Reference

REPORT NO. DOT-TSC-OST-76-60

# TRANSPORTATION STATISTICAL DATA AND INFORMATION

Robert Tap
Alan Kaprelian

U.S. DEPARTMENT OF TRANSPORTATION
TRANSPORTATION SYSTEMS CENTER
KENDALL SQUARE
CAMBRIDGE MA 02142


JUNE 1977
REPRINT
FINAL REPORT



Prepared for

U.S. DEPARTMENT OF TRANSPORTATION
OFFICE OF THE SECRETARY
Office of the Assistant Secretary for Policy, Plans and International Affairs
Office of Transportation Systems Analysis and Information
Washington DC 20590

**Technical Report Documentation Page**

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| DOT-TSC-OST-76-60 | | |

| 4. Title and Subtitle | 5. Report Date |
|---|---|
| TRANSPORTATION STATISTICAL DATA AND INFORMATION | June 1977 Reprint |
| | 6. Performing Organization Code |
| 7. Author(s) | 8. Performing Organization Report No. |
| Robert Tap and Alan Kaprelian | DOT-TSC-OST-76-60 |
| 9. Performing Organization Name and Address<br>U.S. Department of Transportation<br>Transportation Systems Center<br>Kendall Square<br>Cambridge MA 02142 | 10. Work Unit No. (TRAIS)<br>OP735/R7824 |
| | 11. Contract or Grant No. |
| 12. Sponsoring Agency Name and Address<br>U.S. Department of Transportation<br>Office of the Secretary<br>Office of the Assistant Secretary for Policy, Plans and International Affairs<br>Office of Transportation Systems Analysis and Information<br>Washington DC 20590 | 13. Type of Report and Period Covered<br>Final Report<br>July 1975-December 1976 |
| | 14. Sponsoring Agency Code |

15. Supplementary Notes

16. Abstract

The document contains an extensive review of internal and external sources of transportation data and statistics especially created for data administrators. Organized around the transportation industry and around the elements of the U.S. Department of Transportation, it is the most comprehensive single document that reviews transportation data and its history.

| 17. Key Words | 18. Distribution Statement |
|---|---|
| Transportation Data,<br>Transportation Statistics,<br>Data Administration | DOCUMENT IS AVAILABLE TO THE U.S. PUBLIC THROUGH THE NATIONAL TECHNICAL INFORMATION SERVICE, SPRINGFIELD, VIRGINIA 22161 |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 226 | |

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

# PREFACE

A study of current transportation data sources and program used within the Office of the Secretary (OST) for the purpose of identifying and organizating these transportation statistics was performed by the Information Division of the Transportation Systems Center under Project Plan Agreement OP-631*. Sponsorship, overall direction, and guidance, for the study was provided by the Office of Systems Analysis and Information of the Office of the Assistant Secretary for Policy, Plans and International Affairs. Results and conclusions of the study serve to define the scope of transportation statistics available to Departmental analysts.

---

*OP-631: Bibliographic Data File Development.

# TABLE OF CONTENTS

# TABLE OF CONTENTS (CONT'D)

# TABLE OF CONTENTS (CONT'D)

# TABLE OF CONTENTS (CONT'D)

# LIST OF ILLUSTRATIONS

# LIST OF ILLUSTRATIONS (CONT'D)

# 1. CONCLUSIONS AND RECOMMENDATIONS

## 1.1 GENERAL BACKGROUND

The objective of this study, sponsored by the Assistant Secretary for Policy, Plans and International Affairs and performed by the Information Division of the Transportation Systems Center, was to review the current capability within the Office of the Secretary to identify, acquire, and utilize transportation data available for use in analyses of policy issues in the transportation industry. The study showed a highly distributed source of knowledge about what data are available to the Department and exist within the Office of the Secretary. This document summarizes the situation and outlines an approach to developing and maintaining a viable directory of knowledge about such data availability.

This document addresses one part of the problem commonly referred to as the "transportation data problem": "What data is available?" It does not address the issues associated with "Is the Office of the Secretary collecting the right data?", because this question can be best answered only if one knows what data are available to the Department. The document proposes a structured way to scope out the vast array of transportation-oriented-data-collection efforts in both the private and public sectors, and to identify and describe the useful products of these efforts.

The document also identifies, to the extent currently possible, the known major sources of transportation data available to the Office of the Secretary. It concentrates on laying out a structure for relating these sources in order to systematize the data collection efforts for the purpose of identifying data products as a resource for the Office of the Secretary.

## 1.2 CONCLUSIONS

The following major conclusions were reached during the investigation under this review study:

a. The Office of the Secretary overwhelmingly depends upon transportation statistical data gathered by organizations outside DOT. Internal statistical data gathering is important; however, most analyses require a blending of both to illustrate particular issues.

b. Knowledge about the quality of the outside data is dispersed among the various analysts who use the data. Associated evaluations are often subjective and, generally, are not broadly available.

c. There is no unit within the Office of the Secretary that effectively centralizes knowledge about data and data sources. Elements within the Office of the Assistant Secretary of Policy, Plans and International Affairs and within the Transportation Systems Center attempt to provide this knowledge; however, no centalized, structured, and documented source exists.

d. The current situation of disaggregated knowledge about statistical data results in considerable dependence upon individual knowledge of each others efforts in order to be able to search for and identify data sources.

In summary, the current approach to handling statistical data within the Office of the Secretary is distributed and a lack of focus results. No one organization possesses a broad knowledge of the transportation statistics available to and within the Department.

# 2. SUMMARY OUTLINE OF DOCUMENT

## 2.1 THE PROBLEM

National transportation planning and policy analysis is young compared to national economic planning, national defense planning, etc. Because of this, a variety of problems exist in perceiving both the relationships between national planning and policy actions and their impact on society. While actions have been taken by the government prior to this time, these actions have tended to be narrowly defined (regulation, safety standards, etc.) and to be directed to specific sectors (rail industry, aircraft or automotive manufacturers, etc.). The data that are reported today tend to support these narrowly defined and directed activities and only now are the issues of data needs to support national transportation planning and policy analysis coming to the fore.

Prior Secretarial testimony before Congress has stated that...

> "The problem under consideration (the transportation information problem) is of a magnitude and complexity such that a considerable expenditure of effort for a number of years will be required to bring about needed improvements".
>
> "...To begin data collection anew each time a study is required is wasteful both of time and money... Relatively small expenditures on systematically organized and current information can help prevent...errors".
>
> "...No one should think that the development of a useful, operational national transportation statistics program is an easy task or one of short duration. Above all, it must be the objective of such a program to move away from data collection activities that are narrowly oriented to the needs of individual organizations and that have as their end the production of tabular materials and reports of limited usefulness. An effective program should serve the needs of as many

agencies and institutions as possible with the users' needs being the major determinant of content and of organization."

In an effort to further the utility value of data from other organizations the Secretary is directed to... "utilize the data, statistics, and other information available from Federal agencies and other sources to the greatest practicable extent."

An answer to the question, "What data should the Department collect?" requires an objective and orderly method for reviewing the data currently collected by the Department, other Federal agencies, and other sources. This review method should provide an evaluation of the potential users' requirements, as well as an assessment that the current collection cannot yield the required information.

The review of this vast supply of currently-collected data is, in fact, a review of only historical data. Thus, currently-collected data represents a supply to be retailed again and again, alone or in combination with other data, because of the difficulty of going back to collect missing pieces. A new data collection effort is, in effect, a wholesale operation. It must be justified on the basis of the data's retail value from that point forward in time.

This paper presents a framework for organizing Federal (public) transportation statistics to support national planning and policy analysis. It proposes a concept for systematically organizing statistical transportation data as a retailing activity that increases the credibility of statistics from the OST.

The objective of this study of data, data sources, and data availability is the establishment of a statistical

data system viewpoint that will facilitate the rational definition of the dimensions of the transportation-statistical-data universe.

By fitting together the many pieces of transportation and travel data, it becomes possible to develop an intelligently assembled, meaningful body of facts to support national transportation policy analysis.

## 2.2 STRUCTURE OF THE DOCUMENT

### 2.2.1 Objective

The objective of this document is to present the scope of current transportation data collection activities and to relate them to a structure that provides insight into the dimension of the problem. By considering only the retailing aspects of data-collection products, the evaluation of data is approached from the supply side of the supply/demand argument. Projects currently existing for basic data acquisition, the size of the data management problem and the way in which source data relate to the means of supplying vital statistics for creating information is addressed in this document. However, the demand for data not collected is not considered directly in this review. Furthermore, this report does not deal with internal management data, departmental standards for safety and the like, or forecasted data gathered by any means (including opinion surveys). Appendix A provides a more direct definition of data versus information.

### 2.2.2 Data Collection

The data collection processes of the Department are a complex set of interactions between internal and external organizations. For discussion purposes, these processes may be thought of as: 1) a process controlled by an external organization; 2) a process controlled by an internal organization; or, 3) a process controlled by a joint efforts between organizations.

The regulatory agencies, the Department of Commerce, state and private organizations, and others gather data of interest to the Department. In this area, the Department exercises a very limited role to influence change; however, most of the data, which are used within the Department, flow from this process. These data are identified and summarized in Section 3.

Within the Department, the Modal Administrations originate, conduct and largely control a variety of data programs. These programs are identified and summarized in Section 4.

The Office of the Secretary (OST) is largely dependent upon both the Modal Administrations and external organizations for its data. It exercises its role in the process by supplying some data collection funding support and by participating in joint committees, designed to oversee the directions being pursued by the performing organizations. The current OST role is generally limited to output specifications and review functions. Those OST data programs that do exist are reviewed in Section 5. In appendix B, past Congressional testimony on Department information program efforts is reviewed.

### 2.2.3 Data Utilization

Data collected by any private or public program is a resource to be utilized by the collecting organization and any other organization requiring access to the data. The utilization of such a resource, however, is often limited by the media of distribution (a hard copy versus a magnetic tape) and the ability or willingness of the controlling organization to publish data rather than statistics. Many organizations will produce statistics for consumption internally or externally; however, these same organizations are often reluctant to publish the raw data supporting these statistics. Nevertheless, the sharing of data rather than statistics is a growing concept.

The sharing of data, as opposed to statistics, significantly increases the variety of analyses that can be performed; it also adds complexity to the associated information that must be transferred to perform the analysis. For example, if one is to produce statistics from data, then knowledge of the observation and measurement system must be communicated to the analysts along with the data. To assure that the statistics being created are meaningful and to effectively deal with data sharing problems, data dictionaries and supporting documentation are created.

These documents, which represent knowledge about data, would be available to all, and would help reduce the duplication and isolation that now confronts the OST's efforts on vital statistics. While the knowledge about data would be shared, the processing of vital statistics would be distributed and controlled by the various organizations assigned information-reporting functions throughout the Department. The resulting efficiencies would increase Departmental institutional credibility, individual productivity, and resource utilization. The ease by which this knowledge would be shared with other elements would further DOT's standing as a member of the Executive Branch.

# 3. EXTERNAL DATA

## 3.1 CHARACTERIZATION OF DATA SUPPLY SYSTEM

This section focuses on sources of data external to the Department. Such sources produce the bulk of information utilized by the Department in its various analytic, decision-making, and policy formulation activities. The charts presented in support of the discussions illustrate a structure for describing and characterizing the associated transportation statistical data supply system.

The constituent subsections address the major transportation activities involving external sources: aviation, rail, bus, and trucking. Each supporting subsection chart is organized with the major operating elements identified in the left-hand column. These operating elements are the active consumption/production organizations in their respective fields of transportation.

The operating organizations (left-hand column) report various data elements to governmental organizations (reporting forms are identified where known) and to private institutions (e.g., the Reuben H. Donnelley Corp. for the Official Airline Guide) for the purpose of distributing the data to the public. The organizations that receive the data will be referred to as data collection organizations; these organizations are illustrated in each chart in the center column.

The data collection organizations develop products like computer tapes, statistical reports, etc., which they basically wholesale to the public, government, or internal analyst for use in creating information. These products fundamentally make up most of the Department's statistical data base.

Given this arrangement, a more detailed review of two areas is appropriate. One area involves the analysis of the data flow from operating organization to data collection organization to determine the data elements that are reported; the other area involves the analysis of the data products to determine that the collection procedure is producing a satisfactory data flow for the analysts.

Although this section does not present an exhaustive list of external data sources, it does organize and outline a structure for some of the more frequently used data that are available to DOT. This structure can well serve as the basis for cataloging and describing, in detail, the data elements of potential value to the Department. It is important to emphasize that the data selected for representation is that data which are gathered from regularized and repetitive data collection activities and that such data are amenable to time-series analysis, as well as cross-sectional analysis.

## 3.2 AIR TRANSPORTATION DATA

Figure 3-1 depicts major flows of data and information originating from organizations active in air transportation. Note the diversity of the types of data sources. Only some are regulated common carrier airlines. These organizations file required reports with the Civil Aeronautics Board, the agency charged with economic regulation of the industry, and voluntarily provide Reuben Donnelley with schedule information.

Other data are originated by the Federal Aviation Administration, the government agency charged with safety regulation and the control of the use of air space. Other originators of data flows include planning groups, airport operators, airmen and aircraft manufacturers and owners. Their functions encompass (as do those of the FAA) not only the common carrier part of the mode, but also the private, not-for-hire sector of air transportation.

It is also interesting to note that the FAA is not only an originator of data, but also a recipient of data, which it synthesizes into a broad spectrum of publications.

The flows of Figure 3-1 differ from the flows depicted for the rail, intercity bus, and intercity truck modes because Figure 3-1 includes non-commercial activities relevant to air transportation. The flows defined for the other modes include only commercial activity because only very limited reporting requirements are imposed on the non-commercial elements (as in trucking), or reporting

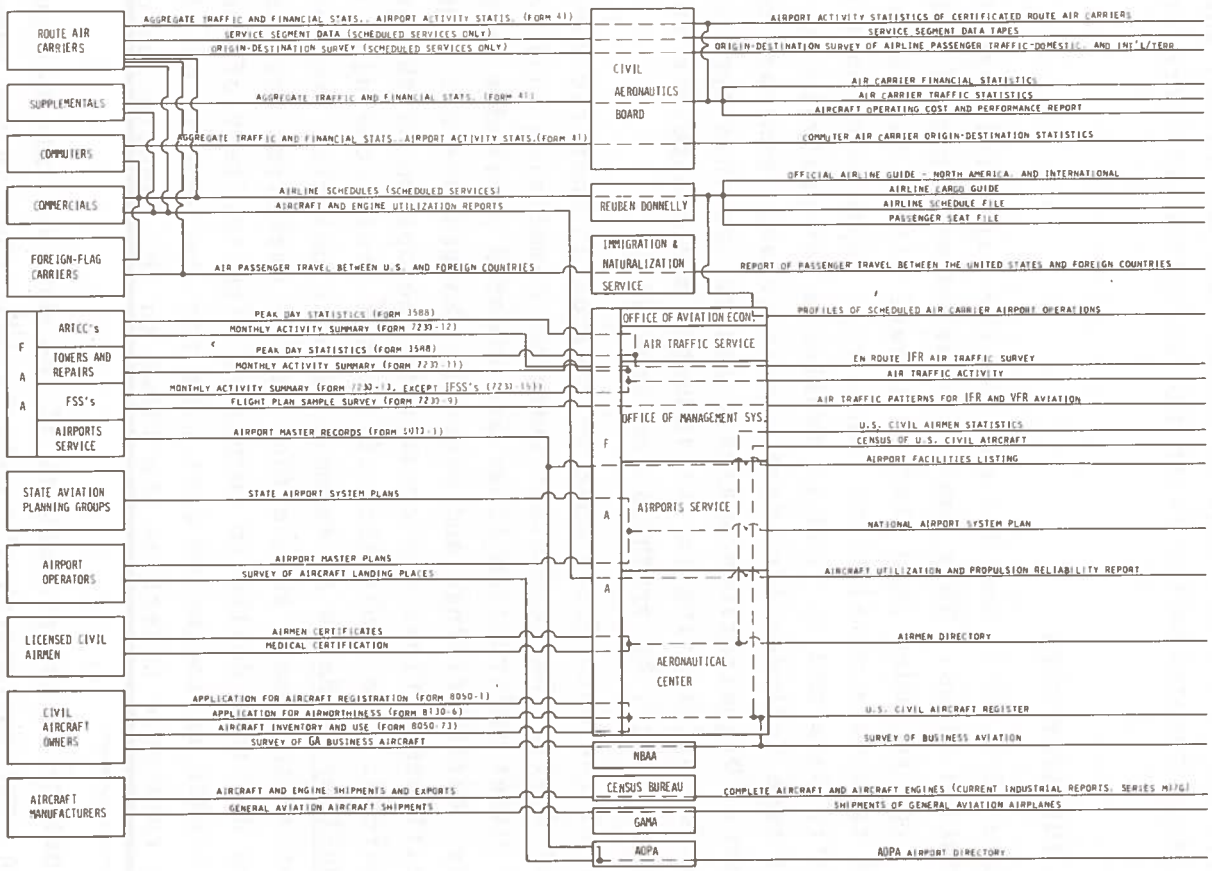



Figure 3-1. Airline Data Flow

requirements are non-existent (as in rail). In the cases of the rail, truck, and intercity bus modes, all flows pertain only to regulated common carriers (private or exempt). Nevertheless, even these activities are commercial in nature and the flows, therefore, are relatively comprehensive (with the exception of intra-state trucking).

## 3.3 RAIL INDUSTRY DATA

Figures 3-2, 3-3, and 3-4 portray the major railroad data and information flows. They can be regarded as being of two types. The first type includes data that originate within the railroad industry and terminate outside it. This type represents most of the rail industry data and is readily available for analysis by users. The second type includes data that originate and terminate outside the industry. Organizations/analysts generating this type use data or information originating in the industry and manipulate it into other formats (e.g., by synthesis or analysis).

Some of the data flows must be created at specific times. Reports filed with the Interstate Commerce Commission fall in this category. Other information flows are created at random intervals. For example, tariff filings and schedule changes are not regularly timed. Still other flows are created at the option of the individual railroad. The publishing of freight train schedules in the Official Railway Guide is an example of a voluntarily created flow. The motivation to create such a flow may be marketing oriented. It may also be due to a desire to cooperate with the rest of the industry to create an inclusive data base for the benefit of all. The latter rationale underlies the flows to the Association of American Railroads.

The nature of the processing of the data or information also varies. In some instances, the data is published in the same "raw" form as that in which it is received. In others (e.g., the construction of price indices), considerable statistical manipulation takes place. Some processing is entirely manual; other processing is performed by computers.

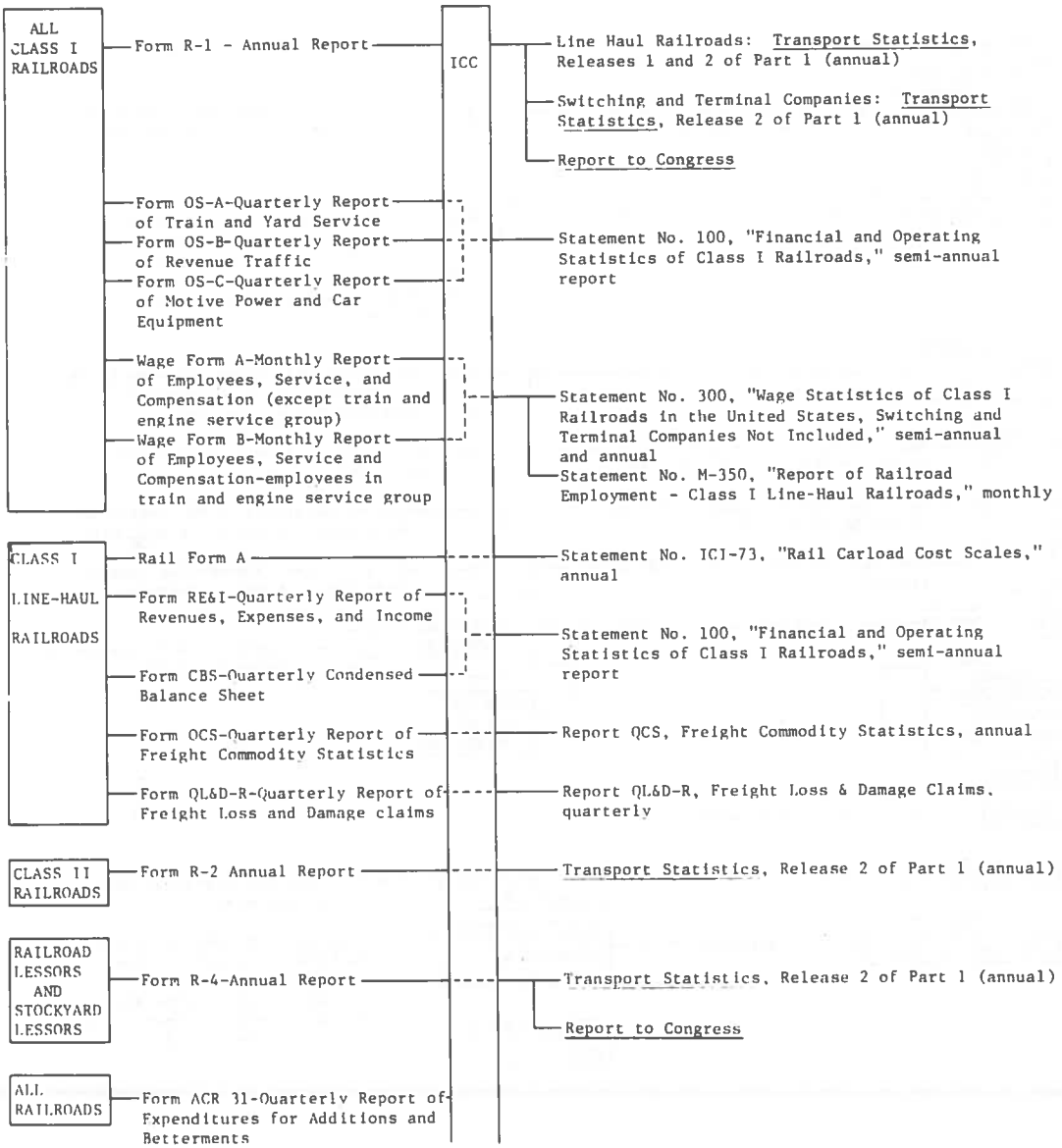



Figure 3-2. Railroads - Major Data and Information Flows

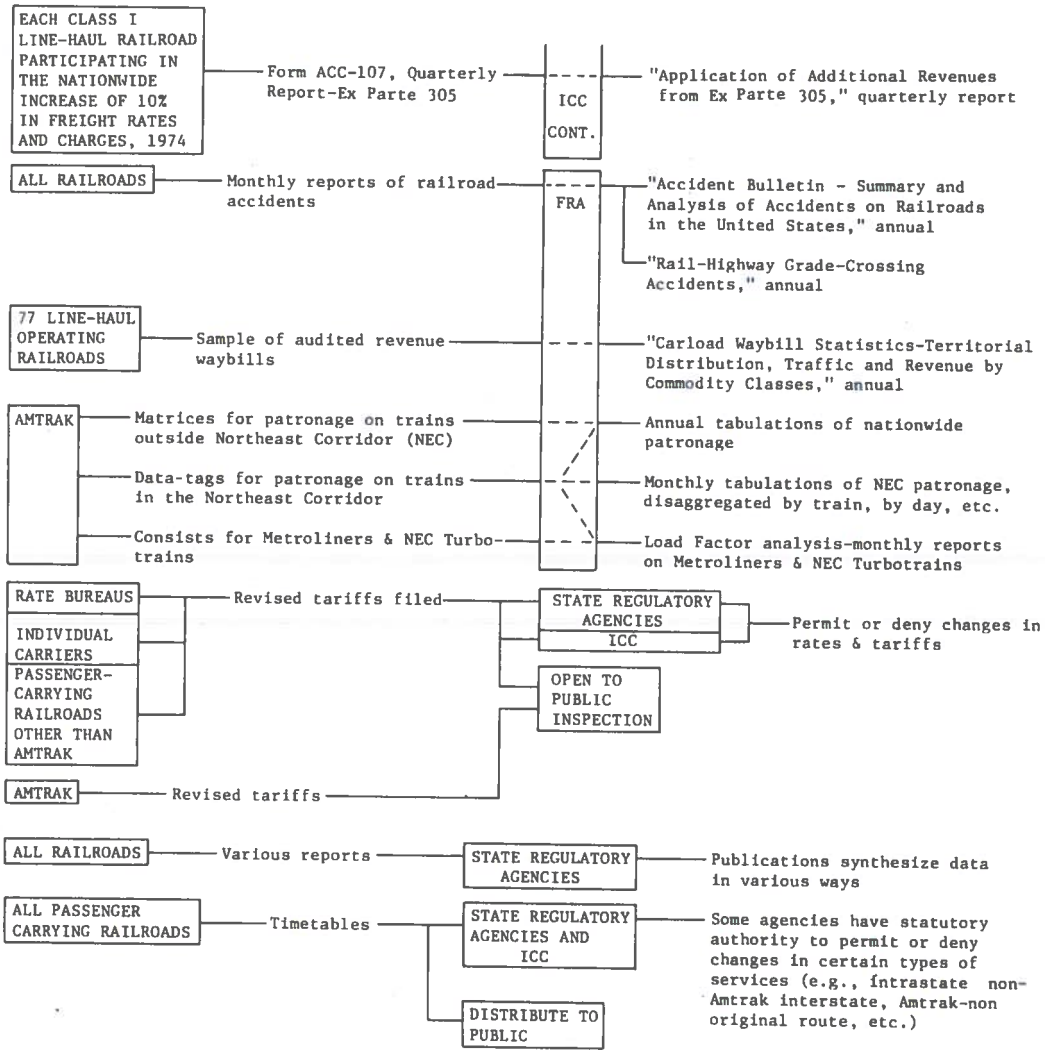



Figure 3-3. Railroads - Major Data and Information Flows (Continued)

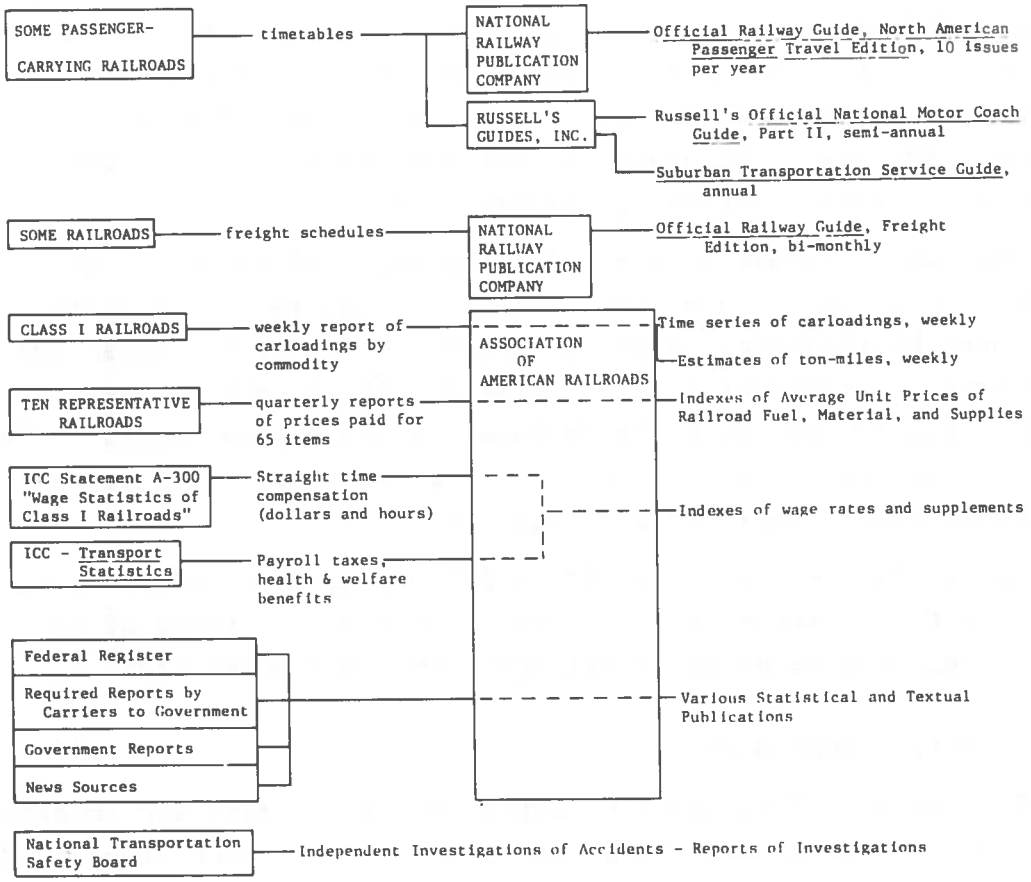



Figure 3-4. Railroads - Major Data and Information Flows (Concluded)

## 3.4 INTERCITY BUS INDUSTRY DATA

The major data and information flows pertaining to the intercity bus industry are shown in Figures 3-5 and 3-6. This industry is relatively small, in terms of total dollar volume, and relatively young, as compared with the railroad industry. These differences in size and "maturity" are reflected in the smaller number of flows that can be identified. They are also reflected in the less precisely defined nature of some of the "raw" data. As a result, intercity bus industry data is less likely to be subjected to sophisicated analysis than that of the larger, more "mature" industries. Moreover, the analysis that does take place is more likely to be manual instead of computerized.

The public nature of the rights-of-way used by bus lines, as distinguished from the dedicated or private nature of the rights-of-way used by railroads, also influences the types of flows that are created. Consequently, some flows are based upon adherence to the same requirements that are followed by all highway users. (For example, statutes pertaining to vehicle size, title certificates, and certain taxes, apply to all vehicles).

As in the case of the railroad data flows noted previously, the intercity bus industry flows can be regular or irregular in their timing, and required or voluntary in their generation.

## 3.5 INTERCITY TRUCK DATA

Figures 3-7, 3-8, and 3-9 portray the major data and information flows created by activities of the intercity truck lines. Because these carrers are motor carriers like the intercity buses discussed in Section 3.4, certain similarities between the data generated by the two modes will be observed. On the other hand, there are many differences in the flows. Some of these differences are based upon the distinctions between moving passengers and moving freight; e.g., in contrast to the bus industry, the trucking industry deemphasizes schedule creation and timetable publishing. Other differences are due to the far greater size (as measured by the dollar volume) of the business transacted by truck lines. These

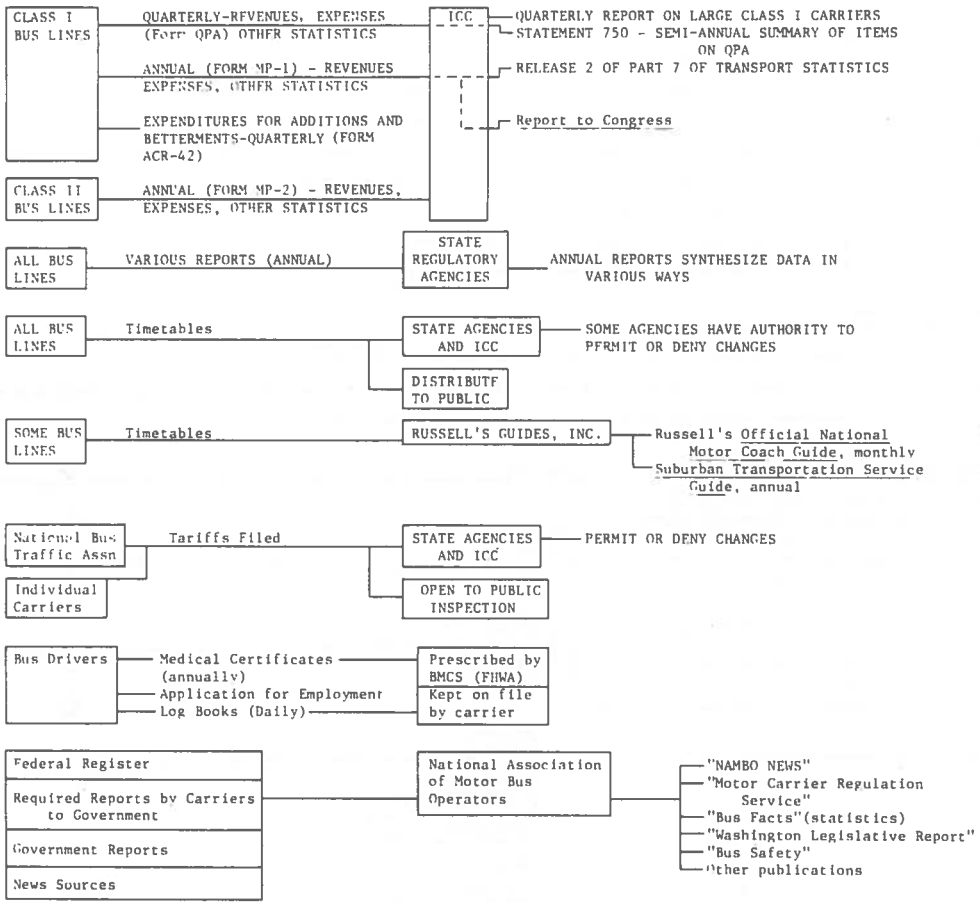



Figure 3-5. Intercity Bus Lines - Major Data and Information Flows

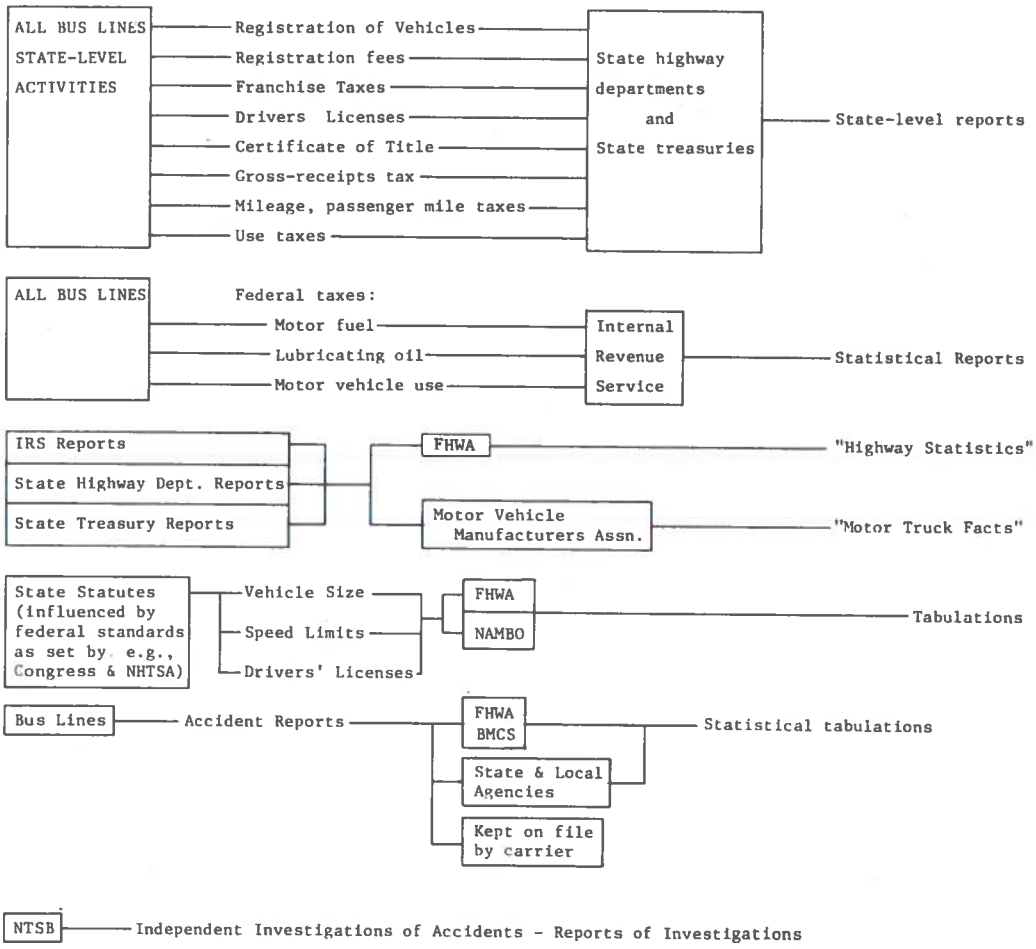



Figure 3-6. Intercity Bus Lines - Major Data and Information Flows (Continued)

factors have resulted in the levying of more reporting requirements upon the trucking industry.

There is also more manipulation of truck line data than bus industry data by non-governmental organizations. For example, the American Trucking Association (ATA) (a trade association) and Trinc's (a division of Dun and Bradstreet) both conduct major programs resulting in the publication of the data reported to the ICC by regulated truck lines. Certain truck lines voluntarily report additional data directly to the ATA for synthesis into a weekly index of tonnage.

Many of the information flows in Figures 3-7, 3-8 and 3-9 are applicable only to truck lines subject to the economic jurisdiction of the ICC or state regulatory agencies. Others, particularly those pertaining to highway usage and highway safety, are applicable to all truck lines.

## 3.6 BART STUDY - EXAMPLE OF URBAN AREA DATA

Because of the massive investment in the Bay Area Rapid Transit (BART) System for the San Francisco Bay Area, a special study was established to understand the relationship between this investment and resulting alterations in the area. Much data are gathered to support this study and is currently available through the Office of the Metropolitan Transportation Commission which has prepared a data dictionary and a modest system for query and retrieval. This section reviews the associated data holdings as an example of local data that could be made available for transportation planning.

The Metropolitan Transportation Commission (MTC) of the San Francisco Area has compiled an extensive data base containing large amounts of data which is presently available for use in transportation planning for assessing the impacts of the Bay Area Rapid Transit (BART).

In order to make the data holdings available to planners and analysts, MTC contracted CACI, Inc., to develop a modest information

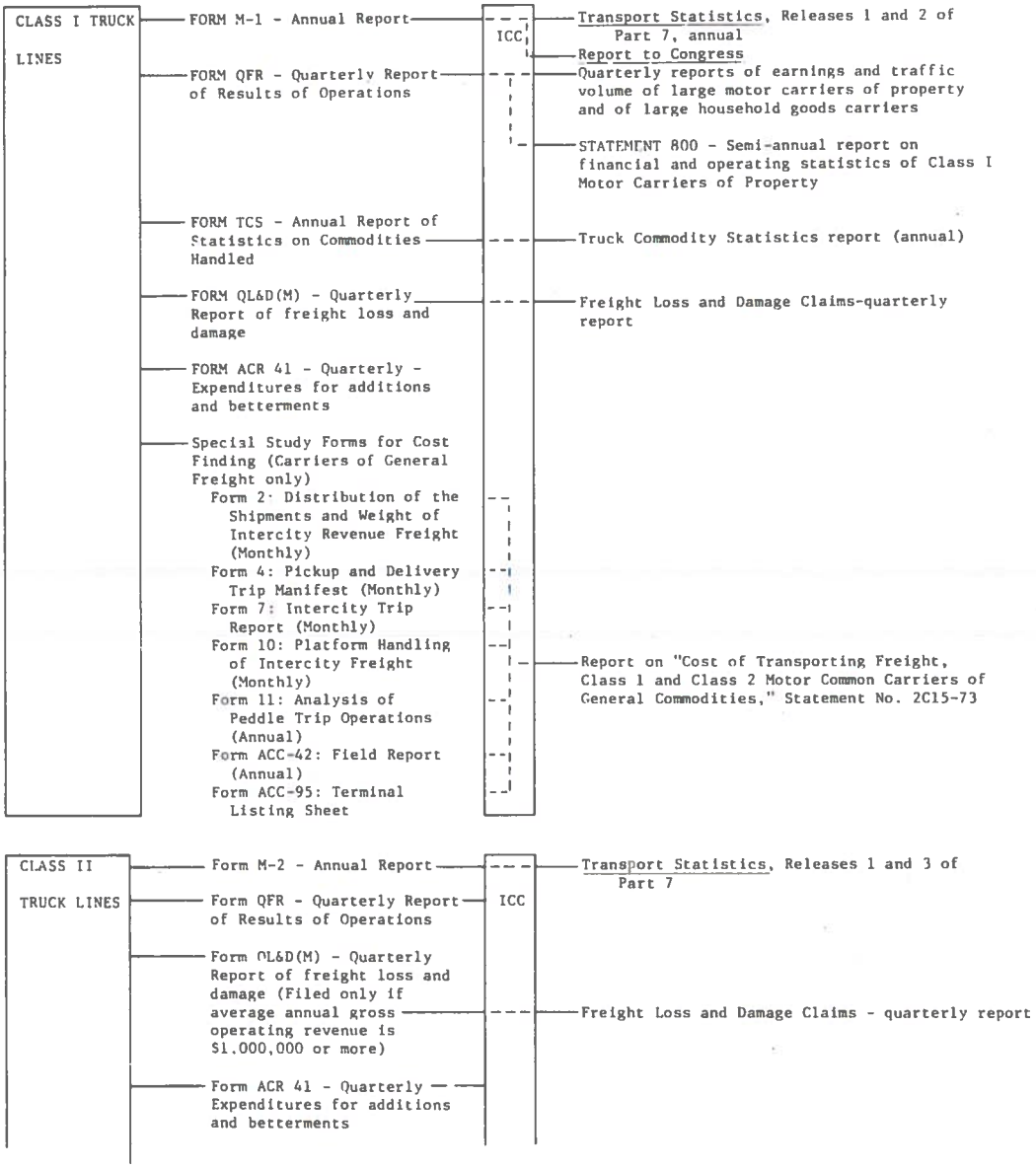



Figure 3-7. Common Carrier Truck Lines - Major Data and Information Flows

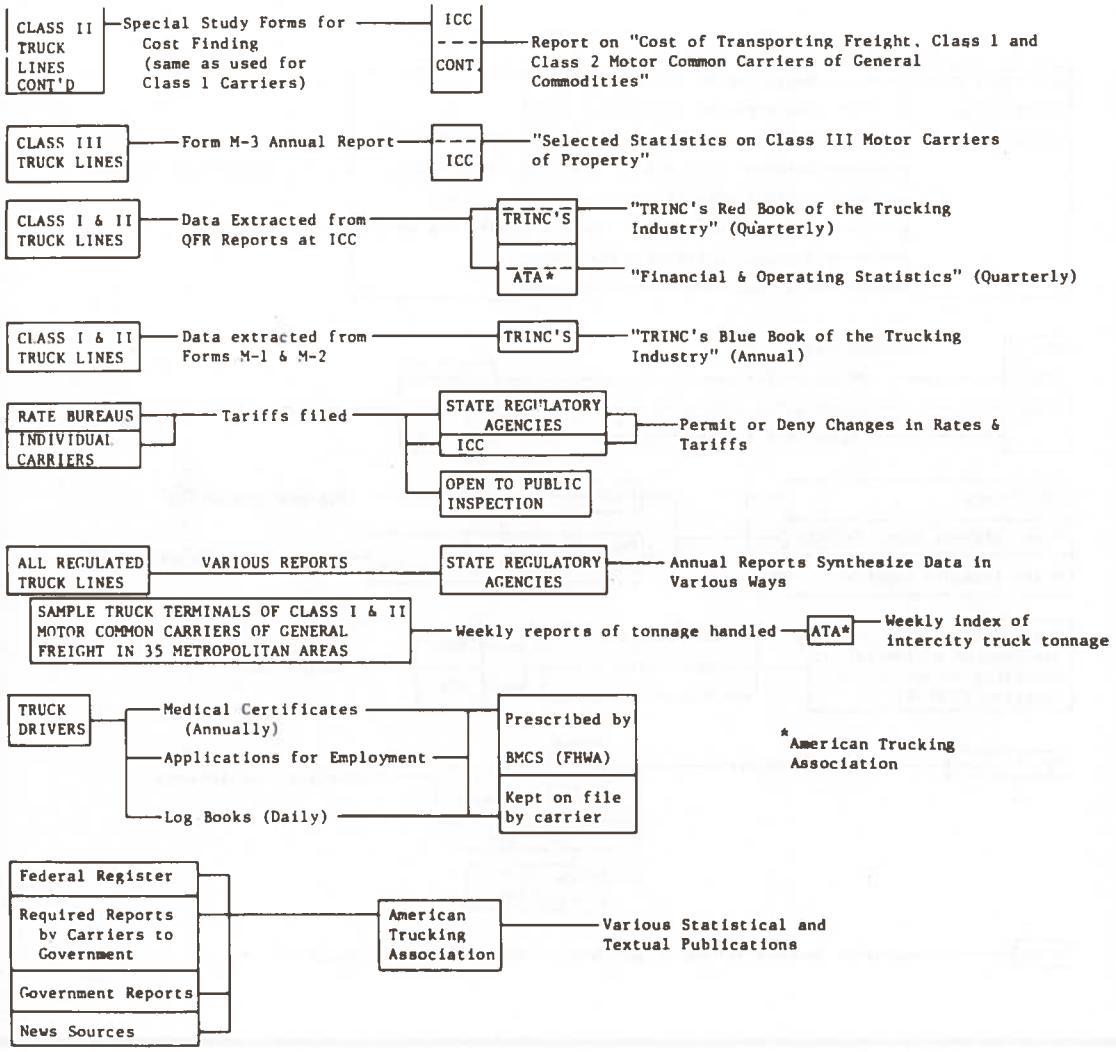



Figure 3-8. Common Carrier Truck Lines - Major Data and Information Flows (Continued)

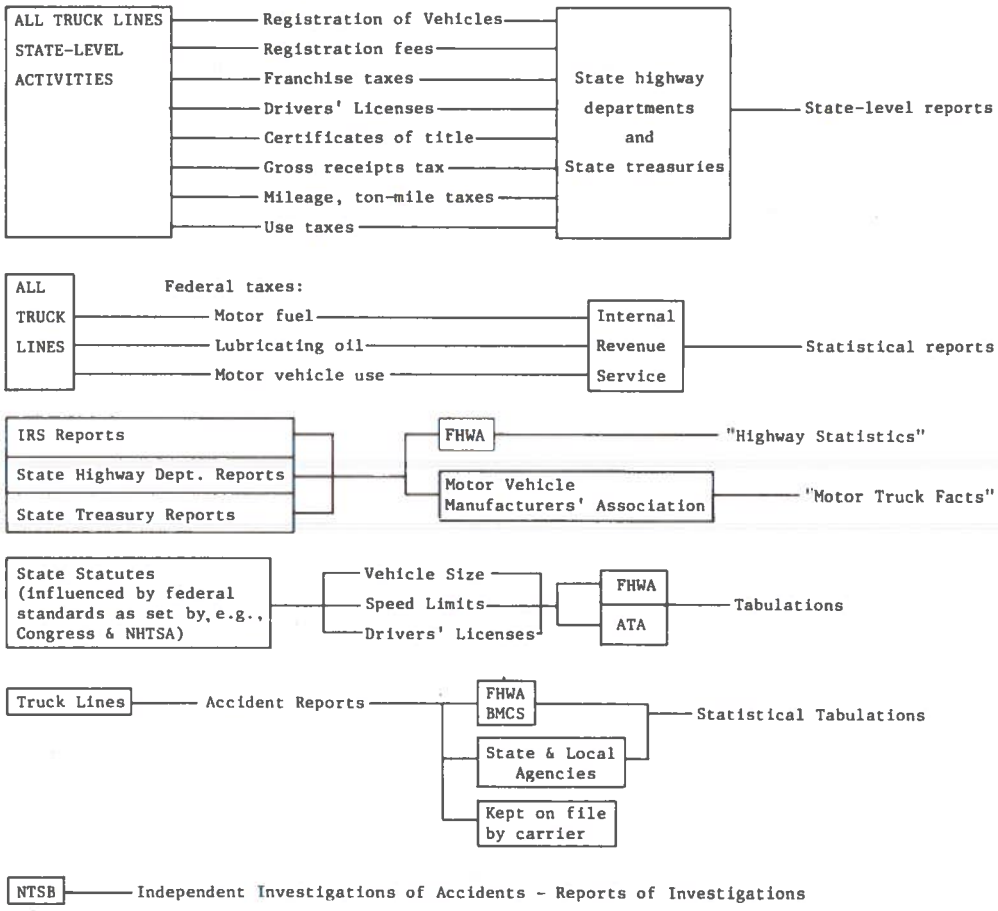



Figure 3-9. Common Carrier Truck Lines - Major Data and Information Flows (Concluded)

retrieval capability. The computer system was designed to give non-programmers the capability to search through a data catalog for elements giving specific descriptors of each data item (see Figure 3-10 for sample entries). Information in the data catalog includes: responsible organizations or individuals, date of creation, readability of data, contact person, physical description, etc., along with a series of computerized keywords (incorporated in several data dictionaries which are not apparent to the user).

The STOFI File Management System is the software language used to retrieve the data. STOFI, which operates in a batch environment, was developed on the Lawrence Berkeley Laboratory computer. After an analyst finds all pertinent data descriptors, the report data catalog (see Figure 3-11) is searched.

If an analyst wishes to produce reports from the data records selected, Quest (a user language) is used. Quest, a software system, has its own unique language which the analysts must learn in order to produce one of a series of "canned reports". MTC has discussed the possibility of converting this batch-operating system to a timesharing mode if the user community expands and needs access from locations using remote access computer terminals.

## 3.7 CALTRANS - EXAMPLE OF STATE TRANSPORTATION DATA

The State of California, Department of Transportation (CALTRANS), possesses one of the most organized state transportation data resources. An organization of people with modest computer support provide analysts, planners, and policy makers with direction and definition in their search for data. This section reviews the details of their index to illustrate the benefits that derive from this concept (which has also been implemented in other states).

CALTRANS Information Division is responsible for maintaining and providing information to analysts, planners, and policy makers within California's Department of Transportation. Recently, the Information Divison instituted the following two procedures to better serve the needs of analysts seeking information:

1.2 BART IMPACT PROGRAM PLANNING AND TECHNICAL PAPER CATALOG 20
BART IMPACT PROGRAM WORKING PAPERS

**BART IMPACT PROGRAM DATA SUMMARY**

DESCRIPTION OF THE DOCUMENT--
THIS WORKING PAPER IS A GUIDE TO DATA COLLECTED FOR USE BY THE BART IMPACT PROGRAM. FOR EACH DATA COLLECTION PROJECT IT CONTAINS--(1) A SHORT SUMMARY OF THE NATURE AND PURPOSE OF THE STUDY, (2) A DESCRIPTION OF EACH TYPE OF DATA COLLECTED AND THE SIZE OF THE DATA SET, (3) THE TIME PERIOD REPRESENTED BY THE DATA, (4) THE GEOGRAPHIC LOCATION OF THE DATA COLLECTION AND (5) A LIST OF THE DATA ITEMS AND REPORTS WHICH ARE PRESENTLY AVAILABLE AND CATALOGUED AT MTC AND THE DATA CATALOG I.D. WORDS BY WHICH THE ITEMS MAY BE REFERENCED.
THE APPENDIX INCLUDES STUDY AREA MAPS, MATRICES OF DATA COLLECTION BY BART STATION AREA, AND THE BART IMPACT DATA CATALOG TABLE OF CONTENTS. 55 PAGES
DOCUMENT ID. WORDS--
**WP 8-1-75**
RESPONSIBLE ORGANIZATIONS OR INDIVIDUALS--
SUSAN BACHMAN
DATE-- JULY 1974
CONTACT PERSON--
SUSAN BACHMAN
PHYSICAL DESCRIPTION--
WORKING PAPER WP8-1-75

PHYSICAL LOCATION--
BART IMPACT DATA LIBRARY
KEYWORDS ARE--
-DATA -BART -GEOGRAPHIC
-OVERALL BART IMPACT -SUMMARY -

**METHODOLOG. APPROACH FOR DEFINING THE GENERALIZED NO-BART ALTERNATIVE**

DESCRIPTION OF THE DOCUMENT--
THE GENERALIZED NO-BART ALTERNATIVE IS THE HYPOTHETICAL TRANSPORTATION SYSTEM JUDGED MOST LIKELY TO HAVE RESULTED IN THE BAY AREA IN FY 1976 IF BART HAD NOT COME INTO EXISTENCE. THE PURPOSE OF THIS WORKING PAPER IS TO EXPLAIN HOW MTC INTENDS TO DEFINE THE GNBA AND THE ROLE TO BE PLAYED IN THIS PROCESS BY FINDINGS FROM THE DECISION HISTORY OF BART. IT IS STRUCTURED IN THREE PARTS--(1) INTRODUCTION TO BASIC CONCEPTS AND CRITERIA, (2) OVERVIEW OF METHODOLOGICAL APPROACH AND GUIDELINES FOR IMPLEMENTATION, AND (3) APPLICATION OF CRITERIA FOR DEFINING THE GNBA. 20 PAGES
DOCUMENT ID. WORDS--
**WP 10-1-75**
RESPONSIBLE ORGANIZATIONS OR INDIVIDUALS--
STEPHEN K. ROSENTHAL
DATE-- OCTOBER 1974

Figure 3-10. Sample Printout of BART Impact Program Data Catalog

**SUM SER 2 POP./EMPLOY./LAND USE PROJECTIONS (S.F. REGION, 1970-2000)**

)ESCRIPTION OF THE DATA SET--
THE ABAG/MTC JOINT LAND USE TRANSPORTATION PLANNING PROGRAM HAS DEVELOPED ALTERNATIVE REGIONAL GROWTH CHOICES FOR THE SAN FRANCISCO BAY REGION BY PREPARING POPULATION, EMPLOYMENT, AND LAND USE PROJECTIONS TO THE YEAR 2000. THE PROJECTIONS INCLUDE THE PREAD CITY-CENTERED CONCEPT OF THE ABAG REGIONAL PLAN. 1970-1990, ENCOMPASSED WITHIN THREE GROWTH ALTERNATIVES, AND THEY PROVIDE A SUBSTANTIAL TECHNICAL BASE FOR PLANNING ALTERNATIVES UPON WHICH A PROCESS OF LOCAL REVIEW AND PARTICIPATION CAN BE INITIATED IN THE FORMATION OF REGIONAL AND LOCAL GROWTH POLICIES.
THE SERIES 2 PROJECTIONS ARE A REFINEMENT OF THE SERIES 1 PROJECTIONS WHICH WERE PRIMARILY INTENDED FOR USE BY THE STATE WATER RESOURCES CONTROL BOARD IN ITS PREPARATION OF A SAN FRANCISCO BAY WATER QUALITY MANAGEMENT PLAN. THEY REFLECT MODELING IMPROVEMENTS AND ARE BASED ON MORE COMPREHENSIVE PLANNING FACTORS AND INFORMATION.
THE FRAMEWORK OF THE STUDY, THE ALTERNATIVES USED, MAJOR PLANNING ASSUMPTIONS, AND THE ZONAL FRAMEWORK ARE DISCUSSED IN THE REPORT. APPENDIX A LISTS THE TECHNICAL WORKING PAPERS WHICH ARE AVAILABLE UPON REQUEST. 65 PAGES

Figure 3-11. Sample Printout of Data Set Description in BART Impact Program Data Catalog

a) A Transportation Planning Support Information System (TPSIS)

b) A TPSIS News Release

The Transportation Planning Support Information System (TPSIS) was developed to provide information and data needed to produce and evaluate transportation plans and environmental impact reports. It was basically designed to build a network of information users, owners/suppliers and delivery systems for communicating the needed data from one to the other. This communication is accomplished by maintaining an alphabetic listing on books, reports, magnetic tape files, programs and computer-based models. Data content includes U.S. Bureau of the Census data files, California Department of Finance population projections, employment data files and income data, to name a few. TPSIS also maintains geographic data (e.g., maps, software, computer-driven plotters) which provide a special capability to individuals doing Socio-Economic or Physical Data analysis. (Figure 3-12 contains a sample of an alphabetic listing).

Analysts seeking transportation information do not have to laboriously search through books of alphabetical listings to obtain information because TPSIS operates two functional systems to provide this information. For analysts preferring large amounts of bibliographic information, a recently purchased Teale Data Computer can produce such listings in a batch mode using a cross-indexing system called KWIC (Key Word In Context). For small amounts on information, a clerk receives and processes requests using a manual microfiche terminal which contains the same dictionary of information as the KWIC system provides.

The CALTRANS Information Division instituted the second procedure, the TPSIS news release, to promote interest and awareness of TPSIS enhancements among analysts within CALTRANS. (Figure 3-13).

The Information Division is also instituting procedures to build future data files and create on-line data bases within TPSIS, as required. CALTRAN's plans to evolve into a more sophisticated environment, as diagrammed in Figures 3-14 and 3-15, are dependent on the degree to which the present system is utilized.

STATISTICAL REPORTS OF THE DEPARTMENT OF PUBLIC WORKS PERTAINING TO DIVISION OF HIGHWAYS DISTRICT-02 LIBRARY D02-0024.000

QUALITY

WATER QUALITY CONTROL PLAN FOR THE NORTH LAHONTAN BASIN, INTERIM DISTRICT-02 LIBRARY D02-0006.005

WATER QUALITY CONTROL PLAN, INTERIM CENTRAL VALLEY, FROM SAN-JOAQUIN DELTA NORTH DISTRICT-02 LIBRARY D02-0006.006

QUARTERLY

MOVING PEOPLE ON URBAN FREEWAYS, REPRINT FROM TRAFFIC QUARTERLY, JULY 1970, DISTRICT-02 LIBRARY D02-0004.030

RANGE

SHORT RANGE TRANSIT PLANNING JULY 1973 (2 COPIES) DISTRICT-02 LIBRARY D02-0014.070

RAPID-TRANSIT

THE POTENTIAL FOR BUS RAPID-TRANSIT. MASS-TRANSIT. DISTRICT-02 LIBRARY. D02-0014.040

RATES

HOW LAND AND FLOOR USUAGE RATES VARY BY INDUSTRY AND SITE FACTORS. ESTIMATING LAND AND FLOOR AREA IMPLICIT IN EMPLOYMENT PROJECTIONS VOLUME I. DISTRICT-02 LIBRARY. D02-0009.000

HOW LAND AND FLOOR USUAGE RATES VARY BY INDUSTRY AND SITE FACTORS. ESTIMATING LAND AND FLOOR AREA IMPLICIT IN EMPLOYMENT PROJECTIONS VOLUME II. DISTRICT-02 LIBRARY. D02-0009.010

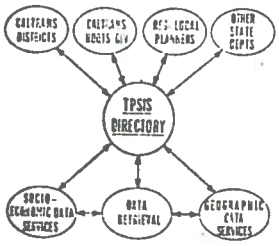
Figure 3-12. Sample Alphabetical Listing from TPSIS

State of California

# Memorandum

To :

Date : July 18, 1975




From : DEPARTMENT OF TRANSPORTATION – Division of Transportation Planning

TPSIS USERS RELEASE NO. 18

Prior to initiating the TPSIS USER RELEASES the Transportation Planning Support Information System (TPSIS) announced in a memo the acquisition and availability of the detail tapes for the Department of Finance population projections from the year 1970 through 2020. Several of you did not receive that memo and may not be aware of the detailed data these files contain.

These computer tapes provide a flexibility not available by using only the printed reports which have arbitrary age groupings and are published only for selected years. The availability of these data on tape makes it possible for the planner to select the data he needs for whatever years or counties he is interested in and the flexibility to combine it with other pertinent data.

There are two sets of files extablished by TPSIS on the Teale Data Center computers. One set covers the projections for the total population. The other set covers the projections for the civilian population only (excluding the military).

There are four separate tape files in each set, each representing a different set of projections criteria for the years 1970-2020. Each file contains 508,776 records. Each record within the magnetic tape file for each set of criteria contains the following data fields: year, county, age (0 through 85 and over), sex and the population projection (see Attachment I for the record layout, the data set names and the tape reel serial numbers).

The following descriptions of the criteria governing the projection of population in each of the four files in each set should help you choose which of them is applicable in your processing:

E-000 – This is the code number of the lowest of the four projections. It assumes a "status quo" growth. That is, the projections were based on a zero net migration to California and an average birth rate of 2.1 children per woman of child-bearing age.

Figure 3-13. Sample of TPSIS News Release

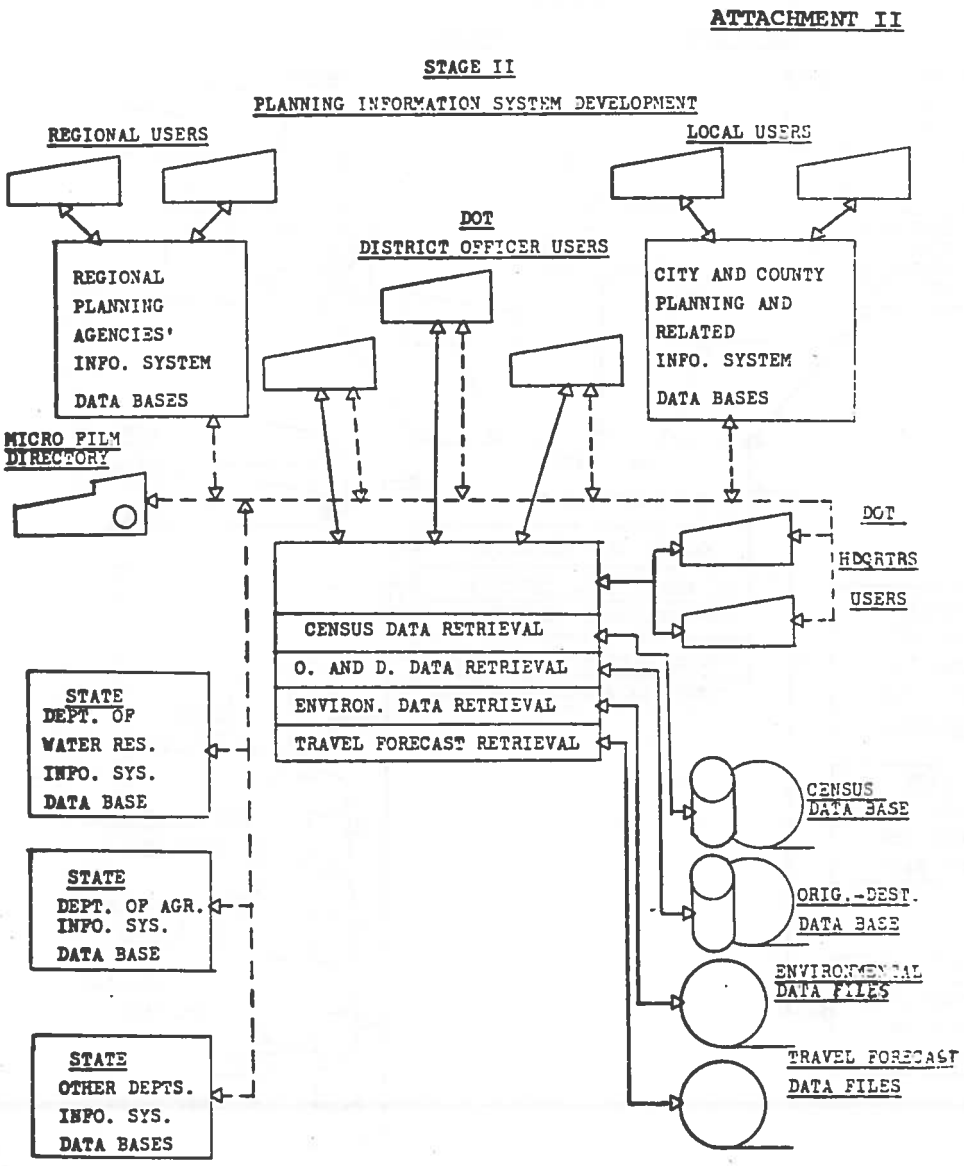



Figure 3-14. TPSIS Development Plan - Stage II

ATTACHMENT III

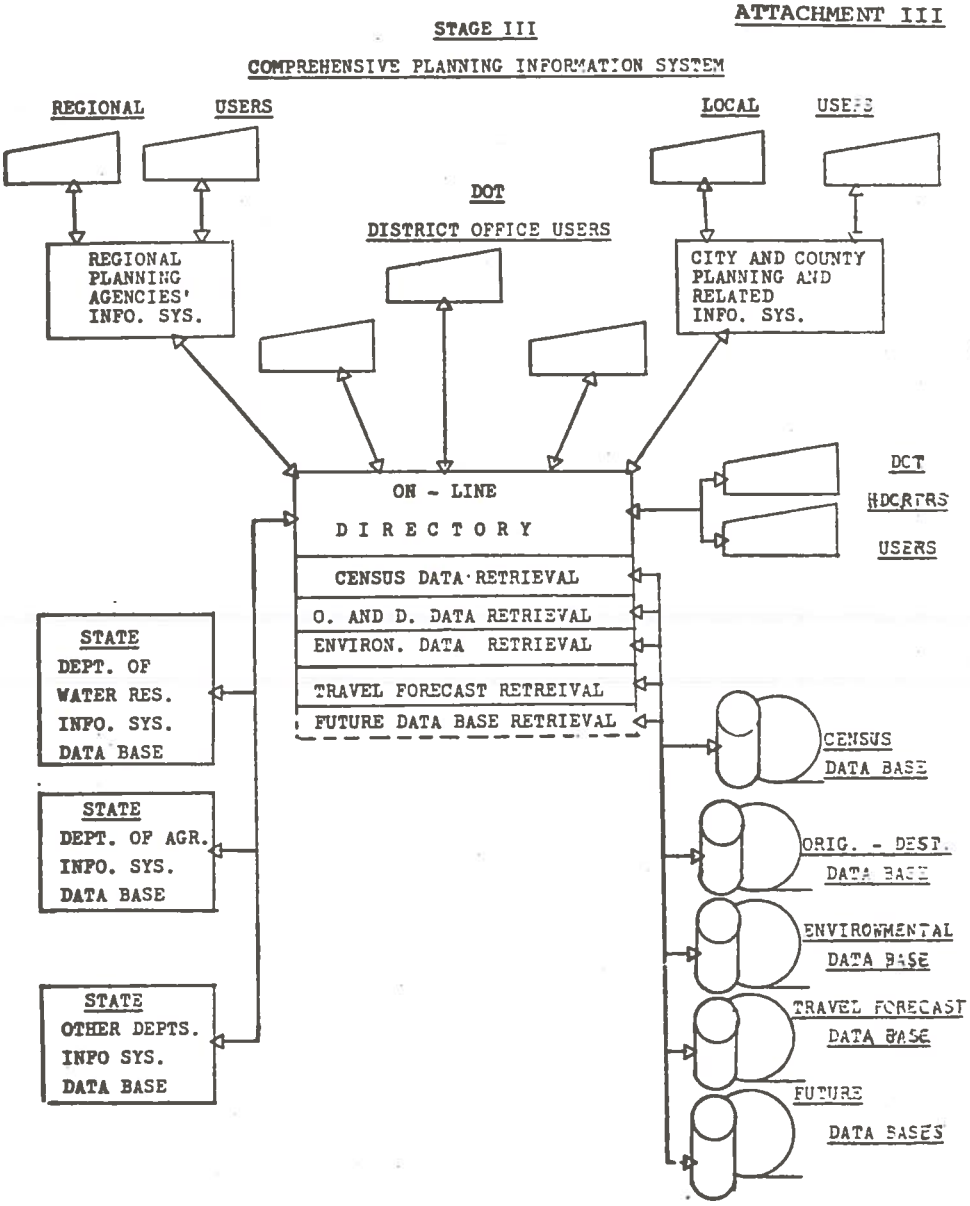



Figure 3-15. TPSIS Development Plan - Stage III

## 3.8 LAWRENCE LIVERMORE LABORATORIES (LLL) - NATIONAL INDEX

As part of its new role in the Energy Research Development Agency, the Lawrence Livermore Laboratories (LLL) began in 1973 to establish a National Index of energy and environmental data bases. Because of its pervasiveness, transportation data bases appear in this nationl index.

In developing the index, LLL conducted a survey in cooperation with other Federal agencies and state and local governments. Private sector universities, professional societies, and industry were included to the extent of a reconnaisance survey. The Federal Agencies included were:

a) The ERDA Laboratories, ANL, BNL, HNL, LASL, LBL, LLL, PNL, SRL,
b) The Federal Energy Administration, FEA,
c) The Environmental Protection Agency, EPA,
d) The U.S. Geological Survey, USGS,
e) The U.S. Bureau of Mines, USBM,
f) National Technical Information Services, NTIS,
g) National Oceanic and Atmospheric Administration, NOAA, and others.

Information on the various existant energy and environmental data bases and calculational models obtained from this survey, as well as earlier and in-progress surveys, will be aggregated to generate the National Index.

A unique aspect of this cumulative survey is its comprehensive coverage, specifying as it does (in technical detail) the different types of variables available in data bases or required by calculational models. Particular attention is given to the geographic and temporal coverage of those variables that lend themselves to nationwide analysis and assessment programs (e.g., the Regional Studies Program and the National Uranium Resources Evaluation Program of ERDA, among others).

The Lawrence Livermore Laboratories (LLL's) Information Research Group (IRG) collected 20,000 different entries, of which 4,000 seemed to be of direct interest to ERDA; approximately 150 entries relate to transportation data and would be of interest to DOT. LLL gathered most of the data entries through a comprehensive questionnaire. Magnetic tapes and printed indexes from EPA, FEA, NOAA, NTIS, USBM, and other federal and state agencies comprised a large portion of the National Energy and Environmental Data Base Survey acquisitions. LLL has a vast amount of computational capability (Figure 3-16) to support the activities of the IRG.

To ensure high reliability and integrity in the ERDA Data Base, the IRG utilized a highly sophisticated data base management language, Master Control, to examine the data base's accuracy. Master Control is a computer language designed to unify storage, validation, manipulation, retrieval and display of information from dissimilar data bases. This language can be used on-line in an interpretive mode or in conjunction with Fortran; it has been successfully used with data bases containing millions of characters of information (Figure 3-17). A unique concordance command within Master Control permitted IRG to produce many useful sets of indexes (e.g., title, author, agency, etc. --- see Figures 3-18 & 3-19). using single and multiple words.

Along with LLL's computational and data base capabilities, IRG has access to its extremely powerful graphical and data communications network capabilities (Figures 3-20 and 3-21). These additional capabilities allow IRG to represent the National Index graphically and statistically and allow the outside user community to access the data base through a computer network.

In summary, of the various data base/computer capabilities reviewed during this study, those of Lawrence Livermore Laboratories were found to be the broadest amd most substantial.

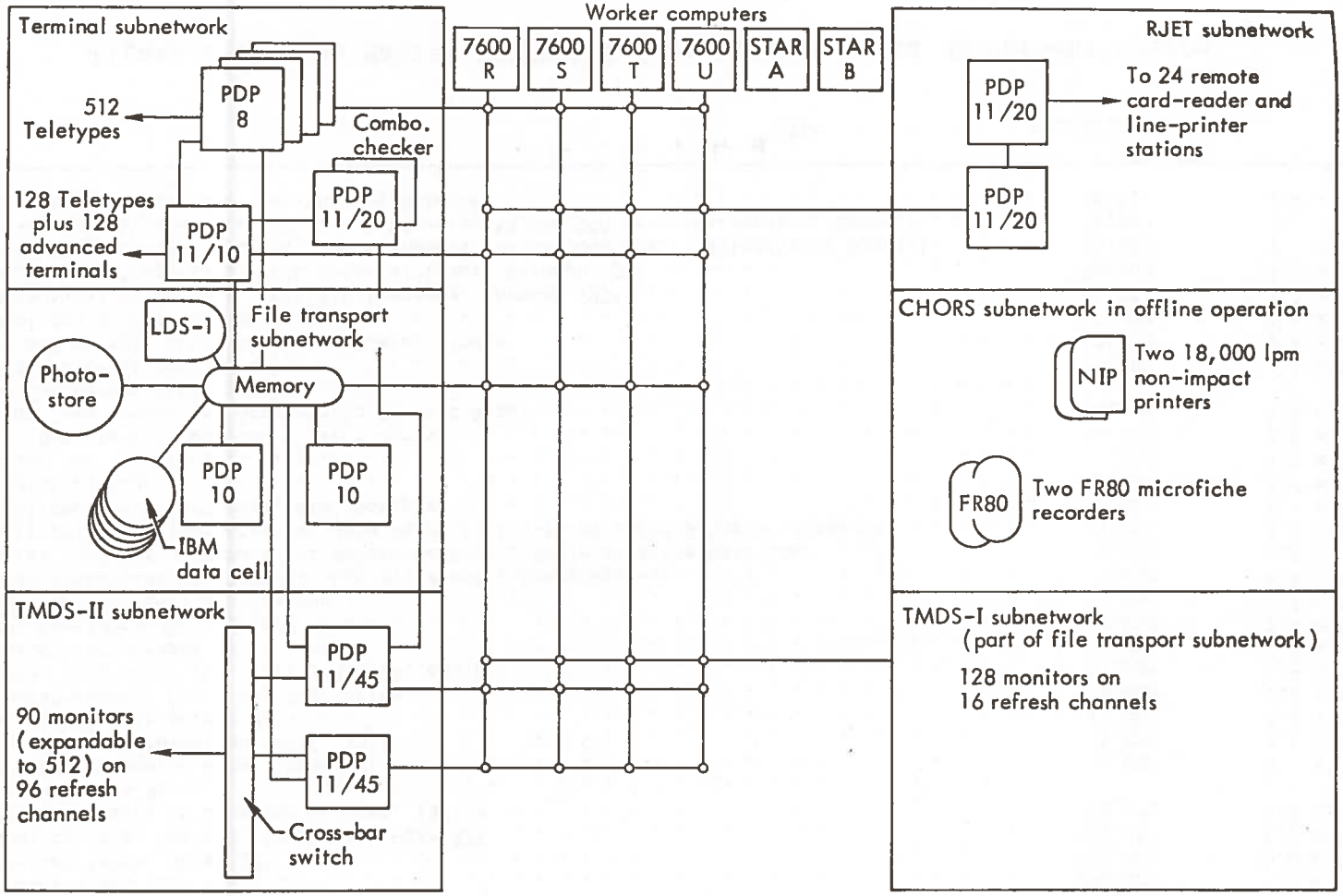



Figure 3-16 Lawrence Livermore Laboratories (LLL) Computational Capability

| | Approx. # of Records | Approx. # of Characters |
|---|---|---|
| Nuclear Science Abstracts, 1968 - | 375,000 | $2.7 \times 10^8$ |
| Nuclear Science Abstracts, Modified Subject Headings (RESPONSA), 1962-1967 | 217,000 | $1.6 \times 10^7$ |
| Air Quality Data, California, 1970 | 171,000 | $4.1 \times 10^8$ |
| Chemical Titles, 1971 - | 270,000 | $1.9 \times 10^8$ |
| Engineering Index, July 1971 - | 260,000 | $5.6 \times 10^6$ |
| Air Quality Data, Basin-2, Bay Area, 1963-1971 | 109,000 | $6.5 \times 10^8$ |
| Searchable Physics Information Notices, 1971 - | 165,000 | $4.5 \times 10^6$ |
| Proprietary Data Bases | | |
| World-Wide Earthquake Data, 1961-1971 | 39,000 | $9 \times 10^7$ |
| Nuclear Safety Information, NSIC, 1971 | 35,000 | $2.5 \times 10^7$ |
| 8-Peak Mass Spectroscopic Data | 17,000 | $2.5 \times 10^7$ |
| Mass Spectroscopic Data, Low Resolution | 8,000 | $1.2 \times 10^8$ |
| Personnel Data Base, LLL, LBL, Site-300, and NTS | 10,000 | $1.9 \times 10^7$ |
| Plutonium Bibliography | 3,500 | $2.5 \times 10^7$ |
| Nuclear Constants Bibliography | 3,400 | $2.4 \times 10^7$ |
| Information Retrieval Bibliography | 3,000 | $2.1 \times 10^6$ |
| Physical Properties of Isotopes, Masses, Binding Energies, etc. | 2,400 | $4.8 \times 10^6$ |
| Technical Information Department Reference Group, Administrative Data Base | 4,000 | $2.9 \times 10^5$ |
| Autistic Children, 165-Parameter Data Base, Institute of Child Behavior Research | 2,400 | $5.8 \times 10^6$ |
| Integral Neutron Experiments, Bibliography | 2,500 | $5 \times 10^6$ |
| Laser Bibliography | 2,000 | $1.5 \times 10^6$ |
| USAEC and LLL Patent Cases | 1,000 | $2.7 \times 10^7$ |
| Textual Data Base (~ 800,000 English words) | 31,102 | $2.5 \times 10^5$ |
| Reaction Rate Constants, Atmospheric Kinetic Data | 2,000 | $4.8 \times 10^5$ |
| Vacuum Technology Bibliography | | |
| Computer Sciences Bibliography | 15,000 | $1.1 \times 10^7$ |
| "Newsline" Monthly LLL Publication Subject Index | 2,000 | $2.4 \times 10^6$ |
| Personal and Administrative Data Bases | 10,000 | $2.2 \times 10^7$ |
| Thermophysical Properties, TPRC Bibliography, through 1971 | 65,000 | $3.7 \times 10^7$ |
| Thermophysical Properties, TPRC Material Index, through 1971 | 180,000 | $4.3 \times 10^7$ |
| Solar System, Position Vectors, Outer Planets, 40,000,000 yrs. (Heliocentric Coord.) | 20,000 | $2 \times 10^7$ |
| Solar System, Position Vectors, Outer Planets, 40,000,000 yrs. (Barycentric Coord.) | 20,000 | $2 \times 10^7$ |
| Computer Performance, LLL Time-Sharing Network | 10,000 | $2.4 \times 10^6$ |
| TOTAL | ~2,000,000 | |

Figure 3-17 The MASTER CONTROL Information and Data Management System

| | Approx. # of Records | Approx. # of Characters |
|---|---|---|
| Material Properties, TRUMP, UCRL-50589 | 2,000 | $4.8 \times 10^5$ |
| Plant Engineering, LLL, Equip. Data Base Maintenance for Work Scheduling | 6,000 | $7.2 \times 10^7$ |
| Car-Pool Matching Personnel Data Base | 5,500 | $9.5 \times 10^7$ |
| Car-Pool Permit Data Base | 800 | $1.9 \times 10^5$ |
| Affirmative Action Data Base | 5,500 | $9.5 \times 10^7$ |
| Continuing Education Data Base, UCRL-51040 | 5,500 | $3.8 \times 10^8$ |
| Biomedical Department, Bibliographic Data Base, UCRL-50163 | 13,000 | $9.6 \times 10^6$ |
| Prediction of the Max. Dosage to Man from Nuclear Device Fallout, UCRL-51013. 1-5 | 5,000 | $2.4 \times 10^6$ |
| Information on Elemental Concentration in Human Organs | 4,800 | $3.3 \times 10^6$ |
| Atomic Masses, Q-Values & Separation Energies | 2,400 | $2.9 \times 10^6$ |
| Evaluated Gamma-Ray Transitions, G.N. Rao | 2,337 | $1.5 \times 10^7$ |
| Experimental Gamma-Ray Transitions, G.N. Rao | 13,297 | $1 \times 10^7$ |
| Radio-Active Decay Gammas, Atomic & Nuclear Data Tables, 13, Feb. 1974 | 19,415 | $1 \times 10^7$ |
| Bibliography in Support of Gamma-Ray Transition,s G.N. Rao, et al. | 8,500 | $5 \times 10^6$ |
| Physical & Chemical Compatibility of Materials | 2,000 | $1 \times 10^6$ |
| Security Department Data Base, Administration of Manuals, etc. | 5,500 | $1.3 \times 10^6$ |
| Budget Office, Miscellaneous Reimbursable Accounts | 2,000 | $4.8 \times 10^5$ |
| Topography of California | 10,000 | $2.4 \times 10^6$ |
| San Francisco Bay Area Atmospheric Pollution Model Data Base, Seven Agencies | 100,000 | $1 \times 10^9$ |
| Fire Prevention | 2,000 | $1.5 \times 10^6$ |
| Bibliography of Geothermal Phenomena, from Mexico Institute of Mining & Technology | 15,000 | $1.5 \times 10^7$ |
| Abstracts on Health Effects of Environmental Pollutants, HEEP, BA-Previews, 1975 - | | |
| Chemical Abstracts REgistry File, Subset of Carcinogenic Substances | 1,000 | |
| Quantitative Data Base of Geothermal Data, USGS* | | |
| Energy-Related Data Base, an Extract from All Other Commercial Data Bases | | |
| Library Book Catalogue, LLL Technical Information Dept. (28,000 are unique)* | 35,000 | |
| Library Document Shelf List, Technical Information Dept. (80,000 are unique)* | 140,000 | |
| INSPEC, Computer and Control Aspects, 1975 -* | | |
| Maintenance and Preventive Maintenance Records of Electronic Instruments* | 30,000 | |
| TOTAL | $\sim$2,500,000 | $\sim 3.0 \times 10^9$ |

*In the process of being merged.*

*The number of proprietary data bases is about 30 and is growing rapidly. Requests can be made with the appppropriate classification and need-to-know.*

Figure 3-17 (Continued)

1 MASTER CONTROL NATIONAL INDEX OF ENERGY AND ENVIRONMENTAL DATA BASES AND MODELS - ERDA/DBER, EDITION 2 (OCT. 75
SUBJECT INDEX BASED ON TITLE, KEYWORDS, AND GEOGRAPHIC DESCRIPTION (DATA BASES)

| TYPE | REF-NUMBER | R-E-L-E-V-A-N-C-E I-N-D-I-C-A-T-O-R | I-N-D-E-X T-E-R-M |
|---|---|---|---|
| | | | (CONT'D) TRANSPORTATION |
| DAT | 60198 | DOT/DF-73-028; *TRANSPORTATION MODELS, *DEVELOPING COUNTRIES, | |
| DAT | 60206 | CEN/DF-73-092; (*GEOCODING, UNITED STATES), *CARGO TRANSPORTATION, | |
| DAT | 60207 | CEN/DF-73-091; (*GEOCODING, UNITED STATES), *CARGO TRANSPORTATION, | |
| DAT | 60210 | ICC/DF-73-001; *COMMON CARRIERS, *MOTOR TRUCKS, *CARGO TRANSPORTATION | |
| DAT | 60211 | DOT/DF-73-020; *AIR TRANSPORTATION, *COMPUTER PROGRAMS, *RENTING, | |
| DAT | 60233 | CEN/DF-73-111; (*GEOCODING, UNITED STATES), *TRUCKS, *CARGO TRANSPORTA | |
| DAT | 60234 | CEN/DF-73-110; (*GEOCODING, UNITED STATES), *TRAVEL, *PASSENGER | |
| DAT | 60239 | CEN/DF-73-034; (*GEOCODING, UNITED STATES), *COMPUTER PROGRAMMING, | |
| DAT | 60252 | ICC/DF-73-003; *RAIL TRANSPORTATION, *REVENUE | |
| DAT | 60263 | BLM/DF-73-011; (*GEOCODING, TOWNSHIPS), *PUBLIC LAND, *LAND TITLES, * | |
| DAT | 60331 | BEA/DF-73-032; (*GEOCODING, UNITED STATES), *COMMODITY MANAGEMENT, | |
| DAT | 60344 | NARS/DF-73-013; (*GEOCODING, MUNICIPALITIES), *AIR TRANSPORTATION, | |
| DAT | 60345 | NARS/DF-73-014; (*GEOCODING, MUNICIPALITIES), *AIR TRANSPORTATION, | |
| DAT | 60346 | NARS/DF-73-015; (*GEOCODING, MUNICIPALITIES), *AIR TRANSPORTATION, | |
| DAT | 60352 | NARS/DF-73-012; (*GEOCODING, MUNICIPALITIES), *AIR TRANSPORTATION, | |
| DAT | 60368 | BEA/DF-73-031; (*GEOCODING, CENSUS REGIONS), *INTERSTATE TRANSPORTATIO | |
| DAT | 60382 | DOT/DF-73-016; *AIRPORTS, *RAPID TRANSIT RAILWAYS, *AIR TRANSPORTATION | |
| DAT | 60386 | COM/DF-73-023; *TRANSPORTATION, *COMPUTER PROGRAMS, *COMPUTERIZED | |
| DAT | 60390 | DOTDF-73-009; (*GEOCODING, UNITED STATES), *MOTOR VEHICLE ACCIDENTS, | |
| DAT | 60398 | DOT/DF-73-015; *CARGO TRANSPORTATION, *ECONOMIC FORECASTING, | |
| DAT | 60401 | DOT/DF-73-012; *CARGO TRANSPORTATION, *ECONOMIC MODELS, *INTERNATIONAL | |
| DAT | 60403 | DOT/DF-73-014; *CARGO TRANSPORTATION, *ECONOMIC FORECASTING, | |
| DAT | 60404 | DOT/DF-73-013; *CARGO TRANSPORTATION, *ECONOMIC FORECASTING, | |
| DAT | 60405 | BEA/DF-73-030; (*GEOCODING, UNITED STATES), *INTERSTATE TRANSPORTATION | |
| DAT | 60407 | CEN/DF-73-027; (*GEOCODING, UNITED STATES), *INTERNATIONAL TRADE, | |
| DAT | 60417 | DOT/DF-73-017; *AIRPORTS, *AIR TRANSPORTATION, *PASSENGER TRANSPORTATI | |
| DAT | 60468 | NARS/DF-73-011; (*GEOCODING, MUNICIPALITIES), *AIR TRANSPORTATION, | |
| DAT | 60492 | NARS/DF-73-019; (*GEOCODING, METROPOLITAN AREA ZONES), *URBAN | |
| DAT | 60493 | NARS/DF-73-018; (*GEOCODING, METROPOLITAN AREA ZONES), *URBAN | |
| DAT | 60494 | NARS/DF-73-017; (*GEOCODING, METROPOLITAN AREA ZONES), *URBAN | |
| DAT | 60495 | NARS/DF-73-016; (*GEOCODING, METROPOLITAN AREA ZONES), *URBAN | |
| DAT | 60506 | DOT/DF-73-029; *AIR TRANSPORTATION, *SHIPPING(MARINE), *HIGHWAY | |
| DAT | 60514 | COM/DF-73-022; *TRANSPORTATION, *NETWORKS, *CLIMATE, *LAND USE, | |
| DAT | 70065 | DOCBOC13; COMMODITY TRANSPORTATION SURVEY; PETROLEUM (PETR); | |
| DAT | 70173 | TRANSPORTATION RESEARCH INFORMATION SYSTEM (TRIS) | |
| DAT | 70180 | DOT6; ENERGY R & D FOR MASS TRANSPORTATION ADMINISTRATION URBAN MASS | |
| DAT | 70271 | NSF21; ENERGY AND FUEL TRANSPORTATION; ELECTRICITY (ELEC); ENERGY | |
| DAT | 20656000 | OPERATION AND MAINTENANCE, MONTHLY OPERATING REPORTS, N.Y.S. | TRANSURANICS |
| DAT | 56207 | ENVIRONMENTAL ASPECTS OF THE TRANSURANICS ENVIRONMENTAL ASPECTS OF | TRANSURANICS ENVIRON |
| DAT | 56207 | ENVIRONMENTAL ASPECTS OF THE TRANSURANICS ENVIRONMENTAL ASPECTS OF | TRAPPED SABLE FISH |
| DAT | 22669000 | STUDY OF TRAPPED SABLE FISH IN MONTEREY BAY; NORTH PACIFIC OCEAN, U.S | TRAUMA SURVEY |
| DAT | 56129 | FSMTS FEDERAL SERVICES MAXILLOFACIAL TRAUMA SURVEY; REPORTING | TRAVEL SURVEY |
| DAT | 70066 | DOCBOC14; NATIONAL TRAVEL SURVEY; ENERGY RELATED (ENRL) | |

Figure 3-18 Sample Extract from National Index of Energy and Environmental Data Bases and Models

MASTER CONTROL ERDA/OBER NATION-WIDE SURVEY OF ENERGY AND ENVIRONMENTALLY RELATED DATA BASES AND MODELS. JUNE 1979.

DATA BASE NAME.....: INVESTIGATION OF BIOLOGICAL CONDITIONS IN MILLON RUN AND PORTIONS OF BELLEVILLE LAKE 82179000

ORGANIZATION..: MI DEPT NAT RES

ABSTRACT:
DATA FILE CONTAINS DISSOLVED OXYGEN AND TEMPERATURE MEASUREMENTS IN THE WATER COLUMN AND SPECIES IDENTIFICATIONS AND COUNTS OF BENTHIC ANIMALS COLLECTED AT 30 STATIONS IN MILLON RUN AND BELLEVILLE LAKE, WAYNE AND WASHTENAW COUNTIES, MICHIGAN. MICHIGAN DEPARTMENT OF NATURAL RESOURCES CONDUCTED THE SURVEY TO EVALUATE EXISTING WATER QUALITY CONDITIONS.

DATES OF DATA.....: 1980107 TO 1980107
GEOGRAPHIC DESCRIPTION..: NORTH AMERICA, U.S., MICHIGAN, WAYNE AND WASHTENAW COUNTIES, MILLON RUN AND BELLEVILLE LAKE

ADDRESS OF CONTACT.....:
ELVIN D. EVANS, AQUATIC BIOLOGIST
WATER RESOURCES COMMISSION, MICHIGAN DEPARTMENT OF NATURAL RESOURCES
STEVENS T. MASON BUILDING
LANSING MICHIGAN
USA 48926

AVAILABILITY.....: ON SITE
STORAGE MEDIUM.....: REPORTS
FILE SIZE.....: 60 BIOLOGICAL AND 60 PHYSICAL OBSERVATIONS ON 9 PAGES OF REPORT
MEASUREMENT STATION....: SHIP
FUNDING OR SPONSOR.....: MICHIGAN
MEASUREMENT LOCATIONS..: 740822
LOCATION UNITS..: WORLD METEOROLOGICAL ORGANIZATION (WMO)
REFERENCES.....: C.M. FETTEROLF, MICHIGAN DEPARTMENT OF NATURAL RESOURCES, BUREAU OF WATER MANAGEMENT, WATER QUALITY APPRAISAL SECTION, BIOLOGICAL SURVEY, REPORT NO. 00000, 1984, 14 P.
REMARKS.....: SAMPLING OCCURRED 8/24-8/28/84 AND 7/1-7/2/84

PARAMETERS MEASURED IN DATA BASE

| PARAMETER | SPHERE | UNITS | AMOUNT | FREQUENCY | METHOD | TECHNIQUE AND REMARKS |
|---|---|---|---|---|---|---|
| TIME | EARTH | YMDL | 9 | | STATION TIME | OBS |
| POSITION | EARTH | DESCRIPTIVE TERMS | 30 | | FIXED POINT | STATIONS<br>REMARKS: MAP |
| BATHYMETRY | WATER | FEET TO TENTHS | 81 | | LEAD LINE | OBS |



| | | | | | | |
|---|---|---|---|---|---|---|
| DEPTH | WATER | FEET TO TENTHS | 63 | | WIRE LENGTH | OBS |
| DISSOLVED OXYGEN | WATER | MG/L | 25 | | TITRATION | OBS |
| TEMPERATURE | WATER | DEG C | 25 | | N | OBS |
| SPECIES DETERMINATION OF BENTHIC ANIMALS | BOTTOM | | 30 | | KEY | OBS<br>REMARKS: MACROINVERTEBRATES COLLECTED |
| COUNT OF BENTHIC ANIMALS | BOTTOM | ORGANISMS OF A SPECIES/SQ FT | 30 | | VISUAL | OBS<br>REMARKS: MACROINVERTEBRATES COLLECTED |

Figure 3-19. Sample Abstract from National Index

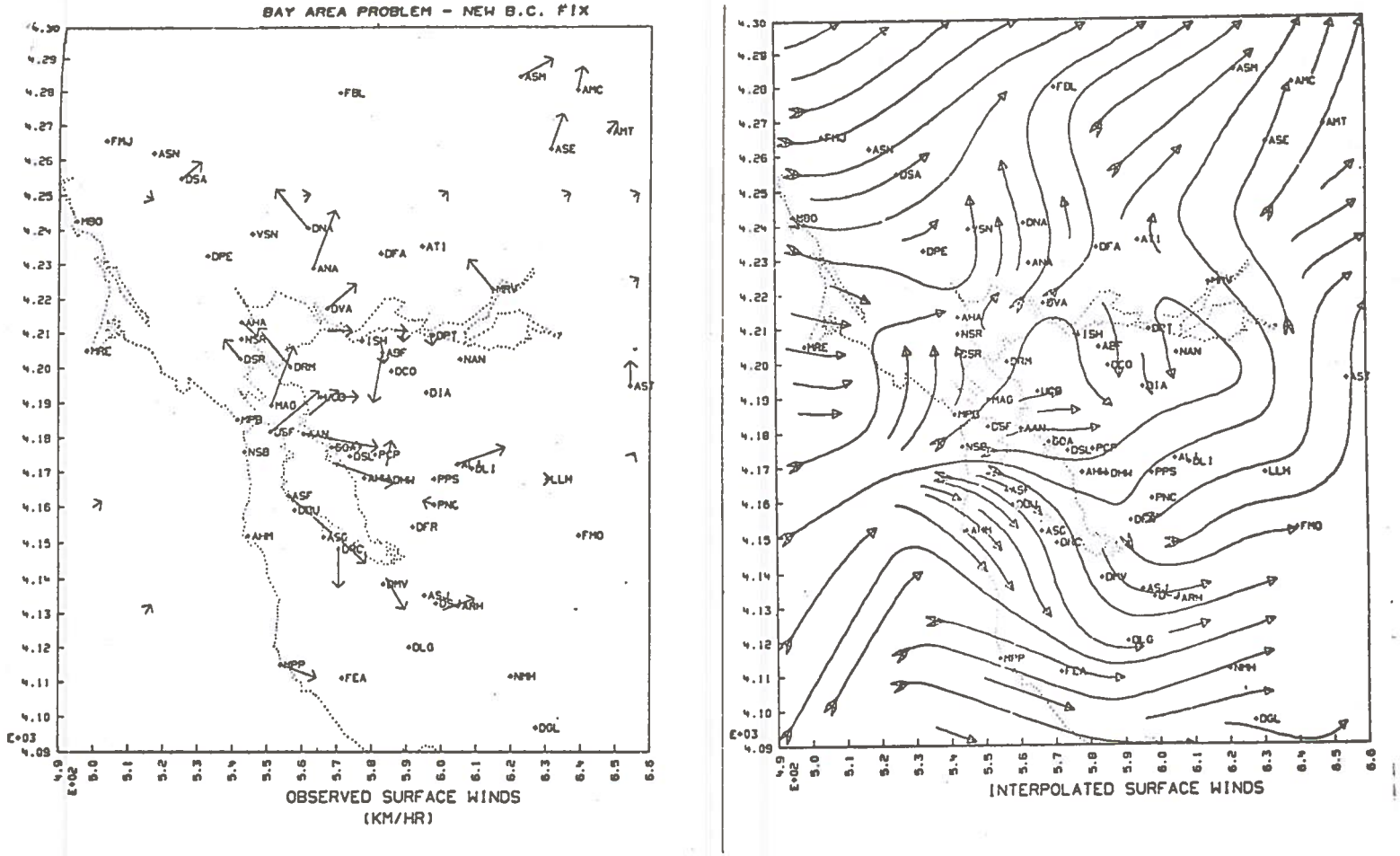



Figure 3-20 Comparison of Observed and Interpolated Surface Winds From the San Francisco Bay Area Pollution Model, Employing a MASTER CONTROL Data Base

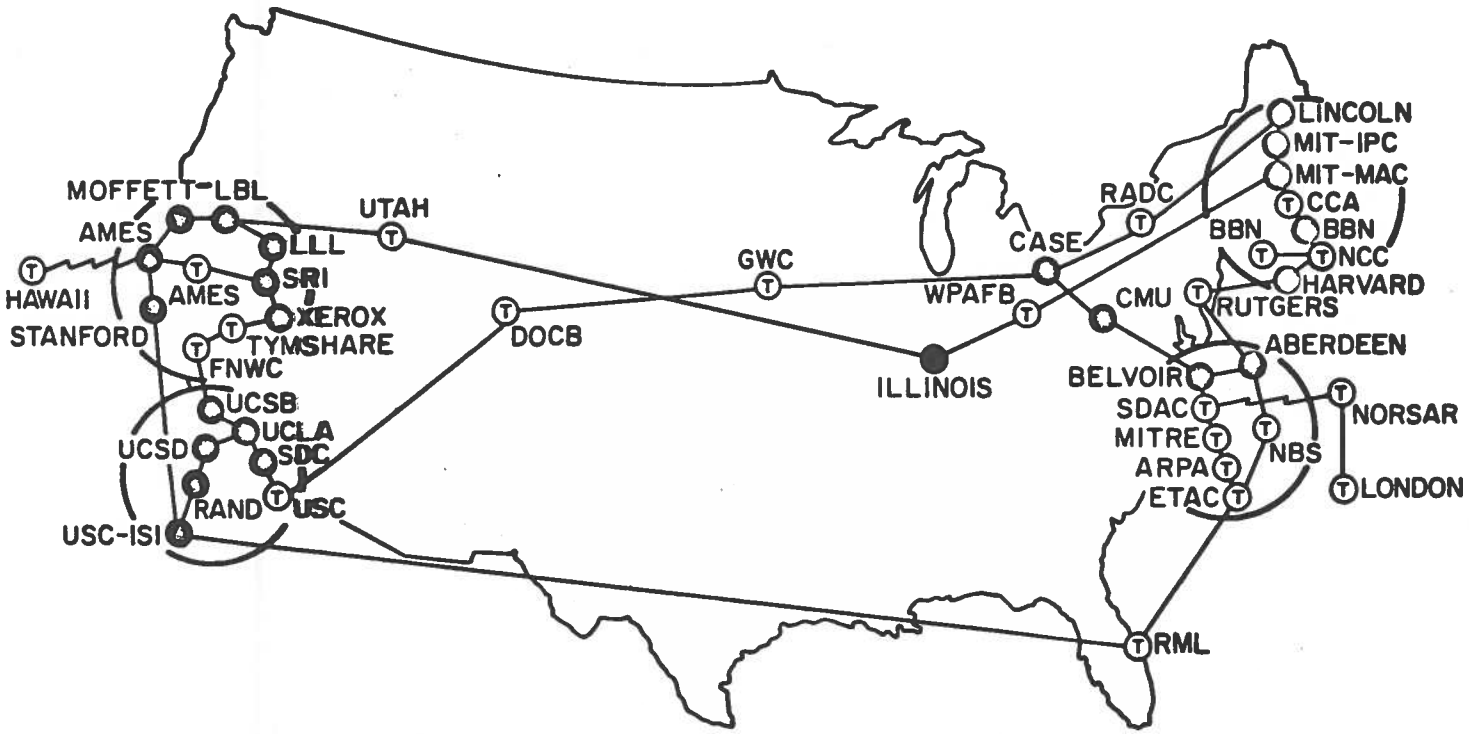



Figure 3-21 ARPA Network, Geographic Map, January 1974

# 4. INTERNAL DOT GENERATED DATA

## 4.1 BACKGROUND

A tremendous amount of statistical data is reported to, or collected by, the modal administrations and other organizations of the DOT. Some of the data are collected pursuant to statutory mandates placed on the respective organizations by Congress. Most of the data serve as the basis for policy formulation and decision-making within the DOT. All of the data sets involved were undoubtedly originally conceived for a specific purpose, most likely to fulfill a monitoring function. However, as the need for vital statistics has increased over the years, these specific-purpose data sets have found increasing use in more general transportation planning and research studies. This section provides a brief description of some of the more pertinent data bases collected and maintained by the modal administrations.

## 4.2 UNITED STATES COAST GUARD (USCG)

### 4.2.1 Motorboat Accident Statistics (MBA)

These data provide for annual statistical summaries on motorboat accidents. They are compiled and photostatically copied to produce the "Coast Guard Boating Statistics" (CG-357) as required by the Federal Boat Safety Act of 1971.

The principal data elements contained in this system are case number, date, state, country, cause, fatalities, injuries, operator age, type vessels and time. Information is maintained in both machine-readable form and in documentation files.

There are thirty-two annual reports generated by this system which categorize boating accidents according to cause, number, people involved, type of vessels involved, money involved, place accident occurred and jurisdiction. (There are no query capabilities.)

### 4.2.2 Merchant Vessel Documentation System

These data provide the Merchant Vessel Documentation Division with the capability to update, maintain and publish the Merchant Vessel Register (CG-408) and to integrate the information into the Merchant Marine Information System.

The principal data elements contained in the system are official register number, vessel name, vessel description data and owner data. Information is maintained in both machine-readable form and in documentation files.

A monthly Merchant Vessel Statistics by Tonnage, an annual Merchant Vessel Statistics by Tonnage Class, an annual Merchant Vessel Register (CG-409), and various processing reports are generated from this system. (There are no query capabilities.)

### 4.2.3 Merchant Seaman Locator System

These data permit the editing and updating of files on Active Seaman, Seaman Reference and Wanted Seaman based on information received from Shipping Articles and Wanted Seaman transactions.

The principal data elements contained in this system are: surname, identification number and official vessel number of each merchant seaman on active voyages; vessel data on merchant ships on active voyages; names of merchant seaman wanted by the USCG or other law enforcement agencies; and supporting personnel data on wanted seaman.

Five daily, four weekly and one monthly reports are generated to show the new seamen wanted in relation to active files, completed voyages, vessel name and locator lists. There is no query capability.

### 4.2.4 Pollution Incident Reporting System II (PIRS)

The Pollution Incident Reporting System II was developed for the Marine Environmental Protection Staff to generate a data base of pollution incidents, responses, and enforcement data.

The principal types of data elements contained in this system are types of pollution incidents, types of responses to these incidents and enforcement data. Examples of these elements are: cases closed without administrative action, cases with Coast Guard action taken or pending, and number of discharges by district each month.

Monthly, quarterly, annual and special-request reports are produced for waterbody reports; number of discharges by district versus month; material reports; reports concerning cases closed, action taken, cases pending, a carried-forward source versus cause reports; system response reports, and performance evaluation reports. There is no query capability.

### 4.2.5 Standardized Aid to Navigation Data Systems (SANDS) - Phase (Work Reporting)

The primary objectives of the new system are to standardize aids-to-navigation equipment terminology, increase the amount of servicing data collected, and reduce the data entry effort on the part of the servicing units.

The principal data elements of this system include on-station inventory of aids-to-navigation equipment, specifications for aid station equipment, aid station support data and historical data of types of services and equipment failures.

The reports produced by this system are: SANDS processing errors, current aid status report, SANDS component equipment listing, work report form, aid/unit assignment listing and unit work projections. Reports are produced daily, weekly, quarterly and monthly. There is no query capability.

### 4.2.6 Operating Facilities Publications System (OPFAC)

This file permits the production of semi-annual reports containing name, geographical location, city/state, OPFAC number, District/OPFAC remarks and missions for USCG boats and aircraft.

The principal data elements contained in this system are: OPFAC name, number, location and missions; boat and aircraft allowances; cutter specifications.

The system produces semi-annual reports of operating facilities of the USCG and a register of USCG cutters.

## 4.3 FEDERAL AVIATION ADMINISTRATION (FAA)

### 4.3.1 Accident, Incident, Violation Information

The basic data are used to identify and monitor airmen involved in accident and/or airmen who have violated Federal Air Regulations. Analyzed to determine facts surrounding accidents and incidents, its use includes determining if FARs have been violated and supporting aviation safety programs in general.

The information is obtained from FAA inspectors, pilots and other crew members, ground crews, passengers and witnesses.

Several periodic publications result from the information. They are Aviation Accident Statistical Summary-General Aviation, Air Taxi Accident Study, Enforcemenet Activity - Air Carrier, Manufacturer, Military (violations), General Aviation Enforcement Activity (violations) and Air Carrier Enforcement History - 5 years (violations).

### 4.3.2 Aircraft Information

The objective is to provide statistical and operational information about the characteristics, ownership and operation of aircraft comprising the U.S. Civil Aircraft Fleet. A secondary objective is to support aviation safety analysis.

Information is derived from aircraft owners, manufactures, operators, banks, etc. and the FAA Flight Standards Inspectors and individuals designated by the FAA to perform certain (inspection and maintenance) functions.

Several periodic reports are published. They are summarized or are part of the Aircraft Registry, Census of U.S. Civil Aircraft, and FAA Statistical Handbook of Aviation.

4.3.3 Airman (Non-Medical) Information

The information provides management and operational statistics relative to certificated airmen. The source also supports aviation safety analysis.

The basic information is developed in the airmen certification process and is obtained from the individual's application for certificate. Surveillance data on the airman and operations is obtained by the local district office during the course of inspections, investigations, etc.

Data are periodically available in the Airman's Directory or as part of the FAA Statistical Handbook of Aviation and U.S. Civil Airmen Statistics.

4.3.4 Aviation Activity Information

The objective is to provide aviation activity information for use in planning, budgeting, and staffing for agency programs and for publishing statistical results.

The data results from the monthly activity reports from each operational FAA facility, from the annual aircraft registration revalidations, and from the enplanement data collection efforts of the Civil Aeronautics Board, the Immigration and Naturalization Service, and the FAA survey.

The data are available in machine readable form as well as in periodical statistical publications like Air Traffic Activity, Military Air Traffic Activity, FAA Statistical Handbook of Aviation, Census of U.S. Civil Aircraft, Airport Activity Statistics of Certificated Route Air Carriers, Commuter Air Carrier Operators, and U.S. Civil Airmen Statistics.

4.3.5 Airports Information

The objective is to develop information to support the Federal involvement in the identification and planning of a National System of Airports, to support through the grants-in-aid airport master planning, development, and operations, to assist the airport ground safety including the Airport Certification Program, and to assure compliance by recipients of Federal airport assistance funds with the conditions of that assistance.

The basic data are derived from a variety of operational activities within the FAA and distributed to a number of offices and administrative activities of the FAA.

The basic periodic publications that result are an annual report of oeprations under the Airport and Airway Development Action, and annual report to Congress, the National Airport System Plan, and Airport Master Records.

## 4.4 FEDERAL HIGHWAY ADMINISTRATION (FHWA)

### 4.4.1 Motor Carrier Accident Reports

All interstate commercial motor carriers subject to the Federal Motor Carrier Safety Regulations report highway accidents in which they are involved in accordance with the requirements of 49 CFR 394. The data from the accident reports are used to create information to refine, update or develop new regulations and to develop implementing programs that promise the best results in terms of eliminating the hazards to safe highway transportation and the prevention of accidents and reduction of their consequences.

A prescribed form is completed by motor carriers showing various kinds of information about the accident. The form has 31 blocks of data with an estimated possible 200 different characteristics.

This data is automatically processed by computer, an annual report is prepared and publicly distributed as well as certain special reports, as needed.

### 4.4.2 Highway Statistics, Annual Publication

The data submitted for this publication permit FHWA to provide statistics on all phases of highway programs at local, state, and Federal levels for planning and administration at all levels.

The data are the basis for construction of annual time series on highway construction, existing mileage, motor vehicles, licensed drivers, motor fuel consumption, travel characteristics, speed trends, construction materials and costs, tax revenues by source and expenditures by purpose for each activity and level of government. Nearly all data are by state, with financial data by political subdivisions within SMSA's and by population groups.

This Highway Statistics publication provides, in a single annual document, data and information used in planning and policy studies, congressional reports on legislative issues, and preparation of testimony both within and outside of government at all levels.

### 4.4.3 National Highway Needs

The objectives of the reporting process are to:

a) Inform the Congress of the extent, condition, and performance of the existing highway network and future highway needs.

b) Develop policy on the direction of the Federal highway program and assess the impacts of different highway investment programs.

The data collected describe the miles of highways by functional classification; physical condition of roadways; highway performance characteristics; highway finance data; current and forecasted travel by highway system and area; and the cost and location of current and future needed highway improvements.

Several models are employed to produce statistics from data in this system:

1) Transportation Resource Allocation Study (TRANS)--assesses the impacts of national transportation investment strategies, mainly in urban areas.

2) National Highway Network Model--assesses the impacts of socio-economic changes on intercity highways.

3) Pardee Model--assesses the environmental impacts of national highway investment strategies.

4) Highway needs sensitivity models--test how highway needs are affected by such variables as performance standards and travel.

5) Highway needs simulation models--simulates highway investment requirements for given costs, performance standards, and travel forecasts.

6) Harris model--evaluates regional economic and environmental impacts of alternative highway systems.

The basic reports produced are:

a) Biennial highway needs reports to Congress. These contain descriptions of existing highway system and performance characteristics, current and future highway needs, benefit/cost analysis, environmental impact and highway finance analyses.

b) Federal Highway Administration policy and proposed legislation.

c) Special reports on highway needs, impacts of alternative highway investment, and various other aspects of the highway program.

## 4.5 FEDERAL RAILROAD ADMINISTRATION (FRA)

### 4.5.1 Waybill Statistics System

The input to this system is compiled from a 1% sample of audited revenue Waybills submitted to the FRA under the terms of the ICC order 49.C.F.R. SEC 1244.

Data are used in Traffic Flow Studies, Commodity Movement Studies, ICC Rate Cases, Revenue Studies, Safety Analyses, and for input to the FRA Railroad Network Model.

Principal data Elements are: Serial & Waybill Numbers, Number of Carloads, Origin & Destination Railroad and Stations,

Rate Types, STCC, Wright & Revenue, Short Line Miles, AAR Car Type, Tons.

The data base is maintained in machine-readable form by the FRA.

"Carload Waybill Statistics - Territorial Distribution-Traffic and Revenue by Commodity Classes" reports are published annually.

System access is through "EASYTRIEVE." Special runs, other than the report cited above, must be approved by Bureau of Economics, Interstate Commerce Commission, Washington, D.C.

### 4.5.2 Accident Incident Reporting System

The Office of the Associate Administrator for Safety of the Federal Railroad Administration (FRA) has the responsibility of receiving, processing and reporting railroad accident/incident statistics.

The purpose of the railroads reporting to the FRA the occupational illness of employees, damage to railroad equipment and structures, and injury to persons (arising from the operation of a railroad) is to carry out the intent of Congress as expressed in the Federal Railroad Safety Act of 1970 and the Accident Reports Act, as amended.

The system allows the Office of the Associate Administrator for Safety to maintain and retrieve Accident/Incident data filed by the railroads.

Accident/Incident data is received on the following forms which are basic inputs to the system.

| Report No. | Report Title | Reporting Frequency |
|---|---|---|
| FRA F6180-54 | Rail Equipment Incident Report | Annually |
| FRA F6180-55 | Railroad Illness and Injury Summary | Monthly |

| Report No. | Report Title | Reporting Frequency |
|---|---|---|
| FRA F6180-57 | Highway Grade Crossing Incident Report | Annually |
| ICC Wage Statistics Form A | Monthly Report of Employees, Service and Compensation | Monthly |
| ICC Bureau of Accounts Form DS-B | Revenue Traffic | Quarterly |
| ICC Bureau of Accounts Form DS-A | Train and Yard Service | Quarterly |
| FRA F6180-56 | Annual Railroad Reports of Manhours by State | Annually |

Principal data elements are: Casualty information, damage costs, location of accident, and train speed, weather, and grade-crossing information.

All Accident/Incident data is maintained in machine-readable format on magnetic disk storage.

The basic statistical reports produced are:

| | Reports Title | Frequency |
|---|---|---|
| 1) | Accident Bulletin publication | Annually |
| 2) | Rail/Highway Grade-Crossing Accidents publication | Annually |
| 3) | Preliminary Report of Railroad Accidents/Incidents and Resulting Casualties, Form FRA F6180-27 | Monthly |
| 4) | Summary of Accidents/Incidents Reported by all Line-Haul and Switching and Terminal Railroad Companies | Monthly |

### 4.5.3 Grade-Crossing Inventory System

This nationwide project for numbering and inventorying Highway Grade-Crossings will be complete in late FY76. The data will be used to isolate apparent accident-contributing characteristics

and to determine cost/benefit ratios for alternative upgrades. The data base will exceed 430,000 records and will contain all pertinent information for each grade-crossing. Information will be maintained in machine-readable form.

The "Grade Crossing Accidents" Bulletin will include inventory data at a later date. The system is being developed using a DBm language entitled INQUIRE.

### 4.5.4 Railroad Locomotive Inspection

These data are collected to maintain records or results of field locomotive inspections. Sources of the data are locomotive inspection records. The reports generated from the data include inspection compliance, locomotive inventory, and FRA inspections.

### 4.5.5 FRA Safety Inspections

These data are collected to compile general statistics on inspection. Data are collected from inspection information reported on Standard Form 1680.1. The reports generated from the data include inspection by railroad, inspection by inspector, and defect matrices.

### 4.5.6 Track Inspection System

This system allows the FRA to maintain and retrieve information pertaining to Track Inspection Reports filed by the FRA and State inspectors.

There are three basic inputs to the track system:

a) Track Inspection Report - FRA 6180-58(A) - filled out by inspectors, describing any track defects.

b) Track Inspection Report FRA 6180-58(A) - copy of the inspectors' report which is given to the railroad. The railroad amends their copy to reflect how the track defect was corrected.

c) Chief Counsel Report - describes claim settlements.

All information is maintained in machine-readable format on magnetic disk storage.

Data elements include: FRA Inspector code, fiscal year, region, railroad, railroad sub-division, violation, location (city, state), track number, mile post of inspection, defect code.

The basic reports generated are:

| | Reports | Frequency |
|---|---|---|
| 1) | Inspectors' Activity Detail Report | Monthly |
| 2) | Inspectors' Activity by State<br>Inspectors' Activity by Railroad<br>Inspectors' Activity by Defects<br>Inspectors' Activity by Region | Monthly |
| 3) | Uncorrected Defect Report | Monthly |
| 4) | Violation Status Report | Monthly |
| 5) | Violation Settlement Report | Monthly |

The track system data is maintained in a Data Base Management format for instantaneous querying using a DBM language entitled INQUIRE.

## 4.6 NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION (NHTSA)

### 4.6.1 National Accident Summary System (NAS/UADT)

Each state submits yearly raw accident data tape (RADT) files with no common record or file structure uniformity. Computer programs are then written/maintained to convert the data to the uniform accident data tape (UADT) format, as specified by NHTSA.

Computer-supported generation and processing of magnetic tape files permits updates to four magnetic tape files (yearly).

Special study reports produced are: Cross tabulations in a variety of specific formats (on request).

### 4.6.2 Fatality Analysis File

The data collected for this file permit analysis of automotive and pedestrain fatalities. The data are computer maintained and operated to produce an analysis report. Output reports are being developed. No query capabilities exist.

### 4.6.3 HSRI Accident Data System

The data collected in this system permit numerous analyses of auto accident causes and related statistics. The data system uses a large integrated computer system (with data bases derived from police data on accidents, as well as data collected by numerous investigation teams) to produce reports. Six analysis subroutines are utilized to generate reports 50-60 times per month.

### 4.6.4 Vehicle-in-Use Questionnaire Processing System

In this system, the data are collected to determine if the type of state motor vehicle inspection has a significant effect on the condition of motor vehicles in the state, if inspection of a few safety-related components affect the overall upkeep of the vehicle, and if differences exist between states in the outage rate of individual vehicle components.

The data files permit analysis of defect counts and other data for each vehicle and its driver for the purpose of developing relationships between defect counts and other data related to the vehicle inspected. It is a computer-maintained and operated system.

Special reports are produced on request.

No query capabilities exist.

### 4.6.5 Vehicle Equipment and Manufacturers Identification System (VEMIS)

The data collected in this system permit identification of vehicle equipment for recall. Data concerning vehicle equipment, manufacturer identification, and automotive safety is taken from input supplied by manufacturers or test laboratories.

The system is computer-maintained and operated and "ad-hoc" reports are generated as required.

No query capabilities exist.

### 4.6.6 Tire Identification and Record Keeping Statistical Applications System

In this system, the data are collected to determine the degree of tire registration effectiveness.

Tire registration information is taken from Compliance Tire Buy Reports (CTBR's) and is used as input into a computerized record system. The data are edited, with valid data passed to the update program and stored. The system contains a generalized retrieval package that can search, correlate and point to any or all data fields. This retrieval package also provides a means for simple calculations (e.g., percentages involved in owner notifications and replaced tires).

The system generates four output reports, as needed:

a) Type visits by Region

b) Type dealer by Region

c) Chi-Square contingency table analysis

d) Significance levels

### 4.6.7 Evaluation Data System for Office of Alcohol Countermeasures

Tabular and statistical data are collected concerning alcohol-related highway incidents. Reports from ASAP serve as input. Two files are maintained and operated by a computer system. Upon request, 50 fixed evaluation reports can be generated.

The system provides general query capabilities (upon request).

### 4.6.8 Motor Vehicle Import Information (MVII) System

System data are collected to keep track of autos entering the U.S. from abroad which are not in compliance with U.S.

standards. Such autos must be corrected within 60 days. The system also generates warning letters to customers.

The files are computer-maintained and operated in batch and on-line modes to produce monthly update of PCI and FCI files.

Basic reports generated using this system are:

a) Report of Preliminary Customs Investigation (PCI)

b) Report of Import Certification Inquiry (ICI)

c) Reports responsive to specific queries.

### 4.6.9 Office of Standards Enforcement Vehicle Selection Matrix (OSE VSM)

Historical data relevant to actual motor vehicle accidents, prior compliance test results, current vehicle owner complaints, and supplementary engineering data (for use in ranking future selection of vehicles for compliance testing) are collected.

The basic system contains a vehicle-selection summary file, an accident data file, CONTAB, a vehicle master file, a sorted priority list, and a buy list.

The system is computer operated and maintained.

Basic reports which can be produced by this system are:

a) Priority list (quarterly)

b) Selection matrix (quarterly)

c) Buy list (prior to commencing FY testing)

There are no query capabilities.

### 4.6.10 VIN Information System (VIN)

This system provides a means of building and maintaining a large-volume data base and an efficient, cost-effective method of extracting and reporting selected elements, as required by users.

The system contains an unrepaired list of VIN's, all query information, query volume data, and query requests.

The system is computer operated and maintained.

Basic reports generated by the system are:

a) VIN Query Status Report (monthly)

b) QREPT (daily)

c) VIN Query Report (daily)

d) VIN Micho-fiche data base (quarterly)

e) VIN Tape data base (quarterly)

The system has capability to handle queries.

## 4.7 URBAN MASS TRANSPORTATION ADMINISTRATION (UMTA)

### Project FARE (Section 15 Reporting System)

The development of the Financial and Accounting Reporting Elements system is currently underway at UMTA. This system, when fully operational, will collect data from transit companies. Work has already been completed on defining the elements to be included in the system. Financial data such as revenues, operating and overhead costs, capital costs, advertising and market costs and operations data such as fleet size, route miles, passengers carried, etc., will be collected.

Presently similar data is collected on a voluntary basis by the American Public Transit Association. However, Section 15 of Public Law 93-503 requires that the Secretary of Transportation collect this data from transit properties. Filing the data will be mandatory as UMTA grants will be denied properties that do not supply the information requested. The system is expected to be operational in about one year.

## 4.8 ST. LAWRENCE SEAWAY DEVELOPMENT CORPORATION (SLSDC)

### Great Lakes - St. Lawrence Seaway System

This system reports vital statistics on the movement of ships through the St. Lawrence locks and the Welland Canal. Data reported

include size and type of vessel, origin, destination (including each eventual port-of-call), and cargo on and off-loaded by commodity.

Use of the data for research purposes is permitted, though some elements of the data set are proprietary in nature. Data on tapes are available from 1964 and in hard copy form from 1969.

## 4.9 MATERIALS TRANSPORTATION BOARD

### 4.9.1 Hazardous Materials Incident Reporting System

In this system, data are collected to produce statistics on the unintentional release of hazardous materials during transportation. This information is compiled in accordance with the requirement levied in the Transportation Safety Act of 1974, Public Law 93-633.

The current system contains information on each reported incident and includes data elements such as the date of the untentional release of hazardous materials, the location, the shipper and carrier, the commodity involved and other information concerning the packaging and nature of the incident.

Periodic outputs are generated which are primarily used within the Office of Hazardous Materials Transportation Operations. However, requests by other agencies, individuals, companies, etc., are filled by the office.

The information produced provides a basis for highlighting problem areas, for pinpointing the need for corrective action, and for providing inputs to the annual report on hazardous materials control (also required by Public Law 93-633).

### 4.9.2 Pipeline Carrier Accident Report, DOT Form 7000-1

In this system, data are collected to support the reporting of failures in liquid pipeline systems when the release of the commodity transported results in any of the following:

a) Explosion or fire not intentionally set by carrier.

b) Loss of 50 or more barrels of liquid.

c) Escape to the atmosphere of more than 5 barrels a day of liquefied petroleum gas or other liquefied gas.

d) Death of any person.

e) Bodily harm to any person resulting in one or more of the following:

   1) Loss of consciousness

   2) Necessity to carry person from the scene

   3) Necessity for medical treatment

   4) Disability which prevents the discharge of normal duties or the pursuit of normal activities beyond the day of the accident

f) Property damage of at least $1,000 to other than the carrier's facilities, based upon actual cost or reliable estimates.

Pipeline Carrier Accident Report, DOT Form 7000-1, contains information as listed below:

1) Name and address of carrier

2) Date, time, and location of accident; part of carrier system involved and physical location

3) Origin of liquid or vapor release

4) Cause of accident

5) Fatalities and injuries of the carrier's employees and non-employees

6) Property damage - items and dollar value

7) Commodity being transported - estimated loss in barrels, year facility installed, and whether there was a fire or explosion.

8) Description of line pipe in detail

9) Explantion, if caused by corrosion

10) Explanation, if caused by equipment rupturing the line

11) Name of company official and title, carrier telephone number and date filed

An Annual Summary of Liquid Pipeline Accidents is reported to the Department of Transportation on DOT Form 7000-1 for calendar year. The report provides the following:

a) Total number of accidents, fatalities, and injuries

b) Total property damage, total carrier damage, and other damage

c) Barrels of commodity lost

d) Total number of accidents, by year, of installation

e) Summary of accidents by commodity, by cause - corrosion; equipment rupturing the line; and by defective pipe seam

f) Summary of accidents by location, state

### 4.9.3 Office of Pipeline Safety Operations Automated Leak and Test Failure Reporting System

The data collected in this system provide causal-related and safety-related information for use in identifying trends and problem areas and for supporting rule-making actions and safety program development. This information is also used to furnish Congress with accident and causality data in accordance with Section 14 of the Natural Gas Pipeline Safety Act.

The system can responsd to inquiry in accordance with specific extraction programs for retrieving elements and combinations of elements. The system contains identification of companies nationwide, with distribution and/or transmission/gathering systems, sizes, materials and ages of gas systems, causes and number of leaks repaired, gas systems on coated or bare pipe, gas system cathodically protected, fatalities and injuries, operator property damage estimates, value of property damage to others (settled), fires and explosions, leak surveys and Cathodic Protection Inspections. Specific incident data shows detailed information such as: when and

where occurred, cause, pipe specifications, when installed, material that failed, fatalities, injuries, estimated operator damage, environment of incident, estimated pressure at time of incident, maximum allowable operating pressure, time until escape of gas stopped, method of leak or failure detection, rupture, ignition, explosion, type of repair, and other utilities contributed or impaired.

Detailed information is included regarding failures due to corrosion, outside forces, construction defects, and material failure.

A quarterly report of pipeline leaks and test failures is printed. An annual summation of all data is printed displaying the occurrences related to the operators affected by the Natural Gas Pipeline Safety Act during the reporting year.

Information is furnished to industry, organizations and operators concerned with pipeline safety upon request.

## 4.10 TRANSPORTATION SYSTEMS CENTER (TSC) - DATA UTILIZATION PROGRAM

### 4.10.1 Air Industry Data Base System

This system provides an easy method to acquire statistics on either airline or route activity. The system enables one to observe the history of performance of either an air route, an airline, or the air industry.

The data, available in machine-readable form, provides a record for each month, for each flight segment of every commercially scheduled airline flight. It describes the originating and destination terminals, details the aircraft configuration, and reports upon utilization of the aircraft (i.e., number of passengers, amount of cargo carried, mail, etc.).

The system provides the basis for generating reports on activities between city pairs by frequency-of-flight, load factors and the number of people traveling. Additionally, airborne and ground delay reports are currently produced by the system for a match mode system which will become an on-line operation.

### 4.10.2 Railroad Industry Data Base

The purpose of this data base is to provide a means of reporting the operational status of the rail industry. The system enables one to observe historical trends and to make short-term estimates of the financial activities.

The data base contains a collection of items from the ICC reports submitted by the railroads concerning the carloading activities of each railroad. The data, available in machine-readable form, consist of a set of records on weekly carloadings for each railroad and 20 developed commercial groups. Additionally, a time-series-expansion-factor data stream permits observation of the factors associated with the carloadings, which, in turn, makes estimates of railroad activities possible.

This system can generate a series of weekly reports estimating railroad activities by financialand performance levels. The system operates in an on-line mode and query may be made by railroad commodity group.

### 4.10.3 Trucking Industry Data Base

The objective of this system is to provide the means to study the history of trucking organizations in the industry. The system enables one to develop time series of the many factors reported to the ICC.

The data, available in machine-readable form, is that which is reported in the quarterly ICC Form QFR and the Annual ICC Forms M-1, M-2, and M-3. These reports detail trucking activities from the corporate point of view. Additional inventory data, including both regulatory and non-regulatory trucking, is available.

The system is currently under development and report formats are not yet defined. The system will permit special reports to be generated through on-line query of the data.

### 4.10.4 Commodity Flow Data Base

The major objective of this activity is to develop a multimode, regional commodity flow data base for use in transportation

traffic and operations analysis. Besides identifying yearly freight-tonnage flows among regions, the data base will describe various shipment characteristics associated with modal movements.

Emphasis will be on constructing commodity-flow data for bulk or unprocessed commodities, and for given manufactured goods which are well-sampled in the CTS. When possible, county-level, regional detail should be maintained, although BEA and state designations should appear in the record to aid in comparisons to other data bases. At minimum, yearly regional tonnage must be reported and average shipment size by mode calculated when feasible. When possible, true (ultimate) origins and destinations of commodities should be identified. Four major modes will be considered; rail, waterborne (shallow and deep draft, separately), pipeline and truck. When available, all rate data on the mode shipment should be included (whether per ton, per ton-mile or per shipment).

Computer analysis of data within this data base will facilitate support to on-going research on transportation forecasting and planning.

### 4.10.5 Interstate Commerce Commission (ICC) Quarterly Data

The data of this data base are collected to enable monitoring of performance and financial results of Class I railroads, of Class I motor carriers (passengers and property), of Class I pipelines, and of Class A and Class B water carriers. The data is maintained in machine-readable form; quarterly and annual publications are provided in printed form.

Data extracted from reports filed with the Interstate Commerce Commission each quarter by the regulated common carriers are as follows:

Railroads:

Number of revenue passengers carried

Passengers carried one mile

Number of revenue tons carried

Tons carried one mile

Operating Revenues-freight, passenger

Operating Expenses and Depreciation-road, equipment

Net Revenue from Railway Operations

Federal Income Taxes

Pipelines:

Number of barrels of oil originated and received from connections

Motor Carriers of Passengers (Bus Lines):

Total Operating Revenues

Ordinary Income Before Income Taxes; Passengers carried; regular route, intercity schedules; regular route, local & suburban schedules; charter or special service

Carriers by Water:

Revenue Tons Carried

Tons Carried One Mile

Freight Revenue

Revenue Passengers Carried

Motor Carriers of Property (Truck Lines) (Collected for 1971-1974 and first quarter of 1975 only):

Ordinary income before income taxes

Tons carried-common carriage

Tons carried-contract carriage

Collected for 1974 and first quarter of 1975 only:

Net profit or loss

The Quarterly Summary of National Transportation System Activity (also includes data provided by FHWA and airline performance data) providing selected statistics from Quarterly Reports submitted to the ICC by Regulated Common Carriers (for use within TSC and TPI only) is published quarterly. This publication contains

data as reported by each carrier with totals for each item, by mode. Revised data is republished each year as an annual publication containing data reported by each carrier each quarter, as well as annual totals.

### 4.10.6 Automotive Data Base

The included data are collected to analyze the 1975 automobile fleet, in terms of vehicle production, fuel economy, emissions, engine characteristics, vehicle price, and performance. The 1975 Automotive Data Base contains 216 types of 1975 model year automobiles (both domestic and imported); approximately 50 attributes are used to describe each vehicle type. Model population is disaggregated by the 50 states and attributes covered include engine characteristics, emission system, transmission, fuel economy, emission level, performance, price, production, inventory, etc.

The data base is used to support quick reaction studies for the OST/TSC Automotive Energy Efficiency Program. The data base is located on the TSC's PDP-10 and is used in conjunction with the System 1022 Data Management System.

### 4.10.7 DOT/TSC Automotive Manufacturing Assessment System

In this system, data are collected to determine the manufacturing and maintenance characteristics of the automotive fleet. The data base contains configurations (broken down by components) for approximately 400 vehicles, including configurations for selected alternative engines (Stirling, gas turbine, Rankine, Wankel). The vehicles are classified by body size. For each vehicle, data collected and maintained are: materials breakdown, labor, manufacturing cost, sticker price, capital investment for tooling and facilities, and the cost of scheduled and unscheduled maintenance.

The outputs of the system are: manufacturing cost, maintenance cost, sticker price, materials breakdown by vehicle, capital investment, and feasible production schedule by years.

The version of the system in operation is an interim system based on the National Academy of Science's manufacturability data base and DOT improvements. The advanced version of the system is under development, and is due to be in operation by July 1976.

## 4.11 JOINT PROJECTS (OST AND OTHER DOT OR NON-DOT ORGANIZATIONS)

### 4.11.1 Journey-to-Work Continuous Survey (In Conjunction with FHWA, UMTA, HUD and Census)

This survey provides for continuous updates of the Journey-to-Work data developed initially by TPI, FHWA and the Census. The purpose of the survey (which supplements the National Annual Housing Survey (NAHS)) is to generate commuter-oriented travel information such as origin and destination locations, trip length and duration, modal selection, change in mode usage and satisfaction with means of transportation to work. Recent surveys have stressed the effects of the Nation's energy and economic dislocations, and of maturing public transportation policies, on travel behavior and attitudes.

### 4.11.2 National Travel Survey (with FHWA, UMTA, and Census)

This survey provides data which permits analytical capability to measure intercity travel demands for small geographic areas and any two selected cities. The program assesses traveler response to changes in service and supports service planning and route selection by carriers. The survey also permits data to identify the service needs by rural areas and small cities.

### 4.11.3 International Air Data Project (with CAB, U.S. Travel and Justice (INS).

See Section 5.2.

### 4.11.4 National Shipper Survey Expansion Project (with Census)

See Section 5.3.

### 4.11.5 Domestic and International Transportation of U.S. Foreign Trade (with COE, MARAD and Census)

Surveys to be performed within this study will identify the "true" origins and destinations of freight movements, the modes utilized (including intermodal movements), the size of shipment, and other characteristics of bulk goods moving in foreign trade. The program will develop a set of data describing bulk commodities for all modes for 1975 and will develop and test a method of obtaining data on a nationwide basis for a continuing transportation data program.

# 5. OST DATA COLLECTION

## 5.1 SUMMARY OF SECTION

This section reviews three efforts of the Assistant Secretary for Policy, Plans, and International Affairs in the area of data collection. These efforts represent the types of efforts historically pursued under the OST Information Program (See Appendix B). They are important because they develop data not otherwise available.

The three collection efforts (in order of discussion) are:

a) International Airline Passenger Data

b) Census of Transportation

c) Journey-to-Work Supplement to the Annual Housing Survey

The International Airlines Passenger Data project supports the analysis of passenger traffic on U.S. to/from foreign destinations by providing the only data that includes details on the foreign flag carriers.

The Expansion to the Census of Transportation (CTS) project encourages studies designed to open the scope of the CTS, one of the Economic Censuses performed each five years and funded directly by Congress.

The funding of the addition of journey-to-work questions provided by TPI to the U.S. Department of Housing and Urban Development's (HUD) Annual Housing Survey permits a pseudo-continuous review of the work-trip habits of the general public.

## 5.2 INTERNATIONAL AIRLINES PASSENGER DATA

### 5.2.1 Background

Beginning in the 1950's, the Office of Management and Budget (OMB) assigned responsibility to the United States Department of Justice, Immigration and Naturalization Service (INS), for the

collection and reporting of data on the entry/exit of citizens and non-citizens through U.S. ports and air-terminals. The data for these air movements are recorded on Form I-92. A blank version of Form I-92, as well as a completed one, are illustrated in Figure 5-1.

Over the last few years, a close working relationship has developed between INS and DOT because the raw data contains records for all carriers, including foreign flag and non-scheduled, down to the specific flight. There is no other data collected by any U.S. agency on foreign flag carriers at this level of detail.

Originally, INS did not keypunch the flight number data. However, because DOT was interested in detail at that level (in order to develop the comprehensive statistics needed for understanding the international air market), DOT and INS joined in a cooperative effort to convert the raw data and produce reports of interest to each party. INS agents collect the basic paper forms; DOT edits the forms, keypunches the data onto cards, and creates several tapes, one of which is returned to INS for its use in producing the required OMB report. DOT produces periodical statistical reports based on this data for consumption by the Civil Aeronautics Board, the Department of Commerce, and DOT itself.

The I-92 data collection and processing project is the only OST data collection program executed by DOT elements where the basic machine-readable data remains at DOT and is available for special processing in response to any additional problems that might arise relative to international air carriers.

### 5.2.2 Raw Data I-92 Forms

The I-92 Form carries the title Aircraft/Vessel Report and is approved by the Office of Management and Budget under OMB No. 43-120497. An I-92 Form is completed for each passenger-carrying flight which departs from, or arrives in, the United States the following types of transport:

a) Private

b) U.S. Military - including charters to military

**AIRCRAFT/VESSEL REPORT**

Form Approved
OMB No. 43-R0497

☐ ARRIVAL
Last Foreign Port _____

☐ DEPARTURE
First Foreign Port _____

| Airline/Vessel (Name and Nationality) | Flight Number | Port of Arr/Dep | Date of Arr/Dep |
|---|---|---|---|
| | | | |

TYPE OF TRANSPORT—CHECK ONE

1. ☐ U.S. military—including charters to military
2. ☐ Commercial—scheduled
3. ☐ Commercial—chartered
4. ☐ Foreign military

Total Passengers

| Do Not Write in These Blocks—For INS Use Only | | |
|---|---|---|
| Passengers Inspected | Passengers Deferred | Deferred Port |
| | | |

Attach CF 7507, ICAO Declaration, or I-418, or List Crew Below.

CREW: Name Status

| FOREIGN PORT AND COUNTRY | PASSENGERS USC | ALIEN | TOTAL |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| TOTAL | | | |

FORM I-92 (Rev. 5-1-72)N (See instructions on reverse of form)
United States Department of Justice
Immigration and Naturalization Service

a. Blank Form - Front

**INSTRUCTIONS FOR IDENTIFYING TYPE OF TRANSPORT**

1. **U.S. MILITARY TRANSPORT**
   U.S. military carriers, including carriers chartered by the military or any other U.S. Government agency, i.e., FAA.

2. **COMMERCIAL TRANSPORT—SCHEDULED**
   All carriers traveling with predetermined time and route schedules.

3. **COMMERCIAL TRANSPORT—CHARTERED**
   All carriers, other than military, without predetermined schedules. Include also "air taxis", or any other transport carrying persons for hire.

4. **FOREIGN MILITARY TRANSPORT**
   Foreign military carriers, including carriers chartered by the foreign military or any other foreign government agency.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

U.S. GOVERNMENT PRINTING OFFICE:1978—O-262-146

b. Blank Form - Back

**AIRCRAFT/VESSEL REPORT**

Form Approved
OMB No. 43-R0497

☐ ARRIVAL
Last Foreign Port _____

☒ DEPARTURE
First Foreign Port GENEVA

| Airline/Vessel (Name and Nationality) | Flight Number | Port of Arr/Dep | Date of Arr/Dep |
|---|---|---|---|
| TWA, USA | 830 | JFK | 29 JUN 75 |

TYPE OF TRANSPORT—CHECK ONE

1. ☐ U.S. military—including charters to military
2. ☒ Commercial—scheduled
3. ☐ Commercial—chartered
4. ☐ Foreign military

Total Passengers: 98

| Do Not Write in These Blocks—For INS Use Only | | |
|---|---|---|
| Passengers Inspected | Passengers Deferred | Deferred Port |
| | | |

Attach CF 7507, ICAO Declaration, or I-418, or List Crew Below.

CREW: Name Status

| FOREIGN PORT AND COUNTRY | PASSENGERS USC | ALIEN | TOTAL |
|---|---|---|---|
| GENEVA, SWITZERLAND | 43 | 4 | 47 |
| VIENNA, AUSTRIA | 44 | 7 | 51 |
| | | | |
| | | | |
| | | | |
| | | | |
| TOTAL | 87 | 11 | 98 |

FORM I-92 (Rev. 5-1-72)N (See instructions on reverse of form)
United States Department of Justice
Immigration and Naturalization Service

c. Completed Form

**AIRCRAFT/VESSEL REPORT** 208

Form Approved
OMB No. 43-R0497

☐ ARRIVAL
Last Foreign Port _____

☒ DEPARTURE
First Foreign Port GENEVA

| Airline/Vessel (Name and Nationality) | Flight Number | Port of Arr/Dep | Date of Arr/Dep |
|---|---|---|---|
| TWA, USA 5381 | 830 | JFK | 29 JUN 75 |

TYPE OF TRANSPORT—CHECK ONE TW

1. ☐ U.S. military—including charters to military
2. ☒ Commercial—scheduled
3. ☐ Commercial—chartered
4. ☐ Foreign military

Total Passengers: 98

| Do Not Write in These Blocks—For INS Use Only | | |
|---|---|---|
| Passengers Inspected | Passengers Deferred | Deferred Port |
| | | |

Attach CF 7507, ICAO Declaration, or I-418, or List Crew Below.

CREW: Name Status O

| FOREIGN PORT AND COUNTRY | PASSENGERS USC | ALIEN | TOTAL |
|---|---|---|---|
| GENEVA, SWITZERLAND 208 | 43 | 4 | 47 |
| VIENNA, AUSTRIA 639 | 44 | 7 | 51 |
| | | | |
| | | | |
| | | | |
| | | | |
| TOTAL | 87 | 11 | 98 |

FORM I-92 (Rev. 5-1-72)N (See instructions on reverse of form)
United States Department of Justice
Immigration and Naturalization Service

d. TSC-Edited Completed Form

Figure 5-1. I-92 Form: Aircraft/Vessel Report

c) Commercial - scheduled

d) Commercial - chartered

e) Foreign military

The form is designed to identify the carrier and to distinguish between crew and commercial passengers. It also lists passengers originating from, or destined to, intermediate stops for multi-stop flights.

During Calendar 1975 the Transportation Systems Center of DOT processed approximately three hundred thousand (300,000) I-92 Forms received from the Department of Justice, Immigration and Naturalization Service (INS).

### 5.2.3 Products

The reports which are produced by TSC from the approximately twenty-five thousand (25,000) monthly raw data forms contain internal air passenger travel information which is assembled by calendar month. Three major report elements are regularly available, specifically:

a) Summary-Travelers by Country

b) Port-to-Port

c) Travel Survey by Carrier, Flight, and Area

The report section entitled Summary-Travelers by Country is divided between arrivals and departures and lists (for each U.S. entry port and for the U.S. in total) the passengers traveling to/from each foreign country. The passengers are displayed as U.S. citizens, aliens, and totals.

The Port-to-Port report section tabulates arrivals and departures by city pair (U.S. port-foreign port) and the following information:

1) Carrier identification code

2) Flight number

3) Number of flights during month reported

4) Separation by scheduled/charter flight

5) Percent citizens by flight number for the month, for both scheduled and charter flights

The Travel Survey by Carrier (Flight and Area) section summarizes information on monthly totals of passengers traveling between U.S. and foreign ports disaggregated by several world areas including:

a) Middle America

b) Carribean

c) South America

d) Europe

e) Africa

f) Middle East

g) Far East

h) South Pacific

i) Canada/Greenland

Totals by country are delineated under each of the above areas, for each airline serving the area. The totals for scheduled and charter flights include the number of flights and the percentages of U.S. citizens and aliens.

In addition to the products cited above, two time-series files---one based on the monthly input data, the other on the processed output data---are stored in machine readable form (tapes or discs) to support additional analyses, as required.

### 5.2.4 Future

Improvements are continually incorporated in the data processing system to provide flexible access to data combinations for analysis purposes. Added statistical methods will be employed to evaluate data quality and consistency.

A proposal is under consideration to merge the Official Airline Guide data with the I-92 data to create a file of foreign carriers similar to that of the CAB Economic Regulation 586 data file.

## 5.3 CENSUS OF TRANSPORTATION

### 5.3.1 Background

The Quinquennial Input-Output and the Economic Census were created by the Department of Commerce over the last forty years to satisfy requirements for national economic planning. The Census of Transportation (CTS) is one of the many economic censuses performed by the Bureau of Census each five years. The first CTS was done in 1963 with a passenger transportation survey. The CTS is actually comprised of three major projects: the National Travel Survey, the Truck Inventory and Use Survey, and the Commodity Transportation Survey.

The National Travel Survey yields statistics showing national and regional passenger transportation patterns and their relationship to socioeconomic and geographic factors. Data are gathered on trips, person-trips, person-nights, and person-miles by origin and destination, purpose of trips and mode of transport employed.

The Truck Inventory and Use Survey accumulates data concerning the Nation's trucking resources, such as the number of trucks (total and classified by physical characteristics), occupational use of trucks, measures of intensity of vehicle utilization, and geographic distribution of vehicles.

The Commodity Transportation Survey collects information on the physical and geographic distribution of commodities shipped by the manufacturing sector of the national economy, measured by tons and ton-miles. The basic information is derived from a probability sampling of bills-of-lading (or other shipping records) obtained from a probability sampling of manufacturing plants. Additionally, "by-product" or special studies resulting from this survey include study of the domestic movement of exports, a survey of small plant

activity (including plants with 10 to 19 employees), and a special study of the printing and publishing industry.

All of the aforementioned projects of the CTS are supported by direct Congressional funding. The role played by OST and the Modal Administrations with respect to this data collection effort is one of setting user requirements for the statistics developed from the data. The requirements have been submitted in a variety of ways, both formal and informal. The basic CTS project responsibility belongs to the Department of Commerce.

In the last few years, efforts by OST, the FHWA and UMTA have been directed toward structuring a better method by which the DOT requirements are developed and accepted by the Bureau of the Census and toward increasing the sample size of each project. The FHWA has sponsored and has performed several experimental surveys to augment the National Travel Survey and the Truck Inventory and Use Survey. The results have been subjected to considerable political/institutional evaluation, although only limited quantitative technical evaluation has occurred because of data coding difficulties.

The major funding effort by OST has been directed at expanding the Commodity Transportation Survey beyond the manufacturing sector. Supported by this funding, a feasibility study was started in 1973 by the Bureau of the Census to do the following:

a) Appraise the feasibility of collecting data on commodity shipment movements in terms of elapsed time, time-in-transit for each haul, and commodity shipment charges.

b) Determine the feasibility of collecting data on shipments from wholesalers; this included searching for acceptable sampling frames and determining the basis for establishing such frames. Subjects of this feasibility study included merchant wholesalers, petroleum bulk stations and terminals, and assemblers of farm products.

c) Determine the feasibility of collecting data on shipments originating in minerial industries (including solids, liquids and gases). Shipments of oil and gas are treated as a special situation in the feasibility study.

d) Appraise the quantity and quality of reliable transportation flow data which could be displayed for the 530 DOT Transportation Traffic Zones. (Consideration was to be made of non-disclosure and sampling reliability constraints.)

The results of these feasibility studies to date have lead OST to fund additional reviews of areas which might be included in an expanded project. When these studies have been completed and results are available which demonstrate clearly what kinds and types of statistics can be produced from the expanded survey, it is expected that Congress will directly fund the expanded CTS.

### 5.3.2 Raw Data of CTS

The National Travel Survey uses a probability sample of about 18,000 households (for 1967). Questionnaires (mailed quarterly to respondents constituting three panels) required a full year's data for each household on passenger traffic flow. Data obtained for each trip included (1) origin and major destination of the trip, (2) month the trip ended, (3) type of transport used, (4) the major reason for the trip and (5) who in the household took the trip. Additional information was obtained concerning the number of nights away from home and the number of nights spent in each state during a trip.

The Truck Inventory and Use Survey is based on a stratified probability sample of about 120,000 trucks selected from about 15 million registrations on file with motor vehicle departments in the 50 states and the District of Columbia. Further stratification is made on the basis of size of state, and vehicle size. Sample listings are prepared from the registry lists provided and the owners are contacted and surveyed. Data reported includes physical characteristics of the vehicle, its use as a truck (for hire, personal, etc.), total vehicle miles travelled, geographic area of operation, and fleet information.

The Commodity Transportation Survey obtains commodity flow data by Transportation Commodity Code (which measures the transportation

and geographic distribution of commodities shipped by manufacturing establishments in the U.S.). A two-stage probability sample is used to collect data. In the first stage some 13,000 plants are sampled and classified into 86 shipper classes. In the second stage, 100 to 200 of the bills-of-lading on file at the plants are sampled. The data collected includes specific facts about the commodity: origin, destination, commodity code, weight, means of transport, and the production area of surveyed plant.

### 5.3.3 Products

The data files of the three surveys discussed above are available from the U.S. Bureau of the Census for the surveys conducted in 1967 and 1972. In addition, the following hard cover summaries are available:

a) National Travel Survey - 1957, 1963, 1967 and 1972

b) Truck Inventory and Use Survey - 1963, 1967 and 1972

c) Commodity Transportation Survey - 1963, 1967 and 1972

### 5.3.4 Future

The methodology used in designing and executing these three major surveys has been continually refined and updated with each census. Comparability of data elements to permit trend analysis has nevertheless been maintained. It is expected that the 1977 census will see both continued improvements in the statistical methodology used and larger samples to yield more reliable estimates.

## 5.4 JOURNEY-TO-WORK SUPPLEMENT TO THE ANNUAL HOUSING SURVEY

### 5.4.1 Background

The objective of the Annual Housing Survey (AHS), which is conducted by the Bureau of the Census for the Department of Housing and Urban Development (HUD), is to provide a current and ongoing series of data on selected housing and demographic characteristics.

This survey was jointly planned by the Census Bureau and HUD in response to a need for frequent and up-to-date data on U.S. housing, considered a prime indicator of the Nation's economic health. Viewing this survey as an excellent vehicle to update the Department of Transportation information on the "Journey-to-Work", DOT joined the Census and HUD effort in 1974. Within DOT, a joint effort is sponsored by OST (TPI), FHWA and UMTA. An improved set of questions on the commuting habits of the household head was developed and incorporated into the survey.

### 5.4.2 Raw Data of the Journey-to-Work (J-T-W) Supplement

The AHS consists of both a national sample and an urban (Standard Metropolitan Statistical Area) sample. The national sample is comprised of 76,000 households, with oversampling in rural areas. The national sample is surveyed every two years.

In addition, an independent sample of households in 60 SMSA's is conducted. This 60 SMSA sample is divided into panels of about 20 SMSA's each, with one panel surveyed every 3 years on a rotating basis. Each panel consists of 4 very large SMSA's with a sample of 15,000 housing units selected in each, and 16 other (large) SMSA's with a sample of 5,000 households in each.

The pertinent socio-economic data gathered in the survey includes household income, household size, age, race and sex of household head; type of structure and property value; availability of telephone; and availability of garage, etc. The journey-to-work component which was developed by the DOT includes the following elements, for the household head:

a) Mode used to commute to work

b) Use of a car for the work trip

c) Number of people in the car

d) Destination of work trip

e) Total trip time

f) Trip distance (one way)

g) Change of mode during last year

h) Satisfaction with means of transportation to work

The first panel on the urban sample (Panel I) was done prior to DOT's participation and does not contain this improved J-T-W section. The second panel on the urban sample (Panel II) was begun in March of 1975 and will soon be completed. The first national sample involving DOT participation began in the fall of 1975 and will be completed in early 1976.

### 5.4.3 Products of Journey-to-Work Supplement

Tabulations have been produced for the earlier version of journey-to-work questions which were asked of Panel I, prior to DOT's involvement. This information, however, is open-ended in that it does not include the destination of trips and modal shift data. Thus, its use in urban research and planning is rather limited. A report on Panel II and the national sample will soon be available as this work incorporating the improved data elements is completed. The reports will summarize the pertinent findings in 27 tables which have been outlined by the participating DOT organizations.

### 5.4.4 Future

Planning is currently underway for the Department's participation in Panel III, as well as the next national sample (to take place in 1977). Also the feasibility of broader involvement by DOT (such as the addition of non-work trips, etc.) is being assessed.

# APPENDIX A - DEFINITIONS

## A.1 DATA ADMINISTRATOR

### A.1.1. Position Summary

The position of Data Administrator is established with responsibility for defining content, structure and research strategies of transportation-related data bases. The incumbent is responsible for maintaining updated documentation that will provide easy access to specific elements of data. Some of the incumbent's functions are directly related to data base effectiveness, control, and toward the creation of an overall data base environment.

### A.1.2 Principal Duties and Responsibilities

Working as a specialist in data administration, the incumbent will establish standards of data collection and data base documentation. In order that the division's data and its use might be defined in a complete and consistent manner, incumbent will assist the Chief, Information Management Branch, in creating a standard methodology of collecting and describing all of the Department's national transportation data bases, and with data bases of Federal agencies of mutual benefit. The incumbent works with other members of the staff in maintaining the integrity and security of data gathered by the division and acts as a consultant in data base usage and methods of accessing data. Further, the incumbent will be responsible for reviewing data base performance and may be required to recommend modifying access capability and/or data structures to satisfy changing requirements of the transportation community.

Working with the Chief, Information Management Branch, incumbent designs transportation data programs, definitions, requirements and specifications for development of data systems through outside contractors, interagency agreements or other cooperative managements, assists in the preparation and maintenance of the

master development plan on an annual basis and provides standards for an overview of the entire data spectrum.

A.1.3 Supervisory Controls

The incumbent is dependent on his own professional knowledge for formulating and carrying through his programs and has technical responsibility for defining content, structure and documentation of data bases. By its nature, work is such that it cannot be reviewed in detail and the Branch Chief must rely heavily on the incumbent's technical judgement and experience.

A.1.4 Selective Factors

The selective factors which are germane to this position are grouped into three mandatory elements and several additional desirable elements. (See CSC Form 972)

a) Mandatory Factors

1. Experience in immediate accessed time sharing systems
2. Experience in administrating the data for at least one data base management

b) Desirable Factors

1. Fortran programming language
2. Experience with PDP-10 and/or compatible computer
3. Scientific oriented applications
4. Business oriented applications
5. Applications programming experience

The availability of computer software for applications in the solution of many computer problems involving large data sets has created the need for a person versed in data base creation and manipulation from an administration viewpoint.

The recent acquisition of TSC of a highly-interactive data

management system (System 1022) for use on the PDP-10 computer has shifted computer processing at TSC away from highly technical programming and towards a user-oriented computer environment. Knowledge of how to create and access large transportation data bases and how to manipulate these data bases via many different available systems is the key to successful use of machine readable information in this environment.

Experience with the several commercially available Data Base Management and computer systems is desirable, as TSC has available to it all of the major systems in current use. Also the ability to distinguish which system is more efficient for a given task is invaluable to any computer application. Prior experience in data administration is a requirement.

## A.2 DATA VERSUS INFORMATION

The purpose of this section is to examine and discuss the relationship between information and data and the process through which knowledge is communicated.

Communication includes all procedures by which one mind affects another. This involves not only written and oral speech, but also music, art and theatre. The word "communication" is derived from the Latin communis, meaning common. In communicating, one tries to establish a commonalty with someone else. Communication involves the sharing of experience with others through formal or informal languages or through signs and symbols. Therefore, when communication takes place, a transference of meaning, or knowledge, occurs.

Information is knowledge, intelligence, facts, or data which can be used, transferred, or communicated. Information has several basic qualities: existence, availability, language or a recognizable representation, and meaning. Information is that which adds to, changes, or repeats a representation of what is known or believed to be known. Information is a human product and, therefore, is subjective at least to the extent that it is constructed by human representational techniques and artifacts. Information for use by decision-makers derive from many sources, including

previous knowledge (conjectures reasoned to be valid), intelligence, political or psychological mood, and, last but not least, the synthesis of quantitative data.

Data elements concern those pieces of information that are presented as having an objective reality; for example, a particular automobile that might be described by class characteristics (an abstraction), or by its unique chassis number. The data elements possess the quality of being actual, can be observed and, most important, measured either qualitatively or quantitatively. Through the assembly of many data elements, the basis for synthesis is formed.

Synthesis is simply the process of combining parts or elements (data) so as to form a whole, thereby providing for the dialectic combination of thesis and antithesis into a higher stage of truth. The foundation of this process is deductive reasoning, the basis of mathematics. The process itself starts with a simple review of the observations (data) and proceeds to orders of complexity limited basically by the quantity and quality of the elements (data) available for application; or, to put it more commonly: "garbage in yields garbage out." The result of the synthesis process is one type of information.

Thus, research uses the methods of synthesis to combine data and create information. The information is communicated to interested parties who determine the truth or knowledge gained from the information. The subjective approval, or disapproval, of the information modifies the data collection process and the synthesis means to increase the quality and to continue the development of new information.

All of the analysts in DOT are engaged in synthesis with little attention given to the management of the data they use. These data can be judged by six parameters:

a) Quantity - measured by the number of documents, pages, words, characters, bits

b) Content - meaning or definition of the data

c) Quality - that which characterizes the completeness, accuracy, relevance, and timeliness of the data

d) Life - total span of time in which value can be derived from the data

e) Structure - format or organization of the data and the logical relationships between elements

f) Language - the symbols, alphabets, codes, and syntax used to express the data.

The process called the management of data, and the work of statisticans and data base administrators, focuses on dealing with these parameters effectively.

# APPENDIX B. THE DOT INFORMATION PROGRAM - AN HISTORICAL AND LEGAL PERSPECTIVE

Transportation data is as old as the nation itself. Elements of the Coast Guard assisting in the development of commerce early in the nation's history participated in the reporting of movements of vessels in and about our shorelines. Much later, with the founding of the Bureau of Public Roads and the Civil Aeronautics Administration, additional transportation elements began collecting and assembling data for the purpose of gaining an insight into their own activities. As the transportation elements grew and the likelihood of consolidating them under one roof (like the DOT) began to materialize again in the late 1950's and early 60's* the need for a basis to correct the widely recognized deficiencies in existing transportation information was noted by the Congress and others.

For example, in 1959 the Subcommittee on Census and Government statistics of the Committee on Post Office and Civil Service of the House of Representatives characterized the general area of transportation statistics as: "One of the most poorly organized of the Federal statistical fields."

In 1960, a special panel of experts in a National Academy of Sciences-National Research Council study** of transportation research developed a set of criteria for the purpose of evaluating the adequacy of current transportation data, as follows:

> "Are available data sufficient to describe the performance of the various transportation activities?
>
> Are they sufficient to enable one to understand the interrelations of the various transport modes with each other, with the economy as a whole, including its international component, and with various regional and industrial sectors of the economy?

*A single transportation department was proposed as far back as 1874.

**Conference on Transporation Research, Woods Hole, Massachusetts, August 1960, Publication 840.

Are they adequate to allow understanding of the effects of transportation upon individual welfare, and the character of rural, urban, and metropolitan life?

Are they adequate to allow an informed assessment of the nature of developments in the transport field of their effects on each of the above mentioned areas?

Are they adequate to permit the formulation and validation of hypotheses which seek to explain the transport operations as a system phenomenon within a socio-economic framework?

Are they adequate to allow for timely and effective planning for defense needs?

Is the data collection system capable of providing information responsive to current and anticipated needs in a timely and efficient manner?"

It was concluded that in important respects, "the data now available are inadequate" to answer any of the above questions affiramatively. The National Academy of Sciences (National Research Council's) panel of experts further concluded that:

> ". . . (transportation) information is not adequate for scientific examination of the transportation system as a whole, nor its relationships to vital economic, social, political, and defense questions. Without such information it is difficult to identify important problems and promising methods for solution. . . The present aggregation of organizations collecting transport data does not now provide the information required for the satisfactory understanding of the transportation system as a whole and its ramifying effects."

In 1962, the Committee on Post Office and Civil Service of the House of Representatives in a report entitled "Improving Federal Transportation Statistics" declared:

> ". . . since there is no Federal agency or even a clearinghouse to fit together the many pieces of transportation and travel information, this fragmentary approach has resulted in duplication and excessive trivia in some systems and complete gaps in others. At present, no one can begin to define the statistical dimension of the transportation universe."

The Committee report identified three major deficiencies in then available transportation information:

"1. It does not cover all commercial transportation -- the principal major gaps are in the intrastate motor carriers and the exempt for-hire and private motor carriage, but there are serious statistical gaps in other modes, also.

2. It does not provide a comprehensive picture. The available data cannot be brought together into a meaningful evaluation of total freight and passenger movement in this country.

3. It does not present significant detail on the movement of passengers and freight, either nationally or in the geographic regions."

In 1965, the Secretary of Commerce in Congressional hearings* stated in regard to a proposed program for the improvement of transportation statistics:

The Committee report stated elsewhere:

". . . it is a truism that no overall national transportation policy can be intelligently developed until a meaningful body of facts is assembled, nor would one expect the uncoordinated and highly competitive segments of the industry to work together voluntarily to fill the statistical voids which, in fact, handicap the entire industry."

". . . it is necessary to acquire knowledge of present transportation patterns and preferences in order to predict future transportation needs and to ascertain the most efficient ways of meeting them. A fundamental requirement in any effort to develop an efficient and productive national transportation system is that the needs of travelers and shippers must be defined and measured. Reason dictates that these tasks must be approached before the Government or private investors commit large amounts of capital.

Presently, seven Federal agencies carry on major transportation data collection programs. A number of other agencies collect transportation statistics in some form. These programs are conducted largely to fulfill the regulatory or operating missions of collecting organizations. It is not suggested that they be replaced or duplicated. It should be recognized however, that in their present form these programs do not meet the requirements of a large number of users, especially those charged with responsibility for public investment policy."

*House Report 89-17, House Committee on Interstate and Foreign Commerce, Sub-Committee on Transportation and Aeronautics, 89th Congress, 1st Session, Commerce Department Research - H.R.5863, June 30, 1965 and Senate Report 89-22, Senate Committee on Commerce Commerce, Sub-Committee on Surface Transportation, 89th Congress, 1st Session, High Speed Ground Transportation, S. 1588, June 16, 1965.

In connection with these Congressional hearings, the Secretary of Commerce submitted extensive written testimony on a National Transportation Statistics Program which included the following statements:

> ". . . The problem under consideration (the transportation information problem) is of a magnitude and complexity such that a considerable expenditure of effort for a number of years will be required to bring about needed improvements.
>
> ". . . The Federal Government spends billions of dollars annually in the transportation field. At the present time information on which rational expenditure decisions can be cased is frequently inadequate. To begin data collection anew each time a study is required is wasteful both of time and money. To make major expenditures without adequate study can lead to errors costing millions of dollars. Improperly located highways or airports without adequate ground access may be considered examples. Relatively small expenditures on systematically organized and current information can help precent such errors.
>
> "No one should think that the development of a useful operational national transportation statistics program is an easy task or one of short duration. Above all, it must be the objective of such a program to move away from data collection activities that are narrowly oriented to the needs of individual organizations and that have as their end the production of tabular materials and reports of limited usefulness. An effective program should serve the needs of as many agencies and institutions as possible, with the users' needs being the major determinant of content and of organization."

With the establishment of the Department of Transportation in October, 1966, the Secretary received authorization to engage in data gathering and statistical activities as follows:

Under Section 4 of the High Speed Ground Transportation Act, 1965, the Secretary of Commerce was authorized to:

> ". . . collect and collate transportation data, statistics, and other information which he determines will contribute to the improvement of the national transportation system. In carrying out this activity, the Secretary shall utilize the data, statistics, and other information available from Federal agencies and other sources to the greatest practicable extent. This data, statistics, and other information shall be

made available to other Federal agencies and to the public insofar as practicable." (PL 89-220, Sec. 4).

Under the Department of Transportation Act, 1966, this legislative authority pertaining to transportation information (contained in Section 4 of the High Speed Ground Transportation Act) was transferred from the secretary of Commerce and vested in the Secretary of Transportation.

Also under the Department of Transportation Act the Secretary of Transportation was charged with the responsibility to:

> ". . . promote and undertake development, collection, and dissemination of technological, statistical, economic, and other information relevant to domestic and international transportation. . ." (PL 89-670, Sec. 4 (a)).

These two legislative provisions provide the Department with a broad statutory basis for the planning and the development of transportation information programs.

To carry out its statutory mandate, the Department established and maintained a separate transportation information planning office to provide the requisite leadership on behalf of the Department in the planning, development and implementation of transportation information programs. This office, known as the Office of Transportation Information Planning, was established within the Office of the Secretary when the Department was formed in April 1967. A predecessor office to the Office of Transportation Information Planning was established in April 1966, by the Secretary of Commerce to develop and implement a national transportation information program pursuant to the authority provided under Section 4 of the High Speed Ground Transportation Act, 1965. The staff and unexpended balances of funds were transferred from this office to the new office created by the Secretary of Transportation.

Initially, the Office of Transportation Information Planning was located within the Office of the Assistant Secretary for Research and Technology. In September, 1968, the Secretary transferred the responsibility for the transportation information functions from the Assistant Secretary for Research and Technology to

the Assistant Secretary for Policy Development and, accordingly, the Office of Transportation Information Planning was transferred to the Office of the Assistant Secretary for Policy Development.

In February, 1969, the Offices of the Assistant Secretary for Policy Development and the Assistant Secretary for International Affairs and Special Programs were combined into the Office of the Assistant Secretary for Policy and International Affairs. The Office of Transportation Information was then located in the Office of the Assistant Secretary of Policy and International Affairs. By September 1969 hearings, the Office was a Division in the Office of Systems Requirements, Plans and Information.

The House Committee on Appropriations requested in their report on the Department of Transportation Appropriation Bill, 1969, a report outlining in specific terms the Department's future plans in the field of transportation information, including the estimated funding and manpower required to carry out these plans. In the hearings, the comment on the Department's transportation information planning activities was as follows:

> "TRANSPORTATION INFORMATION PLANNING . . . .
> No new funds are requested for Transportation Information Planning in Fiscal 1969. The program has progressed slowly and is to be carried on with unobligated funds from fiscal year 1968. Last year, the Committee called on the Department to 'develop a more coherent and effective assignment of the responsibilities within the Office of the Secretary and among the administrations for Transportation Information and statistics functions. There is no evidence that this has been done. The statutory basis for this program is broad in scope, relating to a wide range of statistical and other data collection activities concerning all aspects of transportation. The Committee requests that the Department present a report by January 1, 1969, outlining in specific terms what its future plans are in this field, including the estimated funding and manpower required to carry them out."

On May 29, 1969, Secretary John Volpe delivered to the Honorable Edward P. Boland, Chairman Subcommittee on Department of Transportation Appropriation, the report pursuant to the request contained in House Report 1596, June 27, 1968.

The first opportunity to discuss the report with the committee occurred during the fiscal year 1970 appropriations hearings on September 25, 1969. Testimony given by the Honorable Paul Cherrington, Assistant Secretary of Policy and International Affairs, during those hearings is presented in the excerpts from the published proceedings included below:

Starting with Mr. Cherrington -

This chart covers the structure of our policy and planning research program but on a little different basis. These are the typical places where we get our data—as you know, we are requesting a considerable amount of money for a data program in this area—the classes of information used in policy and planning analyses, and again leading hopefully to policies and plans.

Finally, breaking this box down here into its components, in 1970 this shows the various types of research activity, and here on the chart are the particular studies underway which I have already outlined in specific terms.

TRANSPORTATION INFORMATION PROGRAM

I have also several charts on our transportation information program. This program involves goods movement data, personnel travel data, information on facilities, and it will also include improved methodology for gathering data.

Mr. BOLAND. You are asking for about $2 million in this area?

Mr. CHERINGTON. Yes. That, as you know, is the first installment on this fairly extensive program that was sent down to you in the red book entitled "Transportation Information, a Report to the Committee on Appropriations, U.S. House of Repre-entatives, May 1969."

Mr. BOLAND. What do you anticipate will be the final cost?

Mr. CHERINGTON. The final cost was estimated over a 5-year period at $36 million. Thereafter it would take roughly $6.5 million to maintain the data system per year.

TRANSPORTATION INFORMATION PLANNING

Mr. BOLAND. Place justification pages TR–12 through TR–15 in the record at this point.

(The pages follow:)

**A. Transportation information** ..... $2,000

The programs in this category are designed to produce the transportation data base required for objective and informed planning and policy formulation by Government and Industry. Emphasis is placed on providing the Department with consistent, comprehensive data on passenger and freight movements, transportation facilities and transportation systems.

**1. Goods movement data** ..... 650

Origin and destination data will be gathered on major inter-urban movements of goods. Also, information will be collected which will indicate the commodity carried, as well as he weigh, size, value, distance, time, freight revenue, equipment and type. By gathering this data on truck, rail and air modes, we will obtain the basic information necessary for further research into transportation system requirements. Also methods of obtaining data on goods movement in urban areas will be studied. This program will contain the following specific projects:

(a) Urban goods movement data methods ..... 100

Under this project research will be performed on methods of obtaining data on goods movement in urban areas.

(b) Interurban goods movement data—truck ..... 300

Data for this project, which was an outgrowth of work initiated in fiscal year 1968, will be obtained by combining existing motor freight conference data, a sample of truck license plates and a sample of truck shippers.

(c) Interurban goods movement data—rail ..... 200

Under this project, initiated in fiscal year 1969, data will be obtained by combining a sampling of rail waybills and a sample of rail shippers.

(d) Interurban goods movement data—air ..... 50

Funding for this project, commencing in fiscal year 1970, will provide for initial research on sampling methods. It is anticipated that data will be obtained from air freight bills and air shippers.

2. Person travel data ..... $700

Information on travel by individuals by all modes will be gathered which will include data on trip length, trip frequency, purpose, travel time, travel cost, vacation cost, place of residence, income, family size automobile availability, mode used and an evaluation of transportation service available. These data will be obtained for various national and interurban travel patterns. Such information will permit us to assess the efficiency and adequacy of service in the country as a whole as well as in selected regions and between selected cities. This will help make it possible to recommend alterations to improve the transportation system based on sound statistical information. Also improvement of methods of obtaining personal travel data for urban areas will be explored. The projects in this program are as follows:

(a) Urban person travel data methods ..... 100

This project provides for study to develop improved methods of obtaining data on person travel in urban areas.

(b) Interurban person travel data—all modes national ..... 500

Under this project, travel data will be obtained from a national sample of approximately 10,000 households.

(c) Interurban person travel data—air ..... 100

This project provides for initial data on travel by nonscheduled carriers and general aviation, and for initial studies of a national program for obtaining airport access data.

3. Transportation facilities data methods ..... 300

Initial research and development will be commenced to improve the methods of obtaining and recording comparable data on transportation facilities networks, including highways, bus routes, trucking, subways, commuter rail, freight rail, airports, seaports, bus, rail, and other terminal and transfer facilities. Methods of interrelating facilities data geographically (urban, interurban and international) and across modes will be developed. Data items to be considered will include travel time, delay time, waiting time, travel costs, capacity, traffic volume, schedules, and accident experience on a link-node basis. This program is divided into the following projects.

(a) Urban transportation facilities data methods ..... 100

This project, focusing research into methods of obtaining data on urban facilities, will be initiated in fiscal year 1970.

(b) Interurban transportation facilities data—highway ..... 100

This project provides for initial research on methodology for data collection on the interurban highway network including accident experience, year built, and location for each link of segment in the network. Research will focus on methods by which data are mapped and checked state by state and then converted to machine-readable form by semiautomated methods, including coordinate readers.

(c) Interurban transportation facilities data—rail ..... 100

Initial research on methodology will be commenced to develop a data base on the interurban rail network, including schedules, accident experience, and year built for each link or segment in the network. Research will focus on such methods as those by which data are mapped and checked railroad by railroad and then converted to machine-readable form by semiautomatic methods, including coordinate readers.

4. Systems design and development ..... 350

This provides for the overall design of the transportation data information system including data base, data base management, hardware, software, use of geo-coding and commodity coding systems.

(a) Data systems design ..... 100

This will provide a framework for the Department's transportation data systems. The fiscal year 1970 funding provides for design. Development will be started in fiscal year 1971.

(b) Geocoding and commodity coding systems---------- 250

These are systems for converting diverse locations such as street addresses, intersections, and rural locations, to coordinates; and commodity codes to common form which can be fed into a computer. These systems will permit interrelating data by machine and automatic graphic display.

The geocoding system will be realized by supporting and expanding the system the Bureau of the Census is developing for the 1970 Census of Population and Housing. The commodity coding system will be realized by first assigning standard transportation commodity codes to all national motor freight classification codes, then repeating for codes used by other modes. The amount requested for funding for fiscal year 1970 provides for this research phase as well as partial development work.

Mr. BOLAND. There is $2 million in the research appropriation request for transportation information studies. How much was obligated for this program in 1969 and 1968?

Mr. CHERINGTON. In fiscal year 1968 $401,000 was obligated for contract research and approximately $150,000 for administrative expenses for the transportation information planning program. In fiscal year 1969 $346,000 was obligated for contract research and approximately $150,000 for administrative expenses. Additionally in fiscal year 1969 considerable work was done on the interurban rail and motor freight data projects. These will be under contract shortly and will obligate about $500,000.

Mr. BOLAND. What contracts were let and for what purpose in 1969 for this program?

Mr. CHERINGTON. Contracts totaling $76,000 were let for assistance in the development of the 5-year transportation information program that is described in the report entitled "Transportation information." This is the report that was prepared at the committee's request and that was submitted to the committee in May 1969.

A total of $60,000 was awarded in contracts to study and diagram the movements of people and goods in the transportation system. These studies provide us with a better understanding of the kinds of information that should be obtained and the kinds of measurements that should be provided for in our transportation information programs. An agreement was completed with the Bureau of the Census for the Department's participation in a study of the use of census data. The Department cooperated with several other Federal, State, and local agencies in this project and the cost to the Department in fiscal 1969 was $75,000.

A contract for $85,000 was let to obtain data on the attitudes of people to the present transportation system and data on their views of alternative proposals for improvement of the present system; $50,000 was spent on the purchase of census data for urban areas.

Mr. BOLAND. Describe the kinds of information that were obtained from these studies?

Mr. CHERINGTON. In our studies to develop the Department's 5-year transportation information program, information was obtained on existing transportation statistical programs, deficiencies and gaps in these programs, and critical transportation information requirements for Government and industry.

The studies and diagrams of persons and goods moving through the transportation system concentrated on typical situations, such as a journey from home to work, or a trip from Washington, D.C., to New York. These studies provided information on points at which problems are encountered and at which measurements should be taken for effective analysis of these problems.

Our participation in the census use study has contributed to effecting improvements in the 1970 census, such as in the area of journey to work data. Also it has led to the development of powerful data handling tools, such as a national computer-based geo-coding system for all urban areas.

From the rail freight data project origin-destination data will be obtained on major interurban goods movements by rail for the years 1969 and 1967. The data include commodity carried, weight, distance, time in transit, freight revenue, equipment type and so on. From the motor freight data project we are expecting to obtain similar information on goods movement by trucks for the years 1969 and 1968.

Mr. BOLAND. How will this information help the Department of Transportation make better transportation policy?

Mr. CHERINGTON. Information obtained in connection with the development of the transportation information program has helped the Department formulate the critical data needs for transportation planning and policy development for the Department, industry, and other Government agencies at national, State, and local levels.

The census use study has led to improvements in 1970 census data and will help government and industry at all levels, particularly at the urban level, to make better use of 1970 census data. Data from the 1970 census will be widely used in the development of urban transportation policies for estimating transportation needs, and for estimating probable social and economic impacts of proposed future transportation facilities.

Information on rail and motor freight movements will help in the development of policies on Government and private investment in transportation facilities and equipment, and in the development of policies on user charges, trust funds, regulatory reform, technology, and safety.

### TRANSPORTATION INFORMATION REPORT

Mr. BOLAND. You submitted a report to this committee in May 1969 in response to our request on the transportation information program. That report sets forth a 6-year program totaling $35.6 million including the $2 million in the 1970 budget. Briefly tell the committee what purpose this program will serve, how it will be conducted, and how it will improve transportation in the United States.

Mr. CHERINGTON. The fundamental purpose of the transportation information program is to provide critical information for effective decisionmaking on major transportation matters, both in the public and private sectors. Without adequate information the chances of making costly errors are greatly increased.

The program will utilize existing data sources to the maximum extent practicable. Where existing sources are inadequate, new information sources will be developed. Also research will be undertaken to devise and implement improved methods of data collection and processing.

The transportation information program will help to improve transportation by making a much more complete analysis of planning and policy issues possible.

This program will enable equipment manufacturers, carriers, shippers, and Government officials to make better forecasts of passenger and freight markets including modal and firm shares; better forecasts of the numbers, sizes, and types of equipment needed; better forecasts of transportation demand and of the locations and sizes of facilities required to meet the demand; by this I mean the requirements for highways, airports and rail, water and pipeline facilities. This will permit better estimates of the private and Government investment necessary to finance the required facilities. There are many other areas in which the information program will help to improve transportation.

Mr. BOLAND. Cite some specific programs in the Department that will or might be affected by the results of this work.

Mr. CHERINGTON. The information produced through this program will benefit departmental programs at all levels. The Office of the Secretary would have a much better basis for developing transportation policies and programs to provide for a better coordinated and more efficient national transportation system. Overall transportation planning with attention to intermodal problems could be more readily accomplished. We could also plan more effectively for meeting the requirements created by military emergencies and natural disasters.

In the operating administrations, the Federal Aviation Administration would have a better basis to plan for future requirements in air traffic control and navigation facilities. The Federal Highway Administration would have much better data on important items such as the growth of intercity highway truck operations, intercity auto travel patterns, and the highway access needs of air and rail terminals. The Federal Railroad Administration would have freight flow data, better passenger market data and industrial center growth data for its policy and planning work.

The State and local governments that administer Federal-aid transportation programs would have much of the information required for developing coordinated statewide and urban transportation plans and policies. The newly emerging State departments of transporation already see good data as an important first requirement. Good data would make it easier for these Government agencies to coordinate their planning for urban mass transportation services, highway facilities, and airport development.

I could give you many more examples, but the point is that this program would be very useful for many of the Department's most significant programs.

Mr. BOLAND. Is it intended that all the work be done by contract or will some be done in-house as well: If the work will be split indicate that portion which will be done in-house and that which will be done by contract in 1970.

Mr. CHERINGTON. Basically, the answer to your question is that most of the work provided for in the 5-year $35.6 million program will be done by contracts with consulting firms, universities, nonprofit corporations, and by agreements with Government agencies at Federal, State, and local levels. However, in-house staff leadership and support also are required to administer the contracts and agreements and

provide the initiative essential to implement the transportation information program.

Mr. BOLAND. With regard to the items in the $2 million requested in 1970. How much of this involves the collection of data not already available and how much will simply be collected together from other public or private agencies?

Mr. CHERINGTON. Nearly all of the interurban data projects in the fiscal year 1970 program involve the collection of data not already available from other public or private agencies. These projects provide for interurban goods movement data for truck, rail and air, interurban person travel data air and all modes, and interurban transportation facilities data for highway and rail. We are hopeful that some of the interurban truck freight data can be developed by bringing together data from the traffic studies conducted by the 12 or so motor freight rate conferences. This depends on the cooperation of the motor freight conferences.

It should be made clear that only some of the interurban data projects will yield data as a product of the fiscal year 1970 program; namely, the rail and truck freight flow data projects and the survey of national travel by all modes. For the other interurban data projects, methods developed during the fiscal 1970, and the data will be implemeted in the following years.

The urban transportation data projects in the fiscal year 1970 program are methodological. The purpose here is to assist urban transportation planning agencies to obtain high-quality, relevant, continuing data at low cost. Urban tranportation planning agencies spend about $20 million annually in data collection and processing.

Mr. BOLAND. Cite some specific examples of benefits which we can anticipate if the $2 million you have requested for 1970 is appropiated.

Mr. CHERINGTON. The interurban goods movement and person travel data will provide a basis for better analyses, forecasts and decision-making in the areas of estimating potential freight and passenger markets, purchases of new transportation equipment, investment in terminal modernization, determination of transportation demand, estimation of facilities requirements and studies of public and private financing needed to provide required facilities. The interurban transportation facilities data will be used similarly for better decision-making—particularly to determine gaps between the capacities of present facilities and the capacities required to satisfy estimated future transportation demands.

Improvements in policies and planning areas such as these will result in savings in transportation expenditures. These expeditures run into billions of dollars.

Mr. BOLAND. Once the information program gets set up, what will be the annual cost of collecting the data—maintaining the system?

Mr. CHERINGTON. The annual cost of collecting the data—maintaining the system—will begin at about one-half million dollars in fiscal year 1971 and build up about $6½ million once the information program is set up. This would be at the end of fiscal 1975, according to the Department's report, when all component data systems in the program would be operational.

Mr. BOLAND. What specific kinds of data not already available in one form or another will be collected?

Mr. CHERINGTON. The 5-year transportation information program contains some 30 projects. To answer your question, it would help if we regarded these projects as divided into three groups. The first group would consist of projects that provide for collection of data that has never been nationally available on a regular basis. For example, origin-destination data on intraurban and international person travel by all modes of transportation, origin-destination data on interurban truck, rail, and air freight. I would say that there would be about 10 projects in this group.

The next group would consist of projects based on existing transportation data collection activities. Most of the existing transportation data sources involved only partially meet essential data requirements. The program provides for improving these existing data activities by filling gaps, expansion of coverage, methodological improvement and most importantly by rendering the data from these activities compatible with other data developed from the program. I would say that there would be about 15 projects in this group.

The third group would consist of projects that do not aim to collect data per se. It would include projects whose main purpose is to provide for improvement of the methods used in urban data collection also the system design and development projects whose purpose is to coordinate the program structure so that the data from each of the separate projects are compatible, readily accessible, easy to use and widely available.

Mr. BOLAND. Will the collection be done by your office or by other parts of the Department?

Mr. CHERINGTON. The responsibility for planning, research, design, and initial development of the transportation information program would be in my office. This is the work required to establish each of the programs data projects on an operational basis. But for the products from the program to be useful there must be close working partnerships throughout the planning, research, and development stages of each of the projects with the interested parties in industry and Government-data users, suppliers, and producers.

To achieve these partnerships, working groups will be established. One of their most important tasks would be to determine the collection agency for the ongoing data collection programs. In many cases it would be most appropriate to establish the collection agency in some other parts of the Department. In other cases, it may be more appropriate to establish the collection agency entirely outside of the Department. These matters can be best decided on a case-by-case basis as each project is undertaken, at the same time giving proper attention to the overall efficiencies of these arrangements.

Let me repeat, however, that it will be essential that the separate collection efforts mesh to produce compatible data—to avoid the fragmentation of our present inadequate transportation data activities.

This can only be done through strong coordination and leadership from the Office of the Secretary.

The following excerpts from testimony before the same House Subcommittee on succeeding years illustrates the situation as it moved from 1970 through the 1975 hearings.

On April 21, 1970, Mr. Richard J. Barber, Deputy Assistant Secretary for Policy and International Affairs, gave the following testimony:

TRANSPORTATION INFORMATION

Mr. BOLAND. On page 17, under transportation information, what types of data do you plan to develop? Of what value will this data be and what decisions will be made as a result of your having this information?

Mr. BARBER. You will recall thta last year in response to your request in the report of 2 years ago, as I recall, we submitted to the committee an elaborate transportation information program. That program calls for the expenditure over 5 years of something in excess of $35 million. In fact, it calls for $2 million in fiscal 1970 and $7.4 million in fiscal 1971.

This committee, following considerably the comments by Dr. Cherington in the last hearing, cut the amount back for fiscal 1970 to some $594,000, as I recall.

Consistent with this reduction and contraction in the efforts, we undertook a very pragmatic reassessment of the overall effort, and concluded that where we wanted to put our emphasis in 1970 and 1971 was on pulling together some of these very practical bits of transportation information that now do not exist. We are focusing in three areas—first, on railway bills and rail samples, trying to computerize that information rather than relying on the old hand process.

Second, we are moving into motor truck data, which do not exist now in the United States, but which are vital to us and very important to many of the transport industries.

Our third area, and we have just begun working with the CAB, is to see what kind of information we can get particularly on air freight movement. So, this is the direction in which we are moving, Mr. Chairman.

No other discussion bearing directly on transportation information appears in the FY 1971 hearings.

On May 20, 1971, the Honorable James M. Beggs, Under Secretary and Mr. Robert H. Binder, Deputy Assistant Secretary for Policy and International Affairs provided the testimony as recorded in the following published proceedings.

TRANSPORTATION INFORMATION

Mr. MCFALL. You indicate that you are planning to spend $350,000 to overcome the shortcomings of the 1967 Bureau of Census report on Transportation and Manufacturing. Is the Bureau of the Census not doing a good job? If the census is inadequate, shouldn't they correct it, rather than you?

Mr. BINDER. We are not faulting the work that the Bureau of Census does, but we find in taking their basic work and applying it to our needs, that certain additional information has to be added and certain modifications and adjustments must be made, and it is this work we particularly have in mind.

Mr. BEGGS. If I may add one comment, Mr. Chairman, we asked the Census Bureau in the 1970 census to gather additional data on journey origin and destination to work into one of the various samples that they did. This is above and beyond what they normally do in this census work. It is partly to take care of that additional data that we want to do this.

Mr. MCFALL. Is there another census coming in 1972?

Mr. BEGGS. Yes, sir. There is additional information that they are gathering for us on transportation.

Mr. MCFALL. The question then arises, if they are going to do another census in 1972, why do you need to revise the 1967 census, and will their new census have the same shortcomings as that of 1967?

Mr. BEGGS. There are two different things we are talking about here.

Mr. BINDER. Let me answer the last one first. The 1972 census will produce information in 1972 and thereafter. As I have indicated, one of the principal works we are now addressing is a report which will be transmitted to Congress in 1972. So in that sense, we cannot avoid correcting the 1967 census or adding to that in an appropriate way.

Mr. MCFALL. Will the Department of Commerce have a better census in 1972?

Mr. BINDER. To the extent that the dialog between our staff and Commerce's staff indicates to them that there are changes they could appropriately make, I would be confident they would.

AUTOMATED DATA BASE STUDY

Mr. MCFALL. With reference to the automated data base study, in 1971 your work under this heading was encoding rail and regulated trucking waybill statistics. Does this type of work duplicate efforts such as the 1967 census?

Mr. BINDER. No, sir; not to my knowledge. I think the particular type of information we draw from the waybill studies is not the same type of information which the census is collecting.

Mr. BEGGS. As a matter of fact, I think I am right here, the only place where there is any sampling of waybills is a very small sample that the ICC has been taking. The accuracy of their sample, I think, is open to question.

Mr. MCFALL. Are you going to use this work for anything productive, or are you just collecting data on everything in general?

Mr. BINDER. In a real sense the Information Division of our Office of Systems Planning is a resource, so you cannot say in advance that everything that they are collecting will have a known application. But in another real sense we find that the basic data which is being collected and stored in this division is being drawn upon both for the 1972 needs study, our Northeast Corridor report, and the regulatory work which goes on in other divisions of our Policy Office.

On April 13, 1972, Mr. Robert H. Binder, Deputy Assistant Secretary for Policy and International Affairs, and Dr. Irwin P. Halpern, Deputy Assistant Secretary for Policy and International Affairs gave rather extensive testimony on transportation information, as indicated in the published proceedings, following:

## TRANSPORTATION INFORMATION

Mr. CONTE. What are the notable gaps in personal travel measurements you hope to fill this year under the Transportation information activity?

Dr. HALPERN. Major gaps in personal travel measurements to be filled this year are those of intercity rail and bus passenger activity levels. There are currently no comprehensive data on the level of travel activity served by these modes. These programs will be designed in form and content to complement the air passenger O–D program of the Civil Aeronautics Board. Together these survey programs will constitute a full description of passenger movement by intercity carrier modes. That it, between any two given cities the total account of commercial carrier travel and the proportions attained by each mode will be known for the first time.

There are remaining major gaps, notably in personal auto transportation, and in personal travel demand characteristics, by all modes, into which we will only be able to make small inroads in this fiscal year.

## TRANSPORTATION INFORMATION PUBLICATIONS

Mr. CONTE. Please elaborate on the extensive publications activity you are anticipating next year under this program.

Dr. HALPERN. We feel that the greatest return on the investment made in transportation information will accrue from getting the information out to the users. This includes shippers, carriers, transportation equipment manufacturers, State and local governments, as well as the Department itself. The demand for our information products has been far greater than we expected.

In the remainder of this fiscal year, we expect to produce five major data reports. On the basis of our recent experience, we would expect the demand for them to be high. For example, a recent publication, which we saw as a significant reference document in the area of transportation geographic identification systems serving a limited set of users, has received overwhelming response. Over 300 requests for it have been received in the 2 weeks since its release. Requests came from universities, other Government agencies, air carriers, transportation associations, all modes, manufacturers, financial services, planning authorities, and other allied interests.

Present supplies of the document have already been exhausted. Future runs will be sold through the GPO. We have been reluctant in the beginning stages of the program to put stringent changes on documents which would impede getting them to the public. We have an automated mailing list system in the final stages of development which will insure getting the right documents to the right users.

Costs of publication have been minimal thus far, both in actual dollars and as a percentage of project costs. Publication costs have not exceeded 3 to 5 percent of any project to date. With total expenditures for the year estimated at less than $15,000, it can be expected that in the coming fiscal year, even with nominal charges placed on our products, which include computer tapes and special tabulations, as well as regular reports, that a significant amount of money will be returned to Treasury.

### TRANSPORTATION STATISTICS HANDBOOK

Mr. CONTE. How many years do you expect the development of the Transportation Statistics Handbook to take?

Dr. HALPERN. A first version of the handbook will be available in late April 1972, with a revised version to be published November 1, 1972. Thereafter, an updated version is to be published in November of each year. The Transportation Statistics Handbook must be viewed as an open-ended, continuing activity in which it is our expectation to add improved information as such becomes available.

### INTERACTIVE STATISTICAL SYSTEM

Mr. CONTE. How far along is the interactive system development program?

Dr. HALPERN. Only design development work has been begun on the interactive statistical system to date. However, the major data components of the system have already been developed as part of the statistical handbook. Also, the major computer effort will depend heavily on previously developed, federally owned software systems. These systems are presently being evaluated and tested.

The major effort in fiscal year 1973, then, will be a molding of the extant data base and software. An operative system should be in place by mid-fiscal year 1973.

### NATIONAL TRAVEL SURVEY

Mr. CONTE. When is the national travel survey for 1972 scheduled to begin?

Dr. HALPERN. The national travel survey in its present form is to begin on schedule, late in calendar 1972. However, discussions are currently underway between senior staff in the Bureau of the Census and the Department of Transportation that are intended to lead to extensive changes—for example, increased sample size, broader geographic coverage, greater frequency—in the survey. While the effect of this information division program on the 1972 travel survey will not yet be a major one, it should be noted that the 1972 sample size will be double that used in previous national travel surveys. Moreover, the national travel survey has now been included in the list of surveys that can be conducted on an annual basis.

As indicated earlier, the national travel survey has the potential for being the single most important source of household-based information about non-intraurban travel, an invaluable base for forecasting the demand for travel by mode, evaluating Amtrak performance, and similar analytic purposes. The current discussions between DOT and the Census are expected to lead to a survey design that will not only produce data valuable for these purposes, but also to the large number of non-Federal users of such statistics.

### NATIONAL SHIPPER SURVEY

Mr. CONTE. By how much do you plan to expand the total number of observations made in connection with the national shipper survey?

Dr. HALPERN. The exact extent of the expansion is currently one of the subjects of the technical discussions between DOT and the Census. Sampling experts at the Bureau of the Census are seeking to determine the number of observations required for analysis of, at the very least, State to State commodity movements by six to 10 commodity classes, as well as the sample size required for more fine-grained networks such as the 530-plus node national networks being developed for the Office of Systems Analysis and Information.

In view of the fact that the expansion of the national shipper survey is not only a matter of increasing the number of observations but also one of increasing commodity coverage, the proper balance between sample size and commodity coverage on the one hand, and cost and payoff on the other, is a question requiring the most careful scrutiny. Our crude estimate is that at least an eightfold to tenfold expansion of the number of observations is required. It should be noted also that the national shipper survey is potentially the keystone in any set of national goods movement data, since it is the only base from which an estimate of the national goods movement universe could be derived

### WATERBORNE TRANSPORTATION STUDY

Mr. CONTE. What are the existing documentation systems on which the waterborne transportation study will be based?

Dr. HALPERN. The principal agencies involved in the collection of waterborne transportation data, as well as the publication of statistical reports on the subject, are as follows: The Department of Commerce, (the Bureau of the Census, Maritime Administration, and Office of Business Economics). In the Department of the Army, the Corps of Engineers. Also, the Interstate Commerce Commission. And in the Department of the Treasury, the Bureau of Customs.

The publication of statistics based on the data collected by the above agencies range from periodic reports published, for example, by the Bureau of the Census and the Interstate Commerce Commission, to reports published by other agencies at infrequent intervals as the need and funds permit.

The collection of data for the domestic movement of waterborne commodities is performed primarily by the Corps of Engineers and the Interstate Commerce Commission. Also, it should be noted that DOT is currently cosponsoring a study with the Corps of Engineers to identify the U.S. inland origins and destinations of commodities moving in international trade. The previous study of this sort was last performed in 1959.

The collection of data on the movement of cargoes in foreign trade is conducted primarily by the Bureau of the Census and the Bureau of Customs. The Maritime Administration has been collecting data on container traffic on major ocean trade routes. Ocean fleet inventory statistics are prepared by the Maritime Administration, Coast Guard, and Navy Department.

Information with regard to port activity and capacity is collected by the Corps of Engineers, Maritime Administration, as well as the Port Authorities themselves.

Other organizations, which are users of the statistical reports mentioned above, very often publish their own reports based upon a further analysis of the information contained in the public reports. A few of these organizations are United Nations, World Bank, Agency for International Development.

### INFORMATION DIVISION WORKLOAD

Mr. CONTE. How serious is the current workload problem in the Information Division and how will an additional position affect it?

Mr. BINDER. The current workload in the Information Division is very heavy and projected to get heavier. There is a serious need for specific professional skills to permit the programs of the Division to continue to effectively serve the information needs of Government and industry.

The one position presently in discussion would be filled by a professional survey statistician/statistical analyst. His immediate duties would include design of greater efficiency and cost savings procedures into the rail waybill and air waybill statistics programs. In addition he would serve all of the Depar. ment in our data collection review process which reviews, evaluates, a l enhances the designs for proposed surveys by any unit of the Dep. tment prior to OMB review and clearance. This project, understaffed at present, has added to the efficiency of savings and reporting procedures of all kinds, reduced the reporting burden on the public, and saved budget resources throughout the Department. His presence would permit this project to take on a more active assistance role. In addition, we intend to convert one of the positions which we already have into a technical achive specialist.

STATISTICAL AND INFORMATION NEEDS

Mr. McFALL. On page 16, could you briefly describe for us the process by which you determine what your statistical and informational needs are? How do we know that your data collection program is geared to the needs of those who are making and executing policy?

Dr. HALPERN. There are many elements to the process. A significant one has been our experience in the development of the Transportation Statistics Handbook. There are, unfortunately, many places in that document where the words "not available" appear. We have conducted surveys of available data in the Department and other transportation agencies. In addition evaluation by our information systems people of the analytical requirements for responding to the anticipated policy issues of the future serves to define the structure and priority order of our program.

But most of all, it has been the continuing dialogs with industry and other levels of government that have served to define the program—including the response we have received from the products we have produced, and the direct, unsolicited statements of need we have received.

TRANSPORTATION STATISTICS HANDBOOK

Mr. McFALL. On page TR–19, has the first Transportation Statistics Handbook been published yet? What future year costs do you anticipate for the handbook?

Dr. HALPERN. The first version of the handbook will be received from the printers this week. It will be released soon thereafter. We would expect future year costs to average between $35,000 and $50,000. This would include the costs of updating with new annual statistics, expansion into new areas of reporting, and printing and publishing.

GEOGRAPHIC IDENTIFICATION CODING SYSTEM

Mr. McFALL. Also on page 18, will the geographic identification coding system be a one-time project? Will it be completed in 1973?

Dr. HALPERN. No sir; the full system is expected to be completed sometime in mid fiscal year 1974. However, there will be useful products all along the way. Once completed, there will be continuing maintenance costs associated with incorporating changes that occur in the geographic elements the system describes. We would estimate the maintenance costs to be approximately half the project cost.

FEDERAL TRAFFIC INFORMATION SYSTEM

Mr. McFall. What is the purpose and cost of the Federal traffic information system?

Dr. Halpern. The major purpose of this project is to identify the size of the Federal component in the markets of carriers. In certain sectors we think the component could be significant. Thus, changes in Federal programs could impact heavily on the economic health of portions of the transportation industry. We need to be in a position to anticipate these impacts. We expect the costs of the project to be small as it will consist primarily of assembling and compiling information produced by other agencies.

JOURNEY-TO-WORK SURVEY

Mr. McFall. On page TR–25, how many cities will be used in the journey-to-work survey? How will they be selected? Do you intend to eventually include all cities in this program?

Dr. Halpern. The journey-to-work statistics, as specially developed by the Department from the 1970 census will be available to all cities over 50,000 in population at the cost of processing. In addition the 40 or so cities with populations over 500,000 will be subjected to special analysis based on these statistics. Where applicable, the special analytical tools developed in this project will be made available to cities of less than 500,000 population.

On May 18, 1973, the Honorable John L. Hazard, Assistant Secretary for Policy, Plans, and International Affairs provided testimony, as reflected in the following proceedings excerpts:

GEO-CODING PROJECT

Mr. Conte. What is the status of the geo-coding project? What positive developments from this program can you point to?

Mr. Hazard. Accurate location is the foundation for all transportation analysis and planning. A major methodological problem which has historically faced policy analysis and decisionmakers has been the numerous diverse geographic location coding systems on which empirical data has been and is being collected. The different location coding schemes make it extremely difficult to compare and analyze data collected by different sources. The purpose of the geo-coding project is to devise a system whereby transportation data collected on different location identification systems could be compared and analyzed on a common identification system. We conducted an indepth study and survey of geo-coding systems, held a national geo-coding conference and developed a geographic location converter procedure which permits the comparison and analysis of related data collected by different geographic location coding schemes.

In fiscal year 1974 we want to continue our work in this area by extending the capability of the system to convert other smaller area geographic identification systems and thereby use more effectively a number of existing urban and subcounty data bases which would be very useful in national transportation planning.

## RESEARCH ON PERSON MOVEMENT

Mr. CONTE. You show that your emphasis in the research on person movement will be on the bus passenger. How would you characterize the data that is available on the car, plane and train passenger? Are you going to need funds in the future to develop comparable data on these people?

Mr. HAZARD. Basic data on passenger travel by all modes is either limited or nonexistent. The national travel survey of the U.S. census of transportation is, at the present time, the only source that has the potential for generating data representative of the universe of intercity travel in the United States. Neither the sample size (12,000) of the 1967 survey nor the proposed doubling of that sample for the 1972 survey is sufficient in size to assure origin-destination data sufficiently disaggregated to be useful for determining travel demands and facility requirements at the national level. At least an additional 25,000 households more widely distributed geographically should be added to the sample size. In conjunction to funds available to the U.S. Bureau of Census, it is estimated that an additional $500,000 will be needed to complete the enlarged survey.

A joint program between DOT and Amtrak is being implemented to obtain needed basic data on the level of demand, use of facilities and equipment, origin and destination, and passenger socioeconomic characteristics related to rail travel. Another joint project now under discussion with the CAB would focus on the movement of passengers from true origins to true destinations, both domestically and internationally, as related to air travel demand and usage.

## EXPORT OF BULK COMMODITIES

Mr. CONTE. You show that one of the areas of emphasis in the commodity movements will be that of bulk commodities for export. How many studies on this problem are being conducted by the various Government agencies?

Mr. HAZARD. The DOT recently cosponsored with the Corps of Engineers a study "domestic and international transportation of U.S. foreign trade: 1970" for the purpose of identifying the domestic origin-destinations and transportation characteristics of nonbulk commodities moving in U.S. foreign trade. The study was performed by the Bureau of the Census.

The study to which you refer, "bulk commodity origin-destination study" is now in the planning stage, and is intended to cover the bulk commodities not included in the previously mentioned study. The area of interest will include bulk commodities moving both in the U.S. foreign trade (import-export) as well as our own domestic trade (rail, barge, truck). The emphasis is not on tonnage, which has been thoroughly covered by others, but on the domestic origin-destinations, and the transportation characteristics. Inasmuch as only portions of the above items are collected by various agencies considerable cooperation has been involved. For example, Department of Agriculture has data on terminal storage of grain, but not transportation characteristics to its final domestic destination (flour mill; port of export); the Bureau of Census has foreign trade data to our ocean ports but not to inland domestic points. FHWA has some data about truck type for various commodity groups but not with detailed origins and destinations; Corps of Engineers has data by river port, but not by inland origin-destination. In short, no one has conducted a study focusing on the total movement of bulk commodities. In addition, the survey data items will provide more detailed information on commodities, geographic locations, domestic mode of transportation and length of haul.

Mr. CONTE. How much cooperation is there between you and these other agencies in developing this data? Is the study duplicating anyone else's effort?

Mr. HAZARD. The Corps of Engineers, U.S. Department of the Army, has stated that they would cosponsor and participate with DOT in the funding and design of the study. Funding support will be solicited from other Government agencies—such as ICC, Agriculture, Department of Commerce, and FMC. It has not been decided which agency would conduct the actual survey. All interested agencies are being contacted for their recommendations and data requirements to assist in the final study design. As a whole, the study does not duplicate efforts by other organizations.

On May 1, 1974, the Honorable Robert H. Binder, Assistant Secretary for Policy, Plans, and International Affairs (Designate) gave the testimony included in the following proceedings excerpts.

INFORMATION: THE BASIS FOR POLICY PLANNING AND DEVELOPMENT

Mr. MCFALL. On pages TR-26 and 27, you are requesting $2.7 million for "information—the basis for policy planning and development." Many of the other program activities in your justification relate to data development. How is this program different?

Mr. BINDER. Several of the major issue areas of the transportation policy and planning program have their own specific data development programs. For example, there is outlined in our justification a special $1 million energy data requirements program to support our transportation energy policy activities, and a $1.6 million national transportation study program to supply the Department with State and local level data on the current and estimated future performance of the transportation system. Our $2.7 million information program complements these and other departmental data development activities.

Our information program is geared to the needs of those who are working to develop policy. My staff's evaluation of the analytical requirements for responding to anticipated policy issues has served to define the structure and priority order of our information program. They have related each element of the $2.7 million program to the specific issue area activities in our overall fiscal year 1975 work program.

I will be glad to supplement this information with our plans for the next 12 months for the record.

[The information follows:]

TPI INFORMATION PROGRAM PLANS

Over the next 12 months, the TPI information program will develop data that will bridge a number of major gaps in our knowledge of the Nation's transportation system to include:

Commodity origin destination files for all modes with the individual movements as the unit of observation. The nationwide truck commodity flow study, one of the significant projects of this program area, was designed to bridge the gap of information on the total universe of commodity movement by truck.

The commodity transportation survey expansion project, a contractual agreement between the Bureau of the Census and the Department, involves the expansion into new areas of the existing survey such as collection of transportation data from the mineral industries, and data for shipments from merchant wholesalers, petroleum bulk stations and terminals, and assemblers of farm products. An air cargo origin-destination survey is part of this year's commodity flow program.

Complete person origin-destination files for all carrier modes, with main emphasis on intercity bus survey and Amtrak passenger origin-destination survey.

Beginning of transportation performance monitoring incorporating system performance measure, cost, travel time, delay and schedule adherence indices, and demand monitoring.

A fully operational data manage ment system, with full file management, file retrieval and geographic capabi ity. This project includes the automation of data base for the annually p lished "Summary of National Statistics and the Data Supplement."

LACK OF UNIFORMITY OF DATA

Mr. CONTE. On this same page you discuss the development of a statistical information base as being the fundamental requisite to transportation planning. Is the data to be developed of the sort which will be useful in formulating a national transportation policy?

Mr. BINDER. I think it is fair to say that to the extent we have succeeded in developing that base so far, it has already been helpful. We have definite plans to try to improve it.

One of the basic problems in the statistical area is lack of uniformity of data, both as between States and the Federal Government, and as between parts of the Federal Government. One of our principal objectives reflected in the budget before you now as well as in past years, is to improve this uniformity to make the data easily usable by different parts of the Federal and State governments as well as the local and regional levels.

On May 7, 1975 the Honorable Robert H. Binder, Assistant Secretary for Policy, Plans and International Affairs gave testimony as shown in the following excerpt from the published proceedings:

POLICY INFORMATION BASE

Mr. MCFALL. That is a good idea. I have heard people scream about featherbedding for years. On the one side they say this isn't true. On the other side, they say it is true. Let's find out.

Your largest request is for your "policy information base." We recognize that information is a mandatory requirement for intelligent policy formation, execution, and evaluation. You state on page TR-23 that such an endeavor is an evolutionary process. What are you talking about in terms of future financial requirements for this evolution?

Mr. BINDER. I will want to do more thinking about what future fiscal year requirements may be, Mr. Chairman. By evolutionary I mean that this concept of a policy information base stems from the study which was done for the Department by the McKinsey group, who suggested that we should look at our policy information collectively as to what types of policy information we gather, what types of policy information do we need that we aren't gathering, and where are we overlapping? This is the concept of a policy information base which we described in these pages of the budget. I believe that, as we approach this problem year by year, we will be improving and identifying areas where we need less information and possibly areas where we think we need more. It's evolutionary in that sense.

RELATIONSHIP OF POLICY INFORMATION BASE RESEARCH TO MCKINSEY STUDY

Mr. McFall. Is your request for a policy information base tied to the recommendations made by the McKinsey study.

Mr. Binder. Yes, the need for the policy information base was cogently stated in the McKinsey study. as a major input to the management of the Department's R. & D. and grant programs.

The above was the only discussion that could be found in the FY 76 budget hearings that related directly to TPI's Information Program.

On June 2, 1975, during the 1st Session of the 94th Congress, Congressman Fred B. Rooney of Pennsylvania introduced H.R. 7509 (later H.R. 7778), which was referred to the Committee on Public Works and Transportation. The objective of the bill use to help improve and maintain an adequate transportation system through the systematic collection and publication of transportation statistics, to establish a National Center for Transportation Statistics, and other purposes. A copy of the bill itself and the testimony of John W. Snow, Deputy Under Secretary, U.S. DOT, before the Subcommittee on Transportation and Commerce of the House Committee on Interstate and Foreign Commerce, regarding accounting methods of the Interstate Commerce Commission and other railroad matters follow. The bill did not reach the House floor during the 1st Session.

94TH CONGRESS
1ST SESSION

# H. R. 7778

IN THE HOUSE OF REPRESENTATIVES

JUNE 10, 1975

Mr. ROONEY introduced the following bill; which was referred to the Committees on Public Works and Transportation and Interstate and Foreign Commerce

# A BILL

To help improve and maintain an adequate transportation system through the systematic collection and publication of statistics with respect to the transportation of passengers and freight by railroad and other modes, to establish a National Center for Transportation Statistics, and for other purposes.

*Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,* That this Act may be cited as the "Transportation Statistics Act of 1975".

THE NATIONAL CENTER FOR TRANSPORTATION STATISTICS

SEC. 2. There is hereby established within the Office of the Secretary of Transportation (hereinafter referred to as the "Secretary") a National Center for Transportation

Statistics (hereinafter referred to as the "Center"). The Center shall be headed by a Director (hereinafter referred to as the "Director") who shall be appointed by the Secretary subject to the provisions of title 5, United States Code, governing appointments in the competitive service, and shall be paid in accordance with the provisions of chapter 51 and subchapter III of chapter 53 of such title relating to classification and General Schedule pay rates.

PURPOSES OF THE CENTER

SEC. 3. The purpose of the Center shall be to assure the collection and dissemination of statistics and other data related to transportation by railroad and other modes in the United States and other nations. The Center shall—

(1) be responsible for the collection, collation, and regular reporting of full and complete statistics on the conditions of transportation by railroad and other modes in the United States, and in the foreign commerce of the United States, including but not limited to—

(a) statistics concerning the movement of passengers and freight for hire by railroads and other modes;

(b) statistics concerning the movement of passengers in private motor vehicles and aircraft;

(c) statistics concerning the movement of

freight by persons engaged in proprietary transportation operations; and

(d) statistics reflecting the relative cost and resource effectiveness of passenger and freight transportation by the various modes.

(2) coordinate all Federal activities related to the collection and dissemination of transportation statistics with a view toward assuring the timely collection and prompt publication of comparable statistics concerning railroads and all other modes of transportation;

(3) collect and publish whatever statistics may be necessary to assure a complete statistical picture of transportation in the United States, and between points in the United States and those in foreign countries, while avoiding unnecessary duplication of the statistics collection and publication activities of other Federal agencies;

(4) conduct studies and publish reports on specialized analyses of the meaning and significance of transportation statistics;

(5) assist other Federal agencies and State and local transportation agencies, including regulatory agencies, in improving and automating their statistical and data collection activities;

(6) develop standardized data collection methods to

be used by those required to report transportation statistics to the Center or other Federal agencies; and

(7) review and report on transportation activities in foreign countries.

THE ADVISORY COUNCIL ON TRANSPORTATION STATISTICS

SEC. 4. (a) There shall be an Advisory Council on Transportation Statistics which shall be composed of seven members appointed by the Secretary and the ex officio members listed in subsection (b) of this section. Not more than three of the appointed members of the Council may be representatives of the federally regulated transportation industry.

(b) The ex officio members of the Council, who shall be full voting members, shall be—

(1) the chairmen of the Civil Aeronautics Board, the Federal Maritime Commission, and the Interstate Commerce Commission, or their designees, and

(2) the Director of the Bureau of the Census, or his designee.

(c) Appointed members of the Council shall serve for terms of three years, except that in the case of initial appointments, such members shall serve for such shorter terms as the Secretary determines necessary so that the terms of not more than three members expire in the same calendar year.

(d) The Secretary, or his designee who shall be an

Assistant Secretary, shall serve as the nonvoting presiding officer of the Council.

(e) The Council shall meet at the call of the presiding officer, except that it shall meet at least four times during each calendar year and, in addition, whenever three voting members request that the presiding officer call a meeting.

(f) Six members of the Council, of which at least four are appointed members, shall constitute a quorum of the Council.

(g) The Council shall review general policies for the operation of the Center and shall be responsible for establishing standards to insure that the statistics and analyses disseminated by the Center are of high quality and not subject to political influence.

ANNUAL REPORT

SEC. 5. The Secretary shall include in his annual report to the President and the Congress—

(1) a description of the activities of the Center during the preceding fiscal year and a projection of its activities during the succeeding fiscal year;

(2) estimates of the costs of the projected activities for the succeeding fiscal year; and

(3) a statistical report on the condition of transportation in the United States during the preceding fiscal year.

AUTHORITY OF THE DIRECTOR

SEC. 6. (a) In order to carry out the duties of the Center set forth in section 3 of this Act, the Director may make grants to, and contract with, individuals, private and public institutions, organizations, and agencies, including agencies of the United States.

(b) In order to carry out the provisions of this Act, the Director may require all persons engaged in the transportation for hire of passengers or freight or engaged in the transportation of freight in proprietary operations, in interstate or foreign commerce, to submit such reports and to provide such information concerning such transportation operations as the Center may require. In the case of the refusal to obey an order issued by the Director pursuant to this subsection by any person who resides, is found, or transacts business within the jurisdiction of any district court of the United States, such district court shall, upon petition by the Director, have jurisdiction to issue to such person an order requiring him to comply forthwith. Any failure to obey such an order of the court may be punished by such court as a contempt thereof.

RESPONSIBILITIES OF THE CENTER

SEC. 7. (a) The Center may furnish transcripts or copies of tables and other statistical records to, and make special statistical compilations and surveys for, State or local officials,

public and private organizations, and individuals. The Center shall furnish such special statistical compilations as the Committees on Commerce, Public Works, and Appropriations of the Senate and the Committees on Interstate and Foreign Commerce, Public Works and Transportation, Merchant Marine and Fisheries, and Appropriations of the House of Representatives may each request. Such statistical compilations and surveys, other than those carried out pursuant to the preceding sentence, shall be made subject to the payment of the actual or estimated cost of such work. In the case of nonprofit organizations or agencies, the Director may undertake joint statistical projects, the cost of which shall be shared equitably as determined by the Director so long as the purposes of such projects are otherwise authorized by law.

(b) All funds received in payment for work or services enumerated under subsection (a) shall be deposited in a separate account which may be used to pay directly the costs of such work or services, to repay appropriations which initially bore all of part of such costs, or to refund excess sums when necessary.

(c) The Center shall participate with other Federal agencies having a need for transportation data in forming a consortium with such agencies for the purpose of providing direct joint access to all transportation data received by the Center through automated data processing. The Library of

Congress, the General Accounting Office, and the committees of the Senate and the House of Representatives set forth in subsection (a) of this section shall, for the purposes of this subsection, be considered Federal agencies.

(d) The Center shall, in accordance with regulations prescribed by the Director, provide all interested parties, including public and private agencies and individuals, direct access to data collected by the Center for purposes of research and acquiring statistical information, subject to such conditions as the Director may prescribe in order to assure that information regarding shipments tendered to a common carrier, or transported in a proprietary transportation operation, may not be used to the detriment or prejudice of any shipper or consignee or to disclose his business transactions to a competitor.

AUTHORIZATION FOR APPROPRIATIONS

SEC. 8. There is hereby authorized to be appropriated to the Secretary such funds as may be necessary to carry out the purposes of this Act.

On June 12, 1975 the Honorable John W. Snow, Deputy Undersecretary, U.S. Department of Transportation, before the Subcommittee on Transportation and Commerce of the House Committee on Interstate and Foreign Commerce, regarding accounting methods of the interstate Commerce Commission and other railroad matters:

> "Mr. Chairman and Members of the Subcommittee: Thank you for this opportunity to talk to you today about the problems with the present rail accounting system, a bill to establish a National Center for Transportation Statistics in the Department of Transportation (H.R. 7779), and the availability of materials for rebuilding our Nation's railroads. These are all significant areas of concern, and today I hope to summarize some of out thoughts with respect to each issue. The time has come for significant change in the railroad industry. A few short weeks ago, the President submitted major railroad legislation to Congress, the Railroad Revitalization Act ("RRA"). This legislation is designed to improve the regulatory climate under which railroads operate. A major focus of the bill is greater pricing flexibility. Among the many changes proposed is a provision calling for a new rail accounting system. Secretary Coleman and I wish to commend the Subcommittee and the Committee for so promptly scheduling this hearing and the other rail hearings in the next few months. We are very hopeful that these hearings can extend and significantly add to the dialogue that has been established between the Congress and the Administration with respect to our Nation's rail crisis and lead to the enhancement of sound rail legislation.
>
> The first subject I will address is rail accounting:
>
> The present accounting system is antiquated and leads to serious distortions in the decision-making of rail management and the ICC. Even if the regulatory reforms proposed in the RRA are not adopted -- and we frevently hope that they will be -- accounting reform is needed. But reform of present accounting procedures is even more important if we are to obtain the full benefits of the regulatory breakthrough proposed in the RAA. Although reform of the ICC accounting procedures is not a particularly glamorous subject in itself, it is in many ways the nails and mortar of regulatory reform. As I explained, the RAA does propose a change in the accounting procedures, and we at the Department, and particularly in the Federal Rail Administration, have been working on and studying this problem for about the last two years.

The DOT program was initiated in June 1973 with a contract with TOPS-On Line, a subsidiary of the Southern Pacific Railroad, to evaluate roadway costs. We also have working relationships with the Association of American Railroads and ICC in this area. A second major contract on cars and locomotives should be executed by the end of this fiscal year. Overall, this program will identify the nature of railroad cost variability for pricing, investment, and management control and will develop a standard set of procedures and methodologies which can be used by any railroad. It is the only major current comprehensive cost analysis program of which we are aware.

To show the relationship between accounting and resulatory reform I will have to discuss briefly the major regulatory provisions of the RRA. What we are trying to do is to reintroduce the stimulus and the incentives of the private market place into the railroad industry. To do this, we are proposing giving rail management considerable flexibility in making decisions on how to run their railroads. To put it another way, we are proposing to restrict the regulatory authority of the ICC, whose regulatory policy has made a significant contribution to today's rail crisis.

Pricing flexibility is a key element in the RRA. Faced with the considerable costs and uncertainty of the present regulatory process and the regulatory lag, railroads have been quite hesitant to introduce new rate reductions and innovations and to thereby effectively compete with other modes of transportation. At the same time the Commission has refused to allow the railroads to raise rates, even though the present rates are not compensatory. These unreasonable regulatory restraints are costing the railroads hundreds of millions of dollars each year. Removal of these restraints alone will not "solve" the railroads problems, but it is a critical first step toward a solution.

Specifically, the RRA forbids the Commission from calling a rate decrease unreasonable as long as it is "compensatory". It also forbids the Commission from holding the rates of one carrier up to protect the rates of another carrier -- in other words, it prohibits "umbrella rate making". In addition, the bill would provide that if a carrier has a non-compensatory rate, the carrier could raise that rate to a compen-

satory level without Commission interference. It would also prohibit the Commission from approving non-compensatory rate decreases.

All of these provisions go to the question of ultimate lawfulness of a rate. The bill would also establish a no-suspend zone, in which carriers could -- within certain percentages -- introduce rates without fear of suspension. Finally, the bill would also put a definite time limit on the Commission's hearing process.

How does cost accounting relate to these changes? What the bill does is to carve out a wide area for management discretion, but at the same time provide ample safeguards to ensure that this discretion is not abused. But, if that discretion is to be wisely used, management must have the correct information to relate rates to costs. If the Commission is to prevent a carrier from charging a rate that is not compensatory, it must have specific cost information for that segment of transportation. At the same time, if the Commission is to determine whether a rate increase is reasonable for a specific commodity, it must have specific cost information. Cost information is important in other areas referred to in the RRA. If the Commission is to make reasonable decisions about abandonments, it must know the costs of particular segments. On a much broader level, if policy makers are to make intelligent decisions with respect to restructuring, they have to know which segments are profitable and which are not. Had a proper cost accounting been in place, the study of the Northeast's problems would have been much easier and would have proceeded much faster. The problem with the present ICC accounting system is that it does not properly serve the existing decision-making process at the ICC and it will not serve the type of regulatory system envisioned by the RRA.

The Uniform System of Accounts for Railroads, currently prescribed by the Interstate Commerce Commission (49 CFR Part 1201) is a financial accounting system. The Uniform System of Accounts is useful for aggregate financial comparative purposes between firms, but not for managerial cost control or for pricing purposes. It focuses upon the statements which are prepared for general distribution outside the firm; i.e., the Balance Sheet, now called the Statement of Financial Position, the Income Statement, and the Statement of Retained Income. Even though the system does provide for

selected operating statistics, the emphasis is not upon preparing information for internal management or informed regulatory decision-making. The present system produces only aggregates and averages. It does not efficiently segregate costs with respect to locations or functions, and emphasis on industry averages is consistent with the general rate increase approach, which produces a myriad of distortions, and inconsistent with rate changes on particular commodities. Of course, if a firm wishes to propose an individual rate increase, it can produce a "special cost study". But special studies cost extra money and time and their special preparations extends the time it takes to make regulatory decisions. When the accounting system was first adopted at the ICC it was a model for other industries. But times have changed, and the system is woefully out-of-date. The ICC has established six separate dockets to change the accounting rules in the last thirty-or-so years, and there have been no significant changes. To control costs and to provide timely data on variable expenses for internal management, a cost accounting system is needed for the rail industry.

There must be a complete revision of the Uniform System of Accounts for Railroads and the various formulas used to decide costs. In the RAA we propose a two-year joint effort with the ICC to revise the present accounting system. We call for not just a revision of account titles, contents of some accounts, and statements, but for a total change in emphasis in the uniform system as well. Without a modern accounting system based upon costs, the rail industry will be sorely limited in its ability to develop rational management and pricing behavior.

The present rail accounting system including the cost formulas is deficient in at least the following ways.

1. It does not provide information on cost by location or cost center. Such information is available to other industries; yet, the railroads' accounting deals primarily with firm-wide figures or with averages obtained from these aggregate figures.

2. The system has many accounts which are of a mixed nature; i.e., they contain several separate functional costs lumped together in one account. In account number 402, Train Supplies and Expenses, there is a variety of expenses including car cleaning, car heating and lighting, car inspection and watering, train detour expense, train supplies and other expenses such as apparatus for testing the sight and hearing of enginemen and trainmen and employees' reading and bunk room expense, including pay of attendants, supplies

furnished and laundry. In addition, the same type of expense can be recorded in different accounts. These are a few examples of how costs are jumbled together. These groupings complicate the process of determining how, and with what, rail costs vary.

3. Incremental, differential or variable costs are costs which change in relation to the level of activity. These are the types of cost which rail management needs to evaluate specific courses of action (e.g., development of rate proposals, profitability of segments [branches, perhaps] of the rail plant, capital replacements, and maintenance strategies). Unfortunately, costs of this nature are not available under the existing accounting prescribed by the Uniform System of Accounts. These costs are available to rail management only from special cost studies undertaken by the firm itself which approximate these costs.

4. Cost control is more effective where costs are assigned by responsibility and/or location. Profit control needs revenues separated by major classifications and locations, as well as by type of service rendered.

A fundamental, overriding problem suggested by the foregoing is that the data available in the form prescribed by the Uniform System of Accounts is consistently too broad for direct applicability to questions relating to such specific areas of operations such as the cost and revenue associated with the abandonment of a particular section of track or with a commodity movement between two points. Yet, the majority of the Commission's work deals with these sorts of questions rather than with questions relating to the companies' overall costs and revenues. In such cases, information is either not available in the form needed to directly analyze such questions, or, if it can be made available by special studies prepared for the specific case, it is prepared in an inconsistent manner from case to case and is therefore not usable in intercase comparisons.

The objective of any cost accounting system is to determine the cost of each service provided. In a firm where more than one service is provided, some costs may be assignable to any one service; most, however, are. Modern cost accounting practice is to isolate specific costs associated with each service provided. Prior to World War II and modern computer accounting techniques, formulas were developed at the ICC as a means of making the problem more managable, usually at the cost of considerable accuracy. The process, pioneered by the staff of the Interstate Commerce Commission in the thirties had a modest goal. According to a statement of the ICC Bureau of Statistics to explain the approach,

"...it is desired to obtain costs on a nationwide scale for a large number of point-to-point movements leaving the marginal cases disclosed thereby for further study..." (U.S. Interstate Commerce Commission, Docket 28190 - New Automobiles in Interstate Commerce, Railway Freight Service Costs for the Year 1939, Rail Form B, 1-41 [Washington, D.C.: Interstate Commerce Commission, 1941, p.1].

The product of this limited accounting objective was Rail Form B, which evolved into Rail Form A by 1941 and underwent six revisions, the latest version published in 1973. The relationships of formula were based on data derived during the very prosperous period from 1939-1945. This was also a time of high rail passenger traffic. Although frequently opened for revisions, the form has only changed in detail; the techniques employed have remained relatively unchanged.

The form provides a formula for separating the expenses among various services, yard switching, train switching, running, etc., through a formula utilizing a set of territorial factors developed by the ICC to generate unit costs for each service. The costs are based on fully allocated costs, with a formula adjustment to arrive at variable costs. Individual carrier factors may be substituted for the ICC factors if they are backed up by special studies. It is interesting that by 1973, the sixth edition of Rail From A was still prominently displaying the notice on the cover: "This study issued as information, has not been adopted by the Interstate Commerce Commission."

As developed 34 years ago, the Form represented the state-of-the-art in accounting. Since that time, however, conditions have changed, the state-of-the-art has changed, but the Form A has not significantly changed. A myriad of studies exist which demonstrate that average actual railroad expenses are not consistent with the assumptions of Form A. The form is even less accurate when one realizes that it is not reasonable to expect different railroads' expenses to fit the same formula, but that is the assumption of Form A. Rather than look at individual expenses, we have "wooden" formulas that recite that variable costs equal 80 percent of full costs, plus 50 percent of the return on roads, plus 100 percent of the returns on equipment. Again, the railroad may submit a "special study" to justify a new rate, but the process is time-consuming, costly, and is apt to meet with a heavy presumption pro the Form A formula.

The development of a cost accounting system for the rail industry would permit the development of actual costs for segments of a line, or for functions relating to the line. Such costs could be used for control purposes

(i.e., for responsibility accounting, as well as for decision-making). Cost or profit centers could be established, not unlike those in other industries, to permit management to judge what services, lines or branches cost and which service, line or branch is profitable. Furthermore, it is actual costs, relating to only one item or function, which should be used as data in statistical cost analysis to determine when the service or function is paying its own way.

In summary, the present accounting system relies upon broad aggregates and averages. It does not supply the information needed for today's regulatory system. It will not supply the information needed to make the subtle analyses required in restructuring the Nation's rail system. It will also not provide information for the regulatory system proposed in the RRA. We would hope to find remedies to these problems in the study and new system called for in the RRA.

H.R. 7778, the proposed "Transportation Statistics Act of 1975", would establish a National Center for Transportation Statistics in the Office of the Secretary of Transportation. The Director of the Center would be appointed by the Secretary. The Center would be responsible for : (1) collecting and reporting "full and complete statistics" on the conditions of transportation in the United States and in foreign commerce; (2) coordinating federal collections and dissemination of transportation statistics; and (3) collecting and publishing whatever statistics may be necessary "to assure a complete statistical picture of transportation in the United States."

To accomplish the goals of the Center, the Director would have the authority to require any person engaged in transportation to provide such information as "the Center may require". The Center would be authorized to furnish materials to various State and local governments and would be required to furnish such statistical information as various House Committees shoud request.

At present, the gathering of transportation statistics is divided among the various government agencies involved with transportation or statistics: DOT, ICC, CAB, FMC, Bureau of Census, Corps of Engineers and others. There is overlapping responsibility and authority in some areas and gaps in others. The Office of Management and Budget has attempted to coordinate these activities via an Interagency Committee on Transport Statistics, but the duplications, discontinuity, and gaps still exist. Recent major public issues such as the grain crisis and the rail crisis have highlighted the weaknesses in the availability of transportation statistics.

We agree that a better system for statistics is necessary and we agree that central coordination is necessary if we are to obtain the type of statistics we need to make informed judgments. There are several areas of concern. For example, the roles of the existing agencies in the collection of data vis-a-vis each other and this Department has to be better defined. It is not clear in the bill what is to happen with the existing authority in the various agencies, and this problem is compounded since some of that authority is lodged in the independent agencies.

There are additional problems in how the federal activities relate to existing State programs and in the protection given to proprietary data -- a subject only briefly referred to in the bill. In approaching this subject we must ensure that we do not simply add to overbearing filing requirements already imposed on the industry. We have to ensure that all the data we require is collected in usable form and is used. We do think there is great merit in the object of the bill -- to obtain comprehensive transportation statistics -- and we stand ready to assist in working out an appropriate mechanism to achieve this objective.

The last issue that the Subcommittee has asked me to address -- the availability of materials for rail rehabilitation -- is a very extensive one. In this limited time, I will only be able to lay out certain guideposts. Generally, the sufficiency of rail materials depends on the amount of rehabilitation and the time frame for rehabilitation. We appear to have the mill capacity and the national inventory for short-term programs. For long-term programs, some capacity will have to be developed. But we do have the resources to built the necessary capacity. The United States Railway Association (USRA) and the various steel companies have been actively discussing these needs. I would defer to those parties for details and programs.

Rail is the critical area. At present, we have in inventory, or on order, 3,000 miles of new rail and 2,000 miles of used or re-lay track. The combined production capacity of rolling mills in the United States is estimated at 4,000 miles of track a year. This is the basic supply formula. But in addition, there is the so-called "cascade effect". Normal practice when installing new rails is to save approximately two-thirds of the older rail, and to repair and re-lay this rail in yards and on secondary lines.

As far as demand in concerned, the Department estimates deferred maintenance on a restructured rail system will require 20,000 miles of rail to bring the rail

plant to a normal level of maintenance. In addition, normal yearly requirements amount to about 5,500 miles of track. To round off this picture, the total cost of laying one track mile of new rail, plus 2/3 track mile of old track is approximately $135,000 --- of which 2/3 is material cost, and 1/3 is labor, fringe benefits, tools, machinery, and other expenses.

This concludes my written testimony. I would be happy to answer any of your questions."

# APPENDIX C. DATA BASE ADMINISTRATION REQUIREMENTS DOCUMENT (PROPOSED)

## C.1 INTRODUCTION

### C.1.1 The Objectives of the Data Base Approach

The "data base," from the viewpoint of the TSC Information Program, may be defined as the central repository of a collection of information on all data available to the TSC. The concept of a data base environment then extends to the definition, capture, storage, maintenance, and dissemination of this data for use within the organization in a controlled, consistent fashion. To many organizations, this approach represents a radical departure from present methods of handling data. Typical of many data environments today are:

a) fragmented (independent) application development

b) inadequate user service levels

c) uncontrolled data redundancy

d) inconsistent data definition

e) inconsistent data manipulation

The general objective, then, of the data base approach is the elimination of both data and processing redundancy and inconsistency, and the establishment of a data environment capable of responding to the varied needs of the end users. In essence, the data base approach demands a well defined, centrally administered, set of standards and procedures relative to all aspects of data used by the organization. The data base approach implies a high degree of automation (though not necessarily so), and therefore includes such software elements as an automated data dictionary, a generalized transaction control processor, the application systems, and the generalized data base management system (GDBMS); however, a data base does not imply that the organization will simply acquire a GDBMS from a vendor and "turn it over" to the user groups within the organization. The key element of the data base

environment is that of standards and procedures, and without these standards and procedures, the data base, as described in this operations plan, cannot function.

### C.1.2 Planning for the Data Base Environment

The committment of this type of data base approach is a major organizational decision. Essential to the success of the implementation of the data base environment is full support of the user, management, and data processing groups.

The first step toward the achievement of a data base environment is the acceptance of the statement of objectives (see Section C.1.1, The Objectives of the Data Base Approach, of this appendix), as the goal of this operation plan. These objectives must be further refined and then be communicated to the user, management, and data processing groups. This communication, or education, consists of a full presentation of the goals and concepts of a data base approach, the estimates of costs, resources, and time required to develop a data base environment, and the responsibilities that the users, management, and data processing groups will each have in the data base environment.

Based upon the general acceptance of the goals, concepts, and planned approach, the data base environment Operating Group will be established. From this point on, the Operating Group will develop and execute the discrete tasks necessary to establish the data base environment, including activities relating to the hardware/software components and administrative procedures. At this same time, the TSC User Group will be established to develop user requirements and assist in the assigning of priorities to the development activities including DOT user requirements. In Section C.2, THE FUNDAMENTAL STRUCTURE OF THE OPERATING DATA BASE ENVIRONMENT, each aspect of the planning and implementation process is examined in detail.

### C.1.3 Functional Elements of the Data Base Environment

Aside from the administrative aspects, the data base environment may be defined as consisting of the following elements:

a) transaction controller

b) application software set, or sets

c) generalized data base management system (GDBMS), or systems

d) data base I/O interface (DBIO)

e) data dictionary/directory

Figure C-1 illustrates the interrelationships among these functional components. A description of each component follows.

#### C.1.3.1 The Transaction Controller

The transaction controller has three general functions:

a) handling of transactions entering the TSC data base environment and the distribution of messages/responses flowing from the environment.

b) scheduling and monitoring of application systems and subsystems within the environment.

c) handling of transactions entering the data base environments of vendors to TSC, and the distribution of messages/responses flowing from the vendor's environment.

From the communications aspect, the transaction controller may receive and transmit messages from and to a variety of terminal types, including low speed, non-intelligent terminals (e.g., TTY), intelligent terminals (e.g., Datapoint 2200), mini-computers, medium and large scale computers, or other transaction controllers (as within a distributed network). The transaction controller also typically functions as a message switch for terminal-to-terminal communications. The transaction controller utilizes human interfaces in establishing users in vendor environments.

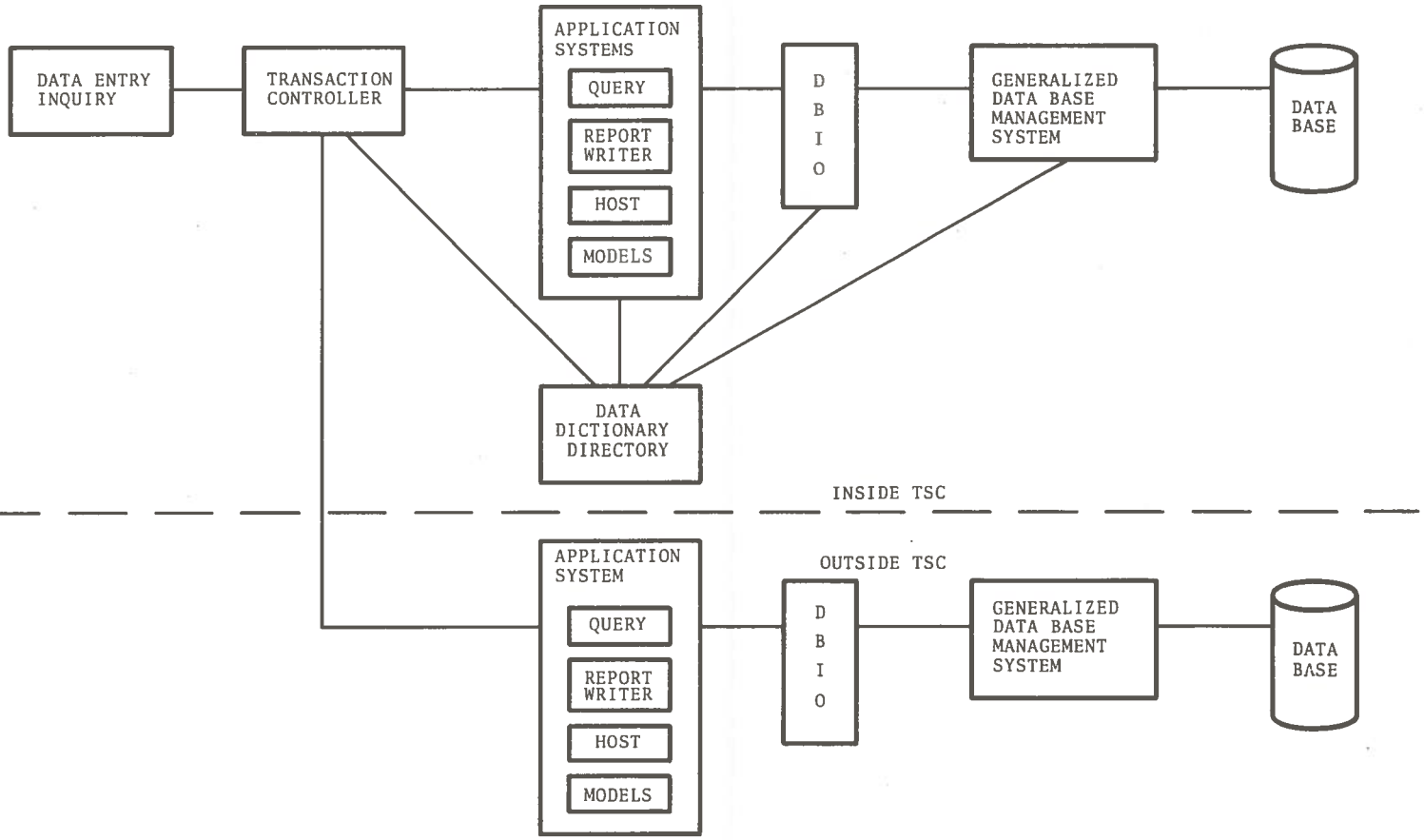



Figure C-1. Functional Elements of the Data Base Environment

### C.1.3.2 The Applications Systems

The applications systems are the problem or model processing elements of the data base environment. Typically, the problem processors are structured from COBOL, FORTRAN, PL/1, Assembler, or other programming languages designed and implemented to handle specific user requirements. These programs communicate with the GDBMS through a language set embedded within (or "hosted" by) the program itself. The language set falls into two categories - the Sub-Schema Data Description Language (DDL) and the Data Manipulation Language (DML).

The Sub-Schema DDL is the language used by the application (user) programmer to describe the subset of the data base that is to be manipulated by the program logic. This subset provides a degree of "program independence" from the data base in two ways. First, the Sub-Schema allows the programmer to view the data base as consisting only of the items, records, and relationships relative to his processing logic. The programmer therefore need not be concerned with the full scope of the data base.

Second, the application programmer need not be concerned with the physical aspects of the data base, such as device type, record blocking, storage strategy, or access methods.

The Sub-Schema concept thus allows the program to be highly (but not entirely) insensitive to data base changes, such as addition/deletion/modification of data items, records, and relationships; and to physical environment changes, such as device changes and modification of access and storage strategies.

The DML is the language set used by the application programmer to manipulate data in the data base. Basically, the DML consists of verbs to locate, retrieve, add, delete, and modify data in the data base. It is designed to be independent of the access methods, storage strategies, and device characteristics at the physical data base level, and therefore protects the problem program from data base changes in these areas.

User requirements may also be supported through higher level languages that are "self-contained" and do not require a COBOL, FORTRAN, etc. expertise to use. These languages are often used by the end-user himself. Such self-contained languages provide basic retrieval and update capabilities, but typically do not support the comprehensive procedural facilities available in compiler level languages. These languages are particularly valuable in supporting the one-time, ad hoc needs of users, especially when turnaround time is critical.

Self-contained languages are placed into two categories - Query and Report Writer. Query Languages generally have an unstructured "natural language" format and are designed for the end user. They are most suited for rapid response, limited output inquiries of an ad hoc nature.

Where output is of higher volume, response time less critical, and display format requirements more structured, a Report Writer Language is preferred over a Query Language. Like the Query Language, the Report Writer Language may be used by the end user. However, because of its more structured syntax and emphasis on report format, it is typically used by technically-oriented support personnel in responding to user requirements. Because it is "shorthand" in nature, the Report Writer allows ad hoc reports to be designed and produced rapidly. Although the cost of the actual report run is typically much higher compared to a "tailored" COBOL or other compiler language program, setup time is greatly reduced - a matter of hours as opposed to perhaps several days.

Regardless of the language type used for requesting data from the data base, all these requests for data are processed by the GDBMS. The actual physical addressing of data within the data base and any logical transformation of data is performed by the GDBMS.

### C.1.3.3 The Data Base Input/Output (DBIO) Interface

The Data Base Input/Output (DBIO) Interface is a frequently used, but optional element of the data base environment. Functionally, the DBIO acts as an intermediate process between the

Application Program or Process (Host, Query, or Report Writer Language) and the DBMS. All "calls," or service requests, from an Application Program are made directly to the DBIO. The DBIO will further process the request, pass the request to the GDBMS, and return all responses, both status and data, from the DBMS to the requesting Application Program.

The need for a DBIO generally depends on the GDBMS selected for a particular data base environment. To a large degree, the DBIO can be viewed as an "enhancement" to the GDBMS, providing for facilities and services not available in the GDBMS. Typical of the major functions a DBIO may perform are the following:

a) functional authorization checking and control (security) at levels of detail greater than provided by the GDBMS.

b) general modular interfacing required in a multi-GDBMS environment, where a single language is used by the programmer and the DBIO interprets the request and reformats it for the appropriate GDBMS.

c) journalizing (logging) of GDBMS calls where the GDBMS does not provide adequate logging.

Another, related function of the DBIO is to provide transparency between Application Programs and the GDBMS. The objective here is to minimize the impact of changing from one GDBMS to another in the event:

1) the GDBMS is replaced by the vendor with another non-compatible GDBMS.

2) the vendor ceases to support the GDBMS altogether (e.g., because of vendor dissolution).

3) a GDBMS is introduced on the market that offers significant advances over the GDBMS then being used.

4) an alternative GDBMS becomes a necessary user requirement.

A third function of the DBIO is to act as a control processor in a distributed data base environment. In this capacity, the DBIO would receive GDBMS calls from an Application Program and deliver the request to the appropriate point in the distributed network, receive any responses from that point, and deliver that response back to the requesting Application Program. The delivery point may be a GDBMS "machine" that will further process the request and determine the location of the needed data or it may be to a specific machine, complete with a full GDBMS capability, in the network. A third potential would be to deliver the request to a transaction controller in another, independent data base environment.

C.1.3.4 The Generalized Data Base Management System (GDBMS)

The GDBMS is the software system that services all requests for data. The GDBMS receives information relative to the content, structure, and physical characteristics from the data base designer via two languages - the Schema Data Description Language (DDL) and the Device Media Control Language (DMCL). Requests for data against the data base may come from a host Data Manipulation Language (DML) call, a Report Writer Language specification, from a Query language command, or from the DBIO.

The Schema DDL is the language used to describe the full scope of the data base to the GDBMS. The GDBMS processes the Schema specification and builds an internal directory[1] containing complete information on:

a) data item names

b) data item physical characteristics

[1]Most GDBMS packages currently available on the market today maintain their own internal data directories. A developing concept is to store all data information in a separate, automated Data Dictionary/Directory System (DD/DS). Under this concept, the data base schema DDL and DMCL would be processed directly by the DD/DS and maintained separately by the DD/DS. At execution time, the GDBMS would extract the necessary mapping data from the DD/DS to process a request. See the discussion on the DD/DS in later paragraphs of this section.

c) record names

d) record relationships

e) record keys and access methods

f) security rules

g) edit/validation rules

Physical device characteristics and physical storage strategies (grouping, page sizes, free space allocation, etc.) are entered via the Device Media Control Language (DMCL).

Once the data base is defined and established, the GDBMS processes all requests against the data base by first verifying the requestor's authorization to manipulate the GDBMS, then relates the Sub-Schema and the request to the data base Schema, extracts the required data from the physical data base, transforms the data according to the Sub-Schema specification and delivers the data back to the requestor.

If a DBIO interface exists, the GDBMS functions as if the DBIO were the requesting Application Program.

C.1.3.5 The Data Dictionary/Directory System (DD/DS)

The DD/DS serves as a repository of data about data. The DD/DS interfaces with the Data Administrator, the Systems Analyst, the Programmer, the User, and the various software elements of the data base environment and is the key element in executing this aspect of the Information Program. The contents of the Data Dictionary/Directory and its interface characteristics are discussed in detail in Section 5, OST DATA COLLECTION.

C.1.4 Data Base Design and Development, and Production Systems

The next task of the data base environment planning and development effort is the integration, testing, and acceptance of the data base environment components. Upon completion of this task, the data base environment becomes production or service oriented. In the production mode, the Operations Group System Manager will coordinate the continuing systems and data base

design functions of the TSC data base environment, according to the standards and procedures established by the Operating Group.

The major procedural characteristic of the data base environment at that time will be the separation of the data base design, the application system design functions, and user activities. The System Manager will continue to be responsible for analysis of the user's needs, as demonstrated through system usage and as identified by the TSC User Group. One facet of this activity will be the continued identification and refinement of data requirements. In the traditional systems design and programming approach, a file design specification to support a given user's requirements would be built. However, in a data base environment, the System Manager will communicate a request for new data to the data base design team, a group reporting directly to the Operating Group.

The data base design team will be responsible for incorporating the requested data into the data base and communicating back to the implementation team all information and documentation necessary to obtain, insert, and then manipulate the requested data.

Described in a different manner and depicted in the flow diagram shown in Figure C-2, the systems analyst proceeds to identify the specific data required to support the user's needs, the edited validation rules for the data, and the processing and response requirements requested by the user to the Systems Manager. The data base design team will incorporate the data into the data base structure, utilizing where possible existing defined data, via the Schema DDL and DMCL. The next step in the data base design process will be the structuring of the Sub-Schema(s) required to provide access to the data. The Sub-Schema(s) will then be made available to the system implementation team via manual (coding form) or automated (DD/DS) means. Finally, after systems design, the Systems Manager will review the systems documentation and approve proceeding to the programming phase.

Further discussion of the data base design and administration functions is given in Section 4, INTERNAL DOT GENERATED DATA.

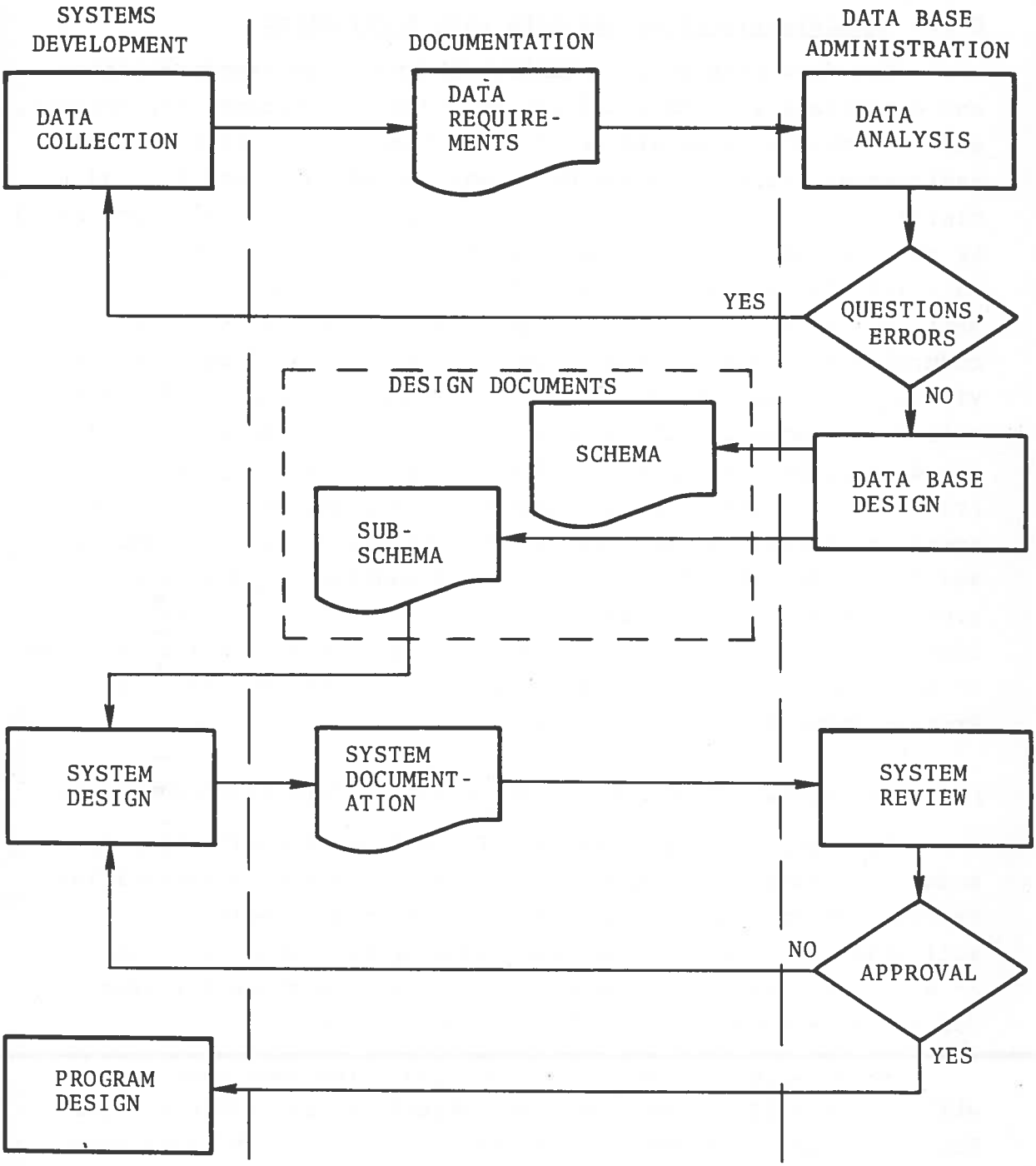



Figure C-2. System Design in a Data Base Environment

### C.1.5 Administration of the Data Base Environment

The data base concept implies the need for standardization and coordination of data definition, capture, storage, maintenance, and dissemination, as discussed in Section C.1.1. In a data base environment, this coordination is manifested in a data base administration function. The Data Base Administrator, as represented by the Systems Manager of the Operating Group, is responsible not only for the coordination and control of the data base environment, but also for the establishment of the standards and procedures for coordination and control. A successful data base environment in the Transportation Systems Center (TSC) will demand extensive communication between the users of the data base and the data processing Operating Group. Within the data processing group itself, clear communication will be required between the operating functional entities of Systems Analysis, Programming, and Operations and the various support functions (such as planning, education, documentation, etc.). Therefore, the data base administration responsibility should be placed as a staff function in an organizational position reporting directly to the Center Program Director for Information.

## C.2 FUNDAMENTAL STRUCTURE OF THE OPERATING DATA BASE ENVIRONMENT

The long range plan for the TSC data base environment is actually a subset of the overall ADP plan for the Transportation Systems Center (TSC). Similarly, the planning process, the activities and tasks involved in planning for the development of a TSC data base environment are the same as those followed for the development of an ADP plan for the TSC.

As a subset of the overall ADP plan, the data base plan will rely heavily on the goals and objectives set forth for the TSC. A formal statement of the objectives of the TSC data base environment may be found in the Program Plan for Management of Transportation Statistical Data and Information (MODS) which, when followed by the formulation and performance of the tasks presented within, will bring the stated goals and objectives to fruition. At this stage in planning process, the

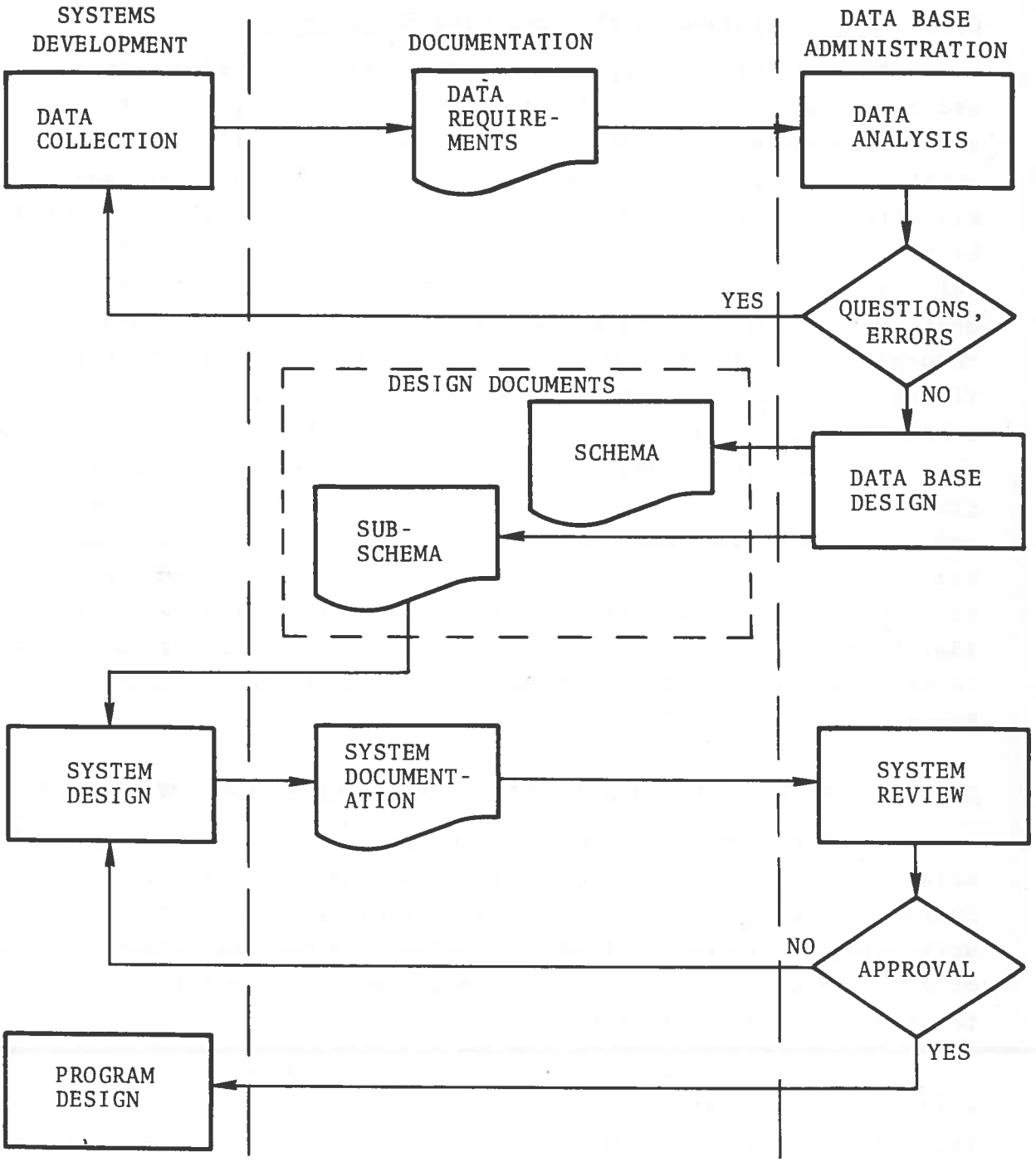



Figure C-2. System Design in a Data Base Environment

### C.1.5 Administration of the Data Base Environment

The data base concept implies the need for standardization and coordination of data definition, capture, storage, maintenance, and dissemination, as discussed in Section C.1.1. In a data base environment, this coordination is manifested in a data base administration function. The Data Base Administrator, as represented by the Systems Manager of the Operating Group, is responsible not only for the coordination and control of the data base environment, but also for the establishment of the standards and procedures for coordination and control. A successful data base environment in the Transportation Systems Center (TSC) will demand extensive communication between the users of the data base and the data processing Operating Group. Within the data processing group itself, clear communication will be required between the operating functional entities of Systems Analysis, Programming, and Operations and the various support functions (such as planning, education, documentation, etc.). Therefore, the data base administration responsibility should be placed as a staff function in an organizational position reporting directly to the Center Program Director for Information.

## C.2 FUNDAMENTAL STRUCTURE OF THE OPERATING DATA BASE ENVIRONMENT

The long range plan for the TSC data base environment is actually a subset of the overall ADP plan for the Transportation Systems Center (TSC). Similarly, the planning process, the activities and tasks involved in planning for the development of a TSC data base environment are the same as those followed for the development of an ADP plan for the TSC.

As a subset of the overall ADP plan, the data base plan will rely heavily on the goals and objectives set forth for the TSC. A formal statement of the objectives of the TSC data base environment may be found in the Program Plan for Management of Transportation Statistical Data and Information (MODS) which, when followed by the formulation and performance of the tasks presented within, will bring the stated goals and objectives to fruition. At this stage in planning process, the

formulation of major project areas provides enough detail to allow the next step of resource estimation and allocation to be performed. In most cases actual implementation will require further detailed breakdowns of the project areas into task activities in moving from the general to the specific. Finally, the tasks must be phased into chronological sequence to insure orderly development of the TSC data base environment within the overall content of the Program Plan. Figure C-3 summarizes these basic steps of the planning process for the TSC data base environment.

The following paragraphs are devoted to detail discussions of the tasks involved in the development of the TSC data base environment, as summarized in Figures C-4 and C5.

### C.2.1 Define Goals

Development of the data base environment is based on the general statement of goals to be achieved as stated in the DOT TSC ADP long-range plan, and the DOT/TSC MODS Program Plan.

### C.2.2 Establish the Data Base Administration (DBA) Function

An integrated data base environment demands thorough internal TSC coordination. Coordination requires standards and procedures and a centralized functional responsibility to administer these standards and procedures. This function is defined as the Data Base Administration function. A key responsibility of the DBA function, then, is the coordination and control of the on-going data base environment. An important aspect of this coordination activity involves the development of the data base environment, including establishment of standards and procedures, analysis of requirements, and the evaluation and selection of the software for data base environment (all of which are subsequent tasks). These functions logically fall within the responsibilities of the DBA function, and therefore the establishment of the DBA function as a responsibility of the System Manager of the Operating Group in the TSC data base environment is essential. Further discussion of the DBA function is presented in Section C.6 of this appendix.

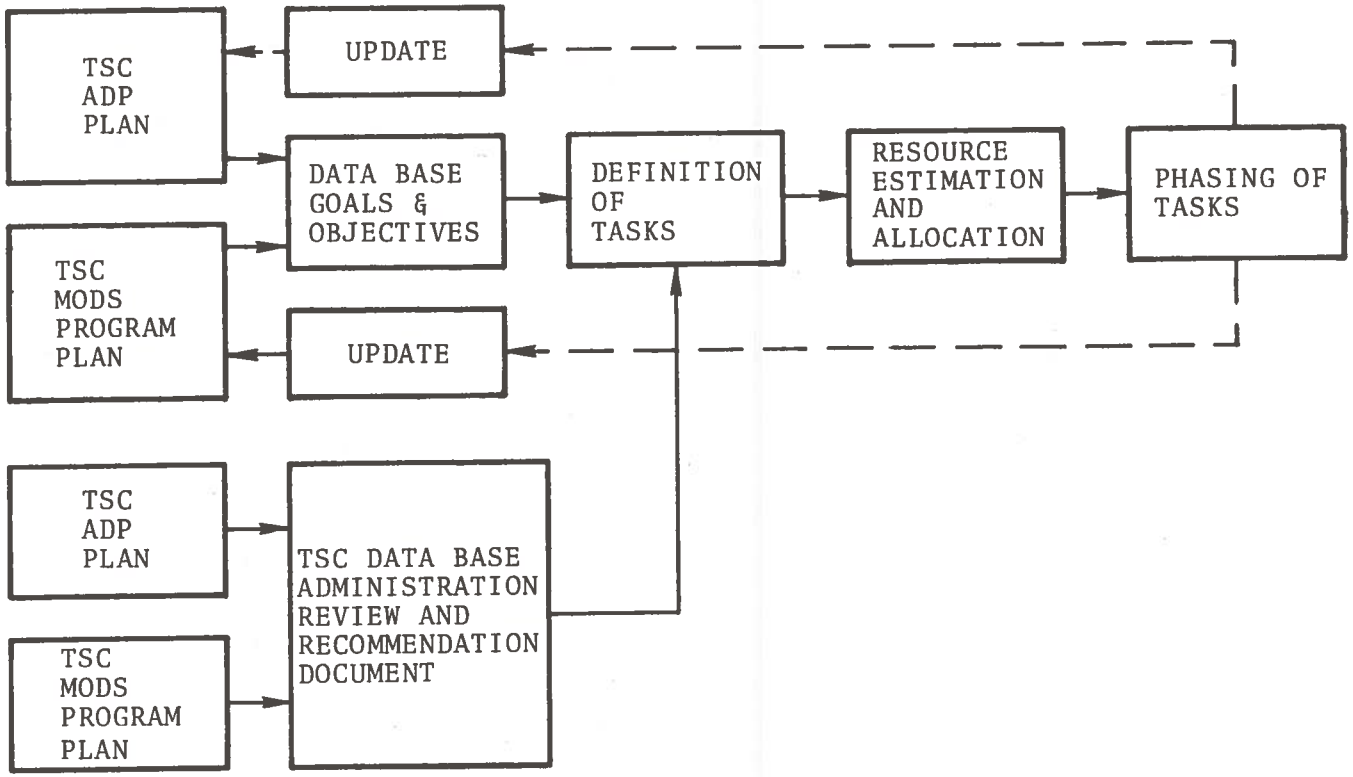



Figure C-3. Planning for the Data Base Environment

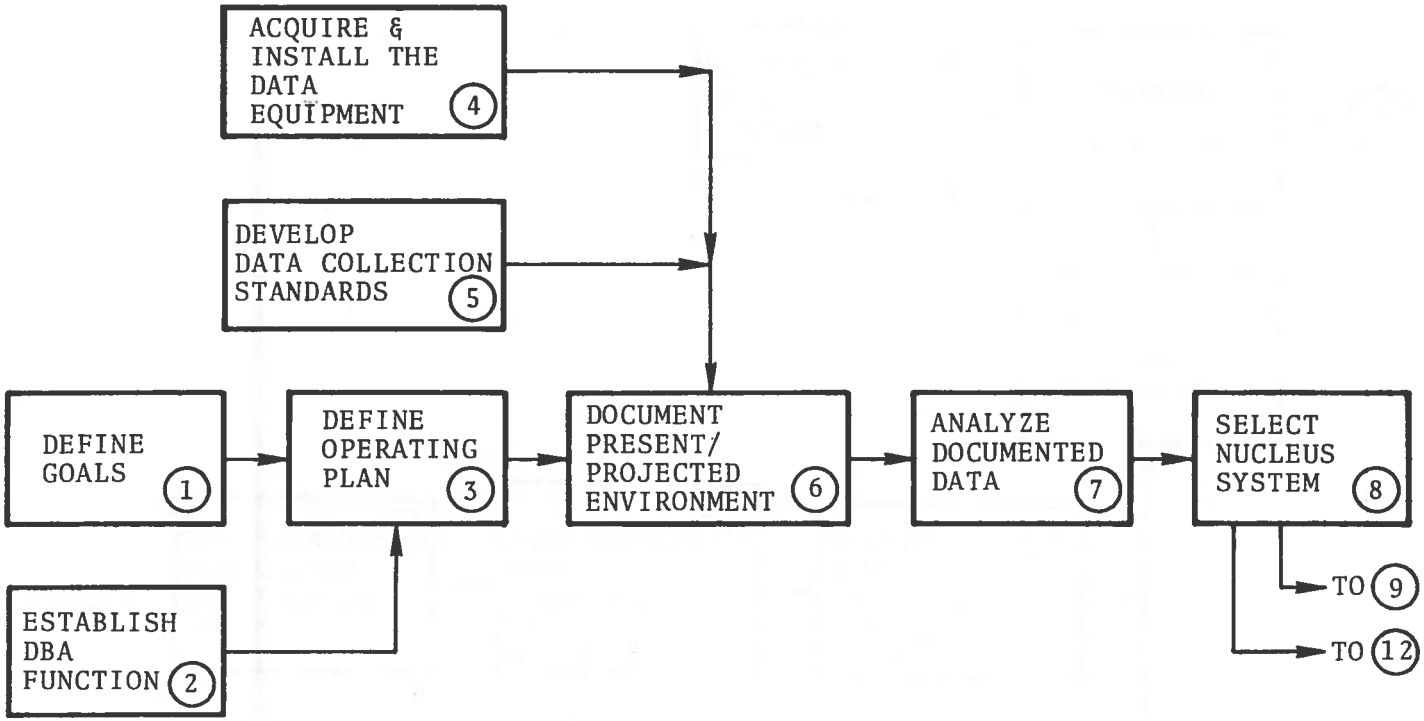



Figure C-4. Technical Development Sequence for the TSC Data Base Environment (Part 1)

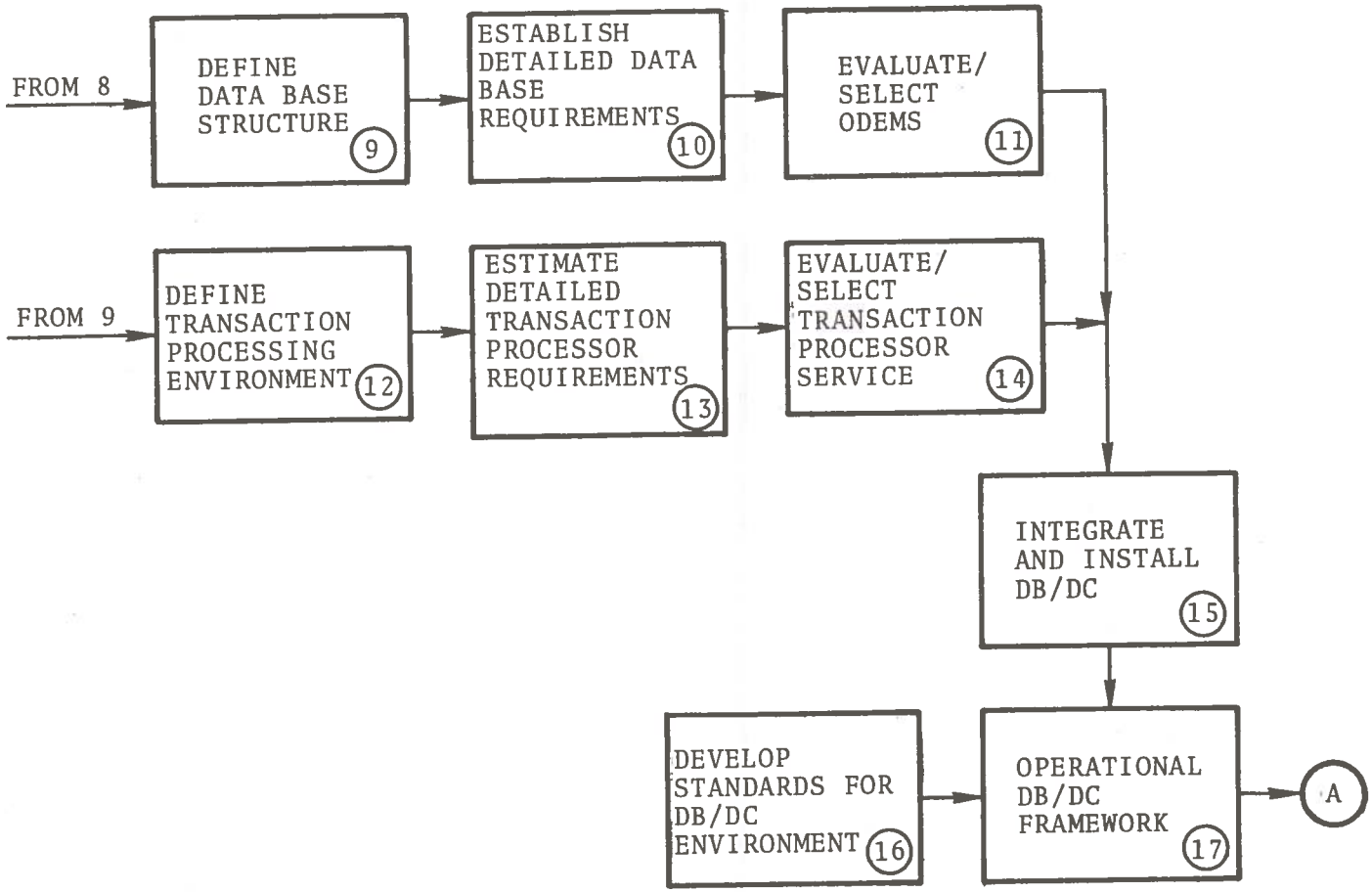



Figure C-5. Technical Development Sequence for the TSC Data Base Environment (Part 2)

### C.2.3 Documentation of the Data Base Administration Requirements

Documentation of the data base administration requirements will address the tasks and phasing required to establish the data base environment, as identified in the definition of goals. These requirements identify the various projects that will result in an operational integrated data base environment. The technical development plan defines specific projects and resources to implement the data base administration requirements.

### C.2.4 Develop Data Collection Standards

A key task in preparing for a data base environment is the identification and documentation of the existing, and the scoping of projected, organizational information needs. This task is the initial effort that is continually refined following establishment of the initial operating capability and is described in greater detail in Section C.4 of this appendix.

### C.2.5 Acquire & Install Data Dictionary

One of the most important tools for the coordination and control of the data base environment is the Data Dictionary. The Data Dictionary, if implemented early, can also be used in the coordination and control of the development process. Thus, the dictionary can be used to facilitate establishment of the standards and procedures to be employed in describing data within the existing and projected environment. These standards and procedures must address the methods of collecting information, the forms to be utilized, the data entities to be described, the data attributes to be described for each entity, and the transactions that operate against these data entities. Establishment of the data dictionary and these standards and procedures is the responsibility of the Systems Manager.

The TSC data dictionary is discussed in greater detail in Section C.5 of this appendix.

### C.2.6 Document the Present and Projected Data Environment

Utilizing the standards and procedures established in the task described above, the Systems Manager will coordinate the process of collecting information about data as it currently exists or is known to be desired. This collection effort may extend at a later date to the total data environment of the DOT or beyond, or to a defined subset of the data environment (e.g., one or more of the functional areas of the DOT). Any effort to initially extend the defined subset of the total data environment will require that this same process be the initial task for all future plans to integrate additional functional areas into the data base environment.

### C.2.7 Analyze Documented Data

Analysis of the data documented in the preceding task will serve to identify common areas of information by use, to identify data redundancy and inconsistency, and to provide a perspective of transaction flow and application interrelationships. This analysis will form the basis for selecting a nucleus application area that will represent the first system development effort under the integrated data base environment. The general processes that are recommended for such an analysis include the following:

a) Cross-reference of data items by file and report

b) Cross-reference of data items by user, including identification of access characteristics (response requirements, access purpose, etc.)

c) Transaction documentation, cross-referenced to files and including both input and output events.

It should be noted that the use of the data dictionary system will assist in the processes of relating data items to files and reports and of relating data items, files, and reports to users/applications.

### C.2.8 Select the Nucleus System

Identification of a nucleus system serves two purposes:

a) defining the initial application(s) to be implemented in the data base environment; and,

b) providing a basis for a detailed requirements analysis in preparation for a selection of a Generalized Data Base Management System (GDBMS) package and a Transaction Processor (TP) package.

Guidelines for selecting a prototypical application should include the following:

1) the selected application should not be overly large or complex.

2) the selected application area should have a potential for a noticeable reduction of data redundancy and inconsistency.

3) the selected application should have interrelationships with other application areas.

4) the selected application should be implementable within a reasonable time frame and involve an area with high visibility within the organization.

Once the nucleus system is selected, efforts in the selection and acquisition of a Generalized Data Base Management System and a Transaction Processor System may begin. Both tasks may proceed in parallel.

### C.2.9 Define the Logical Data Base Structure

Utilizing the data collected and analyzed in earlier phases, the Systems Manager will combine data items into logical groupings and combine these logical groupings, or records, into a general level data base structure, relating the records in the hierarchical and/or network relationships in such a way as to achieve the lowest possible level of data redundancy.

### C.2.10 Establish Detailed Data Base Requirements

Examination of the various transactions and activities imposed against the data base structure will provide the Systems Manager with the data necessary to develop the selection criteria for a Generalized Data Base Management System (GBDMS). Factors to be considered in this analysis include, by transaction or activity:

a) transaction volume/frequency

b) response time requirements

c) access key requirements

d) processing mode (random, sequential, serial)

e) records processed per transaction (one, all, subset)

In balancing the requirements of all transactions against the data base, the Systems Manager will be able to select the optimum access method and storage strategy for each record, or record group, in the proposed data base structure. He should assume at this point of the analysis that all access methods and storage strategies are available. The results of such an analysis will be a desired set of capabilities which can then be translated into selection criteria for a Generalized Data Management System.

### C.2.11 Evaluate and Select a Generalized Data Base Management System

Utilizing the selection criteria developed in the previous task, the Systems Manager can, if necessary, develop a Request for Proposal to be delivered to the various Generalized Data Base Management System vendors, or move to implement on one of the systems already available to TSC. In the selection process, the Systems Manager (or the evaluation team) will take into consideration the alternative of in-house development. In addition to a feature analysis, attention should be given to the following:

a) Ease of implementation

b) Ease of use

c) Resource requirements

d) Vendor Stability

C.2.12 Define Transaction Processing Requirements

The analysis of transaction processing requirements and the selection of a Transaction Processing System may proceed in parallel with the evaluation and selection of a Generalized Data Base Management System. Definition of transaction requirements will, in this phase, include identification of transaction characteristics, analysis of transaction response requirements, and determination of transaction volumes and frequencies.

C.2.13 Establish Detailed Transaction Processing System Requirements

In this task, the requirements of a Transaction Processing System will be defined in detail. These requirements will address the following:

a) Transaction load

b) Priority scheme

c) Scheduling (of message processing programs)

d) Message switching

e) Logging

f) Transaction recovery and security

g) Terminal suport

C.2.14 Evaluate and Select a Transaction Processing System

Utilizing the requirements established in the previous task, a Request for Proposal will be issued to prospective vendors, unless one not currently available to TSC satisfies these requirements. Consideration will also be given to the alternative of in-house development. In addition to a feature analysis, the selection process should also address:

a) Ease of implementation

b) Ease of use

c) Resource requirements

d) Vendor stability

### C.2.15 Integrate and Install the Generalized Data Base Management System and the Transaction Processing System

In this phase, the Generalized Data Base Management System (GDBMS) and the Transaction Processing (TP) System will be integrated, tested and accepted. The Systems Manager will be responsible for establishing the testing procedures and acceptance criteria.

### C.2.16 Develop Standards and Procedures for the Data Base Environment

After selection of the software components for the data base environment, and in conjunction with the integration and installation of these components, the Systems Manager will begin the development of standards and procedures for managing and controlling the environment. These standards and procedures will encompass the following areas:

a) Data collection and analysis - standards and procedures for collecting data on new applications selected for integration into the data base environment and for analyzing this data in relation to the characteristics of the data base environment.

b) Data dictionary - standards and procedures for defining data items, data records, data structures, programs, and systems in the data base environment.

c) Data identification conventions

d) Validation rules

e) GDBMS usage - standards and procedures for using the features of the GDBMS, directed toward the users of the GDBMS.

f) TP System usage - standards and procedures for using the features of the Transaction Processing System, directed toward the users of the Transaction Processing System.

g) Data base design - standards and procedures for use in the design and maintenance of data base structures. These standards and procedures will include:

- access strategies
- storage strategies
- design considerations (structuring)
- restructuring/redefinition

h) User chargeback - standards and procedures for allocating the costs of the data base environment to the user.

i) Security/Integrity - standards and procedures for assigning and controlling data base access authorization.

j) Activity monitoring - standards and procedures for monitoring activity in the data base environment at both the Transaction Processing System and GDBMS levels.

k) Application design specification - standards and procedures for documenting application design specifications.

l) Application documentation - standards and procedures for documenting an operational application.

m) Testing - standards and procedures for testing and accepting new applications in the data base environment.

n) Operations - standards and procedures for operating the data base environment.

o) Environment maintenance - standards and procedures for system software maintenance, including the Operating System, the Transaction Processing System, and the GDBMS.

p) Environment documentation - standards and procedures for documenting the data base environment and its hardware/software components.

Upon completion of the above described tasks, the data base environment may be considered operational. It is recommended that at this point, design, development, and implementation of the nucleus system, identified during the establishment of the data base environment commence. Development of the nucleus application and all future applications within the data base environment will be coordinated by the Systems Manager and will conform to the standards and procedures established under the previous task.

In summary, the fundamental structure to the operating data base environment provides the blue-print for its development and on-going operation. The remaining sections of this appendix provide procedures and standard practices recommended for implementation in the TSC data base environment.

## C.3 DEFINITION OF DATA

A methodology for formally defining TSC data is essential for the coordinated and controlled implementation of a data base environment. The process of defining data will contribute significantly to an understanding of the current TSC environment by providing information relative to the sources of data, the contents of various data collections, and the usage of data. Furthermore, the data definition activity will facilitate the process of projecting and ascertaining long-term information requirements. Finally, the process of data definition itself is an integral element of the system's design effort directed at fulfilling these long-term information requirements.

A distinct advantage for TSC lies in utilizing a uniform methodology for data definition in terms of coordination and control, since the DOT universe of data resources is a somewhat heterogeneous collection of data describing a myriad of different entities and relationships. In some cases actual data collections will exist within the DOT processing environment; in other instances, collections of data will exist externally to the DOT environment and will merely be described and referenced in the data base. Establishing a central control over these resources demands a precise and consistent method for describing and defining the nature and contents of all data resources. Thus, regardless of whether we are dealing with a textual collection of data, a time series, tabular material, or a description of a data collection available elsewhere, uniform procedures for defining the data resources are required. These standards, reflecting the availability and utilization of data, will support the coordination, control data collection, storage, maintenance, and dissemination of data; i.e., the function of data base management.

### C.3.1 Data Entities

The basic foundation for a uniform methodology of data definition is the hierarchy of data entities which comprise the universe of data resources. A data entity is basically a unit of data, formally defined as a named collection of data. This structure of data entities forms the basis for defining collections of data and subsequently referencing them. Definitions for each entity follow below:

a) Data Base: A named collection of units of physical data, organized and related to each other in a specified manner.

b) Data Area: A named sub-division of the physically addressable storage space in the data base, which may contain occurrences of records/sets (or parts thereof) of various types. Also referred to as a Data File or Data Collection.

c) Data Record: A collection of aggregates, groups or segments that are processed by application programs. Named collections of these records are referred to as sets.

d) Data Aggregate: A collection of data elements normally referred to one at a time; also referred to as Data Segment or Data Group.

e) Data Element: Informational units with unique meaning and sub-categories (data items) of distinct units or values; also classes of data, with data items as members of the class (e.g., Date).

f) Data Item: A single named data entity containing no substructure; the lowest level of addressable data in which data value(s) are physically stored. Also referred to as a Data Field (e.g., Mo - Day).

It should be noted that in describing a particular collection of data one need not use each and every level of the data entity hierarchy. In general, from a procedural viewpoint the top and bottom levels of the hierarchy are identified and then appropriate intermediate levels are labeled. A very common example of the data element is found in the use of the "date". Usually, a date is represented as month, day, and year (mmddyy); each containing two digits. Each of these two digit units of data are referred to as data items, whereas the three collectively are referred to as a data element.

### C.3.2 Data Attributes

The attributes, or characteristics, of data entities comprising the data base are used to specifically define the contents of the data base. These data attributes are essentially the contents of the data dictionary. An entry in a data dictionary is comprised of all the data attributes which define a specific data entity. In order to establish the proper

perspective for the ensuing discussion of guidelines and procedures for data collection and analysis, a general classification and description of these data attributes is presented. Subsequently, when presenting the guidelines and procedures relative to the data dictionary, the subject of data attributes will be further expanded.

Data attributes, or entity descriptors, may apply to all or some of the data entities in the hierarchy. For example, the attribute "name" applies to all data entities, whereas the attribute "source" could only apply to the level of the hierarchy being addressed (e.g., data collection). Thus, the following classification of data attributes applies to varying levels of the data entity hierarchy.

a) Identification of Data - This class of attributes provides for unique identification of each data entity. Specifically, names, short designators (labels), textual descriptions and synonym names would be specified.

b) Source of Data - This category relates to the original source of the data entity being described. Organization, documents, and algorithms are examples of sources from which data may be obtained. If a data entity originates from a transaction, then that transaction would be identified.

c) Type of Data - Data entities may be quantitative or qualitative in nature. Data that is textual should be specified as such, as opposed to other data collections which are quantitative measurements and statistics. Types might include: amount(s), code, name, percent, etc.

d) Use of Data - The mode of representation and the media upon which the data entity is stored falls into this class. In this category both internal and external users of the data entity are also recorded. This includes programs, reports and, if appropriate, end users.

e) Qualification of Data - This category indicates the conditions or context in which the data entity is furnished for general use. The accuracy of quantitative entities would be indicated and the time dependencies such as frequency of update or age would be included. Of primary importance also are the accessing rules and edit/validation constraints.

f) Relationships of Data - This category addresses the relationship of each data entity to other entities in the hierarchy. Membership in data bases, data collections, records, groups, etc., would be specified.

In summary, the foundation of TSC guidelines and procedures for the definition and specification of a data base environment is this uniform methodology for describing collections of data. The hierarchy of data entities, and the data attributes used to describe these entities, will provide the basis for the guidelines and procedures for data collection, development of a data dictionary, specifying data structures and for providing administrative control over the data base environment.

## C.4 DATA COLLECTION

The first task in the phased development of a data-base-supported information system at the TSC is the collection of data relative to the current information handling environment. The objectives of data collection will be to:

a) Provide basic information which will facilitate the analysis of the current environment;

b) Provide a basis for determining data relationships -- a pre-requisite to the data base design and structuring process;

c) Provide definitions of data entities which are essential for input to the data dictionary.

Thus, the data collection process is intended to provide a base of information for use in data source identification, systems analysis and data base design. It will further serve to identify possible data redundancies, data sets and their contents, and ultimately the requirements of users and/or systems for access to units of data. As such, the information collected should be considered as necessary to the definition of data base contents, and data base structure.

There are five basic tasks to be performed in the data collection process, as follows:

1) Identification of Data Sources

2) Documentation of Data Sources

3) Data Reduction

4) Data Analysis

5) Assessment of Future Requirements

The remainder of this section is devoted to the presentation of guidelines and procedures for the performance of these five data collection tasks.

### C.4.1 Identification of Data Sources

The first step in collecting data relative to current information handling is the identification of the sources from which TSC information is acquired. The next important step is to identify unique characteristics which may affect data "capture" from these sources and to emphasize the specification of the appropriate data attributes which apply to the adequate description of these sources of data.

The identification of data originating within DOT is facilitated by the greater degree of centralized control which exists, thereby making information about these collections of data more readily available. However, for externally-generated data which is to be captured in total, summarized, or referenced within the TSC Data Base, some difficulty in providing for centralized

control and in specifying some of the data attributes relative to these collections of data can be anticipated. For example, collections of data generated by the use of survey methodology should include the type of methodology employed in order to facilitate comparisons and other types of analysis. Methodology information would be readily available for DOT- conducted surveys. However, for surveys conducted by other agencies or private organizations, methodology information would likely not be so readily available and its acquisition would have to be emphasized. If such qualifying information is not made available to users of the TSC data base, requestors will tend to go to the primary sources for it, thus diminishing the effectiveness of centralized control over transportation data.

### C.4.2 Documentation of Data Sources

The documenation of the data sources identified in the previous step represents the initial phase of the development of a TSC Data Dictionary. This initial phase of Data Dictionary development will take the form of a Data Catalog of major sources of transportation data.

The Data Catalog will be comprised of data attributes describing the transportation data available from various sources. Since the objective of the Data Catalog is to increase visibility and accessibility through centralized control, the data attributes chosen will emphasize those characteristics which contribute to increased visibility and accessibility. Thus, for example, in documenting a source both the name and address of the supplying organization will be included, as will special attributes specifying availability of data collections.

It should be noted that the data attributes of the Data Catalog apply to only one level of the data entity hierarchy. It is in this respect that the Data Catalog represents an initial step toward a full-blown Data Dictionary. The Data Catalog addresses transportation data at the collection/file/area level of the hierarchy, thereby providing,at the earliest possible stage, an index and reference for users of the TSC Data Base.

The data attributes for the Data Catalog are identified in paragraphs C.4.2.1 thru C.4.2.6 to highlight the desirability for this type of information. It is recognized that data for all of these attributes may not be readily available. However, the absence of data for an attribute should not preclude inclusion of the attribute, since for subsequent entries such data may be available.

Paragraphs C.4.2.1 thru C.4.2.6 are organized according to the categories of data attributes presented in Section C.3.

C.4.2.1 Identification of Data

a) Name - The official title of the data collection.

b) Label - The official label assigned to the data collection. This label is usually in the form of an "identification number." It should be noted that there may be more than one label, corresponding, for example to a DOT assigned ID number and a number assigned by the source organization, it is external.

c) Classification - A general coded category specifying the nature of the data collection. For example, if all transportation data were classified into categories of Mode, Source, Geographic Scope, and Commodity (Goods), then a data collection might be classified by using a four part code corresponding to these categories.

d) Abstract - A textual description of the general content of the data collection. The abstract should be worded very carefully so as to avoid ambiguities. Additionally, this attribute should include anything deemed to be unique about the data collection which was not specified in other attributes.

e) TSC Dictionary Reference - A cross-reference to the TSC Data Dictionary entry describing the data collection. The reference should be in TSC standard format/terminology to be developed for the dictionary.

f) EDP Labels - The label assigned to this data collection for the purpose of referring to it in an EDP environment. For example, the Data Set name or the File name.

C.4.2.2 Source of Data

a) Source - The name and address of the organization from which the data collection is available. The entry in this case should be specific enough to allow the user to make direct contact with the source (if this is appropriate).

b) Source Documentation - A reference to, and a description of, documentation supplied by the source organization about the data collection. Such documentation might be relative to the content (data item names & format), physical attributes/characteristics of data items, and various access or usage rules. It should be noted that in essence this entry is a reference to Data Catalog type information available in the source organization.

c) Entry Source - For internal DOT data collections derived from external sources, the entry source document should be identified. For example, if special forms were prepared from externally-provided data, these forms should be identified. The organization or individual responsible for the preparation of these documents should also be specified.

C.4.2.3 Type of Data

This category of data attributes does not apply to the specification at this level of the data entity hierarchy; i.e., data collections. It refers primarily to the lowest three levels of the hierarchy.

C.4.2.4 Use of Data

a) Availability - A statement of availability of the data collection, including limitations and/or constraints.

One type of constraint would be the access rules established to ensure the privacy (unintentional breach) and security (intentional breach) of the TSC data base.

b) Media - A description of the physical storage media in which the data collection resides.

c) Representation - A description of the coding scheme used to represent the information in the physical media. Also of interest would be the amount of space (including record length, block size and format) taken on the media to represent the data collection, since this may have a bearing on the request. Additionally, a description of how the data collection is represented in the output would be useful.

d) Hardware/Software Considerations - A list of the hardware and software factors which may effect the acquisition or use of the data collection. Among these would be the computer on which the data collection was generated, the system software (e.g., OS/370) under which the data collection was created, and any other special software requirements (e.g., data base management system constraints, or query language requirements).

C.4.2.5 Qualification of Data

a) TSC Access Control - A reference to the TSC Standards and Procedures qualifying the usage of the data collection.

b) Access Keys - Identification of the control and/or access keys of the data collection. For example, if a time series of commodity transportation data could be accessed by the type of goods transported, this data item would be identified as an accessing key.

c) Acquisition Cost - An indication of the price or charge for the use of or access to the data collection. If TSC were to charge for their services this would apply to internal as well as external data sources.

d) Content Integrity - An indication of the level of confidence a user might place upon the data collection. It may be advisable to develop a confidence coefficient based on a combination of pertinent factors, such as:

1) collection methodology; e.g., observation, interview, mail questionnaire.

2) scope or completeness of data; e.g., sample versus universe, summary versus detail.

3) validation rules applied.

4) for external sources, an indication of services level and support.

C.4.2.6 Relationships of Data

a) Data Organization - Description of the sequencing and/or partitioning of the data collection. If the data collection is organized as a time series, this would be noted. For large data collections that include indices, these would be described.

b) Data Structure - Description of membership in a data base (external or internal). Some unique aspect of the data structure within the data collection might also be included.

C.4.3 Data Reduction and Analysis

In practice, these last two steps of the data collection process will be performed concurrently. The activities of data reduction and analysis will involve the use of special forms to assist in reducing the information collected about information handling into a manageable form. Thus, proper use of these forms will involve a certain degree of data analysis. Additionally, at this time, as assessment of future TSC information needs would be made. This can be accomplished through intrviews with users and observations of current usage patterns. Thus, the data

collection phase of the data base development effort will provide the foundation for a data-base-supported information system environment based on documented requirements and future needs.

Special forms projected for use in data reduction and analysis activities are discussed in Subsections C.4.3.1 and C.4.3.2, following. The assessment of future needs is covered in Subsection C.4.3.3.

### C.4.3.1 The Data Cross-Reference Matrix Form

This matrix can be used as a tool to assist in the reduction and analysis of data collected about current information handling. The purpose of the form is to provide a convenient vehicle for the identification of the sources, contents and users of all data collections and/or reports in the current transportation data environment at TSC. The form would be filled with information obtained from the Data Catalog and from the Data Dictionary (relating to data aggregates, elements and items).

The information entered on the completed form serves several purposes:

a) Identification of those data entities that appear to be redundant, in that they are substructures of multiple data collections, or components of multiple reports. It is important to recognize, however, that this form provides only a preliminary identification of redundancies (specifically, those data entities identified by a common name). Ultimate determination of redundancy will depend upon a direct comparison of data entity attribute descriptions in the Data Dictionary.

b) Definition of all user and function reference and/or access to data collections and reports for further analysis.

c) Identification of users and/or functional areas that may have access to, or receive, multiple reports containing essentially the same data entities, with perhaps only

a 10-20% uniqueness factor. Often, this discovery may result in a direct reduction in processing required.

d) Identification of multiple users of the same data entities. This will provide an advantage in exercising centralized and coordinated control over the transportation data resources at DOT.

Figure C-6 depicts a typical layout for the Data Entity Cross-Reference Matrix. Note the significance of the intersections of each row and column of the matrix. The existence of an intersection is indicated by filling in the intersect box with a user's code showing who the user is of that data collection (column) containing the particular data entity (row). If a data entiry (or group of data entities) appears in more than one intersection there may be concern for redundancy. The stipulation of multiple users in intersect boxes indicates shared use of data and a possible need for coordination and control. Additional information can be put on this type of form (in the intersect boxes) relative to the functional access and response time requirements for particular users for specific data entities appearing in certain data collections. Considerations relevant to functional access and response time requirements are:

a) Functional Access - The following six functions will be used to indicate various levels of access to data entities and/or their subdivisions.

1) Global - The right to change the contents of any data item, including control keys. Thus, for example, with such access rights the criteria for retrieval for an entire data collection might be altered.

2) Delete - The right to unconditionally delete a record or set of records from the data base. This includes the right of deletion of a unique occurrence, of all its member records, and of all the occurrences of the specific record-type. With this access right a user might delete an entire data collection.

| DATA USERS | TRANSPORTATION DATA | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NETWORK INVENTORIES | CAPITAL EQUIPMENT | FLEET DATA | SERVICE COST INDICES | EMPLOYMENT | DEMOGRAPHIC DATA | ECONOMIC DATA | PASSENGER VOLUME | FREIGHT VOLUME | FARE/ COST DATA | TOTAL REVENUES | SYSTEM OPERATION IMPACTS | PERFORMANCE MEASUREMENTS |
| | S | | | | | | | | | | | | | |
| | TPI | | | | | | | | X | X | X | X | | |
| | TES | | | | | | | | | | | | X | |
| | TST | X | X | X | X | X | | | | | | | | X |
| | TAD | | | | | | | | | | | | | |
| | TSC | X | X | X | X | X | X | X | X | X | X | X | X | X |
| | CG | | | | | | | | | | | | | |
| | FAA | | | | | | | | | | | | | |
| | FRA | | | | | | | | | | | | | |
| | FHWA | | | | | | | | | | | | | |
| | UMTA | | | | | | | | | | | | | |
| | SLSDC | | | | | | | | | | | | | |
| | NHTSA | | | | | | | | | | | | | |

Figure C-6. Tentative Layout of Data Entity Cross-Reference Matrix

3) Remove - The right to remove a single occurrence of unique record-type. Deletion from the data base is not implied.

4) Modify - The right to change the contents or value of data items, epecifically excluding control keys.

5) Add - The right to add a new data entity (occurence of data) to the data base.

6) Inquiry - The right of inquiry (retrieval only) to data entities in the data base.

It should be noted that these six types of functional access are listed from highest to lowest in terms of access rights. Thus, from a procedural viewpoint it is only necessary to specify the highest applicable access right, thus implying the lower types of access rights.

For the TSC data base, it is important to recognize that for transportation information stored and maintained externally, the access rights will usually be limited to inquiry. However, for the reference maintained internally in the TSC data base, it will still be necessary to specify higher levels of access rights. Thus, for example, for Air Cargo, Airport Activity Statistics maintained by the Federal Aviation Administration, the access rights will be inquiry. However, for the reference to this data collection maintained in the TSC data base, the access rights would necessarily be higher.

b) Response Time Requirements - Response time requirements can be indicated for particular users relative to specific data entities by also including this type of information in the intersect boxes of the Data Entity Cross-Reference Matrix. A simple three-level breakdown of possible response requirements seems appropriate for the TSC data base, as follows:

1) Immediate - Less than three minutes. This would generally be considered a real-time requirement and would be quite rare in the TSC environment. It is included for completeness.

2) Rapid - Under thirty minutes. This response requirement would not necessarily indicate an on-line environment. For the TSC it is more likely that this would call for a centralized operation at a remote site for the processing of requests and the dissemination (possibly manually) of information to users.

3) Scheduled - A scheduled response might range from the same day to overnight or even longer.

From a procedural viewpoint, in filling in the Data Entity Cross-Reference Matrix with functional access and response time requirements, it would be useful to use codes as a shorthand method to make these entries. Thus a coded entry might look like, "4-S", (using the level of access of the function and the first letter of the response time) indicating the need to add new record(s) to the data base on a scheduled basis.

C.4.3.2 Data/User Matrix

Another form which will be useful in data reduction and analysis is the Data/User Matrix. This form is essentially an inversion of the Data Entity Cross-Reference Matrix: the users on the horizontal axis (columns) instead of the data collections and/or reports. The body of the matrix (intersect boxes) should be used primarily to indicate the functional/response requirements mentioned earlier. The completed Data/User Matrix will be most useful in providing the data base designer with an overview of user data requirements, (in a consistent, consolidated format,) suitable for performing analyses which can be used in the development of an appropriate data base structure for the TSC environment. In addition, the information obtained will provide the basis for an effective access control procedure (security) for protecting the TSC data base against deliberate or inadvertent access and/or destruction.

### C.4.3.3 Assessment of Future Needs

Possibly the most important aspect of the data collection process is the assessment of future requirements and its effect on the TSC information system. This need for assessing future requirements will be essentially fulfilled by conducting a program of data gathering based on interviews. The objective of the program will be to ascertain the nature and type of service the TSC information system will be required to provide for its users. Good interviewing techniques, of course, are essential to assure that the data gathering process is effective. The following guidelines for use in the interviews have been prepared to assist in achieving such effectiveness.

a) Plan and schedule interviews well in advance.

b) Prepare several key questions prior to the interview.

c) Respect the subjects position, time, experience and opinions.

d) Listen for the significance of what the subject says, not just what he says.

e) Document the interview as soon as possible. Don't write the documentation during the interviews. A generalized form should be utilized to ensure recording of key data.

f) Always be sure that the subject's manager is aware of the occurrence and purpose for the interview.

g) Don't believe everything you hear, particularly a finding which appears significant. Separate, but document, opinion as well as fact.

h) Cross-check by interviewing the "receiver", as well as the "sender", of information.

i) Review or copy significant documents during the session to avoid having to return. Identify documents with subject.

j) Recap your findings with the subject to ensure complete understanding.

k) If working in teams, take the time daily to review your mutual findings. This will help in ensuring complete coverage and cross-checking, and in establishing direction for subsequent interviews.

l) Ask the question, "What would you do if you were higher management?"

m) Don't discuss preliminary conclusions with anyone other than the team.

n) Don't get involved with day-to-day problems.

o) The subject will tell you only what he wants you to know and avoid all areas which may be potentially embarrassing or incriminating. Therefore, he must be assured that all comments are confidential.

p) Deep-seated resentment can be developed against outsiders who imply things, or who convey an attitude of being out to "clean up this mess" -- the subject may have been instrumental in creating the "mess".

### C.4.4 Summary

In summary, the data collection process is the first step in developing a centralized information utility function for the handling of transportation information at TSC. Information gathered in the data collection phase will provide insight into the current environment and help mold the design of the future data supported information system for TSC.

## C.5 DATA DICTIONARY

The data dictionary is designed to be the primary tool to be used by the TSC to achieve centralized control over its data resources. The data dictionary will be the single authoritative source of definitive information on TSC data entities, their use and their interrelationships, in a standard organization and

format. Through implementing the data dictionary early in the development of a TSC data base environment, TSC will have a means of monitoring and controlling data resources without actually having possessed integrated and centralized data itself. Instead, at the outset, information about data entities will be integrated and centralized in a single system and will be available to end users, the data administrator, and systems designers for their respective uses. Subsequently, in later phases, as the TSC data base environment matures, the data dictionary functions will gradually expand to include automatic control and monitoring.

### C.5.1 Functions of the Data Dictionary

The data dictionary will function as a tool in: (1) carrying out the activities of data base development, and (2) serving the central control mechanism in the on-going environment, which will include a requirement to communicate to the outside world information on the size and sources of transportation related data bases through publication of the Data Catalog. Included among the activities which the data dictionary will support are:

a) Data Collection

b) Data Analysis & Reduction

Among the function an on-going data dictionary will be:

c) Data Documentation

d) Data Standardization

e) Data Dictionary Publication

#### C.5.1.1 Supporting Data Collection

The data collection process is predicated on the clear and concise definition of current and future information needs in the TSC environment. This can only be accomplished by using the data dictionary as a tool to maintain clear specifications of the data entities in the TSC environment and the access requirements. In order to design the appropriate flow of information for the future the system designers will rely on the data dictionary to

provide them with specifications and descriptions of all data entities currently in the TSC environment. Thus, the data dictionary should be established as the data collection process begins so that it can serve as the central control point for data description and specification, throughout the entire data base program. Development of the Data Catalog is the initial step in this direction.

C.5.1.2 Supporting Data Analysis and Reduction

The data dictionary will provide TSC analysts and data base designers with a mechanism for detecting inconsistent and/or redundant data entities. In general, it should be noted that there are two types of redundancies. The first is technical redundancy which is knowingly built into data collections because of existing technological tradeoffs and the need to relate sampled data to its universe. Data base management software attempts to minimize this type of redundancy in order to conserve on storage; the data dictionary will serve to maintain consistency between duplicated data entities.

The second type of redundancy, the inadvertent type, should never be tolerated and where possible should always be eliminated when detected. The data dictionary, by providing precise specifications of data entities, will facilitate the comparison of suspected areas of redundancies. Furthermore, the data dictionary should have the capability to highlight these suspected areas. This can be accomplished by utilizing standard conventions for specifying data entity designators and using Key Word Out of Context listings to detect areas of similarity.

C.5.1.3 Data Documentation

In an on-going sense the data dictionary will provide for certain documentation and procedural aspects of data definition. Thus, the data dictionary will be the repository for information concerning all data attributes. Included will be: the sources and users of data collections; the location and availability of data collections; and the documentation of security and access

limitations for data entities. These attributes are mentioned because, as documentation requirements, they are of particular importance in the TSC environment.

Identification of the sources of data entities will be important because they are attributes which almost every user of transportation data will require. Documentation of the users of these same data entities will be equally important to insure against unintentional deletion from the data base. For example, since it is expected that many data entities will be shared by many different users, it will be necessary for the Systems Manager to insure that there is no further need among all users of a particular data collection prior to making a decision to delete.

In addition, in the TSC environment (where physically the data base will be located on many different machines and dispersed over a large geographical area), the locations (accessibility and availability) of data collections will, in themselves, be important attributes to document. Finally, in dealing with transportation data collections which include a wide spectrum of sensitive information, it will be necessary to maintain an accurate account of the security and rules for accessing the data base.

Another aspect of data documentation relates to the data documentation requirements inherent in the systems development process for TSC applications. For each application system designed and implemented in the TSC environment, it will be necessary to document the manner in which the data is used and manipulated in that application (i.e., the data specification aspects of systems documentation). Specifically, this includes record and file layouts, data definition portions of application programs and, in some cases, input and output formats. The data dictionary will provide documentation support in all these areas.

### C.5.1.4 Data Standardization

The need for the establishment and enforcement of standards relating to data usage and responsibility is accentuated in the

TSC environment where extensive sharing of data is expected. Since the users of the TSC data base will be geographically dispersed and the types of data accordingly diversified, standardization will be required for formats, the meanings of terminology, and the use of codes simply to understand and use the material.

One important aspect of data standards sometimes overlooked relates to data editing and validation standards. In dealing with the proposed TSC-type data base in which input sources are primarily external, it will be imperative to establish stringent edit and validation requirements to insure maximum integrity of the data base content. If these edit and validation rules are not uniformly applied to all data base inputs (no matter what is the source), the consistency of the content will suffer. The data dictionary should be used to record these standards for editing and validating input. In later stages of data base development the data dictionary can be interfaced directly with the software programs performing edit-and-validate functions, therebe ensuring maximum control over the input process.

Using the data dictionary to ensure the consistent usage of redundant data entities will be another aspect of data standardization. As previously mentioned, redundancy may be tolerated due to certain technological trade-offs. However, it then becomes imperative that the use of the duplicated data entities be consistent. This can be accomplished with the data dictionary, by using cross-referencing and/or version designation.

### C.5.1.5 Data Dictionary Publication

Although the data dictionary will be part of the data base system itself and will exist in both machine readable and hard copy form, its publication will be a necessary part of communicating throughout TSC the size, scope, and specific holdings, rights, and access rules for use by analysts and other interested parties. The publication will take two forms. A frequently issued data catalog and its supplements will provide the basis for analysts to monitor and review holdings in a general way. The data catalog will include the basic contents of the data dictionary

described in Section C.5.2 in an abstracted narrative style. The details of the data dictionary will be published annually and distributed as appropriate.

### C.5.2 Contents of the Data Dictionary

The attributes of data entities make up the contents of the Data Dictionary. These attributes describe each data entity to the extent necessary to serve the functions of the Data Dictionary. Subsections C.5.2.1 thru C.5.2.7 review the data attributes to be included in the TSC Data Dictionary. The subsections are organized according to the classification of data attributes presented in Section C.3.

#### C.5.2.1 Identification of Data

a) Label - A unique identifier differentiating the data entity from all others in the dictionary. The label will usually be in the form of an "identification number." Careful attention should be given in designing this number, specifically with regard to:

   1) expansion

   2) coding structure

   3) readily recognizable and easily transmitted by humans.

   This label will be used in all referencing to Data Dictionary entities.

b) Name - The short (sometimes official) title or designation used by general users of the data entity. Examples of such names are the COBOL name of a data item or the title of a data collection.

c) Description - A textual description of the general content and purpose of the data entity. The description should be worded carefully to avoid ambiguities. Additionally, the description should be used to note unique aspects of the data entity not covered elsewhere

in the dictionary. For example in the case of data items that contain codes, a description of each code value should be provided. If the code list is extensive, the entry should indicate where code values are explained (e.g., data set containing a table).

d) Designator - A short user-oriented identifier constructed from a controlled list of key-words. This is the description that will provide a rudimentary indication of possible redundancies and/or inconsistencies. The Data Dictionary system should provide a Key-Word Out-of-Context (KWOC) index of these key words, as related to dictionary labels, so that an analyst looking at a proposed new element and knowing only its general meaning can choose some key-words and use the index to determine whether the element has been previously defined. Another use could be to group data entities with equal or similar sets of key-words (one-match, two-matches, etc.) and then to analyze them for redundancy and/or inconsistency. Still another use could be for a spontaneous user of the TSC data base to apply the same technique to find the description of a certain data entiry, knowing only its general meaning.

e) Synonym - A pointer (using the label) to another data entity described in the dictionary that has the same meaning, but a different label. It is very likely that the two entity descriptions will have identical or similar designators, but this fact in itself is not sufficient to assume that the synonyms are the same. Assume, for example, that the designators for two elements are the same, but the source or frequency of update or the input validation criteria are different.

f) Version - A number assigned when a new entity description is created with the same meaning and label as a previously defined entity. For example, a new version of a census survey having slight differences in composition from a previous version that may eventually

replace the currently used version. The combination of label and version is used to identify a specific data entity in the dictionary.

g) Generation - An indication of the number of history back-up versions (i.e., grandfather, father, son) maintained for the data entity.

h) Location - An indication of where the data entiry is stored. Geographic system and device location should be noted.

i) Classification - A general coded category specifying the nature of the data entity. For example, if all transportation data were classified into categories of Mode, Source, Geographic Scope, and Commodity (Goods), then a data entity might be classified by using a four part code corresponding to these categories.

j) EDP Labels - The label assigned to this data entity for the purpose of referring to it in an EDP environment. For example the Data Set name or the File Name.

C.5.2.2 Source of Data

a) Source - The name and address of the organization from which the data entity is available. The entry in this case should be specific enough to allow the user to make direct contact with the source (if this is appropriate).

b) Source Documentation - A reference to, and a description of, documentation supplied by the source organization about the data entity. Such documentation might be relative to the content (data item names & format), physical attributes/characteristics of data items, and various access or usage rules. It should be noted that, in essence, this entry is a reference to Data Dictionary type information available in the source organization.

c) Entry Source - For internal TSC data entities derived from external sources, the entry source document should be identified. For example, if special forms were prepared from externally provided data, these forms should be identified. The organization or individual responsible for the preparation of these documents should also be specified.

d) Content Responsibility - The name of the individual or organizational unit responsible for the data entry (relative to this data entity) at its source. This attribute is useful to the Systems Manager in establishing and maintaining data integrity at the source level.

e) Derivation Algorithm - A description of how a calculated data item (an item that is derived from another item), is obtained and which data items are involved in the calculation. The description can be free-text or in a formal programming language (e.g., PL/I).

C.5.2.3 Type of Data

a) Data Type - An indication of the type of data entity (i.e., quantitative or qualitative). A suggested list of data types follows:

1) Name - Data, whether alphabetic or numeric, which identifies specific entities (as opposed to CODES which identify classifications of entities). This would include air carrier names or numbers, geographical codes, flight numbers, originating and destinating airports.

2) Code - Data which identifies classifications of entities such as modes of transportation, transaction codes, and standard industrial codes.

3) Count - Number of or quantity (including fractions and percentages) of anything except dollar or monetary amounts.

4) Amount - Quantity of dollars or other monetary amounts.

5) Date - Date, whether year-month-day, or month-day-year, or month, or day-month, or year, or only the last two digits of year, etc.

6) Text - Data of relatively informative and undefined content, such as "remarks" or "descriptions" or "free-form narrative."

7) Flag - A code expressed as a bit or byte and limited to two conditions (e.g., yes/no, presence/absence, on/off, etc.).

8) Control - Information used primarily for control of other information during processing (e.g., delimeters and carriage control characters).

9) Constant - Data which has a constant value always associated with it (e.g., a message).

C.5.2.4 Use of Data

a) Media - A description of the physical storage upon which the data entity resides (e.g., card, disk, tape, etc.).

b) Representation - A description of how the data entity is represented on physical storage and how it is represented to users. A number of sub-attributes may be specified here.

1) Length - The length of the data entity, expressed as "number of characters."

2) Mode - The internal representation (e.g., bit string, packed decimal, etc.).

3) Justification - right or left. Includes also padding, if present.

4) Picture - A description how the data entity is represented for display purposes, including edit marks.

5) Decode Method - If the data entity is encoded, encrypted, or compacted, a description of the decoding scheme. For security purposes, a reference to where and how the decoding scheme can be found might be used.

6) Format - The block size and format (e.g., fixed or variable length).

c) Volume of data - An indication of the amount of data in the data entity. This may be expressed in terms of number of characters, or by some other standard means (e.g., number of records).

d) Growth Factor - An indication of the expected growth of the data entity, or the specified period of time of its reporting frequency.

e) Security - A description of the level of security to be exercised relative to the availability of the data entity to users. This should be described in two parts.

1) Access Level - A code indicating one of six access levels (previously defined in Section C.4).

2) Security Level - A code to indicate the level of confidentiality or secrecy for the data entity (i.e., who may have the defined level of access).

Thus, the combination of Access Level and Security Level defines by whom, and how, the data entity may be accessed.

f) Frequency of Use - An indication of how often the data entity is referenced by users.

g) Users - A list of the users of the data entity both internal and external. A primary user should be designated if possible. Programs should also be considered as users so that a list of programs referencing the data entity should be included. Additionally, reports in which the data entity appears should be specified.

h) Hardware/Software Considerations - A list of the hardware and software factors which may affect the acquisition or use of the data entity. Among these would be the computer on which the data entity was generated, the software (e.g., OS/370) under which the data entity was created, and any other special software requirements (e.g., a generalized data base management system constraint, or query language requirements).

C.5.2.5 Qualifications of Data

a) DOT Access Control - A reference to the DOT Standards and Procedures qualifying the usage of the data entity.

b) Access Keys - Identification of the control and/or access keys of the data entity. In the case of a data item which is itself a key, it would be so indicated.

c) Index - An indication of the type of index which may exist for the data entity. If the data entity is part of an index, it would be so indicated.

d) Edit Rules - A description of the edit rules applied to the data which comprises or updates this data entity. For example, if certain contents are mandatory, or if value ranges or default values exist, these would be described.

e) Precision - A description of the precision of numeric values, such as decimal placement and statistical confidence limits.

f) Unit - The unit of measure for numeric values, such as pounds, inches, dollars, etc.

g) Status - An indication of the status of the data entity, in the TSC environment. For example, in cases of shared data where multiple users must concur with a single representation, four levels of status might be used.

1) proposed

2) concurred

3) approved

4) effective

It is suggested that a date be used to indicate, "as of when," for the specified status.

h) Frequency of Update - An indication of how often the data entity is updated. A date of last update should be provided.

i) Acquisition Cost - An indication of the price or charge for the use of, or access to, the data entity. If TSC were to charge for their services, this would apply to internal as well as external data sources.

j) Content Integrity - An indication of the level of confidence a user might place upon the data entity. It may be advisable to develop a confidence coefficient based on a combination of pertinent factors, such as:

1) collection methodology (e.g., observation, interview, mail questionnaire).

2) scope or completeness of data (e.g., sample versus universe, summary versus detail).

3) for external sources, an indication of services level and support.

It should be noted that other factors might be included in the confidence coefficient which were already specified in other attributes, such as edit rules applied.

C.5.2.6 Relationships of Data

a) Organization - An indication of the data organization used for the data entity (e.g., indexed, sequential or random). If a generalized data base management system (GDBMS) is used, the data organization of the GDBMS should be identified.

b) Sort Sequence - A list of the data items upon which the sequence of the data entity is maintained.

c) Sequence - An indication (by label and name) of the sequential position that the data entity occupies in the hierarchy of data entities (e.g., the sequential position of a data element in a specified data record).

d) Data Structure - A description of the membership of the data entity with all other entities in the data entity hierarchy (e.g., for data bases and data collections, a description of the data structure of these data entities would be specified.

In addition to these data attributes, there are others which relate to administrative control of a Data Dictionary entry for a specific data entity. These entries are to be used by the Systems Manager in establishing and maintaining control over the Data Dictionary itself. Among these are:

1) Preparer - The name of the individual who prepared the entry.

2) Content Responsibility - The name of the individuals responsible for the integrity of the entry (approval).

3) Last Change Date - The date that the last change was made to the entry.

As previously mentioned, it should be recognized that the Data Dictionary data attributes apply to varying levels of the data entity hierarchy. As a practical matter, it should be noted that the Data Dictionary will be organized around the lowest level of the data entity hierarchy (i.e., the item). This is to facilitate referencing and to simplify the organization of the dictionary.

### C.5.2.7 Other Entity Considerations

Besides the data entity hierarchy, there are several other important entities, in the TSC environment, which should be considered within the context of the Data Dictionary. These relate to the data flow and procedures which surround the TSC data environment. The most important entities which should be considered are as follows:

a) Data inputs (transactions, source documents)

b) Data outputs (reports)

c) Data processes (programs, application systems)

d) Data Users (internal, external)

These entities represent forms which the data entities themselves have taken prior to and after they have been in the Data Base, the processes applied to the data entities, and the users the data entities are meant to serve. In structuring the list of data attributes, we have included these data entity relationships as part of the description of the data entity. Thus, in the class of attributes "Source of Data," there is a description of the Data Inputs. Alternatively, one could define these entities in the Data Dictionary by making entries for Data Inputs, Data Outputs, etc. These entries would then be related to entries for for data entities through appropriate declarations in the "Relationships of Data" class of atrributes.

Regardless of which approach is preferred, it is important to recognize that these relationships between data entities and the other entity considerations described above do exist and should be recorded in the Data Dictionary.

### C.5.3 Example Descriptions of a DOT Data Collection

For purposes of clarity and demonstration, a typical set of Data Dictionary entries is shown here, as it might appear for the National Geocoding Coding Converter File. The examples are meant to illustrate the content, and not the form, of the description, and are, therefore, not a complete description of the file. The file is described by definitions at three levels of the data entity hierarchy.

a) Data Item

b) Data Record

c) Data Collection

Two examples are shown at the first, or Data Item, level and one each at the Record and Collection levels in the four examples which follow.

EXAMPLE 1. DATA ITEM LEVEL: "SERIAL NUMBER"

Identification of Data

1. Label - IT-001 (Note, this is just an arbitrarily chosen number. The TSC will decide upon an appropriate coding scheme for its Data Dictionary.)
2. Name - COBOL name is SER-NUM
3. Description - Six-digit numeric record identification.
4. Designator - Item which is name of number of record of Geocode File 1.

5, 6, 7, 8, 9, 10 - N/A

Source of Data

1, 2 - N/A

3. Entry Source - Machine generated.
4. Content Responsibility - Machine generated.
5. Derivation Algorithm - sequentially assigned.

Type of Data

1. Data Type - Name of Number.

Use of Data

1. N/A
2. Representation -
   a) Length - 6 characters
   b) N/A
   c) Justification - right, zero padded.
   d, e, f, - N/A

3, 4, 5, 6, 7, 8 - N/A

Qualification of Data

1. N/A
2. Access Keys - Access key of the file is this item.
3. N/A
4. Edit Rules - Numeric only

5, 6, 7, 8, 9, 10 - N/A

Relationship of Data

1, 2 - N/A

3. Sequence - Position 1-6 in RE001.
4. Data Structure - IT001 is number of RE001.

EXAMPLE 2. DATA ITEM LEVEL: "NAME OF STATE"

Identification of Data

1. Label - IT002
2. Name - COBOL name is STATE-NM
3. Description - Full alphabetic spelling for names of the 50 United States and the District of Columbia.
4. Designator - Item of name of states of U.S.A.

5, 6, 7, 8, 9, 10 - N/A

Source of Data

1, 2 - N/A

3. Entry Source - Punched card input.
4. Content Responsibility - Federal Information Processing Standard Publication 6-2, County and County Equivalents of the States of the United States, Sept. 15, 1973. Issued by National Bureau of Standards.
5. N/A

Type of Data

1. Data Type - Name

Use of Data

1. N/A
2. Representation
   a) Length - 20 characters
   b) N/A
   c) Justification - left, padded blanks
   d, e, f - N/A

3, 4, 5, 6, 7, 8 - N/A

Qualifications of Data

1. N/A
2. Access Keys - IT001
3. N/A
4. Edit Rules - All alphabetic

5, 6, 7, 8, 9, 10 - N/A

Relationships of Data

1, 2 - N/A

3. Sequence - Position 7-26 in RE001
4. Data Structure - IT002 is a member of RE001

EXAMPLE 3. DATA RECORD LEVEL: GEOCODING CONVERTER FILE I

Indentification of Data

1. Label - RE001
2. N/A
3. Description - Records for each of the counties and county equivalents in the United States, containing the geographical codes for Regions, States, Areas, County Sectors and Counties.

4. Designator - N/A

5, 6, 7, 8, 9, 10 - N/A

Source of Data - N/A

Type of Data - N/A

Use of Data - N/A

Qualification of Data

1. N/A

2. Access Keys - IT-001

3, 4, 5, 6, 7, 8, 9, 10 - N/A

Relationships of Data

1, 2, 3, - N/A

4. Data Structure - RE-001 is a member of DC-001
RE-001 contains IT-001 thru IT-065.

EXAMPLE 4. DATA COLLECTION LEVEL-GEOCODING CONVERTER FILE I

Identification of Data

1. Label - DC-001

2. Name - DOT National Geocoding Converter File 1.

3. Description - A cross-reference tool that provides a means to convert from one geocoding system to another.

4, 5, 6, - N/A

7. Generation - No back-up.

8. Location - DOT Transportation Systems Center, Tape Library.

9. Classification - Code Table.

10. EDP Labels - OS/360, GEDFIL1. REVNEX. DYYMMOO.aaa

Source of Data

1. Source - Available from Information Division, Transportation System Center, Kendall Square, Cambridge, Mass. 02142.
2. Source Documentation - Report available from source describing structure and content, Report #R4805/OP 409.
3. N/A
4. Content Responsibility - Information Division at Source.
5. N/A

Type of Data - N/A

Use of Data

1. Media - Tape, 1600 bpi 9 track
2. Representation
   a) N/A
   b) Mode - EBCDIC
   c, d, e - N/A
   f) Format - Block Size = 3330, Fixed length Records - 185 characters
3. Volume - 4,000 records
4. Growth Factor - None
5. Security - Available to the public
6. N/A
7. Users - Users of geographical data
8. Hardware/Software - OS/360

Qualification of Data

1. DOT Access Control - None
2. Access Keys - IT001

3,4,5,6,7,8,9,10 - N/A

Relationships of Data

1. Organization - Sequential by IT001
2. Sort Sequence - IT001
3. N/A
4. Data Structure - DC001 Contains RE001

It should be noted that many of the data attributes marked as optional (O) were left blank in the four examples above. This illustrates the fact that, even though some information is unavailable, the total Data Dictionary entry should still be made and updated at a later time.

### C.5.4 User Interfaces with the Data Dictionary

Certain issues derive from Data Dictionary interface considerations. There will be three classes of Data Dictionary users in the TSC environment. Additionally, in the future, automated systems which will use the Data Dictionary will represent another class of user interfaces. The following is a description of these users and interface issues ensuing from their involvement.

a) Data Processing Users - The individuals involved in planning, designing and implementing data processing systems in TSC will use the Data Dictionary in the course of carrying out their functions. The systems planner will use data dictionary information to assist him in identifying information flow and relationships. The data attributes relating to Source of Data, and Relationships of Data will be particularly useful in this regard.

   The system designer will look to the Data Dictionary for support in establishing particular data requirements, such as frequency of update, volume, and security constraints. The data attributes relating to Use of Data and Qualification of Data will be particularly useful to the systems designer.

The applications analyst/programmer will use the Data Dictionary as a source of guidelines in designing his program and as a means of documentation support. The data attributes relating to Identification of Data, Use of Data and Qualification of Data will all be of interest to the analyst/programmer involved in systems implementation.

b) Data Base Systems Manager (DBSM) - The TSC Data Base Systems Manager (discussed in greater detail in Section C.7) will use the Data Dictionary as a primary tool in performing his functions. The DBSM activities of data definition, data base design, data base documentation will be directly supported by the Data Dictionary by the capabilities it provides as a repository of information about the TSC data base. Data definitions will be entered directly into the dictionary, as well as the data base design. Moreover, the dictionary serves the DBSM as a design aid in highlighting inconsistency and/or redundancy of data entities. Other functions of the DBSM (i.e., responsibility for maintaining data base integrity, control over data base access and data base standards) will also be served by the dictionary.

However, it is important to recognize that the Data Dictionary is only a tool which the DBSM will use. Having the dictionary alone will not insure that these functions will be performed. It remains the DBSM's responsibility to exercise control and utilize the dictionary to manage the TSC data base environment.

c) End-user - The end-user of transportation information in the TSC environment will utilize the Data Dictionary as an aid in locating more precisely the information which is required from the data base. By concentrating on the data attributes relating to Identification of Data, the end-user will be assisted in narrowing down

the scope of his request --- thus allowing the data base to be more responsive to his needs. Use of the DESIGNATOR in this regard in conjunction with Key Word Out of Context (KWOC) listings will be particularly helpful to the end-user. As the dictionary facility evolves and becomes more sophisticated, it will be possible to provide end-users with a query capability designed to allow browsing through dictionary entries from a terminal.

d) Automated-Use - Additional control over the TSC data base environment can be achieved by providing automated interfaces between the Data Dictionary and

   1) Compilers

   2) Data Base Management Systems (DBMS)

   These interfaces will depend heavily on the specific nature of the software involved, thus precluding any specific discussion of how they would operate. Nevertheless, in general, an interface with language could provide better control over data definition and program access to the data base. An interface to the DBMS can facilitate greater data base integrity and a degree of program data independence (i.e., the data base can be changed without affecting the porgram).

   The necessity for these interfaces with automated users is usually not required until later stages when the data base environment is fully operational. However, if these interfaces are not planned for and implemented in the initial stages, it will be extremely difficult to do so at a later time. Thus, it is recommended that provisions for interfaces to automated users be incorporated into the overall design of the data dictionary.

For the first three classes of user, interfaces with the Data Dictionary will take the form of various types of reports and listings which can be produced using the dictionary files. These reports and/or listings generally will contain different

combinations of, and crosslistings of, Data Dictionary content. Basically, there are three types of reports/listings that are required by dictionary users:

1) Glossary Reports - Glossary reports will be organized around attribute(s) and list the value(s) for those attribute(s) for all entries at a particular level of the data entity hierarchy.

2) Catalog Reports - Catalog reports will be organized around data entities and provide a listing of all the attributes for that particular entity, or set of entities.

3) Control Listings - These listings will be used by the DBSM to maintain control over the Data Dictionary. A journal of all transactions updating the dictionary and an error listing are fundamental tools of control required by the DBSM.

It should be noted that a listing of all the attributes for all the entities in the dictionary, represents a dump of the entire contents of the dictionary. If this listing were organized by attribute (one would very rarely do this), it would be called a Glossary; it it were organized by entity it would be categorized as a Catalog of the entire contents of the Data Dictionary. The DBSM will require a complete Catalog of this nature.

Other users of the dictionary (data processing users and end-users), will require a capability to select, for specific Catalog Reports, precisely which data dictionary entries are to be listed. This type of selective retrieval from the dictionary will make it easier to use and thus promote its acceptance among users.

The Glossary reports will be particularly useful in analyzing data base content, and in data base design. An example of these reports is the Key-Word-Out-of-Context (KWOC) listing. This report is a listing of the data attribute, DESIGNATOR, alphabetically by key-words, cross-indexed against the entity LABEL.

Other attributes such as NAME and DESCRIPTION could be included as well. The KWOC listing could also be used by the end-user to locate relevant entries in the dictionary for a particular request he may have in mind.

It should be recognized that much of the flexibility and ease of use that is being emphasized for the Data Dictionary user-interface, will only be possible with an automated implementation of the dictionary. This issue of a manual versus automated dictionary is discussed in the following section devoted to implementation of the dictionary.

### C.5.5 Data Dictionary Implementation

There are a number of design issues which must be considered when approaching the implementation of the Data Dictionary in the TSC environment. Among these issues, the following three areas should be given special attention in the planning stages of development:

a) Automated versus Manual Data Dictionary

b) Design of the Data Dictionary data base

c) Guidelines for Data Dictionary data capture

#### C.5.5.1 Automated versus Manual Data Dictionary

The TSC Data Dictionary will be implemented as an automated system. An automated TSC Data Dictionary system can be based on a purchased off-the-shelf software package, a number of which are currently available from reputable software vendors. A partial list of available data dictionary systems and their respective vendors are:

a) IMS Dictionary System - IBM

b) LEXICON - Arthur Anderson & Co.

c) UCC-10 - University Computing

d) Data Catalogue - Synergetics

Alternatively, the TSC data dictionary could be developed in-house. Governing the make-versus-buy decision will be an evaluation of each prospective package against the specified requirements, as presented herein and as perceived by TSC personnel. Regardless of the outcome, however, careful attention must be given to the relationship between the automated data dictionary and the DBMS. Given the wide diversity of types of data collections to be stored in the TSC data base, and recognizing the unpredictability of user requests, it would be desirable to maintain maximum flexibility with regard to the use of various types of DBMS systems. Thus, in choosing the type of automated dictionary to implement, it is recommended that care be taken not to limit TSC to the use of a single DBMS.

The decision to automate the TSC Data Dictionary is based on the following considerations:

1) Volume - Given the amount of detail required in the TSC Data Dictionary, when the volume of data dictionary entries[1] exceeds approximately 1,000, and the rate of update is more than 20%, the use of automation should definitely be considered.

2) Need for Analysis & Design Support - The TSC must consider the need for Data Dictionary support in the Data Collection phase (see Section C.4). Data analysis and data base design can be significantly enhanced through use of a dictionary, especially in the TSC environment. If this type of support is required, an automated dictionary should be considered.

[1] An entry is defined as a specification of all the data attributes for a specific data entity.

### C.5.5.2 Design of the Data Dictionary Data Base

The design of the data organization for the TSC dictionary is an important aspect of dictionary implementation. If the dictionary is automated and purchased as a package, no issue will exist. However, in the event TSC develops its own automated dictionary, the following two important design issues should be considered.

a) The basic design of the dictionary data organization should be developed around the data item. This approach will facilitate flexibility and expandability because the data item is the smallest common denominator among all data entities.

b) The design should take into account the concept of variable/non-variable data attributes. Non-variable attributes do not change from one use of the data item to another (i.e., DESCRIPTION). Variable attributes can change from one use of a data element to another (i.e., REPRESENTATION). Utilizing this dichotomy of attributes in dictionary data organization design will minimize redundancy by storing non-variable attributes only once. Furthermore, data specifiers will have the ability to define similar data entities more easily by referencing non-variable attributes already defined elsewhere.

### C.5.5.3 Guidelines for Data Dictionary Data Caputre

The methodology employed in capturing data for the TSC Data Dictionary will affect the integrity of the dictionary, as well as the effort's cost. Essentially, the objectives of the methodology used will be to reduce errors and increase accuracy of definition and specification. The following actions will be implemented to achieve these objectives in the TSC environment:

a) Special forms will be designed for use in data capture. Preformatted input will minimize data entry errors in the automated system.

b) Data specialists will be trained in screening input and refining definitions, descriptions and specifications to improve the integrity of the dictionary.

c) Formal procedures will be established for the process of data capture, including review, edit checks, and approvals at appropriate levels.

In summary, the Data Dictionary will play an important role in providing for centrally-controlled data resources in the TSC environment. Early implementation of the dictionary can be beneficial in this regard, if only by providing an authoritative control source for information about TSC data resources.

# APPENDIX D. POSSIBLE DOT SOURCES OF DATA

The Information Division, Transportation Systems Center, conducted a review of the DOT Organizational Manual with a view to identifying possible statistical data sources within DOT. The review team relied heavily upon descriptions of division functional responsibilities in identifying organizations to be studied as possible sources; however, special assistants and branches with "promising" titles were also included in the review.

Table D-1 identifies the various organizations revealed in the study. Table D-1 includes the main portions of the identified organizations' functional statements as they relate to information.* While no perfect classification scheme can be established to mark the bounds of responsibilities for any organization, the twenty-four organizations identified can be divided into seven groups differentiated by main mission, as follows:

a) one organizational unit concerns itself with the broadest and most general definition of information;

b) six units deal directly with data, statistics, and compilation;

c) three units distribute transportation research information;

d) six units provide information systems for management data;

e) four units operate ADP hardware;

f) two units provide ordinary library services; and

g) two units concentrate on software development.

*Responsible personnel who may be contacted and their telephone numbers are also included for reference convenience.

# TABLE D-1. DOT AGENCIES AND THEIR DESCRIPTIONS

| | Organization | Name | Telephone | Description |
|---|---|---|---|---|
| Information-General | OST<br>TPI-10<br>Office of Transportation Systems Analysis and Information | G. Wiggins | (202)426-4168 | Provides policy oversight and coordination of transportation statistical and data collection activities of the Department. To collect and analyze policy-related data on performance needs and capabilities of the transportation system of the Nation, to specify requirements for transportation data, and to develop and direct programs to satisfy these requirements. |
| Statistical Compilation and Reporting | OST<br>TSC-220<br>Information Division | R. Tap | (617)494-2654 | Responsible for the development and maintenance of the DOT transportation information base through such processes as data collection, analysis, management, interpretation, establishment of logical data standards, and distribution of transportation-related data and statistics. |
| | FRA<br>RA-53<br>Information and Analysis Division | C. Braddock | (202)426-2920 | Provides qualitative and quantitative analytic support services throughout the administrative. Provides assistance in bringing computers to bear on problems, including time-sharing, software development and the establishment of technical information programs. Coordinates studies of rail transportation data needs and problems. |
| | NHTSA<br>N43-33<br>Information Systems Division | R. Schweitz | (202)426-4844 | Acquires, indexes and announces availability of technical information related to vehicle and traffic safety. Prepares special summaries based upon research findings. |
| | FHWA<br>HHP-46<br>Highway Statistics Division | A. French | (202)426-0180 | Compiles, evaluates, and publishes pertinent statistical data on all phases of highway development and finance. |
| Transportation Research Information | OST<br>TST-25<br>R&D Resources Management Division | A. Hoshovsky | (202)426-0975 | Develops and maintains those information systems required for implementation of the responsibilities of the Office of R&D Plans and Resources (principal advisor in the planning, coordination,budgeting and resources management of DOT's R&D programs). |
| | OST<br>TSC-151<br>Office of Technology Sharing | R. Giangrande | (617)494-2486 | Performs as the designated focal point within the Department of Transportation for the exchange of information, ideas, experience and technology with the national transportation community. Manages DOT Technology Sharing Program and its subordinate project areas. |

# TABLE D-1. CONTINUED

| | Organization | Name | Telephone | Description |
|---|---|---|---|---|
| Transportation Research Information (Continued) | UMTA<br>UTM-1<br>Office of Transit Management | B. Barkley (Acting) | (202)426-9157 | Manages the UMTA Research Information and associated projects for the accumulations, storage, and dissemination of research project information. The Transit Research Information Center has been established to maintain a collection of all UMTA-sponsored reports and studies and to provide information concerning their availability and findings. |
| Management Data Information Systems | OST<br>TAD-25<br>Information Systems Division | W. Barney | (202)426-4735 | Provides leadership; develops standards for DOT data processing standards. Coordinates development of overall concepts and procedures for the establishment and maintenance of departmental data banks and comprehensive data indexes. |
| | OST<br>TSC-810<br>Management Services Division | F. Huron | (617)494-2254 | Provides central coordination and processing support to current and budget year technical program planning. Coordinates and manages institutional planning. Formulates and submits budgets, and recommends and manages funds allocations. Analyzes use of resources against plans, and makes appropriate recommendations. Analyzes requirements for and controls development of new or improved internal automated management information systems. Provides library and technical reference services, information research and analysis, publications processing and special documents preparation. |
| | USCG<br>G-FIS/84<br>Information Systems Division | Captain Bauman | (202)426-2437 | As part of the Office of the Comptroller, provides computer system support to fulfill management information requirements for Coast Guard operating and support programs. |
| | FAA<br>AMS-20<br>Office of Management Support - Information Department | F. Osgood | (202)426-3350 | As part of Office of Management Systems, provides FAA Management Information.<br>(Note: also responsible for FAA operational statistics). |
| | FAA<br>AMS-200<br>Information and Statistics Division | L. Williams | (202)426-3165 | Provides data processing, other than a traffic operational data processing as part of Office of Management Systems. |
| | UMTA<br>UAD-50<br>Management Information Division | W. Pirkle | (202)426-9537 | As part of Office of Program Planning directs and coordinates the collection, analysis and dissemination of information required for developing, evaluating and appraising the UMTA programs. |

TABLE D-1. CONTINUED

| | Organization | Name | Telephone | Description |
|---|---|---|---|---|
| ADP Hardware Systems Management | OST<br>TSC-840<br>Data Services Division | R. Gaudet | (617)494-2344 | Provides management and operation of all TSC computer facilities; systems analysis and programming services, acquisition of ADP hardware, software, and external ADP services. Centrally coordinates all matters relating to ADP equipment, services and management with external organizations. |
| | FAA<br>AMS-300<br>Data Systems Division | E. Hootenboom | (202)426-8067 | As part of Office of Management provided data processing and motion picture services to Washington headquarters. |
| | FAA<br>H65-20<br>Computer Services Division | K. Kohler | (202)426-1746 | As part of general services plans, develops, and administers the agency's data processing requirements with the Department's and Federal Government policies and systems. |
| Library Services | OST<br>TAD-49<br>Library Services Division | M. Helvestime | (202)426-2565 | As part of Administrative Operations, provides library and information retrieval services. |
| | NHTSA<br>N48-50<br>Office of Administration Services | | (202)426-2987 | Maintains a collection of research documents and related materials on the subject of motor vehicle safety and traffic safety. Plans, establishes, and maintains systems for the storage retrieval and dissemination of publications in various format. Develops information tools such as indexes, catalogs, and bibliographies. |
| Software Development | UMTA<br>URD-70<br>Software Systems Division | R. Dial | (202)426-9271 | No specific statement in O.M. (Known to be system/software development.) |
| | OST<br>TSC-620<br>Information Services Division | D. VanMeter | (617)494-2646 | Develops computer technology in the areas of digital, analog and hybrid devices and the methodology for their application to all transportation modes; software engineering, specialized input/output requirements; computation systems, architecture, operation and analysis; computer-driven displays and computer/communication system interfaces. |
| | | | | |

TABLE D-1. CONCLUDED

| Organization | Name | Telephone | Description |
|---|---|---|---|
| OST<br>S-81<br>Office of Public Affairs | J. Stafford | (202)426-4531 | The Assistant Director for Information in the Office of Public Affairs promotes public awareness and understanding of the role of the transportation in the U.S.A., the nation's transportation needs and problems and the Department's plans and accomplishments in the response to such needs. |
| OST<br>TES-1<br>Office of Environment Safety and Consumer Affairs | J. Connors | (202)426-4474 | In representing the Secretary as the principal DOT contact in the areas of environment, safety, security, facilitation and consumer affairs, the Assistant Secretary excercises executive direction over offices which must develop data and information on the status of their respective area of responsibility. This involves maintaining an overview of all Departmental programs in these areas, evaluating their effectiveness, etc. All offices in this organization excercise data and information functions but there is no single point of contact within the Assistant Secretary's office. |
| OST<br>TST-20<br>Office of R&D Plans and Resources | L. Greene | (202)426-4211 | Develops and maintains information systems required to act as the principal advisor in the planning, coordinating, budgeting and resources management for the Department's R&D programs and R&D facilities. Robert Paullin, 426-4224 (R&D Resources Management Division, TST-25) is the principal contact on ongoing information work. |

# APPENDIX E - TSC DATA TAPE LIBRARY SYSTEM
## (ROLES AND RESPONSIBILITIES DESCRIPTION)

### E.1 OBJECTIVES

#### E.1.1 General Objectives

To structure and install a comprehensive procedural and physical process which efficiently (cost minimization) regulates the acquisition and utilization of machine readable and other data at TSC.

#### E.1.2 Background

The development of socio-economic projects at TSC has accelerated the requirements for the use of statistical data in support of these projects proportionately. Since there has been no control mechanism for specified acquisition, delivery, and sharing of data to the staff except for the legal procurement process, it is only natural that each staff member with data requirements develops differing data needs with varying characteristics. Furthermore, upon receipt of data, staff members have no reason to offer their holdings to the TSC analyst community for wider use. As a result, there has been an accumulation of data at TSC which is not of lasting availability and tends to be discarded after the initial utilization. A further circumstance of current data acquisition procedures involve the repeated purchase of machine readable data which is not compatible with the TSC PDP-10 computing system.

#### E.1.3 Specific Objectives

To replace the uncertain process which now exists in acquisition of machine readable statistical data, a new approach is offered with specific objectives.

a) Single point inspection of data requirements and specifications to minimize duplicate data or data processing incompatibilities with the TSC computing system.

b) Central receivership of delivered data for recording, copying and distributing to TSC analysts.

c) Separate but coordinated functions for archieves and working data base libraries.

d) Standards for machine readable data to facilitate automated accessibility and computing services.

e) Directory services to assist analysts in reviewing current holdings and determining approaches to satsify requirements.

It is proposed that a system as outlined herein and currently in limited operation be completed and installed at the earliest practicable time.

## E.2 SYSTEM OVERVIEW

The beginning of any analysis at TSC concerns itself with two problems simultaneously. One is the appropriate structure for a model of the problem and the other is the identification and acquisition of data to support the proposed analysis. In this instance, it is assumed that the data problem is large enough to warrant ADP in the exercise of the analysis.

The objective of the data tape library is to provide a central location and system for supplying the analyst with data through operations that approximate any good source library. The analyst in this problem situation reviews the holdings of the data tape library and makes a determination of the availability of the desired data. If it is available, the user would behave as described later in this section.

If the data were not in the holdings, as perceived by the analyst, the analyst would generate a brief document establishing the data requirements desired. Coordinating with the Information Division (Code 220) would reinforce the earlier decisions of the analyst or reveal options from the present holdings that were not obvious in the initial review by the analyst. Based upon the results of the coordination efforts, the desired data would be

obtained in its currently best form. Inadequate collection activities will be documented by Code 220 for use in establishing Department needs for data collection, and the efforts of Code 220 in establishing the best avenue for obtaining the data in the most logically consistent manner will be added to the physical standards of the TSC Data Services Division (Code 840).

Receipt of the data within TSC will always be through the Library operated by the Information Services Branch (Code 813). There, a formal log-in will be accomplished including archiving of the hard copy supporting documents that describe the data. The Library will forward the data to Code 840 and update the records that describe the TSC current holdings.

Finally, the analyst will be advised of the receipt of the data and periodic bulletins will be issued to all of TSC to review the tape library holdings.

The data tape library system consists of five operating functional activities that enable any user of information stored on magnetic tape at TSC to seek out, borrow and use data tapes through a systematic procedure. This system has been set up to alleviate the data take handling problems that have plagued the center in the past; e.g., data tapes were kept in different offices throughout the center without a record of where they were stored, data tapes have been also found without labels or documentation to support them. In order to correct these and other data tapes handling problems, the data library was conceived; its five activities are described in the following paragraphs and illustrated in the overview diagram of Figure E-1.

### E.2.1 Data Acquisition

An analyst who wishes to acquire data on magnetic tapes can do so after two activities are accomplished:

a) A review of the holdings in the library to determine whether the data elements requested presently exist on one or more tapes. If so, then arrangements are made for access to these data elements. If not, the next activity is pursued.

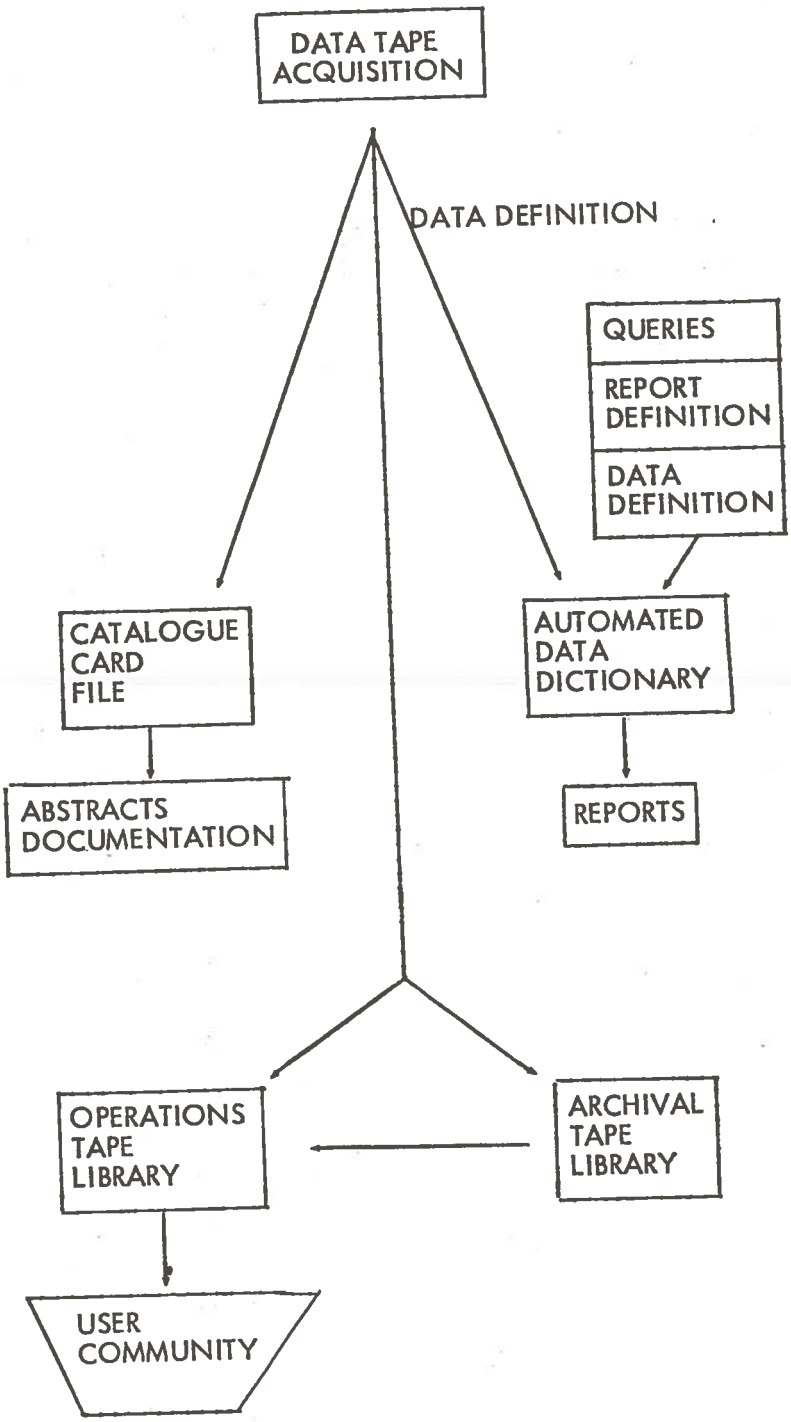



Figure E-1. TSC Data Tape Library System

b) The analyst develops formal specifications with a request for data tapes. These are reviewed and requested tapes made compatible with the TSC computer or in a form that the computer utility programs can convert efficiently.

### E.2.2 Catalog Card File

An organized card file will be maintained that will contain all pertinent information relating to physical format characteristics; e.g., 800 B.P.I., 9 TRK, BCD, etc. Accompanying the card will be a logical record layout defining and describing each element of the actual data; e.g., company name, 81-109 char. pos., alpha field.

### E.2.3 Automated Data Dictionary

An on-line computer microfiche system will enable analysts seeking information pertaining to specific data elements stored on the data tapes the ability to view catalogue cards, abstracts, record layouts and examples of the data. The procedure of using a key word searching technique and on-line microfilm display terminal will help determine whether tapes stored in the data tape library can be used in conjunction with the analysts needs.

### E.2.4 Operations Data Tape Library

The operations data tape library is located near the computer facility and is used by computer programmers and analysts. Tapes in this library are write protected and assigned tape numbers. Analysts can have data tapes mounted from the library on tape drives by calling computer operations and specifying a library tape number.

### E.2.5 Archival Tape Library

All of the tapes in the archival tape library are labeled and recorded in the card catalogue and the abstracts and record layouts are microfilmed and stored in the on-line microfilm terminal.

Tapes stored in the Archival Data Tape Library can be used by the analyst/computer programmer community with TSC by acquiring lending copies. Historical tapes are only used for production of lending copies under careful control of the data tape library operations. Because there exists certain legal restrictions to the use of some of the data tapes stored in the library, authorization will have to given before restricted tapes can be utilized. With the constant growth of the tapes submitted to the data tape library the above procedures for qualifying analysts' requests for certain types of data tapes and tracking of data tapes will become more beneficial to TSC.

## E.3 OPERATIONAL RESPONSIBILITIES

In order to make the TSC Data Tape Library system function as proposed, the cooperation of many divisions with TSC is required. The major institutional responsibilities involved in the library are detailed as follows:

### E.3.1 Review of Data Requirements - Information Division, 220; Information Services, 813

As analysts' new data requirements arise, the holdings of the TSC Archival Data Tape Library are reviewed to ascertain if the required data set is already in-house. The on-line microfiche system will be used to determine what data elements are available to satisfy an analyst's requirements. In most straightforward cases the data would be in the library in an acceptable machine readable form and would be ready for loan to the requesting analyst. Code 220's responsibilities are associated with keeping the library holdings current by continuously recommending appropriate systems (Data Resources, Inc., National Bureau of Economic Research), or data set to be added or deleted from the holdings.

In the event that the data is not in the Archival Data Tape Library the analyst would be asked to document the requirement for data. This requirements document would be reviewed by Code 220 which would advise the analyst of the availability of the data for purchase or the necessity of building the data set if none exists.

If the data already exists and can be bought commercially or from another agency, Code 220 will review and concur on the purchase order. If it is available in-house, Code 220 will assist in creating the new structure through all appropriate means. If it is not available at all, Code 220 will gather the requirement as evidenced for a data collection program which might include such data.

E.3.2 Establishment of Technical ADP Standards - Information Division, 220; Data Services Division, 840

During the procurement of data, personnel of the Data Services Disivion will advise the analyst and Code 220 of the best logical specifications for the data tapes being procured. Logical standards for data elements will be set by Code 220 in all cases. Additionally, physical standards for data processing will be set by Code 840 in all cases. Code 840 will keep the Information Division informed of the most appropriate conversion programs, available systems, etc. and their utility to the technical community. Code 220 will keep the Data Service Division informed of the most appropriate logical representations to well-known data elements and of available conversion programs.

E.3.3 Creation of Data Sets - Information Division, 220; Contractor

If the data set must be created, the analysts' requirements will be reviewed by Code 220 to ascertain the feasibility of constructing the data set in-house or of procuring the services of a contractor to create the data. In the case of TSC creating the data set, Code 220 will be the performing organization in most cases. It is anticipated that the work will be conducted under TSC policies regarding Inter-divisional Agreements. In the case of an outside contractor the 220 staff will assist the technical initiator in the preparation of the procurement package, evaluation of contractors and consulting in the technical aspects of data collection and processing.

E.3.4 Archival Data Tape Library Maintenance - Information Services, 813; Information Division, 220

The library will be operationally maintained by the professional library staff in Code 813. They will be responsible for receiving and logging in all data sets, preparation of statistical abstracts, record formats, etc. and maintenance of the on-line microfiche system. Also, all data tape purchase orders will flow through the Archival Data Tape Library. This is to insure that the library staff is aware of all data acquisitions and that the data tape(s) are delivered directly to the library for processing prior to use. Code 220 will review the process and capabilities of this activity for recommendation and funding improvements.

E.3.5 Creation of Archival Copies - Data Services Division, 840; Analysts

Once a data set has been purchased or created the process of delivery to the Archival Data Tape Library, logging in and the creation of an archival PDP-10 compatible copy will take place. Code 813 will transmit the tapes to Code 840 and inform the analyst of the data's availability. It is the responsibility of the user to work with Code 840 in the creation of PDP-10 readable data tapes. The user will be informed of the necessary conversion costs which must be incurred in spinning a PDP-10 compatible copy and will initiate an account number with Code 840 so that the conversion can be made. In those cases where multiple tapes are being procured, it will not be cost effective to convert all tapes upon receipt. In this situation tapes will be converted as they are used with the cost of conversion being borne by the first user with a requirement to access that data set.

E.3.6 Provide Machine Operations and Data Access Control - Data Services Division, 840; Information Division, 220

The Data Services Division's prime responsibility, of course, is to provide the various computing facilities available to TSC. The analyst may have a data tape mounted for use on his computer account on the PDP-10 or, after referral to an outside vendor such

as DRI or NBER by Code 220, he may work with Code 840 to gain access to one of these systems. The Data Services Division maintains the TSC Operations Data Tape Library which contains PDP-10 compatible tapes in current use by the TSC community.

As the Archival Data Tape Library holdings are updated and archive copies stored, it is anticipated that analysts would find it unnecessary to store full data sets in the TSC Operations Data Tape Library once the work is completed. This will reduce instances where individuals store seldom used data tapes in personal bins, thereby taking up valuable storage space.

Code 220 has final access release control for those tapes which are legally not in the public domain. Code 840 will check each user of these tapes for proper authorization.

### E.3.7 Data Acquisition Announcements - Information Services, 813; Information Division 220

Following the creation of an archival copy for the library the data set will be added to the list of TSC Data Acquisitions which will be printed periodically by Code 813 and distributed throughout TSC by Code 220.

In summary, Table E-1 depicts the institutional responsibilities involved in the library system.

This section has outlined the general flow of activities and the responsible division of labor associated with them. Figure E-2 provides the reader with a comprehensive look at how the activities interralate with one another. The next two sections deal more specifically with the operating characteristics of the library and its ability to respond to user requirements.

## E.4 USER-LIBRARY INTERFACE

Assuming that a member of the TSC facility has developed the need for a specific data base, and has determined, after consultation with the library personnel, that the data base is not presently available in-house, it is his responsibility, in cooperation with the library, to locate a source from which that data base may

TABLE E-1. LIBRARY SYSTEM INSTITUTIONAL RESPONSIBILITIES

| Element | Major Responsibility | Minor Responsibility |
|---|---|---|
| Data User | • Structure Data Requirements<br>• Enlist Support of 220, 813, 840 | |
| Code 220 | • Legal Access Rights<br>• Establish and Maintain Logical Standards<br>• Review Data Requirements<br>• Create Data Sets<br>• Technical Consulting | • Support Library Maintenance<br>• Support Data Dissemination |
| Code 813 | • Maintain Archival Library<br>• Prepare and Disseminate Data Acquisition Announcements | • Review Data Requirements |
| Code 840 | • Establish and Maintain Physical ADP Standards<br>• Create Archival Copies<br>• Provide Machine Operations and Data Access Control | • Support Creation of Data Sets |

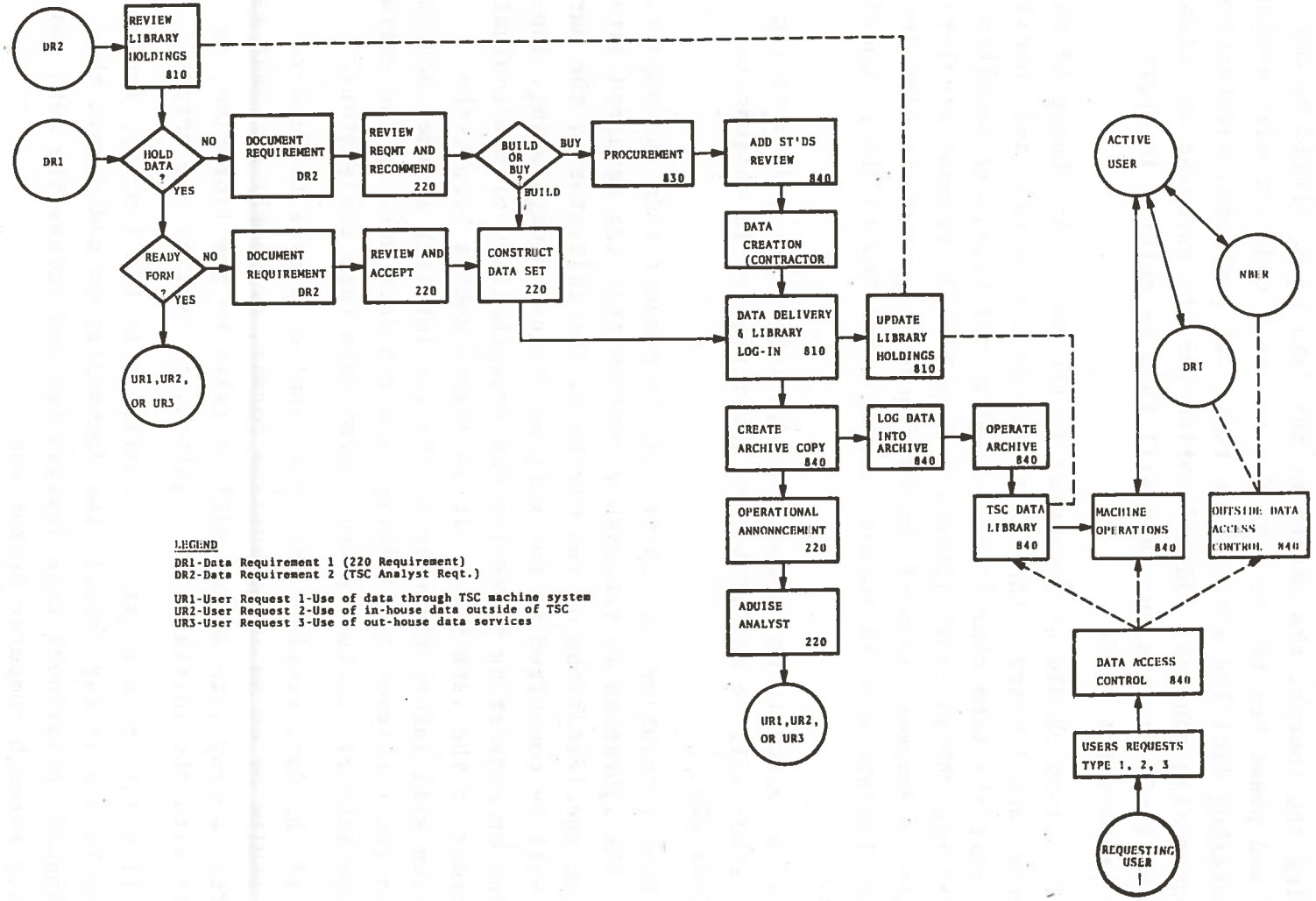



Figure E-2. Proposed Scheme for Managing Data at TSC

be obtained. If such a source if found, all pertinent information regarding the source, the nature of the data base, should be documented and presented to Code 220 for review. Code 220 will assist in a make/buy dicision advising on technical problems, alternatives, and contracting sources. Justification for its purchase and allocation of funds for the purchase will then be entered in appropriate procurement forms.

This form should be presented to the person in charge of the reference tape-library. The librarian shall ascertain and specify on the form the tape characteristics required to assure compatibility with the PDP-10 capabilities, and explicitly request complete and current documentation of the data base. Appropriate signatures will be obtained and the purchase order submitted to the purchasing office.

If a source is not found, the development of a data collection project will be undertaken by the initiator in conjunction with Code 220.

Upon receipt by the library of the product from the supplier, unless the librarian is reasonably certain that the acquired data meet the specifications of the purchase, the initiator of the purchase will be consulted at the various stages of cataloging, duplicating or converting to verify the acceptability of the content and format of the data base. If no discrepancies occur, the librarian will inform the user of the availability of the documentation in the reference tape-library and the data base in the operating tape-library, including the appropriate tape designators.

If the data received from the supplier proves to be defective, faulty or of an inappropriate format, or simply not what was expected, appropriate action will be taken by the librarian, in concert with the initiator is required, to rectify the errors.

If a data base is already available in the library, anyone wishing to use it may consult the documentation and obtain the coordinated operational tape designators and access the data base directly through computer operations.

### E.4.1 Depositing a Data Base in the Data Tape Library System

If a member of the TSC staff has a data base that is not already resident in the reference tape-library and if that data base is judged by Code 220 to be of sufficient value to be placed in the library, the tape-library will accept the data tapes provided that adequate documentation accompanies the tapes. The library is not a repository for useless, undocumented, tapes.

In fact, in the startup phase, a continuing effort will be made to transfer control of data bases presently in-house to the reference tape-library. No interruption in the present users' operation should be anticipated. Such transfer principally involves paperwork with one exception: the user must supply the documentation. The operational tapes will be assigned new codes and be listed as elements of the Operating Data Tape Library, rather than the personal property of the present user. The master tapes will be cataloged into the Archival Data Tape Library.

### E.4.2 Ownership and Access Restrictions

All data based purchased or deposited in the library are understood to be the custodial property of the library facility: the master tapes and documentation reside in the Archival Data Tape Library; the operational tapes in the Operating Data Tape Library. Neither the tapes nor the documents are to be removed and sequestered in a private holding. It is assumed that the initiator of a purchase has priority in availability of any of the elements. Such priority, however, should not, under normal circumstances, be of any significance since duplication of the documents is ordinarily a simple matter, and machine-access to the operational tapes by all who have a legitimate use should negate any priority claims. Duplication of the machine-readable tapes is at the option of the user, provided originator-imposed restrictions, if any, are understood and observed.

## E.5 SUMMARY

The Archival Data Tape Library includes the following major activities:

a) To establish, develop, and maintain a central computerized data base library with the facilities for the shortage of magnetic tapes, discs, and computer printouts.

b) Coordinate all requests for the acquisition of computerized data.

c) Acquire, document, and maintain all appropriate writeups for the cataloging and descriptive documentation of record layouts, computer processing specifications, and general information including history of the data base.

d) Coordination with the TSC ADP I/O unit of all necessary functions in the processing cycle, e.g., tape conversion and backup tape creation.

e) Assist users of the Archival Data Tape Library in locating and referencing desired data or recommending other sources for the same.

f) Operation of automatic microfilm and data retrieval system.

In order for the Archival Data Tape Library to be an effective tool of management it must maintain efficiency in all aspects of its operation. It must function with a minimum of complexity from the initial acquisition phase through the final data processing stages. Outlined below are basic operations within the ADTL.

1) Log Procedure. A single card record is completed for each unit of data received. The information posted on the card summarizes the nature of the data and includes:

   a) Control number

   b) Tape number or other ID

   c) Title

   d) Source of data

   e) Creation date and chronological span of data

f) Number of tracks

g) Character set, e.g., EBCDIC

h) Density

i) Tape parity

j) Blocking factor

k) Number of characters per record

l) Data packed or unpacked

m) Individual contact or TSC source

The reverse side of the card reflects users of the particular document (tape) and dates of usage.

2) Description Briefs. The development and maintenance of summary descriptions provides additional detail of the data file. Included would be statistical variables within the file and their organization and/or location in the file. The descriptions are prepared after careful review of existing documentation. If the documentation is incomplete contact is made with the TSC source of the data or the originating organization for the establishment of basic computer configuration, record descriptions and layouts.

3) Documentation. This stage entails the development and maintenance of physical files of all available written reports, history, and writeups for each data base. These documents detail the statistical nature and properties of the data contained in the data base. Also described are the details of the computer specifications for processing the data and outlines of the structure of the file. Relevant historical information detailing the early development of the data base may also be included.

4) ADP Coordination. All necessary coordination with the TSC ADP I/O unit will be carried out. These functions include requests for tape conversion, backup tapes, and requests for computer-generated reports on tape user activitity.

5) Bulletin. A Bulletin of Recent Acquisitions is published periodically and distributed in the Center. It lists and describes, in as much detail as possible, the current data bases available in the Archival Data Tape Library.

6) Computerized Microfilm System.

   a) Development of a terminology list

   b) The installation of an on-line computer microfilm system is planned for the ADTL.

      The user, with the aid of a computer, searches the microfilm file using combinations of terms and phrases. For example, the user may request, via terminal, a display of all CAB "freight and damage" entries. The computer will search and retrieve the desired information for display. The data or information seen on the display can also be converted to hardcopy.

*U.S. GOVERNMENT PRINTING OFFICE: 1977-701-345/89



100Copies

U. S. DEPARTMENT OF TRANSPORTATION
TRANSPORTATION SYSTEMS CENTER
KENDALL SQUARE, CAMBRIDGE, MA. 02142

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300


POSTAGE AND FEES PAID
U. S. DEPARTMENT OF TRANSPORTATION
518