.

Figure 2: Natural gas energy hierarchy

Figure 3: Petroleum energy hierarchy

Table 2: Field Name List and Definitions

Field Name	Definition	Length (Bytes) ^a
Advisor name:	Name of an energy advisor to a governor.	30
Associated crude storage ID:	See Facility Codes above.	16
Associated facility ID:	See Facility Codes above.	16
Associated product storage ID:	See Facility Codes above.	16
Category Code	See Energy Codes above.	8
Category Name	The name of an energy category.	30
City:	City as would appear in mailing address.	30
Company ID:	Unique identifier for each company mentioned in the database.	16
Company name:	The name of a company.	30
Computed quantity:	An amount of energy calculated from other data elements in the database. See reported quantity.	9
Contact person:	The person to call or write at facility, state office, etc.	30
Cost:	The <u>total</u> cost for transportation or storage of an energy type. Note the implied difference between cost and price.	5
Country code:	See Geographic Codes above.	8
Country name:	The name of a country.	20
County code:	See Geographic Codes above.	4
County descriptor:	See Geographic Codes above.	4
County name:	The name of a county or parish.	30
Crude bulk terminal ID:	See Facility Codes above.	16
Crude storage ID:	See Facility Codes above.	16
Cushion capacity:	In underground gas storage, the amount of gas that can be removed from storage only with special permission. It is required for pressure maintenance.	9
Demand category:	A category of energy usage by a large end-use consumer. Can be "space heating and miscellaneous thermal," "process steam," "direct heat," "petrochemical feedstock."	3

^aWe assume packed decimal storage for all quantity-related and ordinal fields with two decimal digits per byte and alphanumeric for the remainder. Thus, for any packed decimal field, the number of digits is twice the number of bytes.

Table 2 (cont.)

Field Name	Definition	Length
Demand facility ID:	See Facility Codes above.	16
Demand sector:	One of the demand sectors. Can be "residential/commercial," "utilities," "transportation," "industrial," "military," "residential," "commercial."	3
Demand subsector:	One of the demand subsectors. Can be "residential," "commercial."	3
Director name:	The name of the director of a regional energy commission.	30
Distillation capacity:	The maximum amount of crude a refinery can distill per unit time.	9
Domestic crude production ID:	See Facility Codes above.	16
Energy category:	See Energy Codes above.	1
Energy code:	See Energy Codes above.	1
Energy group:	See Energy Codes above.	1
Energy level:	See Energy Codes above.	1
Energy sector:	See Energy Codes above.	1
Energy type:	See Energy Codes above.	4
Energy usage:	The percentage of energy usage corresponding to a SIC code or FERC priority.	2
Facility function:	See Facility Codes above.	2
Facility ID:	See Facility Codes above.	16
Facility identifier:	See Facility Codes above.	4
Facility level:	See Facility Codes above.	1
Facility name:	The name of a facility	30
Facility sector:	See Facility Codes above.	1
Facility type:	See Facility Codes above.	4
FERC priorities:	The FERC priorities that apply to gas usage by a large end-use consumer.	2
Foreign country ID:	See Facility Codes above.	16
From date:	The beginning date of a reporting period.	7

Table 2 (cont.)

Field Names	Definition	Length
Geographic code:	See Geographic, Facility, and Link Codes above.	8
Geographic descriptor:	See Geographic Codes above.	7
Geographic level:	See Geographic Codes above.	1
Group code:	See Energy Codes above.	8
Group name:	The name of an energy group.	30
Head name:	The name of a head of a state energy department.	30
Heat value:	The heat value of an energy type.	4
Input facility ID:	The facility ID of the facility at the origin of a transportation link.	16
Input facility type:	See Link Codes above.	4
Lag time:	The length of time required for traversing a link (days).	1
Large consumer ID:	See Facility Codes above.	16
Latitude:	The latitude of an object or point on a boundary. A negative value indicates a break in a boundary.	9
Length:	The length of a transportation link (miles).	3
Link energy form:	See Link Codes above.	4
Link ID:	See Link Codes above.	26
Link identifier:	See Link Codes above.	4
Link level:	See Link Codes above.	1
Link mode:	See Link Codes above.	1
Link name:	The name of a transportation link.	30
Link #:	The order of a link within a route, numbering from the origin.	1
Link type:	See Link Codes above.	14
LNG terminal ID:	See Facility Codes above.	16
Location:	A shorthand field.	114
Longitude:	The longitude of an object or point on a boundary. A negative value indicates a break in a boundary.	9

Table 2 (cont.)

Field Name	Definition	Length
Max drawdown rate function:	A function that gives the maximum rate at which an energy type can be removed from storage, as a function of the amount of the energy type in storage.	30
Max heat value:	The upper limit on the heat value of an energy type.	4
Max percent sulfur:	The upper limit on the percentage of sulfur in an energy type.	2
Max pour point:	The upper limit on the pour point of an energy type.	3
Max production capacity:	The upper limit on the amount of an energy type a facility can produce.	9
Max specific gravity:	The upper limit on the specific gravity of an energy type.	3
Min heat value:	The lower limit on the heat value of an energy type.	4
Min percent sulfur:	The lower limit on the percentage of sulfur in an energy type.	2
Min pour point:	The lower limit on the pour point of an energy type.	3
Min production capacity:	The lower limit on the amount of an energy type a facility can produce.	9
Min specific gravity:	The lower limit on the specific gravity of an energy type.	3
Natural gas production ID:	See Facility Codes above.	16
Nature of operation:	The type of a transmission and distribution company. Can be "interstate," "intrastate."	2
Non-political area name:	The name of a non-political area: oil production region, natural gas production region, demand sector area.	30
Number of categories:	The number of categories in an energy sector.	1

Table 2 (cont.)

Field Name	Definition	Length
Number of groups:	The number of energy groups in an energy category.	1
Number of links:	The number of links making up a route.	1
Number of sectors:	The number of energy sectors in the energy hierarchy.	1
Number of types:	The number of individual energy types in an energy group.	2
Non-political area code:	See Geographic Codes above.	16
Office or building #:	The office or building number as would appear in mailing address.	30
On date:	The date on which a snapshot is made.	7
Output facility ID:	The facility ID of the facility at the end of a transportation link.	16
Output facility type:	See Link Codes above.	4
Overlap:	The relation between two overlapping geographic areas. Can be "contains," "same," "intersects," "contained by."	8
Percent owned:	The percentage of one company owned by another.	2
Percent sulfur:	The percentage of sulfur in an energy type.	2
Phone #:	Area code and local phone number.	10
Point #:	The order of a point on the boundary of a geographic region, going clockwise, starting with the northernmost, westernmost point.	4
Pour point:	The pour point of an energy type.	3
Price:	The total price for a quantity of energy. This can be a unit price.	3
Processing capacity:	The maximum amount of wet gas a gas plant can process	9

Table 2 (cont.)

Field Name	Definition	Length
Processing plant ID:	See Facility Codes above.	16
Product bulk terminal ID:	See Facility Codes above.	16
Product storage ID:	See Facility Codes above.	16
Production capability:	The capability a refinery has to do other processes besides distillation, such as "re-forming," "catalytic cracking."	9
Production capacity:	Amount of energy type that can be handled or produced at a facility.	9
Q-field:	Shorthand field.	36
QP-field:	Shorthand field.	42
Refinery ID:	See Facility Codes above.	16
Region code:	See Geographic Codes above.	8
Region descriptor:	See Geographic Codes above.	1
Region name:	The name of a super-state geographic region.	30
Region office name:	The name of a regional office of a company.	30
Reported quantity:	An amount of energy reported to the database from outside. See computed quantity.	9
RES function:	One of the RES functional sectors. Can be "extraction," "refining and conversion," "transport and storage," "central station conversion," "transmission, distribution, and storage," "decentralized conversion," "utilizing device," "energy services."	8
Route name:	The name of a route entering or leaving a facility, such as "main 1," "alternate 2."	30
RL-field:	Shorthand field.	27
RN-field:	Shorthand field.	31
S-field:	Shorthand field.	28

Table 2 (cont.)

Field Name	Definition	Length
Sector code:	See Energy Codes above.	8
Sector name:	The name of an energy sector.	30
Sendout capacity:	The amount of an energy type that can leave a facility in a given period.	9
Service type:	Type of gas service. Can be "firm," "interruptible."	4
SIC codes:	SIC codes that apply to a large end-use consumer.	3
Specific gravity:	The specific gravity of an energy type.	3
State code:	See Geographic Codes above.	8
State descriptor:	See Geographic Codes above.	2
State name:	The name of a state.	30
Storage capacity:	The maximum amount of an energy type that can be stored at a facility.	9
Storage type:	The type of storage at a gas supplemental storage facility. Can be "LNG," "propane," or "SNG."	3
Strategic reserve ID:	See Facility Codes above.	16
Street address:	The address as would appear on mailing label.	30
Supplemental storage ID:	See Facility Codes above.	16
Switching capability:	The percentage of primary fuel usage that can be replaced by alternate fuel.	2
T and D facility ID:	See Facility Codes above.	16
TA (temporal aggregation):	The length of time a reporting period covers, such as "month," "quarter," "year."	1
To date:	The ending date of a reporting period.	7

Table 2 (cont.)

Field Name	Definition	Length
Transportation junction ID:	See Facility Codes above.	16
Type code:	See Energy Codes above.	8
Type name:	The name of an individual energy type.	30
Type table name:	The name of the energy type table for an energy group.	30
Underground storage ID:	See Facility Codes above.	16
Units:	The units for a quantity capacity or rate field.	3
Universal Code:	See Energy Codes above.	8
Well type:	The type of well natural gas is produced from. Can be "gas well" or "oil well."	3
Working capacity:	In underground gas storage, the amount of gas that can be easily removed from storage.	9
Zip:	Zip code.	5

III. MAPPING AND SIZE CALCULATIONS

One function of the ultimate EEMIS database is to provide an up-to-date, easily accessible database of energy flow/cost information for the U.S. energy system. This database will permit the surveillance and prediction of short-term energy supply-demand dislocations and provide an information framework for examining alternative energy solutions--possibly providing the optimal solution, if an objective can be quantified.

Formation of this database is a three-step design process, as illustrated below:

The design of the database first requires an examination of the appropriate portion of the physical world that is to be emulated. The database must accurately capture and represent physical reality. This preparation requires a systematic exposition of the characteristics that determine a physical entity. Consider, for example, an oil refinery. What characteristics make it a unique facility? Such factors as geographic location, owning company, distillation capacity, and refining capabilities all contribute to distinguish this refinery from others. We may also want to know if the refinery has connections to or impacts on other facilities, and, if so, how it affects and is affected by them. If the refinery goes off-line for maintenance or repair, for example, how will the distribution of oil in the area serviced by the refinery be changed? It is these kinds of questions that we are trying to answer by capturing essential information in the physical system.

In EEMIS, we need to be concerned with the representation of energy at a facility level and with a transportation network that accurately captures the flow of energy in the United States. Many of the critical energy-related

questions relate to the transportation, handling, and storage of energy. The abstraction of the appropriate physical system as a data structure is represented by the arrow labeled "abstraction process 1" in the diagram. We want to capture the essential information about energy components in the physical system and present it in a tabular or list form that implies a hierarchical organization. At the same time, we want to be able to present the information in such a way that questions from casual users of the system can be answered in a meaningful way. Several previous reports^{2-4,8} have discussed this information in some detail.

Once the data organization has been established, a second abstraction process takes place whereby the representative data element lists are developed into a structure or model that incorporates the conceptual view of the information as it appears to the user. This conceptual model reflects relationships that exist between the actual physical characteristics as they exist in the real world. If certain characteristics rely upon others to be present before they exist, this dependency should be reflected in the conceptual design. For example, quantity sent by source by route exists only if a quantity was shipped from a given source. In this process, the translation of lists of characteristics and relationships into abstract models of the real world must carefully capture the necessary information and interrelationships.

This second abstraction process has been accomplished in EEMIS by the development of the conceptual model⁶ based on the sector structures.⁸ The process of relating the necessary information is reflected in the three-dimensional hierarchical presentation of data within the conceptual model. The presentation is three dimensional because it incorporates not only the two-dimensional physical hierarchical dependencies but also the time dependencies (multiple entries at several levels). The questions of completeness, capturing all the necessary data of the sector structures, preservation of hierarchies, and capturing the relationships among the data elements are addressed within this section by the development of mapping algorithms that cross-reference the conceptual structures and hierarchies by the individual elements in the data element lists.

The third step in the process of database design is the systematic encoding of the conceptual designs into whatever medium is used to represent them.

For EEMIS, physical encoding involves rendering the conceptual models into ADABAS schemata. These schemata represent the templates for loading data into the database system. It is here that such concrete considerations as size of data items and representation of data become important. While this process is just beginning to be considered, it represents the final and perhaps most important part of the design: a careful analysis of the translation of the conceptual model to the physical system. Is the system implementable, given the computer storage and retrieval limitations of existing systems? If not, what compromises are necessary? Is the system sufficiently valuable after the necessary compromises to continue the implementation? Are there alternatives to the compromises?

THE MAPPING ALGORITHM

We shall now detail the mapping process, showing how the association of data elements with conceptual model segments and fields demonstrates completeness, and how the hierarchy of the data structure is preserved in the element list. The illustrative model for our discussion is the Crude Oil Bulk Terminal (PBT), which appears on page 74 of the Stony Brook report,⁶ and is reproduced, with additions, in the following pages. It is typical of the conceptual models in the database structure, and the mapping process is readily applicable to it. The appropriate data element structure appears on page 2 of the EEMIS Structure report.⁸

We first consider the mapping of the data element structure against the conceptual model to derive the "Segment/Field Cross-Reference Chart." We begin by specifying the chart number, name, and page in Ref. 6 in the chart header. In this example we get the header: "II. "CRUDE OIL BULK TERMINAL" (page 74). The Roman numeral chart number refers to the data element list describing the particular chart. The chart name is the name of the particular facility chart under consideration. The location is the page reference to the Stony Brook Report.⁶

Proceeding with the actual mapping, we begin with the first data element described in the data element structure. A singleton field that is unqualified by any additional data appears in the root (first) segment either as a separate field or as a field in a segment suspended from the root. The element location

for example, is mapped to the field "location" appearing in the root segment named PBT. The outline letters and numbers are copied directly into the Segment/Field Cross-Reference Chart, followed by the corresponding field or segment name. In the Segment/Field Cross Reference Chart, fields are designated by lower case and segments by upper case letters. The entry for the location element is therefore: "A 1. Expansion of location macro-field in PBT."

A macro-field is a shorthand field that actually consists of several fields. It has been coded in shorthand for ease of conceptual model design. These fields are listed below.

Location: street address, city, state, zip, phone #, contact person, latitude, longitude.

Q-field (quantity): from date, to date, reported quantity, computed quantity, units, TA (Time Aggregation).

QP-field (quantity/price): Q-field plus price.

S-field (snapshot) on date, reported quantity, computed quantity, units.

RN-field (route name): route name, number of links.

RL-field (route links): link #, link ID.

Any element that is qualified by the word "by," such as Quantity by Crude Type, refers to a field that occurs below the root segment. The level of the segment is determined by the number of "bys" that occur in the element list. Quantity by Crude Type for instance, appears in the segment TYPE RECEIVED in the conceptual model, which hangs off the segment TYPE (the element is reference by crude oil type) which in turn hangs from the root.

The remaining fields are encoded similarly. The full Segment/Field Cross Reference Chart follows.

II. CRUDE OIL BULK TERMINAL (page 74)

- A 1. Expansion of location macro-field in PBT
- 2. Company ID

- B.1. TYPE RECEIVED
 - 2. SOURCE RECEIVED/PRICE
 - 3. ROUTE RECEIVED
 - 4. SOURCE RECEIVED/PRICE
 - 5. Refer to cross-reference chart for TRANSPORTATION NODES
 - 6. TYPE RECEIVED PROJECTION for (B1) and SOURCE RECEIVED PROJECTION (B2).
- C 1. TYPE ON HAND
 - 2. Storage capacity in PBT (root segment)
- D 1. TYPE SENT
 - 2. DESTINATION SENT/PRICE
 - 3. ROUTE SENT
 - 4. DESTINATION SENT/PRICE
 - 5. Refer to cross reference chart for TRANSPORTATION NODES
 - 6. TYPE SENT PROJECTION for (D1) and DESTINATION SENT PROJECTION for (D2).

This completes the formation of the Segment/Field Cross-Reference Chart. We now proceed to insert the cross-reference identifiers into the conceptual model.

To map the data element list to the conceptual model proceed from top to bottom, this time transferring the outline "location" into the actual field or segment as it appears in the conceptual model. The field "location" in the conceptual model would, therefore, contain the code II A 1, meaning it was cross-referenced to chart II, level A, element 1. Those elements qualified by the word "by," such as Quantity "by Crude Type," are encoded in the same manner. The key field of the segment they are contained in is designated with the phrase "by 'qualifier'," where the 'qualifier' represents the qualifier name (here, the words "Crude Type"). The remaining fields are similarly encoded.

The end result of this mapping process is three cross-listed sections: data element structure against the conceptual model resulting in the Segment/Field Cross-Reference Charts; conceptual model by data element, resulting in the labeled forms of the models; and conceptual model by cross-reference chart where labels can be checked against outlines for data correspondence. The result of doing this mapping is the verification that all data elements do correspond to locations within the conceptual models, and therefore the mapping is complete.

Preservation of hierarchy can be demonstrated in a similar fashion. We utilize the data element sector structure chart and the conceptual model again. Using the conceptual model, label all links from the first segment below the root to other segments with a '1'. Links from those segments to segments below them in the hierarchy are labeled '2', etc. A detailed example is discussed below. Eventually all links except those from the root to the first level of segments in the hierarchy will be labeled with a number, indicating the level of origin of the link. If more than one segment hangs from the root, this labeling remains consistent since the level numbers are associated with the links between segments, not the segments themselves.

We now annotate the EEMIS data element list. Proceed by labeling all occurrences of the word "by" in a given element description in ascending order from left to right. The result of this procedure is that all indirect references within the element list (i.e., all information referenced via another piece of information) are labeled by the level of the indirect references. Not all data elements have explicit indirect references. Some are implied by the nature of the element. For example, consider element B5 in the Crude Oil Bulk Terminal structure, "Transportation Links Available" (Input). In this example, Transportation Links Available refers to all incoming links immediately connected to the facility whether in use or not. This information can be found in the transportation NODES section of the transportation system. On the other hand, there are segments, PBT SOURCE ROUTE LINKS, that appear in the conceptual model but that have no counterpart in the data element list. PBT SOURCE ROUTE LINKS, for example, would imply a data element entry "transportation links by crude type by source by route." The latter segment has been introduced to identify and define routes in use and several good alternative candidates in cases of disruptions.

With the labeling completed, we now equate the data element list to the conceptual model by the level numbers at each point in the structure. Thus, any level-one "by" maps to links descending from the segment PBT TYPE, any level-two "by" to PBT SOURCE and PBT DEST, and so forth. Taking any segment in the model and following a labeled link, one proceeds through increasing levels of indirect reference by following ascending order links, and proceeds through decreasing levels of indirect reference along descending order links. In the data element list, going from left to right through ascending labeled "bys" increases indirect references and depth in the hierarchy (although the actual

value contained at the terminal of the path is the first value specified in the particular element), and going from right to left decreases indirect references.

By way of example, consider element B3, "Quantity by Crude Type by Source by Route." The first "by" is labeled 1, so we are along a link descending from the segment PBT TYPE. The next "by" is labeled 2, and it is by Source, so we are along one of the three links below PBT SOURCE. The last "by" is labeled 3, so the segment is one of the two below PBT ROUTE FROM SOURCE. A moment's thought shows it to be PBT ROUTE RECEIVED, a quantity shorthand field.

The result of this process is a labeled data element list, corresponding to levels of indirect reference in the hierarchy, and a labeled conceptual model, with the links between segments labeled by ascending numbers as depth in the structure increases. The above procedure verifies that the hierarchy is truly preserved by the conceptual design.

The following three pages illustrate the entire process for crude oil bulk terminal facilities. The first page is the labeled conceptual model, showing the element mapping and link labeling. The second page is the cross-reference chart from the data-element list to the conceptual model, listing these segments and fields that map the data. The last page is the data element list, with the labeling associated with the "bys" by level of indirect reference.

CRUDE OIL BULK TERMINAL

II. CRUDE OIL BULK TERMINAL (PAGE 74)

- A
 - 1. Expansion of location macro-field in PBT
 - 2. Company ID
- B
 - 1. TYPE RECEIVED
 - 2. SOURCE RECEIVED / PRICE
 - 3. ROUTE RECEIVED
 - 4. SOURCE RECEIVED / PRICE
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE.
 - 6. TYPE RECEIVED PROJECTION FOR (B1) and SOURCE RECEIVED PROJECTION for (B2).
- C
 - 1. TYPE ON HAND
 - 2. Storage capacity in PCS
- D
 - 1. TYPE SENT
 - 2. DESTINATION SENT / PRICE
 - 3. ROUTE SENT
 - 4. DESTINATION SENT / PRICE
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE
 - 6. TYPE SENT PROJECTION FOR (D1) and DESTINATION SENT PROJECTION for (D2).

II. Crude Oil Bulk Terminal

A. Characteristics

1. Location--defines unique facility
2. Controlling company ID

B. Input

1. Quantity (M) by¹ crude type
2. Quantity (M) by¹ crude type by² source (e.g., domestic crude oil production facility, or foreign country)
3. Quantity (M) by¹ crude type by² source by route
4. Delivered price (M) by¹ crude type by² source
5. Transportation links available
6. Short-term receipt projections (M) by¹ crude type by² source

C. Storage

1. Quantity (MS) by¹ crude type
2. Capacity

D. Output

1. Quantity (M) by¹ crude type
2. Quantity (M) by¹ crude type by² destination (e.g. refinery)
3. Quantity (M) by¹ crude type by² destination by³ route
4. Price (M) (charged by terminal) by¹ crude type by² destination
5. Transportation links available
6. Short-term shipment projections (M) by¹ crude type by² destination

IV. CROSS-REFERENCE MAPS AND SIZES

The following section consists of the Segment/Field Cross Reference Charts for the conceptual models that comprise the database. Each model is described by four charts:

1. The cross-referenced and labeled conceptual model
2. The Segment/Field Cross-Reference Chart
3. The labeled data element list
4. Formula for calculating storage requirements

The storage requirements formulae were derived from the conceptual models. They represent the amount of storage (in bytes) that the data associated with a particular model will occupy in the EEMIS system.

DOMESTIC CRUDE OIL PRODUCTION

Xref, Chart 6 & 7

I. DOMESTIC CRUDE OIL PRODUCTION FACILITY (Page 69)

- A
 - 1. Domestic crude production ID of root segment PDC. Refer to section 2 of the Stony Brook report for subfield breakdowns and encoding.
 - 2. COMPONENT COMPANIES
- B
 - 1. TYPE PRODUCED / PRICE
 - 2. TYPE PRODUCED / PRICE
 - 3. Production capacity in TYPE
 - 4. TYPE PRODUCED PROJECTION
- C
 - 1. Refer to cross-reference chart for ASSOCIATED CRUDE STORAGE ID.
 - 2. Refer to cross-reference chart for ASSOCIATED CRUDE STORAGE ID.
- D
 - 1. TYPE SENT
 - 2. DESTINATION SENT / PRICE
 - 3. ROUTE SENT
 - 4. DESTINATION SENT / PRICE
 - 5. Refer to cross reference chart for NODE in the transportation portion.
 - 6. TYPE SENT PROJECTION for (D1) and DESTINATION SENT PROJECTION for (D2).

I. Domestic Crude Oil Production Facility

A. Characteristics

1. Domestic crude region ID	}	defines unique facility
2. Company ID		

B. Production

1. Quantity (M) by ¹crude type
2. Wellhead price (M) by ¹crude type
3. Daily production capacity by ¹crude type
4. Short-term production projections (M) by ¹crude type

C. Storage

1. Quantity (MS) by ¹crude type
2. Capacity

D. Output

1. Quantity (M) by ¹crude type
2. Quantity (M) by ¹crude type by ²destination (e.g., refinery, bulk terminal)
3. Quantity (M) by ¹crude type by ²destination by ³route
4. Price (M) (charged by production facility--exclusive of transportation costs) by crude type by destination
5. Transportation links ¹available ²
6. Short-term projections (M) by ¹crude type by ²destination

Note: (M) stands for monthly totals, historical data for 5 years, and projected data for 1 year.
(MS) stands for end-of-month snapshots, 5-year historical data.

DOMESTIC CRUDE OIL PRODUCTION

Definitions

f - number of crude production companies
c - number of component companies
t - number of types
h - number historic periods
p - number of projected periods
d - number of destinations
rd - number of routes to destination
L - number of links/per route

Computer Storage Formula

$$\begin{aligned}
 f & \cdot \left\{ \left| \text{PDC} \right| + \right. \\
 & \quad c \left| \text{COMPONENT COMPAN} \right. \\
 & \quad t \left[\left| \text{PDC TYPE} \right| + \right. \\
 & \quad \quad h \left| \text{PDC TYPE PRODUCED/} \right| + p \left| \text{PDC TYPE PRODUCED} \right| + \\
 & \quad \quad \quad \text{PRICE} \quad \quad \quad \text{PROJ} \\
 & \quad \quad h \left| \text{PDC TYPE SENT} \right| + p \left| \text{PDC TYPE SENT PROJ} \right| + \\
 & \quad d \left(\left| \text{PDC DEST} \right| + \right. \\
 & \quad \quad h \left| \text{PDC DEST SENT/} \right| + p \left| \text{PDC DEST SENT} \right| + \\
 & \quad \quad \quad \text{PRICE} \quad \quad \quad \text{PROJ} \\
 & \quad \quad rd \left\langle \left| \text{PDC ROUTE TO DEST} \right| + \right. \\
 & \quad \quad \quad L \left| \text{PDC DEST ROUTE} \right| + h \left| \text{PDC ROUTE SENT} \right| \left. \right\rangle \left. \right\} \\
 & \quad \quad \quad \text{LINKS}
 \end{aligned}$$

CRUDE OIL BULK TERMINAL

PBT

II. CRUDE OIL BULK TERMINAL (PAGE 74)

- A
 - 1. Expansion of location macro-field in PRT
 - 2. Company ID
- B
 - 1. TYPE RECEIVED
 - 2. SOURCE RECEIVED / PRICE
 - 3. ROUTE RECEIVED
 - 4. SOURCE RECEIVED / PRICE
 - 5. Refer to cross reference chart for NODE in the transportation portion.
 - 6. TYPE RECEIVED PROJECTION FOR (B1) and SOURCE RECEIVED PROJECTION for (B2).
- C
 - 1. TYPE ON HAND
 - 2. Storage capacity in PCS
- D
 - 1. TYPE SENT
 - 2. DESTINATION SENT / PRICE
 - 3. ROUTE SENT
 - 4. DESTINATION SENT / PRICE
 - 5. Refer to cross-reference chart for NODE in the transportation portion.
 - 6. TYPE SENT PROJECTION FOR (D1) and DESTINATION SENT PROJECTION for (D2).

II. Crude Oil Bulk Terminal

A. Characteristics

1. Location--defines unique facility
2. Controlling company ID

B. Input

1. Quantity (M) by¹ crude type
2. Quantity (M)¹ by crude type² by source (e.g., domestic crude oil production facility, or foreign country)
3. Quantity (M) by¹ crude type by² source by route
4. Delivered price (M) by¹ crude type by² source
5. Transportation links available
6. Short-term receipt projections (M) by¹ crude type by² source

C. Storage

1. Quantity (MS) by¹ crude type
2. Capacity

D. Output

1. Quantity (M) by¹ crude type
2. Quantity (M) by¹ crude type by² destination (e.g. refinery)
3. Quantity (M) by¹ crude type by² destination by³ route
4. Price (M) (charged by terminal) by¹ crude type by² destination
5. Transportation links available
6. Short-term shipment projections (M) by¹ crude type by² destination

Crude Oil Bulk Terminal (PBT)

Definitions

- f - number of facilities
- t - number of energy types
- s - number of source facilities
- d - number of destinations
- rs - number of routes from source
- rd - number of routes from destination
- L - number of links in a routes
- h - number of historical records
- p - number of projection records

Computer Storage Formula

$$\begin{aligned}
 f . & \left\{ \left| \text{PBT} \right| \right. \\
 & + t . \left[\left| \text{PBT TYPE} \right| + h . \left(\left| \text{PBT TYPE ON HAND} \right| + \left| \text{PBT TYPE REC} \right| \right) \right. \\
 & + p . \left| \text{PBT TYPE REC PROJ} \right| + h . \left| \text{PBT TYPE SENT} \right| + p . \left| \text{PBT TYPE SENT PROJ} \right| \\
 & + s . \left(\left| \text{PBT SOURCE} \right| + p . \left| \text{PBT SOURCE REC PROJ} \right| + h . \left| \text{PBT SOURCE REC/PRICE} \right| \right. \\
 & + rs . \left\langle \left| \text{PBT ROUTE FROM SOURCE} \right| + p . \left| \text{PBT ROUTE REC} \right| \right. \\
 & + L . \left| \text{PBT SOUCE ROUTE LINKS} \right| \left. \right\rangle \\
 & + d . \left(\left| \text{PBT DEST} \right| + p . \left| \text{PBT DEST SENT PROJ} \right| + h . \left| \text{PBT DEST SENT/PRICE} \right| \right. \\
 & + rd . \left\langle \left| \text{PBT ROUTE TO DEST} \right| + p . \left| \text{PBT ROUTE SENT} \right| \right. \\
 & + L . \left| \text{PBT DEST ROUTE LINKS} \right| \left. \right\rangle \left. \right\}
 \end{aligned}$$

STRATEGIC PETROLEUM RESERVE

**SPR types only

III. STRATEGIC PETROLEUM RESERVE (Page 76)

- A
 - 1. Expansion of location macro-field in PSR
 - 2. Company ID
- B
 - 1. TYPE RECEIVED
 - 2. SOURCE RECEIVED / PRICE
 - 3. ROUTE RECEIVED
 - 4. SOURCE RECEIVED / PRICE
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE.
 - 6. TYPE RECEIVED PROJECTION for (B1) and SOURCE RECEIVED PROJECTION for (B2).
- C
 - 1. TYPE-ON-HAND
 - 2. Maximum drawdown rate in TYPE
 - 3. Storage capacity in TYPE
- D
 - 1. TYPE SENT
 - 2. DESTINATION SENT / PRICE
 - 3. ROUTE SENT
 - 4. DESTINATION SENT / PRICE
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE.
 - 6. TYPE SENT PROJECTION for (D1) and DESTINATION SENT PROJECTION for (D2).

III. Strategic Petroleum Reserve

A. Characteristics

1. Location--defines unique facility
2. Company ID

B. Input

1. Quantity (M) by ¹ crude type
2. Quantity (M) by ¹ crude type by ² source (e.g., foreign country)
3. Quantity (M) by ¹ crude type by ² source by ³ route
4. Delivered price (M) by ¹ crude type by ² source
5. Transportation links available
6. Short-term receipt projections (M) by ¹ crude type by ² source

C. Storage

1. Quantity (MS) by ¹ SPR type
2. Maximum drawdown rate by ¹ SPR type
3. Capacity by ¹ SPR type

D. Output

1. Quantity by ¹ SPR type
2. Quantity (M) by ¹ SPR type by ¹ destination (e.g., refinery)
3. Quantity (M) by ¹ SPR type by ² destination by ³ route
4. Price (M) (charged by ¹ SPR Office) by ² SPR type by ³ destination
5. Transportation links available
6. Short-term shipment projections (M) by ¹ SPR type by ² destination

STRATEGIC PETROLEUM RESERVE

Definitions

f - number of SPR facilities
t - number of types
s - number of sources
d - number of destinations
rs - number of routes to source
rd - number of routes to destination
L - number of links/route
p - number of projected records
h - number of historic records

Computer Storage Formula

$$\begin{aligned}
 f & \left\{ \left| \text{PSR} \right| + \right. \\
 & \quad t \left[\left| \text{PSR TYPE} \right| + \right. \\
 & \quad \quad h \left| \text{PSR TYPE REC} \right| + p \left| \text{PSR TYPE REC PROJ} \right| + \\
 & \quad \quad h \left| \text{PSR TYPE SENT} \right| + p \left| \text{PSR TYPE SENT PROJ} \right| + \\
 & \quad \quad h \text{ PSR TYPE-ON-HAND} + \\
 & \quad \quad s \left(\left| \text{PSR SOURCE} \right| + \right. \\
 & \quad \quad h \left| \text{PSR SOURCE REC PRICE} \right| + p \left| \text{PSR SOURCE REC PROJ} \right| + \\
 & \quad \quad rs \triangleleft \text{PSR ROUTE FROM SOURCE} \right| + \\
 & \quad \quad L \left| \text{PSR SOURCE ROUTE LINKS} \right| + h \left| \text{PSR ROUTE REC} \triangleright \right) + \\
 & \quad \quad d \left(\left| \text{PSR DEST} \right| + \right. \\
 & \quad \quad h \left| \text{PSR DEST SENT/PRICE} \right| + p \left| \text{PSR DEST SENT PROJ} \right| + \\
 & \quad \quad rd \triangleleft \text{PSR ROUTE TO DEST} \right| + \\
 & \quad \quad L \left| \text{PSR DEST ROUTE} \right| + h \left| \text{PSR ROUTE SENT} \triangleright \right) \left. \right\} \\
 & \quad \quad \text{LINKS}
 \end{aligned}$$

OIL REFINERY

PRE

Xref Chart 6 & 7

refinery ID	facility name	location \$ IVA1	company ID \$ IVA2	associated crude storage ID IVC1,IVC2	associated product storage ID	
distillation capacity IVA3	min percent sulfur \$ IVA5.....	max percent sulfur \$	min specific gravity \$	max specific gravity \$	min pour point \$	max pour point \$
min heat value \$	max heat value \$IVA5	units	PRE OTHER CAPABILITIES IVA4

IV. OIL REFINERY (Page 78)

- A
 - 1. Expansion of location macro-field in PRE
 - 2. Company ID in PRE
 - 3. Distillation capacity in PRE
 - 4. OTHER CAPABILITIES
 - 5. Minimum and maximum of % sulfur, specific gravity, pour point, and heat value in PRE
 - 6. Minimum production capacity in TYPE
 - 7. Maximum production capacity in TYPE and REMAINING PRODUCTIVITY
- B
 - 1. TYPE RECEIVED
 - 2. SOURCE RECEIVED / PRICE
 - 3. ROUTE RECEIVED
 - 4. SOURCE RECEIVED / PRICE
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE.
 - 6. TYPE RECEIVED PROJECTION for (B1) and SOURCE RECEIVED PROJECTION for (B2)
- C
 - 1. Refer to cross-reference chart for ASSOCIATED CRUDE STORAGE ID and ASSOCIATED PRODUCT STORAGE ID.
 - 2. Refer to cross-reference chart for ASSOCIATED CRUDE STORAGE ID and ASSOCIATED PRODUCT STORAGE ID.
- D
 - 1. TYPE SENT
 - 2. DESTINATION SENT / PRICE
 - 3. ROUTE SENT
 - 4. DESTINATION SENT / PRICE
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE.
 - 6. TYPE SENT PROJECTION for (D1) and DESTINATION SENT PROJECTION for (D2).

IV. Oil Refinery

A. Characteristics

1. Location--defines unique facility
2. Controlling company ID
3. Distillation capacity
4. Other processing capacities (e.g., reforming, catalytic cracking)
5. Minimum and maximum values for each crude attribute (e.g., sulfur content, specific gravity)
6. Minimum production for each product type
7. Maximum production possible for each product type and total remaining capacity for other product types

B. Input

1. Quantity (M) by ¹ type for all petroleum groups
2. Quantity (M) by ¹ type by ² source (e.g., domestic facility, bulk terminal, foreign country)
3. Quantity (M) by ¹ type by ² source by ³ route
4. Delivered price (M) by ¹ type by ² source
5. Transportation links available
6. Short-term receipt projections (M) by ¹ type by ² source

C. Storage

1. Quantity by ¹ type (MS)
2. Capacity by ¹ group

D. Output

1. Quantity (M) by ¹ type
2. Quantity (M) by ¹ type by ² destination (e.g., demand facility, foreign country)
3. Quantity (M) by ¹ type by ² destination by ³ route
4. Price (M) (charged by refinery) by ¹ type by ² destination
5. Transportation links available
6. Short-term shipment projections (M) by ¹ type by ² destination

Petroleum Refinery (PRE)

Definitions

- f - number of refineries
- o - number of other capabilities
- t₁ - number of input types
- t₂ - number of product types
- s - number of sources
- d - number of destinations
- rs - number of routes from source
- rd - number of routes to destinations
- L - number of links in a route
- h - number of historical records
- p - number of projection records

Computer Storage Formula

$$\begin{aligned}
 & f \left\{ \left| \text{PRE} \right| + o \left| \text{PRE OTHER CAPABILITIES} \right| \right. \\
 & + t_1 \left[\left| \text{PRE TYPE} \right| + h \left| \text{PRE TYPE REC} \right| + p \left| \text{PRE TYPE REC PROJ} \right| \right. \\
 & + s \left(\left| \text{PRE SOURCE} \right| + h \left| \text{PRE SOURCE REC/PRICE} \right| \right. \\
 & + p \left| \text{PRE SOURCE REC PROJ} \right| + rs \triangleleft \left| \text{PRE ROUTE FROM SOURCE} \right| \\
 & + h \left| \text{PRE ROUTE REC} \right| \\
 & + L \left| \text{PRE SOURCE ROUTE LINKS} \right| \triangleright \left. \right] + t_2 \left[\left| \text{PRE TYPE} \right| \right. \\
 & t_{2-1} \left| \text{PRE REMAINING CAPACITIES} \right| + h \left| \text{PRE TYPE SENT} \right| \\
 & + p \left| \text{PRE TYPE SENT PROJ} \right| + d \left(\left| \text{PRE DEST} \right| + h \left| \text{PRE DEST SENT/PRICE} \right| \right. \\
 & + p \left| \text{PRE TYPE SENT PROJ} \right| \\
 & + p \left| \text{PRE DEST PROJ} \right| + rd \triangleleft \left| \text{PRE ROUTE TO DEST} \right| \\
 & + h \left| \text{PRE ROUTE SENT} \right| + L \left| \text{PRE DEST ROUTE LINKS} \right| \triangleright \left. \right\}
 \end{aligned}$$

OIL PRODUCT BULK TERMINAL

V. OIL PRODUCT BULK TERMINAL (Page 81)

- A
 - 1. Expansion of location macro-field in PPT
 - 2. Company ID in PPT
- B
 - 1. TYPE RECEIVED
 - 2. SOURCE RECEIVED / PRICE
 - 3. ROUTE RECEIVED
 - 4. SOURCE RECEIVED / PRICE
 - 5. Refer to cross reference chart for TRANSPORTATION NODE.
 - 6. TYPE RECEIVED PROJECTION for (B1) and SOURCE RECEIVED PROJECTION for (B2).
- C
 - 1. TYPE-ON-HAND
 - 2. Storage capacity in TYPE
- D
 - 1. TYPE SENT
 - 2. DESTINATION SENT / PRICE
 - 3. ROUTE SENT
 - 4. DESTINATION SENT / PRICE
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE.
 - 6. TYPE SENT PROJECTION for (D1) and DESTINATION SENT PROJECTION for (D2).

V. Oil Product Bulk Terminal

A. Characteristics

1. Location--defines unique facility
2. Controlling company ID

B. Input

1. Quantity (M) by¹ product type
2. Quantity (M) by¹ product type² by source (e.g., refinery, bulk terminal, foreign country)
3. Quantity (M) by¹ product type by² source by³ route
4. Delivered price (M) by¹ product type by² source
5. Transportation links available
6. Short-term receipt projections (M) by¹ product type by² source

C. Storage

1. Quantity (MS) by¹ product type
2. Capacity by¹ group

D. Output

1. Quantity (M) by¹ product type
2. Quantity (M) by¹ product type² by destination (e.g., demand facility, foreign country)
3. Quantity (M) by¹ product type by² destination by³ route
4. Price (M) (charged by terminal) by¹ product type by² destination
5. Transportation links available
6. Short-term shipment projections (M) by¹ product type by² destination

OIL PRODUCT BULK TERMINAL

Definitions

f - number of product bulk terminals
t - number of types
s - number of sources
d - number of destinations
rs - number of routes from source
rd - number of routes to destination
L - number of links/routes
h - number of historic records
p - number of projection records

Computer Storage Formula

$$\begin{aligned}
 f \{ & | \text{PPT} | + \\
 & t \left[| \text{PPT TYPE} | + \right. \\
 & \quad h | \text{PPT TYPE-ON-HAND} | + h | \text{PPT TYPE REC} | + p | \text{PPT TYPE REC PROJ} | + \\
 & \quad h | \text{PPT TYPE SENT} | + p | \text{PPT TYPE SENT PROJ} | + \\
 & \quad s \left(| \text{PPT SOURCE} | + \right. \\
 & \quad \quad h | \text{PPT SOURCE REC/PRICE} | + p | \text{PPT SOURCE REC/PROJ} | + \\
 & \quad \quad rs \triangleleft | \text{PPT ROUTE FROM SOURCE} | + \\
 & \quad \quad \quad L | \text{PPT SOURCE ROUTE LINKS} | + L | \text{PPT ROUTE REC} \triangleright \left. \right) + \\
 & \quad d \left(| \text{PPT DEST} | + \right. \\
 & \quad \quad h | \text{PPT DEST REC/PRICE} | + p | \text{PPT DEST REC PROJ} | + \\
 & \quad \quad rd \triangleleft | \text{PPT ROUTE TO DEST} | + \\
 & \quad \quad \quad L | \text{PPT DEST ROUTE LINKS} | + h | \text{PPT ROUTE SENT} \triangleright \left. \right) \left. \right\}
 \end{aligned}$$

CRUDE OIL STORAGE

PETROLEUM PRODUCT STORAGE PPS

VI. CRUDE OIL STORAGE (Page 71)

This is a facility associated with those facilities requiring storage of crude oil prior to processing into products or shipment. These facilities that refer to this one are:

DOMESTIC CRUDE OIL PRODUCTION FACILITY

OIL REFINERY

The parent facility and this facility are linked by containment of the facility ID of the parent in the associated storage facility as a field, and containment of the storage facility's ID in the parent as a field. This association allows one to "catalogue" all associated storage facilities to a given refinery or production facility.

The data elements in question are the C 1 and C 2 elements dealing with storage. C 1, Quantity by Type, is cross-referenced by the TYPE-ON-HAND segment, while C 2, Capacity, is cross-referenced by the field Storage Capacity in the segment TYPE.

VII. PETROLEUM PRODUCT STORAGE (Page 71)

This is an associated facility to those facilities requiring storage of petroleum products prior to shipment or use. Those facilities that refer to this one are:

OIL REFINERY

The parent facility and this facility are linked by containment of the facility ID of the parent in the associated storage facility as a field, and containing the storage facility's ID in the parent as a field. This association allows one to "catalogue" all associated storage facilities to a given refinery.

The data elements in question are the C 1 and C 2 elements dealing with storage. C 1, Quantity by Type, is cross-referenced by the TYPE-ON-HAND segment, while C 2, Capacity, is cross-referenced by the field Storage Capacity in the segment TYPE.

CRUDE OIL STORAGE

Definitions

f - number of associated crude oil storage facilities

t - number of types

h - number of historical records

p - number of projected records

Computer Storage Formula

$$f \left\{ | PCS | + \right. \\ \left. t \left[| PCS TYPES | + \right. \right. \\ \left. \left. h | PCS TYPE-ON-HAND | + h | PCS TYPE REC | + \right. \right. \\ \left. \left. h | PCS TYPE SENT | \right] \right\}$$

PETROLEUM PRODUCT STORAGE

Definitions

f - number of associated product storage facilities

t - number of types

h - number of historical records

p - number of projected records

Computer Storage Formula

$$f \left\{ \left| \text{PPS} \right| + \right. \\ \left. t \left[\left| \text{PPS TYPE} \right| + \right. \right. \\ \left. \left. h \left| \text{PPS TYPE-ON-HAND} \right| + h \left| \text{PPS TYPE REC} \right| + \right. \right. \\ \left. \left. h \left| \text{PPS TYPE SENT} \right| \right] \right\}$$

NATURAL GAS PRODUCTION

I. NATURAL GAS PRODUCTION (Page 83)

- A 1. Natural gas production ID of NPR. Refer to section 2 of the Stony Brook report for subfield breakdowns and encoding.
- 2. COMPONENT COMPANIES
- B 1. PRODUCED / PRICE
- 2. WELL PRODUCED
- 3. PRODUCED / PRICE
- 4. Production capacity in NPR
- 5. PRODUCED PROJECTION
- C 1. SENT
- 2. DESTINATION SENT / PRICE
- 3. ROUTE SENT
- 4. T AND D SENT / PRICE
- 5. DESTINATION SENT / PRICE
- 6. T AND D SENT / PRICE
- 7. Refer to cross reference chart for TRANSPORTATION NODE.
- 8. T AND D SENT PROJECTION

I. Natural Gas Production

A. Characteristics

1. Natural gas production region ID	}	defines unique facility
2. Company ID		

B. Production

1. Quantity (M)
2. Quantity (M) by ¹ well type (gas well or oil well)
3. Wellhead price (M)
4. Daily production capacity
5. Short-term production projections (M)

C. Output

1. Quantity (M)
2. Quantity (M) by ¹ destination (processing plant)
3. Quantity (M) by ¹ destination by ² route
4. Quantity (M) by ¹ purchasing T&D facility
5. Price (M) (charged by producing facility) by ¹ destination
6. Price (M) (charged by producing facility) by ¹ purchasing T&D facility
7. Transportation links available (gathering lines).
8. Short-term shipment projections (M) by ¹ purchasing T&D facility

Note: (M) monthly totals, historical data for 5 years, projected data for 1 year.
(MS) end-of-month snapshots, 5 year historical data.

NATURAL GAS PRODUCTION

Definitions

f - number of natural gas production facilities
 c - number of component companies
 w - number of well types
 d - number of destination
 td - number of F and D facilities
 rd - number of routes to dest
 L - number of links/route
 h - number of historic records
 p - number of projected records

Computer Storage Formula

$$\begin{aligned}
 f \{ & \{ \text{NPR} \mid + \\
 & c \mid \text{NPR COMPONENT COMPANIES} \mid + h \mid \text{NPR PRODUCED/PRICE} \mid + \\
 & p \mid \text{NPR PRODUCED PROJ} \mid + h \mid \text{NPR SENT} \mid + \\
 & w \mid \text{NPR WELL TYPE} \mid + h \mid \text{NPR WELL PRODUCED} \mid + \\
 & d \left[\mid \text{NPR DEST} \mid + \right. \\
 & \quad h \mid \text{NPR DEST SENT/PRICE} \mid + \\
 & \quad rd \left(\mid \text{NPR ROUTE TO DEST} \mid + \right. \\
 & \quad \quad L \mid \text{NPR DEST ROUTE LINKS} \mid + h \mid \text{NPR ROUTE SENT} \mid \left. \right) \\
 & \quad td \left(\mid \text{NPR T AND D} \mid + \right. \\
 & \quad \quad \left. \left. \left. h \mid \text{NPR T AND D} \mid + p \mid \text{NPR T AND D} \mid \right) \right) \right] \} \\
 & \quad \quad \quad \text{SENT/PRICE} \quad \quad \quad \text{SENT PROJ}
 \end{aligned}$$

GAS PROCESSING PLANT

**Light products types only

II. GAS PROCESSING PLANT (Page 85)

- A
 - 1. Expansion of location macro-field in NPP
 - 2. Company ID in NPP
 - 3. Processing capacity in NPP
- B
 - 1. TYPE RECEIVED
 - 2. SOURCE RECEIVED / PRICE
 - 3. ROUTE RECEIVED
 - 4. SOURCE RECEIVED / PRICE
 - 5. Refer to cross reference chart for TRANSPORTATION NODE.
 - 6. SOURCE RECEIVED PROJECTION
- C
 - 1. TYPE-ON-HAND
 - 2. Storage capacity in TYPE
- D
 - 1. TYPE SENT
 - 2. DESTINATION SENT, / PRICE
 - 3. ROUTE SENT
 - 4. DESTINATION SENT / PRICE
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE.
 - 6. DESTINATION SENT PROJECTION

II. Gas Processing Plant

A. Characteristics

1. Location--defines unique facility
2. Controlling company ID
3. Capacity (thru-put) per unit time

B. Input (wet gas)

1. Quantity (M)
2. Quantity (M) by¹ source (producing region) facility ID
3. Quantity (M) by¹ source by² route
4. Delivered price (M) by¹ source
5. Transportation links available
6. Short-term receipt projections (M) by¹ source

C. Storage of natural gas liquids (NGL)

1. Quantity (MS) by¹ type
2. Capacity by¹ group

D. Output

1. Quantity (M) by¹ type
2. Quantity (M) by¹ type by² destination (T&D facility, refinery, bulk terminal)
3. Quantity (M) by¹ type by² destination by³ route (primary alternative)
4. Price (M) by¹ type by² destination
5. Transportation links available
6. Short-term shipment projections (M) by¹ type by² destination

GAS PROCESSING PLANT

Definitions

f - number of gas processing plants
 t - number of types
 s - number of sources
 d - number of destinations
 rs - number of routes from source
 rd - number of routes to destination
 L - number of link/route
 h - number of historic records
 p - number of projected records

Computer Storage Formula

$$\begin{aligned}
 & f \left\{ \left| \text{NPP} \right| + \right. \\
 & \quad t \left[\left| \text{NPP TYPE} \right| + \right. \\
 & \quad \quad h \left| \text{NPP TYPE-ON-HAND} \right| + h \left| \text{NPP TYPE REC} \right| + h \left| \text{NPP TYPE SENT} \right| + \\
 & \quad \quad s \left(\left| \text{NPP SOURCE} \right| + \right. \\
 & \quad \quad \quad h \left| \text{NPP SOURCE REC/PRICE} \right| + p \left| \text{NPP SOURCE REC PROJ} \right| + \\
 & \quad \quad \quad rs \left\langle \left| \text{NPP ROUTE FROM SOURCE} \right| + \right. \\
 & \quad \quad \quad \quad L \left| \text{NPP SOURCE ROUTE LINKS} \right| + h \left| \text{NPP ROUTE REC} \right| \left. \right\rangle \\
 & \quad \quad d \left| \text{NPP DEST} \right| + \\
 & \quad \quad \quad h \left| \text{NPP DEST SENT/PRICE} \right| + p \left| \text{NPP DEST SENT PROJ} \right| + \\
 & \quad \quad \quad rd \left\langle \left| \text{NPP ROUTE TO DEST} \right| + \right. \\
 & \quad \quad \quad \quad L \left| \text{NPP DEST ROUTE LINKS} \right| + h \left| \text{NPP ROUTE SENT} \right| \left. \right\rangle \left. \right\}
 \end{aligned}$$

LNG TERMINAL

III. LNG TERMINAL (Page 87)

- A
 - 1. Expansion of location macro-field in NLG
 - 2. Company ID in NLG
- B
 - 1. TYPE RECEIVED
 - 2. SOURCE RECEIVED / PRICE
 - 3. ROUTE RECEIVED
 - 4. SOURCE RECEIVED / PRICE
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE.
 - 6. SOURCE RECEIVED PROJECTION
- C
 - 1. TYPE-ON-HAND
 - 2. Storage capacity in TYPE
 - 3. Send out capacity in TYPE
- D
 - 1. TYPE SENT
 - 2. DESTINATION SENT / PRICE
 - 3. ROUTE SENT
 - 4. DESTINATION SENT / PRICE
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE.
 - 6. DESTINATION SENT PROJECTION

III. LNG Terminal*

A. Characteristics

1. Location--defines unique facility
2. Controlling company ID

B. Input

1. Quantity (M) by¹ type (LNG, natural gas)
2. Quantity (M) by¹ type by² source, (e.g., foreign country facility)
3. Quantity (M) by¹ type by² source by³ route
4. Delivered price (M) by¹ type by² source
5. Transportation links available
6. Short-term receipt projections (M) by¹ type by² source

C. Storage

1. Quantity (MS) by¹ type in storage
2. Capacity by¹ group
3. Sendout capacity (rate) by¹ group

D. Output

1. Quantity (M) by¹ type
2. Quantity (M) by¹ type by² destination (e.g., T&D facility)
3. Quantity (M) by¹ type by² destination by³ route
4. Price (M) by¹ type by² destination
5. Transportation links available
6. Short-term shipment projections (M) by¹ type by² destination

*Importing and exporting LNG terminals are accounted for in this structure.

Importing terminal: Import and store LNG, convert LNG to natural gas, and send out natural gas via pipeline.

Exporting terminal: Receive natural gas from producing facility, convert natural gas to LNG and store, and export LNG via tanker.

LNG TERMINAL

Definitions

f - number of LNG terminals
 t - number of LNG types
 s - number of LNG sources
 d - number of LNG destinations
 rs - number of LNG routes from source
 rd - number of LNG routes to destination
 L - number of LNG links/route
 h - number of LNG historic records
 p - number of LNG projected records

Computer Storage Formula

$$\begin{aligned}
 f & \left\{ \left| \text{NLG} \right| + \right. \\
 & t \left[\left| \text{NLG TYPE} \right| + \right. \\
 & \quad h \left| \text{NLG TYPE-ON-HAND} \right| + h \left| \text{NLG TYPE SENT} \right| + h \left| \text{NLG TYPE REC} \right| + \\
 & \quad s \left(\left| \text{NLG SOURCE} \right| + \right. \\
 & \quad \quad h \left| \text{NLG SOURCE REC/PRICE} \right| + L \left| \text{NLG SOURCE REC/PROJ} \right| + \\
 & \quad \quad rs \left\langle \left| \text{NLG ROUTE FROM SOURCE} \right| + \right. \\
 & \quad \quad \quad L \left| \text{NLG SOURCE ROUTE LINKS} \right| + h \left| \text{NLG ROUTE REC} \right\rangle + \\
 & \quad d \left(\left| \text{NLG DEST} \right| + \right. \\
 & \quad \quad h \left| \text{NLG DEST SENT/PRICE} \right| + p \left| \text{NLG DEST SENT PROJ} \right| + \\
 & \quad \quad rd \left\langle \left| \text{LG ROUTE TO DEST} \right| + \right. \\
 & \quad \quad \quad L \left| \text{NLG DEST ROUTE LINKS} \right| + h \left| \text{NLG ROUTE SENT} \right\rangle \left. \right] \left. \right\}
 \end{aligned}$$

UNDERGROUND STORAGE

NUS

IV. UNDERGROUND STORAGE (Page 89)

- A
 - 1. Expansion of location macro-field in NUS
 - 2. Company ID in NUS
 - 3. Maximum drawdown rate function in NUS
- B
 - 1. RECEIVED
 - 2. SOURCE RECEIVED
 - 3. ROUTE RECEIVED
 - 4. OWNING COMPANY RECEIVED
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE.
- C
 - 1. ON-HAND / STORAGE COST
 - 2. OWNING COMPANY ON-HAND / STORAGE COST
 - 3. Cost in ON-HAND / STORAGE COST
 - 4. Cost in OWNING COMPANY ON-HAND / STORAGE COST
 - 5. Working capacity in NUS
 - 6. Cushion capacity in NUS
 - 7. Maximum drawdown rate in NUS
- D
 - 1. SENT
 - 2. DESTINATION SENT
 - 3. ROUTE SENT
 - 4. OWNING COMPANY SENT
 - 5. Refer to cross-reference chart for TRANSPORTATION NODE.

IV. Underground Storage

A. Characteristics

1. Location--defines unique facility
2. Controlling company ID
3. Drawdown rate as a function of gas in storage

B. Inputs

1. Quantity (M)
2. Quantity (M) by ¹ source
3. Quantity (M) by ¹ source by ² route
4. Quantity (M) by ¹ owning company ID (T&D facility)
5. Transportation links available

C. Storage

1. Quantity (MS)
2. Quantity (MS) by ¹ owning company ID
3. Price (M) charged for storage
4. Price (M) charged for storage by ¹ owning company ID
5. Working gas capacity
6. Cushion gas capacity
7. Maximum drawdown rate

D. Outputs

1. Quantity (M)
2. Quantity (M) by ¹ destination
3. Quantity (M) by ¹ destination by ² route
4. Quantity (M) by ¹ owning company ID
5. Transportation links available

UNDERGROUND STORAGE

Definitions

f - number of underground storage facilities
 c - number of owning companies
 s - number of sources
 d - number of destinations
 rs - number of routes from source
 rd - number of routes to destination
 L - number of links/route
 h - number of historic records
 p - number of projections

Computer Storage Formula

$$\begin{aligned}
 f & \left\{ \left| \text{NUS} \right| + \right. \\
 & \quad h \left| \text{NUS ON HAND/STORAGE COST} \right| + h \left| \text{NUS REC} \right| + h \left| \text{NUS SENT} \right| + \\
 & \quad c \left(\left| \text{NUS OWNING CO} \right| + \right. \\
 & \quad \quad h \left| \text{NUS OWNING CO ON HAND/} \right| + h \left| \text{NUS OWNING CO SENT} \right| + \\
 & \quad \quad \text{STORAGE COST} \\
 & \quad \quad h \left| \text{NUS OWNING CO REC} \right| \left. \right) + \\
 & \quad s \left(\left| \text{NUS SOURCE} \right| + \right. \\
 & \quad \quad h \left| \text{NUS SOURCE REC} \right| + \\
 & \quad \quad rs \left\langle \left| \text{NUS ROUTE FROM SOURCE} \right| + \right. \\
 & \quad \quad \quad L \left| \text{NUS SOURCE ROUTE LINKS} \right| + h \left| \text{NUS ROUTE REC} \right| \left. \right\rangle \\
 & \quad d \left(\left| \text{NUS DEST} \right| + \right. \\
 & \quad \quad h \left| \text{NUS DEST SENT} \right| + \\
 & \quad \quad rd \left\langle \left| \text{NUS ROUTE TO DEST} \right| + \right. \\
 & \quad \quad \quad L \left| \text{NUS DEST ROUTE LINKS} \right| + h \left| \text{NUS ROUTE REC} \right| \left. \right\rangle \left. \right\}
 \end{aligned}$$

SUPPLEMENTAL FUEL STORAGE

V. SUPPLEMENTAL FUEL STORAGE FACILITY (Page 91)

- A 1. Expansion of location macro-field in NSS
- 2. Storage type in NSS
- 3. Company ID in NSS
- B 1. TYPE RECEIVED
- 2. SOURCE RECEIVED / PRICE
- 3. ROUTE RECEIVED
- 4. SOURCE RECEIVED / PRICE
- 5. Refer to cross reference chart for TRANSPORTATION NODE.
- C 1. TYPE-ON-HAND
- 2. Storage capacity in TYPE
- 3. Maximum drawdown rate in TYPE
- D 1. TYPE SENT
- 2. DESTINATION SENT / PRICE
- 3. ROUTE SENT
- 4. DESTINATION SENT / PRICE
- 5. Refer to cross reference chart for TRANSPORTATION NODE.

V. Supplemental Fuel Storage Facility

A. Characteristics

1. Location	}	defines unique facility
2. Storage type (LNG, propane, SNG)*		
3. Controlling company ID		

B. Inputs

1. Quantity (M) by ¹ type
2. Quantity (M) by ¹ type by ² source (T&D facility, refinery, bulk terminal)
3. Quantity (M) by ¹ type by ² source by ³ route
4. Delivered price (M) by ¹ type by ² source
5. Transportation links available

C. Storage

1. Quantity (MS) by ¹ type
2. Capacity by ¹ group
3. Maximum drawdown rate by ¹ group

D. Output

1. Quantity (M) by ¹ type
2. Quantity (M) by ¹ type by ² destination (T&D facility)
3. Quantity (M) by ¹ type by ² destination by ³ route
4. Price (M) by ¹ type by ² destination
5. Transportation links available

*Synthetic natural gas (SNG) is produced on site from naphtha.

SUPPLEMENTAL FUEL STORAGE

Definitions

f - number of supplemental fuel storage facilities
 t - number of types
 s - number of sources
 d - number of destinations
 rs - number of routes from source
 rd - number of routes to destination
 L - number of links/route
 h - number of historic records
 p - number of projection records

Computer Storage Formula

$$\begin{aligned}
 f & \left\{ \left| \text{NSS} \right| + \right. \\
 & t \left[\left| \text{NSS TYPE} \right| + \right. \\
 & \quad h \left| \text{NSS TYPE-ON-HAND} \right| + h \left| \text{NSS TYPE REC} \right| + h \left| \text{NSS TYPE SENT} \right| + \\
 & \quad s \left(\left| \text{NSS SOURCE} \right| + \right. \\
 & \quad \quad h \left| \text{NSS SOURCE REC/PRICE} \right| + \\
 & \quad \quad rs \triangleleft \left| \text{NSS ROUTE FROM SOURCE} \right| + \\
 & \quad \quad \quad L \left| \text{NSS SOURCE ROUTE LINKS} \right| + h \left| \text{NSS ROUTE REC} \right| \triangleright + \\
 & \quad d \left(\left| \text{NSS DEST} \right| + \right. \\
 & \quad \quad h \left| \text{NSS DEST SENT/PRICE} \right| + \\
 & \quad \quad rd \triangleleft \left| \text{NSS ROUTE TO DEST} \right| + \\
 & \quad \quad \quad L \left| \text{NSS DEST ROUTE LINKS} \right| + h \left| \text{NSS ROUTE SENT} \right| \triangleright \left. \right) \left. \right\}
 \end{aligned}$$

NTD

TRANSMISSION AND DISTRIBUTION FACILITY NTD

NTD cont.

VI. TRANSMISSION AND DISTRIBUTION (T&D) FACILITY (Pages 93 & 94)

- A. 1. Company ID in NTD
- 2. Nature of operation in NTD
- 3. STATES SERVED
- B 1. RECEIVED
- 2. SOURCE RECEIVED / PRICE
- 3. ROUTE RECEIVED
- 4. SELLING FACILITY / RECEIVED PRICE
- 5. SOURCE RECEIVED / PRICE
- 6. SELLING FACILITY / RECEIVED PRICE
- 7. Refer to cross reference chart for TRANSPORTATION NODE.
- 8. RECEIVED PROJECTION for (B1), SOURCE RECEIVED PROJECTION for (B2), ROUTE RECEIVED PROJECTION for (B3), and SELLING FACILITY RECEIVED PROJECTION for (B4).
- C 1. TYPE-IN-STORAGE
- D 1. SENT
- 2. END USER SENT
- 3. DESTINATION SENT / PRICE
- 4. ROUTE SENT
- 5. SERVICE SENT / PRICE
- 6. DESTINATION CURTAILMENT
- 7. SERVICE CURTAILMENT
- 8. ALTERNATE TYPE USED
- 9. Switching capability in ALTERNATE TYPE
- 10. DESTINATION SENT / PRICE
- 11. SERVICE SENT / PRICE
- 12. SENT PROJECTION for (D1), END USER SENT PROJECTION for (D2), DESTINATION SENT PROJECTION for (D3), ROUTE SENT PROJECTION for (D4), SERVICE SENT PROJECTION for (D5), DESTINATION CURTAILMENT PROJECTION for (D6), and SERVICE CURTAILMENT PROJECTION for (D7).

VI. Transmission and Distribution (T&D) Facility

A. Characteristics

1. Company ID--defines unique facility.
2. Nature of operation (interstate or intrastate)
3. States served

B. Inputs

1. Quantity (M)
2. Quantity (M) ¹by source (processing plant, T&D facility, storage facility, foreign country, or LNG terminal)
3. Quantity (M) ¹by source ²by route
4. Quantity (M) ¹by selling production facility
5. Delivered price (M) ¹by source
6. Delivered price (M) ¹by selling production facility
7. Transportation links available
8. Short-term projections (M) for item numbers 1, 2, 3, and 4

C. Storage

1. Quantity (MS) ¹by type (stored in other facilities)

D. Output

1. Quantity (M)
2. Quantity (M) delivered to end-use customers
3. Quantity (M) ¹by destination (T&D facility, storage facility, foreign country, or demand facility)
4. Quantity (M) ¹by destination ²by route
5. Quantity (M) delivered to end-use customers ¹by type of service (firm or interruptible)
6. Quantity (M) of gas curtailments ¹by demand facility
7. Quantity (M) of gas curtailments ¹by type of service

8. Quantity (M) of alternate fuels used by alternate fuel type (e.g., No. 2 fuel oil, propane, bituminous coal) for all end-use customers
9. Existing capability (% gas displaced) for switching to alternate fuels by alternate fuel for all end-use customers
10. Delivered price (M) by destination
11. Delivered price (M) to end-use customers by type of service
12. Short-term projections (M) for item numbers 1 through 7

Transmission and Distribution (T&D) Facility

Definitions

- f - number of T&D facilities
- st - number of states served
- t - number of storage types
- sf - number of selling facilities
- s - number of sources
- d - number of destinations
- rs - number of routes to source
- rd - number of routes to destination
- L - number of links in a route
- n - number of service types
- a - number of alternate energy types
- h - number of historical records
- p - number of projection records

Computer Storage Formula

$$\begin{aligned}
 f . \quad & \left\{ \left| \text{NTD} \right| + st \left| \text{NTD STATES SERVED} \right| + h \left| \text{NTD REC} \right| + p \left| \text{NTD REC PROJ} \right| \right. \\
 & + h \left| \text{NTD SENT} \right| + p \left| \text{NTD SENT PROJ} \right| + h \left| \text{NTD END USER SENT} \right| \\
 & + p \left| \text{NTD END USER SENT PROJ} \right| \\
 & + t \left(\left| \text{NTD STORAGE TYPE} \right| + h \left| \text{NTD TYPE-IN-STORAGE} \right| \right) \\
 & + sf \left(\left| \text{NTD SELLING FACILITY} \right| + h \left| \text{NTD SELLING FAC/REC/PRICE} \right| \right. \\
 & + p \left| \text{NTD SELLING FAC/REC/PROJ} \right| \left. \right) . \\
 & + n \left(\left| \text{NTD SERVICE TYPE} \right| + h \left| \text{NTD SERVICE SENT/PRICE} \right| \right. \\
 & + p \left| \text{NTD SERVICE SENT PROJ} \right| + h \left| \text{NTD SERVICE CURTAILMENT} \right| \\
 & + p \left| \text{NTD SERVICE CURTAILMENT PROJ} \right| \left. \right) \\
 & + a \left(\left| \text{NTD ALTERNATE TYPE} \right| + h \left| \text{NTD ALTERNATE TYPE USED} \right| \right) \\
 & + c \left[\left| \text{NTD SOURCE} \right| + h \left| \text{NTD SOURCE REC/PRICE} \right| + p \left| \text{NTD SOURCE REC} \right| \right. \\
 & \quad \left. \text{PROJ} \right]
 \end{aligned}$$

$$+ \text{rs} \left(\left| \text{NTD ROUTE FROM SOURCE} \right| + \text{h} \left| \text{NTD ROUTE REC.} \right| \right.$$

$$+ \text{p} \left| \text{NTD ROUTE REC PROJ} \right|$$

$$+ \text{L} \left| \text{NTD SOURCE ROUTE LINKS} \right| \left. \right) \left. \right]$$

$$\text{d} \left[\left| \text{NTD DEST} \right| + \text{h} \left| \text{NTD DEST SENT/PRICE} \right| \right.$$

$$+ \text{p} \left| \text{NTD DEST SENT PROJ} \right| + \text{h} \left| \text{NTD DEST CURTAILMENT} \right|$$

$$+ \text{p} \left| \text{NTD DEST CURTAILMENT PROJ} \right|$$

$$+ \text{rs} \left(\left| \text{NTD ROUTE TO DEST} \right| + \text{h} \left| \text{NTD ROUTE SENT} \right| \right.$$

$$+ \text{p} \left| \text{NTD ROUTE SENT PROJ} \right| + \text{L} \left| \text{NTD DEST ROUTE LINKS} \right| \left. \right) \left. \right\}$$

DEMAND FACILITY
FOR DRC, DUT, DTR, DIN, DMI, DRE, DCO

DEM

**natural gas only

I. DEMAND FACILITY (Page 97)

- A. 1. Demand facility ID of DEM. Refer to section 2 of the Stony Brook report for subfield breakdowns and encoding.
- 2. Demand sector in DEM
- B 1. TYPE RECEIVED
- 2. SOURCE RECEIVED / PRICE
- 3. ROUTE RECEIVED
- 4. SERVICE RECEIVED / PRICE
- 5. TYPE CURTAILMENT
- 6. SERVICE CURTAILMENT
- 7. ALTERNATE TYPE USED
- 8. SOURCE RECEIVED / PRICE
- 9. SERVICE RECEIVED / PRICE
- 10. Refer to cross-reference chart for TRANSPORTAION NODE.
- 11. TYPE RECEIVED PROJECTION for (B1), SOURCE RECEIVED PROJECTION for (B2), ROUTE RECEIVED PROJECTION for (B3), SERVICE RECEIVED PROJECTION for (B4), TYPE CURTAILMENT PROJECTION for (B5), and SERVICE CURTAILMENT PROJECTION for (B6).
- 12. Switching capability in ALTERNATE TYPE

I. Demand Facility

A. Characteristics

1. State ID	}	defines unique facility
2. Demand sector (residential, commercial, industrial, transportation, electric utility, residential/commercial, military)		

B. Input

1. Quantity (M) by ¹ fuel type (e.g., natural gas, No. 2 fuel oil, bituminous coal)
2. Quantity (M) by ¹ fuel type by ¹ source
3. Quantity (M) by ¹ fuel type by ² source by ³ route
4. Quantity (M) of natural gas by ¹ source by ² type of service (firm or interruptible)
5. Quantity (M) of gas curtailments
6. Quantity (M) of gas curtailment by ¹ type of service
7. Quantity (M) of alternate fuel usage with quantity (M) of primary fuel
type replaced by alternate fuel type by primary fuel type replaced
8. Price (M) by ¹ fuel type by ² source
9. Price (M) for natural gas by ¹ source by ² type of service
10. Transportation links available
11. Short-term projections (M) for item numbers 1 thru 6
12. Existing capability (% fuel displaced) for switching by ¹ primary fuel
types by each alternate fuel type

Note: For the interim system the facility level demand facility is at the state level of aggregation.

DEMAND FACILITY

Definitions

f - number of demand sectors
t - number of types
a - number of alternate types
s - number of sources
rs - number of routes from source
v - number of service types
L - number of links/route
h - number of historic records
p - number of projected records

Computer Storage Formula

[illegible]

****natural gas only**

II. LARGE END-USE CONSUMER (Page 99)

- A
 - 1. Expansion of location macro-field in LAR
 - 2. Company ID in LAR
 - 3. Demand section in LAR
 - 4. FERC PERCENTAGES
 - 5. SIC PERCENTAGES
- B
 - 1. RECEIVED
 - 2. TYPE RECEIVED
 - 3. SOURCE RECEIVED / PRICE
 - 4. SERVICE RECEIVED / PRICE
 - 5. DEMAND RECEIVED
 - 6. ROUTE RECEIVED
 - 7. SERVICE CURTAILMENT
 - 8. ALTERNATE TYPE USED
 - 9. SOURCE RECEIVED / PRICE
 - 10. SERVICE RECEIVED / PRICE
 - 11. Refer to cross-reference chart for TRANSPORTATION NODE.
 - 12. RECEIVED PROJECTION for (B1), TYPE RECEIVED PROJECTION for PROJECTION for (B4), DEMAND RECEIVED PORJECTION for (B5), ROUTE RECEIVED PROJECTION for (B6), SERVICE CURTAILMENT PROJECTION for (B7), and ALTERNATE TYPE USED PROJECTION for (B3).
 - 13. Switching capacity in ALTERNATE TYPE
- C
 - 1. TYPE-ON-HAND
 - 2. Storage capacity in TYPE

II. Large End-Use Consumers

A. Characteristics

1. Location--defines unique facility
2. Controlling company ID
3. Demand sector (commercial, transportation, industrial, electric utility, military)
4. List of FERC priorities with associated percentages of gas usage
5. List of 2-digit SIC codes with associated percentages of total fuel usage

B. Input

1. Quantity (M) for all fuel sectors
2. Quantity (M) by ¹fuel type (e.g., natural gas, No. 2 fuel oil, bituminous coal)
3. Quantity (M) by ¹fuel type by ²source
4. Quantity (M) of natural gas by ¹source by ²type of service (firm or interruptible)
5. Quantity (M) by ¹fuel type by ²demand category (space heating and miscellaneous thermal, process steam, direct heat, or petrochemical feedstock)
6. Quantity (M) by ¹fuel type by ²source by ³route
7. Quantity (M) of gas curtailments by ¹type of service
8. Quantity (M) of alternate fuel usage with quantity (M) of primary fuel type replaced by alternate fuel type by primary fuel type replaced

12
9. Delivered price (M) by ¹fuel type by ²source
10. Delivered price (M) of natural gas by ¹source by ²type of service
11. Transportation links available
12. Short-term projections (M) for item numbers 1 thru 8
13. Existing capability (% fuel displaced) for switching to alternate fuel types for each primary fuel type

C. Storage

1. Quantity (MS) by ¹ fuel type
2. Capacity by ¹ group

LARGE END-USE CONSUMER

Definitions

- f - number of large consumers
- t - number of types
- sc - number of SIC codes
- fe - number of FERC priorities
- a - number of alternate types
- s - number of services
- rs - number of routes from source
- v - number of service types
- c - demand categories
- L - number of links/route
- h - number of historic records
- p - number of projected records

Computer Storage Formula

$$\begin{aligned}
 f \{ & | \text{LAR} | + \\
 & h | \text{LAR REC} | + p | \text{LAR REC PROJ} | + \\
 & sc | \text{LAR SIC PERCENTAGES} | + fe | \text{LAR FERC PERCENTAGES} | + \\
 & \quad \text{CURTAILMENT} \quad \text{PROJ} \\
 t [& | \text{LAR TYPE} | + \\
 & h | \text{LAR TYPE-ON-HAND} | + h | \text{LAR TYPE REC} | + \\
 & p | \text{LAR TYPE REC PROJ} | + \\
 & a (| \text{LAR ALTERNATE TYPE} | + \\
 & \quad h | \text{LAR ALTERNATE TYPE USED} | + p | \text{LAR ALTERNATE TYPE} |) + \\
 & \quad \text{USED/PROJ} \\
 c (& | \text{LAR DEMAND} | + \\
 & h | \text{LAR DEMAND REC} | + p | \text{LAR DEMAND REC PROJ} |) + \\
 s (& | \text{LAR SOURCE} | + \\
 & h | \text{LAR SOURCE REC} | + p | \text{LAR SOURCE REC PROJ} | + \\
 v < & | \text{LAR SERVICE} | + \\
 & h | \text{LAR SERVICE REC/PRICE} | + p | \text{LAR SERVICE REC PROJ} | + \\
 & h | \text{LAR SERVICE} | + p | \text{LAR SERVICE CURTAILMENT} | > + \\
 & \quad \text{CURTAILMENT} \quad \text{PROJ} \\
 rs < & | \text{LAR ROUTE FROM SOURCE} | + \\
 & L | \text{LAR SOURCE ROUTE LINKS} | + h | \text{LAR ROUTE REC} | + \\
 & p | \text{LAR ROUTE REC PROJ} | >] \}
 \end{aligned}$$

FOREIGN COUNTRY**

**Individual facility and national level only

I. FOREIGN COUNTRY (Page 102)

A 1. Foreign country ID in XFC

2. Company ID in XFC

B 1. TYPE RECEIVED

2. SOURCE RECEIVED / PRICE

3. ROUTE RECEIVED

4. SOURCE RECEIVED / PRICE

5. Refer to cross-reference chart for TRANSPORTATION NODE.

6. TYPE RECEIVED PROJECTION for (B1), SOURCE RECEIVED PROJECTION for (B2), and ROUTE RECEIVED PROJECTION for (B3).

7. SOURCE ROUTE IN-TRANSIT

C 1. TYPE SENT

2. DESTINATION SENT / PRICE

3. ROUTE SENT

4. DESTINATION SENT / PRICE

5. Refer to cross-reference chart for TRANSPORTATION NODE.

6. TYPE SENT PROJECTION for (B1), DESTINATION SENT PROJECTION for (B2), and ROUTE SENT PROJECTION for (B3).

7. DESTINATION ROUTE IN-TRANSIT

I. Foreign Country

A. Characteristics

1. Country ID	}	defines unique facility
2. Company ID		

B. Input

1. Quantity (M) by ¹ type
2. Quantity (M) by ¹ type by ² source (production facility, bulk terminal, T&D facility)
3. Quantity (M) by ¹ type by ² source by ³ route
4. Price (M) by ¹ type by ² source
5. Transportation links available
6. Short-term projections (M) for item numbers 1 thru 3
7. Quantity (MS) in transit by ¹ type by ² source by ³ route

C. Output (to USA)

1. Quantity (M) by ¹ type
2. Quantity (M) by ¹ type by ² destination (refinery, bulk terminal, T&D facility)
3. Quantity (M) by ¹ type by ² destination by ³ route
4. Price (M) by ¹ type by ² destination
5. Transportation links available
6. Short-term projection (M) for item numbers 1 thru 3.
7. Quantity (MS) in transit by ¹ type by ² destination by ³ route

FOREIGN COUNTRY

Definitions

f - number of foreign countries
 t - number of types
 s - number of sources
 d - number of destinations
 rs - number of routes from source
 rd - number of routes from destination
 L - number of links/route
 p - number of projected records
 h - number of historic records

Computer Storage Formula

$$\begin{aligned}
 & f \left\{ \left| \text{XFC} \right| + \right. \\
 & \quad t \left[\left| \text{XFC TYPE} \right| + \right. \\
 & \quad \quad h \left| \text{XFC TYPE REC} \right| + p \left| \text{XFC TYPE REC PROJ} \right| + \\
 & \quad \quad h \left| \text{XFC TYPE SENT} \right| + p \left| \text{XFC TYPE SENT PROJ} \right| + \\
 & \quad s \left(\left| \text{XFC SOURCE} \right| + \right. \\
 & \quad \quad h \left| \text{XFC SOURCE REC/PRICE} \right| + p \left| \text{XFC SOURCE REC PROJ} \right| + \\
 & \quad \quad rs \triangleleft \left| \text{XFC ROUTE FROM SOURCES} \right| + \\
 & \quad \quad \quad L \left| \text{XFC SOURCE ROUTE LINKS} \right| + h \left| \text{XFC SOURCE ROUTE-IN-} \right| + \\
 & \quad \quad \quad \text{TRANSIT} \\
 & \quad \quad \quad h \left| \text{XFC ROUTE REC} \right| + p \left| \text{XFC ROUTE REC PROJ} \right| \triangleright \left. \right) \\
 & \quad d \left(\left| \text{XFC DEST} \right| + \right. \\
 & \quad \quad h \left| \text{XFC DEST SENT/PRICE} \right| + p \left| \text{XFC DEST SENT PROJ} \right| + \\
 & \quad \quad rd \triangleleft \left| \text{XFC ROUTE TO DEST} \right| + \\
 & \quad \quad \quad L \left| \text{XFC DEST ROUTE LINKS} \right| + h \left| \text{XFC DEST ROUTE-IN-TRANSIT} \right| + \\
 & \quad \quad \quad h \left| \text{XFC ROUTE SENT} \right| + p \left| \text{XFC ROUTE SENT PROJ} \right| \triangleright \left. \right) \left. \right\}
 \end{aligned}$$

TRANSPORTATION LINKS

LINK

\$ - physical links only

- A
 - 1. Company ID in LINK
 - 2. Input facility ID in LINK
 - 3. Output facility ID in LINK
 - 4. Link ID of LINK. Refer to section 2 of the Stony Brook report for subfield breakdowns and encoding.
 - 5. Length in LINK
 - 6. Lag time in LINK
 - 7. Subfield of LINK ID
 - 8. Link ID of LINK. Refer to section 2 of the Stony Brook report for subfield breakdowns and encoding.
- B
 - 1. Capacity in LINK
 - 2. TYPE TRANSPORTED / TRANSPORTATION COST
 - 3. Cost in TYPE TRANSPORTED / TRANSPORTATION COST
- C
 - 1. TYPE IN-TRANSIT

II. TRANSPORTATION JUNCTION FACILITY (Page 104)

Although Transportation Junction Facility is technically a facility, it is conceptually a component of the transportation network and for clarity of discussion appears here.

I. Transportation Links

A. Characteristics

1. Company ID
2. Input facility ID
3. Output facility ID
4. Mode of transportation
5. Length
6. Lag time
7. Link type (megalink or detailed)
8. Energy sector or category

B. Transportation of fuel

1. Capacity
2. Quantity (M) by ¹ type
3. Transportation cost (M) by ¹ type

C. Fuel in Transit

1. Quantity (MS) by ¹ type

II. Transportation Junction Facility

A. Characteristics

1. Location
2. Company ID

B. Input

1. Transportation links available

C. Output

1. Transportation links available

F. List of subsidiary company IDs and % ownership

G. List of owning company IDs and % ownership

TRANSPORTATION JUNCTION

TJN

transportation junction ID IIA2	facility name	location IIA1	company ID IIA3
------------------------------------	------------------	------------------	-----------------------

IIB and IIC are described in the Transportation Node Chart

III. TRANSPORTATION NODES

Transportation nodes provide a uniform representation of every facility in the database for network analysis purposes. This record is used to store all transportation links entering or leaving a facility in LINKS IN and LINKS OUT. It also stores total flow information for the facility and flow information broken down by links. Transportation nodes do not explicitly appear in the EEMIS sector structures.

TRANSPORTATION NODES

TRANSPORTATION JUNCTION

Definition

t - number of transportation junctions

Computer Storage Formula

$$t \cdot | \text{TJN} |$$

TRANSPORTATION LINKS

Definitions

L - number of links

t - number of link types

h - number of historic fields

Computer Storage Formula

$$L \cdot \left\{ | \text{LINK} | + t \left(| \text{LINK TYPE} | + h \left| \text{LINK TYPE IN TRANSIT} \right| + h \left| \frac{\text{LINK TYPE TRANSPORTATION}}{\text{TRANSPORTATION COST}} \right| \right) \right\}$$

TRANSPORTATION NODE

Definitions

n - number of nodes
 t - number of energy types
 i - number of incoming links
 o - number of outgoing links
 it - number of incoming links for each energy type
 ot - number of outgoing links for each energy type
 h - number of historic records
 p - number of projected records

Computer Storage Formula

$$\begin{aligned}
 n \{ & | \text{ NODE } | + \\
 & i | \text{ NODE LINKS IN } | + o | \text{ NODE LINKS OUT } | + \\
 & t [| \text{ NODE TYPE } | + \\
 & \quad h | \text{ NODE TYPE-ON-HAND } | + h | \text{ NODE TYPE REC OR PRODUCED } | + \\
 & \quad h | \text{ NODE TYPE SENT } | + \\
 & \quad it (| \text{ NODE TYPE LINKS IN } | + \\
 & \quad \quad h | \text{ NODE LINKS REC } |) + \\
 & \quad ot (| \text{ NODE TYPE LINKS OUTS } | + \\
 & \quad \quad h | \text{ NODE LINK SENT } |)] \}
 \end{aligned}$$

COMPANY DESCRIPTION RECORDS

Each company mentioned in the EEMIS data base is described by the following conceptual record.

COMPANY

company ID A	company name B	contact person C1	phone # C2	office or building # C3	street address C4
city C5	state C6	zip C7			

COMPANY DOE CONTACT

contact person E1	phone # E2	office or building # E3	street address E4	city E5
state E6	zip E7			

COMPANY REGIONAL OFFICES

regional office name D1	contact person D1	phone # D2	office or building # D3	street address D4
city D5	state D6	zip D7		

COMPANY SUBSIDIARIES

F

company ID	percent owned
------------	---------------

COMPANY OWNERS

G

company ID	percent owned
------------	---------------

COMPANY DESCRIPTION RECORDS (Page 56)

- A 1. Company ID in COMPANY
- B 1. Company name in COMPANY
- C 1. Contact person in COMPANY
 - 2. Phone # in COMPANY
 - 3. Office or building # in COMPANY
 - 4. Street address in COMPANY
 - 5. City in COMPANY
 - 6. State in COMPANY
 - 7. Zip in COMPANY
- D 1. Contact person in COMPANY REGIONAL OFFICES
 - 2. Phone number in COMPANY REGIONAL OFFICES
 - 3. Office or building number in COMPANY REGIONAL OFFICES
 - 4. Street address in COMPANY REGIONAL OFFICES
 - 5. City in COMPANY REGIONAL OFFICES
 - 6. State in COMPANY REGIONAL OFFICES
 - 7. Zip in COMPANY REGIONAL OFFICES
- E 1. Contact person in COMPANY DOE CONTACT
 - 2. Phone number in COMPANY DOE CONTACT
 - 3. Office or building number in COMPANY DOE CONTACT
 - 4. Street address in COMPANY DOE CONTACT
 - 5. City in COMPANY DOE CONTACT
 - 6. State in COMPANY DOE CONTACT
 - 7. Zip in COMPANY DOE CONTACT
- F 1. COMPANY SUBSIDIARIES
- G 1. COMPANY OWNERS

Company File

A. Company ID

B. Name

C. National Office

1. Name of contact
2. Phone Number
3. Office or building number
4. Street address
5. City
6. State
7. Zip code

D. Regional Offices

1. Names of contact
2. Phone numbers
3. Office or building number
4. Street address
5. City
6. State
7. Zip Code

List of names and addresses
to contact for each region.
First listed is primary contact.

E. DOE Contacts

1. Name of contact
2. Phone number
3. Office of building number
4. Street address
5. City
6. State
7. Zip code

List of names and addresses
to contact for each region.
First listed is primary contact.

COMPANY DESCRIPTION RECORDS

Definitions

- c - number of companies
- r - number of regional offices
- s - number of company subsidiaries
- o - number of company owners

Computer Storage Formula

$$C \quad \left\{ | \text{COMPANY} | + | \text{COMPANY DOE CONTRACT} | \right. \\ \quad + r \quad | \text{COMPANY REGIONAL OFFICES} | + s \quad | \text{COMPANY SUBSIDIARIES} | \\ \quad \left. + o \quad | \text{COMPANY OWNERS} | \right\}$$

RES Functional Hierarchy

The purpose of this hierarchy is to associate facilities and links with the different RES functions.

"XXX" "XXX"
[1] and [2]
expand as in energy hierarchy

Demand Sector Hierarchy

The purpose of this hierarchy is to associate facilities and links with different demand sectors.

RES FUNCTION

Definitions

r - number of RES functions
fr - number of facility types for each RES function
Lr - number of link types for each RES function
f - number of facilities of each type
L - number of links of each type

Computer Storage Formula

$$r \left\{ \begin{aligned} &| \text{RES FUNCTION} | \\ &+ fr \left[| \text{RES FACILITY TYPES} | + f | \text{RES FACILITIES} | \right] \\ &+ Lr \left[| \text{RES LINK TYPES} | + L | \text{RES LINKS} | \right] \end{aligned} \right\}$$

DEMAND SECTOR

Definitions

d - number of demand sector
 s - number of subsectors/demand sector
 fd - number of facility types per demand sector
 fs - number of facility types per demand subsector
 Ld - number of link types per demand sector
 Ls - number of link types per demand subsector
 f - number of facilities of each type
 L - number of links of each type

Computer Storage Formula

$$\begin{aligned}
 & d \left\{ \left| \text{DEMAND SECTOR} \right| + \right. \\
 & \quad fd \left(\left| \text{DEMAND SECTOR} \right| + f \left| \text{DEMAND SECTOR} \right| \right) + \\
 & \quad \quad \text{FACILITY TYPES} \quad \quad \quad \text{FACILITIES} \\
 & \quad Ld \left(\left| \text{DEMAND SECTOR} \right| + L \left| \text{DEMAND SECTOR LINKS} \right| \right) + \\
 & \quad \quad \quad \text{LINK TYPES} \\
 & \quad s \left[\left| \text{DEMAND SUBSECTOR} \right| + \right. \\
 & \quad \quad fs \left(\left| \text{DEMAND SUBSECTOR} \right| + f \left| \text{DEMAND SUBSECTOR} \right| \right) + \\
 & \quad \quad \quad \text{FACILITY TYPES} \quad \quad \quad \text{FACILITIES} \\
 & \quad \quad Ls \left(\left| \text{DEMAND SUBSECTOR} \right| + L \left| \text{DEMAND SUBSECTOR} \right| \right) \left. \right\} \\
 & \quad \quad \quad \text{LINK TYPES} \quad \quad \quad \text{LINKS}
 \end{aligned}$$

Conceptual Record for Energy Forms

One purpose of this record is to express the relationship among sectors, categories, groups and individual types in the energy hierarchy. It is also used to store lists of facilities and links that deal with a given type, group, etc.

ENERGY FORMS

"XXX"

1 expands to

"XXX" FACILITY TYPES

"XXX" FACILITIES

"XXX"

2 expands to

"XXX" LINK TYPES

"XXX" LINKS

ENERGY FORMS

Definitions

u - the number of universal forms (=1)
s - the number of energy sectors
ft - the number of facilities of each energy type
Lu - the number of link types associated with universal form
Ls - the number of link types associated with energy sectors
Lc - the number of link types associated with energy categories
Lg - the number of link types associated with energy groups
Lt - the number of link types associated with energy types
fu - the number of facilities associated with the universal form
fs - the number of facilities associated with energy sectors
fc - the number of facilities associated with energy categories
fg - the number of facilities associated with energy groups
ft - the number of facilities associated with energy types
L - the number of links of each link type

Computer Storage Formula

$$\begin{aligned}
 u \cdot & \left\{ \left| \text{ENERGY FORMS} \right| + \right. \\
 & f_u \left(\left| \text{UNIVERSAL FACILITY TYPES} \right| + f \left| \text{UNIVERSAL FACILITIES} \right| \right) + \\
 & L_u \left(\left| \text{UNIVERSAL LINK TYPES} \right| + L \left| \text{UNIVERSAL LINKS} \right| \right) + \\
 & s \left[\left| \text{ENERGY SECTOR} \right| + \right. \\
 & \quad f_s \left(\left| \text{SECTOR FACILITY TYPES} \right| + f \left| \text{SECTOR FACILITIES} \right| \right) + \\
 & \quad L_s \left(\left| \text{SECTOR LINK TYPES} \right| + L \left| \text{SECTOR LINKS} \right| \right) + \\
 & \quad c \left(\left| \text{ENERGY CATEGORY} \right| + \right. \\
 & \quad \quad f_c \left(\left| \text{CATEGORY FACILITY TYPES} \right| + f \left| \text{CATEGORY FACILITIES} \right| \right) + \\
 & \quad \quad L_c \left(\left| \text{CATEGORY LINK TYPES} \right| + L \left| \text{CATEGORY LINKS} \right| \right) + \\
 & \quad g \left\{ \left| \text{ENERGY GROUP} \right| + \right. \\
 & \quad \quad f_g \left(\left| \text{GROUP FAC TYPES} \right| + f \left| \text{GROUP FACILITIES} \right| \right) + \\
 & \quad \quad L_g \left(\left| \text{GROUP LINK TYPE} \right| + L \left| \text{GROUP LINKS} \right| \right) + \\
 & \quad \quad t \left\{ \left| \text{ENERGY TYPE} \right| + \right. \\
 & \quad \quad \quad f_t \left(\left| \text{TYPE FAC TYPES} \right| + f \left| \text{TYPE FACILITIES} \right| \right) + \\
 & \quad \quad \quad L_t \left(\left| \text{TYPE LINK TYPES} \right| + L \left| \text{TYPE LINKS} \right| \right) \left. \right\} \left. \right\}
 \end{aligned}$$

CONCEPTUAL RECORD GEOGRAPHIC HIERARCHY
GEOGRAPHIC HIERARCHY Domestic or Foreign Countries File (Page 36)

The Crude Oil Production Region File, Natural Gas Production Region File, Foreign Country File, State File, and Region File have been combined in the conceptual model design into the geographic hierarchy. In doing this, the design better reflects the relationship between the various geographic regions represented in the EEMIS system, and integrates in one conceptual record all geographic information in the database.

GEOGRAPHIC HIERARCHY (Page 62)

- A 1. The IDs, names, latitudes and longitudes, etc., for the Domestic Crude Oil and Natural Gas Production Region Files are specified in the segments from NON-POLITICAL AREA down.
- B 1. Foreign Country ID and Name are specified in GEOGRAPHIC HIERARCHY
- C 1. State ID and Name are specified in STATE
2. State latitude and longitude pairs are in STATE BOUNDARY
- D 1. 6 in the state file are in the segment GOVERNOR'S ENERGY ADVISORY
- E 1. 6 in the state file are in the segment STATE ENERGY OFFICE

The entries under region file are encoded in the data element list as the segments REGION, REGION BOUNDARY, and REGIONAL ENERGY COMMISSIONS.

Domestic Crude Oil Production Region File

- A. Domestic crude oil production region ID
- B. Name
- C. List of county IDs
- D. List of latitude, longitude pairs
- E. Ordered list of company IDs

Natural Gas Production Region File

- A. Natural gas production region ID
- B. Name
- C. List of county IDs
- D. List of latitude, longitude pairs
- E. Ordered list of company IDs

Foreign Country File

- A. Foreign country ID
- B. Name

Region File

- A. Region ID
- B. Name
- C. List of state IDs
- D. List of latitude, longitude pairs
- E. List of regional energy commissions
 - 1. Name
 - 2. Name of director
 - 3. Phone number
 - 4. Office number
 - 5. Street address
 - 6. City
 - 7. State
 - 8. Zip code

State File

- A. State ID
- B. Name
- C. List of latitude, longitude pairs
- D. Governor's office
 - 1. Name of Energy Advisor to Governor
 - 2. Phone number
 - 3. Office or building number
 - 4. Street address
 - 5. City
 - 6. Zip code
- E. Energy Office
 - 1. Name of department head
 - 2. Phone number
 - 3. Office of building number
 - 4. Street address
 - 5. City
 - 6. Zip code
- F. EEC Contact
 - 1. Name
 - 2. Phone number
 - 3. Office or building number
 - 4. Street address
 - 5. City
 - 6. Zip code

GEOGRAPHIC HIERARCHY

Definitions

- x - number of foreign countries
- r - number of regions
- s - number of states/region
- c - number of counties/state
- p - number of non-political areas
- fu - number of facility types in U.S.
- fr - number of facility types in each region
- fs - number of facility types in each state
- fc - number of facility types in each county
- fp - number of facility types in each non-political area
- Lu - number of link types in U.S.
- Lr - number of link types in each region
- Ls - number of link types in each state
- Lc - number of link types in each county
- Lp - number of link types in each non-political area
- f - number of facilities of each type
- L - number of links of each type
- b - number of points in a boundary
- e - number of energy commissions in each region
- i - number of geographic areas intersecting each
non-political area

Computer Storage Formula

$$\begin{aligned}
 X & \{ \{ \text{GEOGRAPHIC HIERARCHY} \mid + \\
 & \quad fu \left(\mid \text{COUNTRY FACILITY TYPES} \mid + f \mid \text{COUNTRY FACILITIES} \mid \right) + \\
 & \quad Lu \left(\mid \text{COUNTRY LINK TYPES} \mid + L \mid \text{COUNTRY LINKS} \mid \right) + \\
 & \quad p \mid \text{NON POLITICAL AREA} \mid + \\
 & \quad \quad b \mid \text{NON POLITICAL AREA BOUNDARY} \mid + i \mid \text{NON POLITICAL INTER-SORTING AREAS} \mid \\
 & \quad \quad fp \left(\mid \text{NON POLITICAL FACILITY TYPES} \mid + f \mid \text{NON POLITICAL FACILITIES} \mid \right) \\
 & \quad \quad Lp \left(\mid \text{NON POLITICAL LINK TYPES} \mid + L \mid \text{NON POLITICAL LINKS} \mid \right) \\
 & \quad r \left[\mid \text{REGION} \mid + \right. \\
 & \quad \quad b \mid \text{REGION BOUNDARY} \mid + e \mid \text{REGIONAL ENERGY COMMISSION} \mid + \\
 & \quad \quad fr \left(\mid \text{REGION FACILITY TYPE} \mid + f \mid \text{REGION FACILITIES} \mid \right) \\
 & \quad \quad Lr \left(\mid \text{REGION LINK TYPES} \mid + L \mid \text{REGION LINKS} \mid \right) + \\
 & \quad \quad s \left(\mid \text{STATE} \mid + \mid \text{STATE ENERGY OFFICE} \mid + \mid \text{GOVERNORS ENERGY ADVISOR} \mid + \right. \\
 & \quad \quad + b \mid \text{STATE BOUNDARY} \mid + \\
 & \quad \quad fs \left(\mid \text{STATE FACILITY TYPES} \mid + f \mid \text{STATE FACILITIES} \mid \right) + \\
 & \quad \quad Ls \left(\mid \text{STATE LINK TYPES} \mid + L \mid \text{STATE LINKS} \mid \right) + \\
 & \quad \quad c \{ \mid \text{COUNTY} \mid + \\
 & \quad \quad \quad b \mid \text{COUNTY BOUNDARY} \mid + \\
 & \quad \quad \quad fc \left(\mid \text{COUNTY FACILITY TYPES} \mid + f \mid \text{COUNTY FACILITIES} \mid \right) \\
 & \quad \quad \quad Lc \left(\mid \text{COUNTY LINK TYPES} \mid + L \mid \text{COUNTY LINKS} \mid \right) \} \} \}
 \end{aligned}$$

V. CASE STUDIES - STORAGE REQUIREMENTS

To provide a further consideration of the database size requirements, we present four case studies with actual size estimates drawn from field length estimates provided. Each study will consist of a set of variable definitions, size estimate formula with values given above the field or segment in question, and a value for the final size, which will be broken down into actual data size, implementation data size, and percentage of overhead. Where more than one number appears in parenthesis in the size estimates (for example, 62 + 28), the first value represents overhead in redundant key storage, etc., and the second number represents actual data.

Company Description Records

<u>Definitions</u>	<u>Example Parameters</u>
	<u>Values</u>
c - number of companies	300
r - number of regional offices	4
s - number of company subsidiaries	2
o - number of company owners	1

Computer Storage Formula

$$\begin{aligned}
 &300 \quad (0 + 173) \quad (16 + 127) \\
 &C \quad \{ |COMPANY| + |COMPANY DOE CONTRACT| \\
 &\quad \quad (16 + 167) \quad 2 \quad (16 + 18) \\
 &\quad + r \quad |COMPANY REGIONAL OFFICES| + s \quad |COMPANY SUBSIDIARIES| \\
 &\quad \quad 1 \quad (16 + 18) \\
 &\quad + o \quad |COMPANY OWNERS| \}
 \end{aligned}$$

	<u>Bytes</u>	<u>Words</u>
Total Storage Requirements	$3.33 \cdot 10^5$	$8.325 \cdot 10^4$
Actual Data Requirements	$2.946 \cdot 10^5$	$7.365 \cdot 10^4$
Overhead Requirements	$3.84 \cdot 10^4$	$9.6 \cdot 10^3$
Percentage of Overhead	13.0	

Transmission and Distribution (T&D) Facility

<u>Definitions</u>	<u>Example Parameter Values</u>
f - number of T&D facilities	1000
st - number of states served	4
t - number of storage types	2
sf - number of selling facilities	1
s - number of sources	2
d - number of destinations	10
rs - number of routes to source	1
rd - number of routes to destination	1
L - number of links in a route	1
n - number of service types	2
a - number of alternate energy types	2
h - number of historical records	60
p - number of projection records	12

Storage Requirements

1000 (0 + 178) 4 (16 + 8) 60 (16 + 36) 12 (16 + 36)
 f . { | NTD | + st | NTD STATES SERVED | + h | NTD REC | + p | NTD REC PROJ |

 60 (16 + 36) 12 (16 + 36) 60 (16 + 36)
 + h | NTD SENT | + p | NTD SENT PROJ | + h | NTD END USER SENT |

 12 (16 + 36)
 + p | NTD END USER SENT PROJ |

 2 (16 + 1) 60 (17 + 28)
 + t [| NTD STORAGE TYPE | + h | NTD TYPE-IN-STORAGE |]

 1 (16 + 16) 60 (32 + 42)
 + sf [| NTD SELLING FACILITY | + h | NTD SELLING FAC/REC/PRICE |]

 12 (32 + 36)
 + p | NTD SELLING FAC REC PROJ |] .

 2 (16 + 4) 60 (20 + 42)
 + n [| NTD SERVICE TYPE | + h | NTD SERVICE SENT/PRICE |

 12 (20 + 36) 60 (20 + 36)
 + p | NTD SERVICE SENT PROJ | + h | NTD SERVICE CURTAILMENT |

 12 (20 + 36)
 + p | NTD SERVICE CURTAILMENT PROJ |]

 2 (16 + 1) 60 (17 + 36)
 + a [| NTD ALTERNATE TYPE | + h | NTD ALTERNATE TYPE USED |]

 2 (16 + 1) 60 (32 + 42) 12 (32 + 36)
 + s [| NTD SOURCE | + h | NTD SOURCE REC/PRICE | + p | NTD SOURCE REC |
 PROJ

$$+ \quad \begin{matrix} 1 & (32 + 31) & 60 & (62 + 36) \\ \text{rs} & \left(\left| \text{NTD ROUTE FROM SOURCE} \right| + \left| \text{NTD ROUTE REC} \right| \right) \end{matrix}$$

$$+ \quad \begin{matrix} 12 & (62 + 36) \\ \text{p} & \left| \text{NTD ROUTE REC PROJ} \right| \end{matrix}$$

$$+ \quad \begin{matrix} 1 & (62 + 27) \\ \text{L} & \left| \text{NTD SOURCE ROUTE LINKS} \right| \end{matrix} \left. \vphantom{\begin{matrix} 1 & (62 + 27) \\ \text{L} & \left| \text{NTD SOURCE ROUTE LINKS} \right| \end{matrix}} \right]$$

$$\begin{matrix} 10 & (16 + 16) & 60 & (32 + 42) \\ \text{d} & \left[\left| \text{NTD DEST} \right| + \left| \text{NTD DEST SENT/PRICE} \right| \right] \end{matrix}$$

$$+ \quad \begin{matrix} 12 & (32 + 36) & 60 & (32 + 36) \\ \text{p} & \left| \text{NTD DEST SENT PROJ} \right| + \left| \text{NTD DEST CURTAILMENT} \right| \end{matrix}$$

$$+ \quad \begin{matrix} 12 & (32 + 36) \\ \text{p} & \left| \text{NTD DEST CURTAILMENT PROJ} \right| \end{matrix}$$

$$+ \quad \begin{matrix} 1 & (32 + 31) & 60 & (62 + 36) \\ \text{rs} & \left(\left| \text{NTD ROUTE TO DEST} \right| + \left| \text{NTD ROUTE SENT} \right| \right) \end{matrix}$$

$$+ \quad \begin{matrix} 12 & (62 + 36) & 1 & (62 + 27) \\ \text{p} & \left| \text{NTD ROUTE SENT PROJ} \right| + \left| \text{NTD DEST ROUTE LINKS} \right| \end{matrix} \left. \vphantom{\begin{matrix} 12 & (62 + 36) & 1 & (62 + 27) \\ \text{p} & \left| \text{NTD ROUTE SENT PROJ} \right| + \left| \text{NTD DEST ROUTE LINKS} \right| \end{matrix}} \right\}$$

NTW

	<u>Bytes</u>	<u>Words</u>
Total Storage Requirments	$2.5 \cdot 10^8$	$6.24 \cdot 10^7$
Actual Data Requirements	$1.29 \cdot 10^8$	$3.02 \cdot 10^7$
Overhead Requirements	$1.21 \cdot 10^8$	$3.28 \cdot 10^7$

Percentage of Overhead 93.4

Crude Oil Bulk Terminal (PBT)

<u>Definitions</u>	Example Parameter <u>Values</u>
f - number of facilities	300
t - number of energy types	3
s - number of sources	3
d - number of destinations	3
rs - number of routes from source	1
rd - number of routes from destination	1
L - number of links in a route	2
h - number of historical records	60
p - number of projection records	12

For purposes of illustration, the formula for size has been expanded by fields, to show the size contribution of each field to the formula.

Computer Storage Formula

$$\begin{aligned}
 & 300 \quad (0 + 176) \\
 & f . \left\{ \left| \text{PBT} \right| \right. \\
 & + \quad 3 \quad \left[\left| \text{PBT TYPE} \right| + h . \left(\left| \text{PBT TYPE ON HAND} \right| + h \left| \text{PBT TYPE REC} \right| \right) \right. \\
 & + \quad 12 \quad \left(17 + 36 \right) \quad 60 \quad \left(17 + 36 \right) \quad 12 \quad \left(17 + 36 \right) \\
 & + \quad p . \left| \text{PBT TYPE REC PROJ} \right| + h . \left| \text{PBT TYPE SENT} \right| + p . \left| \text{PBT TYPE SENT PROJ} \right| \\
 & + \quad 3 \quad \left(17 + 16 \right) \quad 12 \quad \left(33 + 36 \right) \quad 60 \quad \left(33 + 42 \right) \\
 & + \quad s . \left(\left| \text{PBT SOURCE} \right| + p . \left| \text{PBT SOURCE REC PROJ} \right| + h . \left| \text{PBT SOURCE REC/PRICE} \right| \right. \\
 & + \quad 1 \quad \left(33 + 42 \right) \quad 12 \quad \left(63 + 31 \right) \\
 & + \quad rs . \left\langle \left| \text{PBT ROUTE FROM SOURCE} \right| + p . \left| \text{PBT ROUTE REC} \right| \right. \\
 & + \quad 2 \quad \left(63 + 27 \right) \\
 & + \quad L . \left| \text{PBT SOUCE ROUTE LINKS} \right| \left. \right\rangle \\
 & + \quad 3 \quad \left(17 + 16 \right) \quad 12 \quad \left(33 + 36 \right) \quad 60 \quad \left(33 + 42 \right) \\
 & + \quad d . \left(\left| \text{PBT DEST} \right| + p . \left| \text{PBT DEST SENT PROJ} \right| + h . \left| \text{PBT DEST SENT/PRICE} \right| \right. \\
 & + \quad 1 \quad \left(33 + 42 \right) \quad \left(63 + 31 \right) \\
 & + \quad rd . \left\langle \left| \text{PBT ROUTE TO DEST} \right| + p . \left| \text{PBT ROUTE SENT} \right| \right. \\
 & + \quad 2 \quad \left(63 + 27 \right) \\
 & + \quad L . \left| \text{PBT DEST ROUTE LINKS} \right| \left. \right\rangle \left. \right\}
 \end{aligned}$$

EXPANSION OF SIZE FORMULA BY FIELDS FOR CRUDE BULK TERMINAL

$$\begin{aligned}
 f \left\{ \right. & \begin{array}{l}
 \begin{array}{ccc}
 16 & 30 & 30 \\
 \text{CRUDE BULK TERMINAL ID} & + & \text{PBT FACILITY NAME} & + & \text{PBT STREET ADDRESS}
 \end{array} \\
 + & \begin{array}{ccccc}
 20 & 2 & 5 & 10 & 30 \\
 \text{PBT CITY} & + & \text{PBT STATE} & + & \text{PBT ZIP} & + & \text{PBT PHONE \#} & + & \text{PBT CONTACT} \\
 & & & & & & & & \text{PERSON}
 \end{array} \\
 + & \begin{array}{cccc}
 8 & 9 & 16 & 9 \\
 \text{PBT LATITUDE} & + & \text{PBT LONGITUDE} & + & \text{PBT COMPANY ID} & + & \text{PBT STORAGE} \\
 & & & & & & \text{CAPACITY}
 \end{array} \\
 + & \begin{array}{ccc}
 3 & 16 + 1 = 17 & 16 + 1 + 7 = 24 \\
 \text{PBT SC UNITS} & + & t \left[\begin{array}{l} \text{PBT BTID ENERGY CODE} \end{array} \right] + p \left(\begin{array}{l} \text{PBT BTID ECOD ONDATE} \\ \text{TOH} \end{array} \right)
 \end{array} \\
 + & \begin{array}{ccc}
 9 & 9 & 3 \\
 \text{PBT REP QUANT TOH} & + & \text{PBT COMP QUANT TOH} & + & \text{PBT UNITS TOH}
 \end{array} \\
 + & \begin{array}{ccc}
 16 + 1 + 14 = 31 & 9 & 9 \\
 \text{PBT BTID ECOD FTDATE TR} & + & \text{PBT REP QUANT TR} & + & \text{PBT COMP QUANT TR}
 \end{array} \\
 + & \begin{array}{ccc}
 3 & 1 & 16 + 1 + 14 = 31 \\
 \text{PBT UNITS TR} & + & \text{PBT AGGREGATE TR} & + & \text{PBT BTID ECOD FTDATE TRP}
 \end{array} \\
 + & \begin{array}{ccc}
 9 & 9 & 3 \\
 \text{PBT REP QUANT TRP} & + & \text{PBT COMP QUANT TRP} & + & \text{PBT UNITS TRP}
 \end{array} \\
 + & \begin{array}{ccc}
 1 & 16 + 1 + 14 = 31 & 9 \\
 \text{PBT AGGREGATE TRP} & + & \text{PBT BTID ECOD FTDATE TS} & + & \text{PBT REP QUANT TS}
 \end{array} \\
 + & \begin{array}{ccc}
 9 & 3 & 1 \\
 \text{PBT COMP QUANT TS} & + & \text{PBT UNITS TS} & + & \text{PBT AGGREGATE TS}
 \end{array} \\
 + & \begin{array}{cc}
 16 + 1 + 14 + 31 & 9 \\
 \text{PBT BTID ECOD FTDATE TSP} & + & \text{PBT REP QUANT TSP}
 \end{array}
 \end{array}
 \right.
 \end{aligned}$$

$$\begin{array}{r}
 \begin{array}{ccc}
 9 & 3 & 1 \\
 + \mid \text{PBT COMP QUANT TSP} \mid + \mid \text{PBT UNITS TSP} \mid + \mid \text{PBT AGGREGATE TSP} \mid \mid \\
 \\
 16 + 1 + 16 = 33 & 16 + 1 + 16 + 14 = 47 & 9 \\
 + s \left(\mid \text{PBT BTID ECOD SOURCE} \mid + p \left\{ \mid \text{PBT BTID ECOD FID} \mid + \mid \text{PBT REP QUANT SRPR} \mid \right. \right. \\
 \quad \text{FACILITY ID} & \quad \text{FTDATE SRPR} & \\
 \\
 \begin{array}{ccc}
 9 & 3 & 1 \\
 + \mid \text{PBT COMP QUANT SRPR} \mid + \mid \text{PBT UNITS SRPR} \mid + \mid \text{PBT AGGREGATE SRPR} \mid \\
 \\
 3 & 16 + 1 + 16 + 14 = 47 & 9 \\
 + \mid \text{PBT PRICE SRPR} \mid + \mid \text{PBT BTID ECOD FID FTDATE SRP} \mid + \mid \text{PBT REP QUANT SRP} \mid \mid \\
 \\
 \begin{array}{ccc}
 9 & 3 & 1 \\
 + \mid \text{PBT COMP QUANT SRP} \mid + \mid \text{PBT UNITS SRP} \mid + \mid \text{PBT AGGREGATE SRP} \mid \\
 \\
 16 + 1 + 16 + 30 = 63 & 2 \\
 + rs \left\{ \mid \text{PBT BTID ECOD FID SOURCE RTE NAME} \mid + \mid \text{PBT SOURCE NUMBER OF LINKS} \mid \right. \\
 \\
 16 + 1 + 16 + 30 + 14 = 77 & 9 \\
 + p \triangleleft \mid \text{PBT BTID ECOD FID SRN FTDATE RS} \mid + \mid \text{PBT REP QUANT RR} \mid \\
 \\
 \begin{array}{ccc}
 9 & 3 & 1 \\
 + \mid \text{PBT COMP QUANT RR} \mid + \mid \text{PBT UNITS RR} \mid + \mid \text{PBT AGGREGATE RR} \mid \triangleright \\
 \\
 16 + 1 + 16 + 30 + 2 = 65 & 26 \\
 + \triangleleft \mid \text{PBT BTID ECOD FID SRN SRC LINK} \mid + \mid \text{PBT SOURCE LINK ID} \mid \triangleright \mid \mid \\
 \\
 16 + 1 + 16 = 33 & 16 + 1 + 16 + 14 = 47 \\
 + d \left(\mid \text{PBT BTID ECOD DEST FACILITY ID} \mid + p \left\{ \mid \text{PBT BTID ECOD FID FTDATE DSPR} \mid \right. \right. \\
 \\
 \begin{array}{ccc}
 9 & 9 & 3 \\
 + \mid \text{PBT REP QUANT DSPR} \mid + \mid \text{PBT COMP QUANT DSPR} \mid + \mid \text{PBT UNITS DSPR} \mid \\
 \\
 1 & 3 & 16 + 1 + 16 + 14 = 47 \\
 + \mid \text{PBT AGGREGATE DSPR} \mid + \mid \text{PBT PRICE DSPR} \mid + \mid \text{PBT BTID ECOD FID FTDATE DSP} \mid
 \end{array}
 \end{array}$$

Petroleum Refinery (PRE)

<u>Definitions</u>	<u>Example Parameter Values</u>
f - number of refineries	300
o - number of other capabilities	2
t ₁ - number of input type	2
t ₂ - number of product type	3
s - number of sources	2
d - number of destinations	5
rs - number of routes from source	2
rd - number of routes to destinations	2
L - number of links in a route	2
h - number of historical records	17
p - number of projection records	12

300 (0 + 244) 2 (16 _ 21)

$$\begin{aligned}
 & f \left\{ \left| \text{PRE} \right| + o \left| \text{PRE OTHER CAPABILITIES} \right| \right. \\
 & \quad 2 \quad (16 + 22) \quad 60 \quad (17 + 36) \quad 12 \quad (17 + 36) \\
 & + t_1 \left[\left| \text{PRE TYPE} \right| + h \left| \text{PRE TYPE RECEIVED} \right| + p \left| \text{PRE TYPE REC PROJ} \right| \right. \\
 & \quad 60 \quad (17 + 36) \quad 12 \quad (17 + 36) \\
 & \quad + h \left| \text{PRE TYPE SENT} \right| + p \left| \text{PRE TYPE SENT PROJ} \right| \\
 & \quad 2 \quad (17 + 36) \quad 60 \quad (33 + 42) \quad 12 \quad (33 + 36) \\
 & + s \left(\left| \text{PRE SOURCE} \right| + h \left| \text{PRE SOURCE RECEIVED/PRICE} \right| + p \left| \text{PRE SOURCE REC PROJ} \right| \right. \\
 & \quad 2 \quad (33 + 36) \quad 60 \quad (63 + 36) \\
 & + rs \left\langle \left| \text{PRE ROUTE FROM SOURCE} \right| + h \left| \text{PRE ROUTE REC} \right| \right. \\
 & \quad 2 \quad (63 + 36) \\
 & \quad + L \left| \text{PRE SOURCE ROUTE LINKS} \right| \left. \right\rangle \\
 & \quad 5 \quad (17 + 16) \quad 60 \quad (33 + 42) \quad 12 \quad (33 + 36) \\
 & + d \left(\left| \text{PRE DEST} \right| + h \left| \text{PRE DEST SENT/PRICE} \right| + p \left| \text{PRE DEST SENT PROJ} \right| \right. \\
 & \quad 2 \quad (33 + 36) \quad 60 \quad (63 + 36) \\
 & + rd \left\langle \left| \text{PRE ROUTE TO DEST} \right| + h \left| \text{PRE ROUTE SENT} \right| \right. \\
 & \quad 2 \quad (63 + 36) \\
 & \quad + L \left| \text{PRE DEST ROUTE LINKS} \right| \left. \right\rangle \left. \right] \\
 & \quad 3 \quad (16 + 22) \quad 2 \quad (17 + 13) \\
 & + t_2 \left[\left| \text{PRE TYPE} \right| + t_2^{-1} \left| \text{PRE REMAINING PRODUCTIVITY} \right| \right] \left. \right\}
 \end{aligned}$$

PRE

	<u>Bytes</u>	<u>Words</u>
Total Storage Requirements	$7.95 \cdot 10^7$	$2.0 \cdot 10^7$
Actual Data Requirements	$4.47 \cdot 10^7$	$1.12 \cdot 10^7$
Overhead Requirements	$3.48 \cdot 10^7$	$8.7 \cdot 10^6$

Percentage of Overhead 128.4

SUMMARY OF STORAGE CONSIDERATIONS

It is apparent, from just the case studies, that vast amounts of storage will be required for the EEMIS database. One "conservative" estimate places the amount of storage at about 1 billion words. This massive storage requirement will require careful formulation and analysis of existing storage algorithms and devices before the data base is implemented. A short table of some common storage devices, together with the number required to store the above estimate is given below.

<u>Device</u>	<u>Capacity</u>	<u>No. Required</u>
3330-I	100 million bytes	10
3330-II	200 million bytes	5
3350	317.5 million bytes	4
3850 MSS	50-500 billion bytes	1

(Mass Storage System)

Clearly we cannot expect a system to have ten 3330-Is around just for EEMIS (although a 3850, if available, would be eminently suitable). Further consideration must be given to the ultimate storage configuration for the database to ensure proper operation of the entire DBMS.

REFERENCES

1. Brown, G.D., System 370 Job Control Language, Wiley, New York, 1977.
2. Croteau, K. and Kydes, A. S., EEMIS Sector Correspondence to Conceptual Model Design; BNL Informal Report, June 1979.
3. D'Acerno, J. and Hermelee, A., Compatability of EIA Databases With EEMIS Data Requirements, BNL 51109, Dec. 1979.
4. D'Acerno, J., and Hermelee, A., Physical Aspects of the U.S. Oil and Gas Systems, BNL 51076, Nov. 1979.
5. Date, C. J., An Introduction to Database Systems, 2nd ed., Addison-Wesley, Reading, MA, 1975.
6. Croteau, K., Heller, J., Krawiec, T., Maier, D., Russell, B., and Wolk, P., Final Report: The Conceptual Model for EEMIS, Dept. of Computer Sci., SUNY, Stony Brook, June 1979.
7. Martin, J., Computer Data-Base Organization, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ 1977.
8. Technology Assessment Group, Phase I Report: Alternative Data Structures Report for EEMIS Program Office, Energy Information Administration, BNL 26206 (Limited Distribution), April 1979.
9. Shipman, D.W., DAPLEX a functional database model, Presented at 1979 ACM SIGMOD Conference on Management of Data, Boston, May-June 1979; to appear in ACM Transactions on Database Systems.

U.S. GOVERNMENT PRINTING OFFICE 614-090-#31