社会開発協力部報告書

No. 2

THE KINGDOM OF THAILAND
MINISTRY OF INTERIOR
DEPARTMENT OF TOWN AND COUNTRY PLANNING

CITY PLANNING MANUAL

VOLUME VII DATABASE MANAGEMENT SYSTEM

THE STUDY ON APPLIED TECHNOLOGY FOR MAKING CITY PLAN

JANUARY 1989

JAPAN INTERNATIONAL COOPERATION AGENCY

THE KINGDOM OF THAILAND
MINISTRY OF INTERIOR
DEPARTMENT OF TOWN AND COUNTRY PLANNING

CITY PLANNING MANUAL

VOLUME VII DATABASEMANAGEMENTSYSTEM

THE STUDY ON APPLIED TECHNOLOGY FOR MAKING CITY PLAN

JANUARY 1989

JAPAN INTERNATIONAL COOPERATION AGENCY

VOLUME VII

DATA BASE MANAGEMENTS SYSTEM

(Table of Contents)

DETAILED TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

CHAPTER 1
INTRODUCTION

At present, DTCP has a variety of problems regarding the management of data in the city planning process. As a result, much of the data collected are not effectively integrated into the plan making efforts. The JICA Study Team has pointed this out in its earlier report (Interim I), and suggested that establishing an integrated data management system for city planning, is an essential task for improving the rationality and efficiency of the planning process in DTCP. A Data Base Management System (DBMS) based on modern computer technology has thus been proposed.

This Manual presents explanatory material on DBMS.

CHAPTER 2

PURPOSE OF DBMS IN CITY PLANNING

CHAPTER 2

PURPOSE OF DBMS IN CITY PLANNING

A DBMS does not consist merely of a computer (hardware) and electronically stored data. It is rather a "system" of process and organization which collectively serves the planners for their information requirements in making a variety of planning decisions. A plan is a reflection of such judgments, and it is clear that these judgments should be based on an internally consistent information system, therefore the proposed DBMS is an indispensable catalyst to the realization of an integrated data management system.

The development of the DBMS, then, has to be promoted simultaneously in the following areas:

(1) Standardization of the City Planning Process: Standardization of data formats, process and software through the improvement of internal information network in DTCP;

(2) Development of Applications Software: Statistical processing, analytical models and simulation techniques;

(3) Introduction of Computer System: Hardware improvement and expansion of computer network.

From a technical point of view, it is very important to consider what the system will do. If the purpose definition is ambiguous, many changes may have to be made during the system development.

Therefore, the purpose should be clearly defined. Regarding the Data Base Management System, it goes without saying that DBMS has the power to improve the existing system, besides, DBMS could create a new information network and more detailed standard city planning procedures.

In addition, the DTCP has been struggling to prepare general plans. Therefore, to cope with is situation, it is gravely necessary to improve and strengthen the existing systems as soon as possible for the purpose of carrying out these plans effectively and smoothly. Taking into consideration the environment of DBMS, the purpose should be defined as follows:

(1) To improve the existing systems as soon as possible.

(2) To facilitate the smooth and effective formulation of city plans.

(3) To create a new information network and a more detailed standard planning procedure.

CHAPTER 3

CONCEPT OF DBMS IN CITY PLANNING

CHAPTER 3

CONCEPT OF DBMS IN CITY PLANNING

3.1 Standard City Planning Procedure

City planning activities require an internally consistent data management system. The first step in the development of this system demands the standardization of the plan making process in which data collection and subsequent data processing procedures are synchronized into a mutually reinforcing flow of activities. This concept is shown in Fig. 3.1.1.

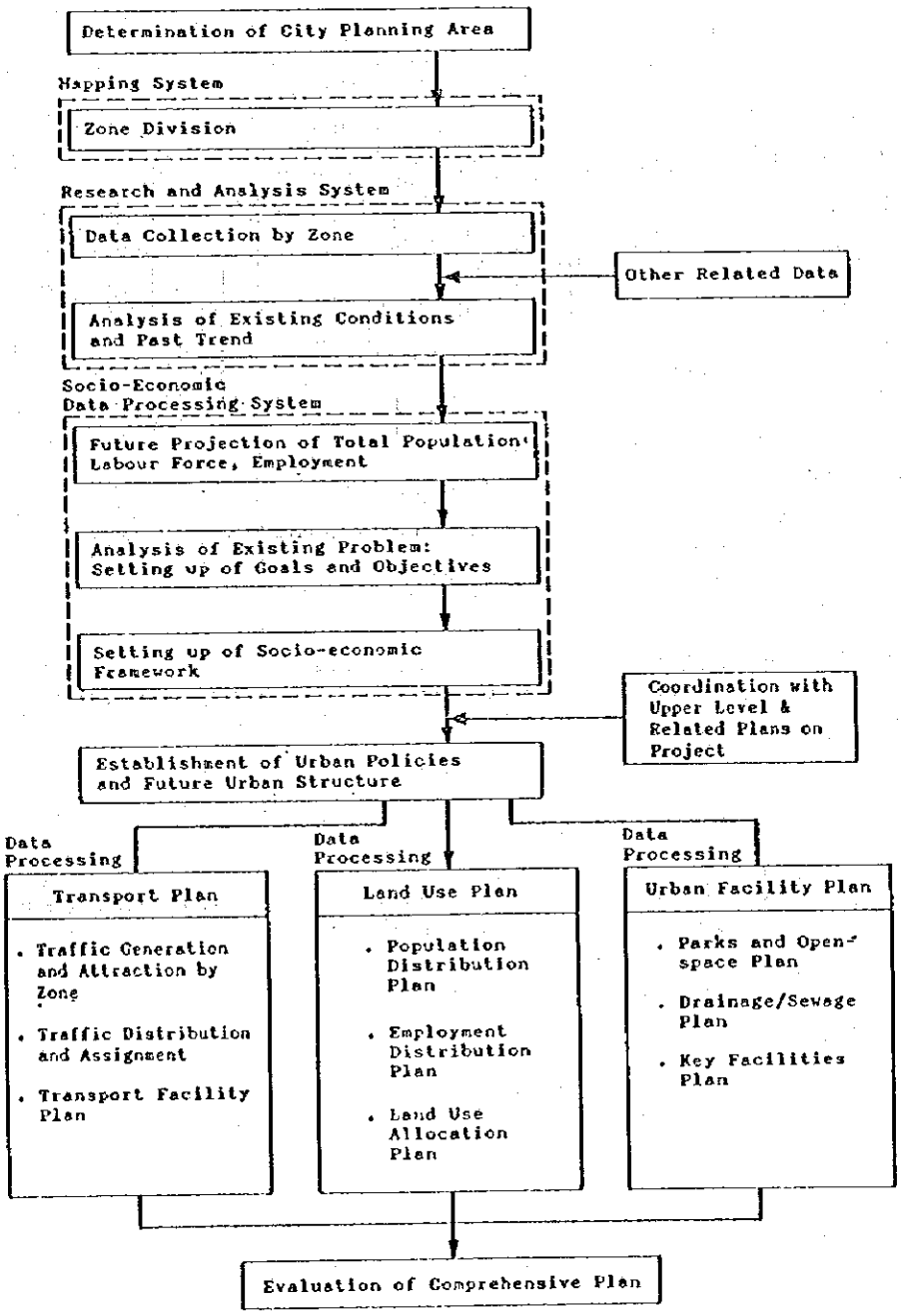
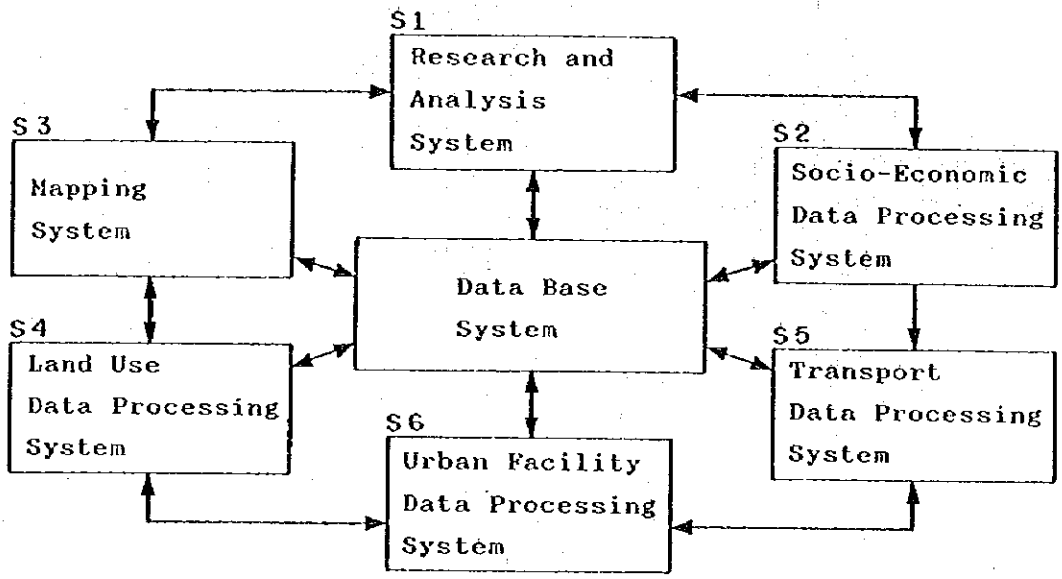


Fig. 3.1.1 STANDARD CITY PLANNING PROCEDURE

3.2 Concept of DBMS

3.2.1 Concept of DBMS

DBMS aims at the establishment of a system of software to manage storage, retrieval, and updating the records in the data base. It is a highly structured file* that attempts to provide all the data allocated to certain subject and allow programmes to use the system. The concept of DBMS as it is applied to the standardized city planning process, can be shown in Fig. 3.2.1.



* Data Base System - Data input, File Production, Data Research, Refference File, Data Conservation and Maintenance

Fig. 3.2.1 OUTLINE OF SYSTEMS FOR MAKING CITY PLAN

3.2.2 The File Concept

A file is a collection of information or data. The diagram shows the structure and contents of three files. Each file belongs to a different section of a corporation.

- The socio-economic section file contains information on population;
- The land-use section file contains information on the various land-uses; and
- The transportation section file contains information on trip production.

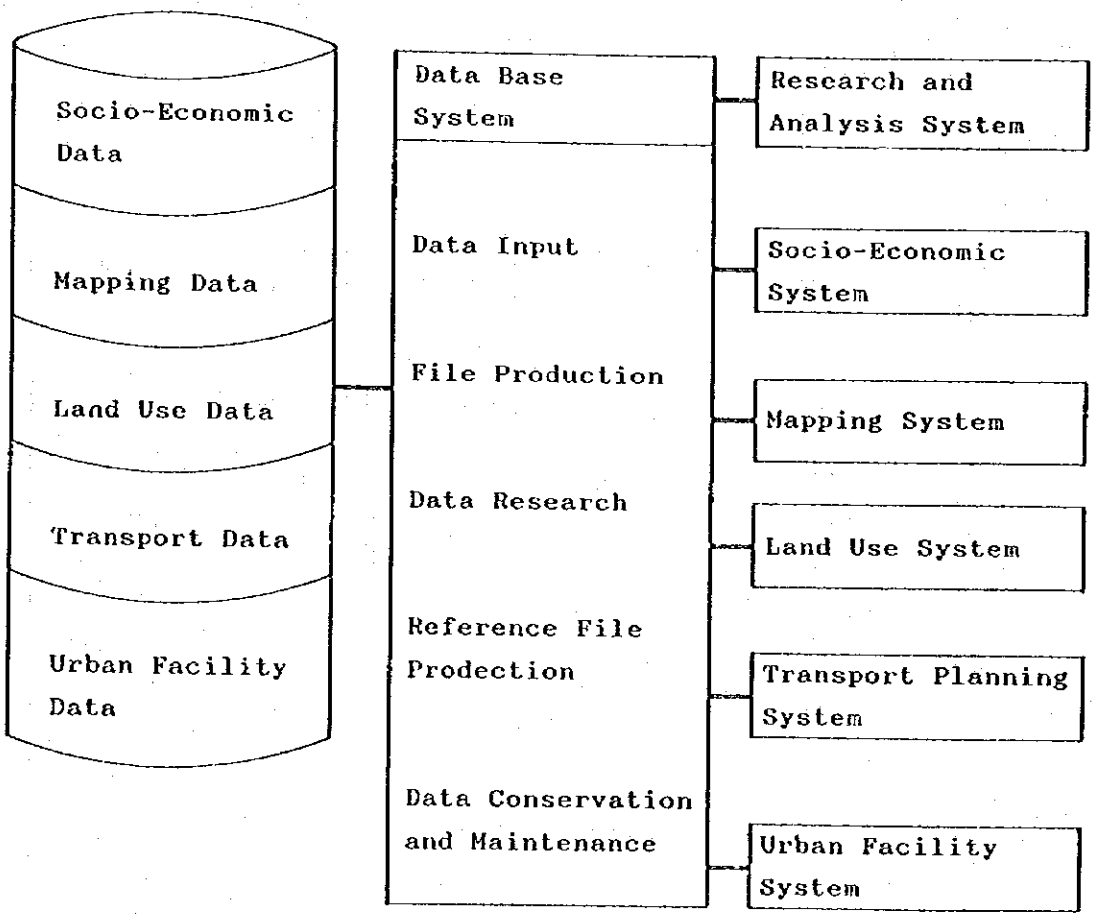


Fig. 3.2.2 CONCEPT OF DATA BASE SYSTEM

Socio-Economic Section
<SCEC. file>

No.	Zone No.	Population
1	0001	100
2	0002	150
3	0003	200
4	1001	100
5	1002	300
6	2001	450

Transportation Section
<TRAN. file>

No.	Zone No.	Population
1	0001	150
2	0002	250
3	0003	300
4	1001	120
5	1002	450
6	2001	600

Land-use Section
<LANU. file>

No.	Zone No.	Residence	Commercial	Industrial
1	0001	10	5	0
2	0002	20	5	10
3	0003	30	10	0
4	1001	10	2	20
5	1002	15	5	0
6	2001	40	4	0

Fig. 3.2.3 FILE CONCEPT

3.2.3 The Data Base Concept

A data base is a collection of related and linked information.

The diagram illustrates a data base with three data base files. There are three points to be noted:

Tabular format
Each file contains data in a tabular format.

Data consistency
Data consistency is achieved because there is minimal duplication of data.

Easy access to data
The three files have the zone number (No.) field in common. When two or more files have fields in common, they may be easily accessed at the same time to retrieve or update information on specified item in all the related files of the data base.

Data Base

No.	Zone No.	Socio-Economic Transportation		Land Use		
		Population	Trip Production	Residence	Commercial	Industrial
1	0001	100	150	10	5	0
2	0002	150	250	20	5	10
3	0003	200	300	30	10	0
4	1001	100	120	10	2	20
5	1002	300	450	15	5	0
6	2001	450	600	40	4	0

Fig. 3.2.4 THE DATA BASE CONCEPT

CHAPTER 4

SYSTEM DESIGN (SOFTWARE)

CHAPTER 4

SYSTEM DESIGN (SOFTWARE)

During the system design, decisions are made about the following components of the new system:

(1) Definition of system function
(2) Input and output items
(3) Codes
(4) File layouts
(5) Input and output layouts
(6) Programme specifications

4.1 Organization of Software System

The design of DBMS should be considered in terms of data availability, accuracy/efficiency of data processing which is highly instrumental to the standardization of city planning process.

From the computerization point of view, as discussed in section 3-2-1, data processing system consists of some functions such as statistical processing programmes, analytical applications and simulation programmes.

The organization of these software system is shown in Fig. 4.1.1.

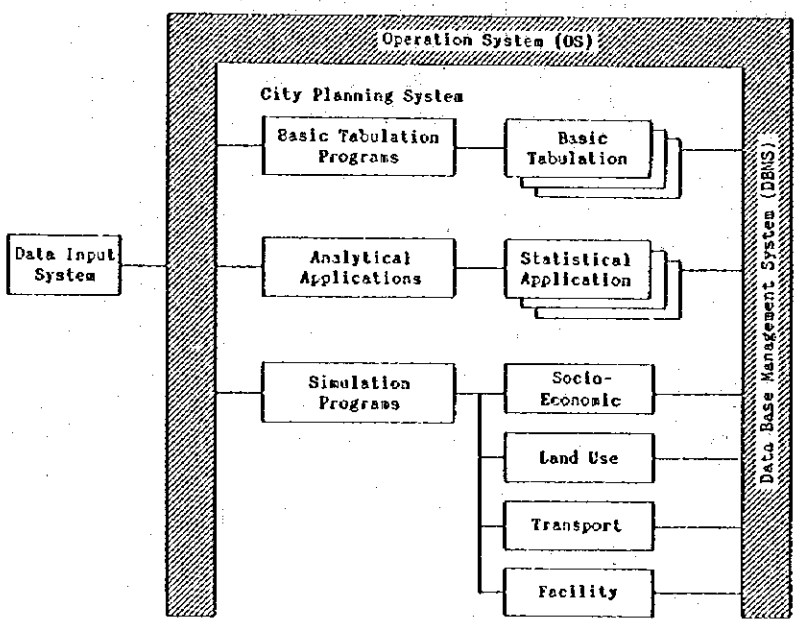
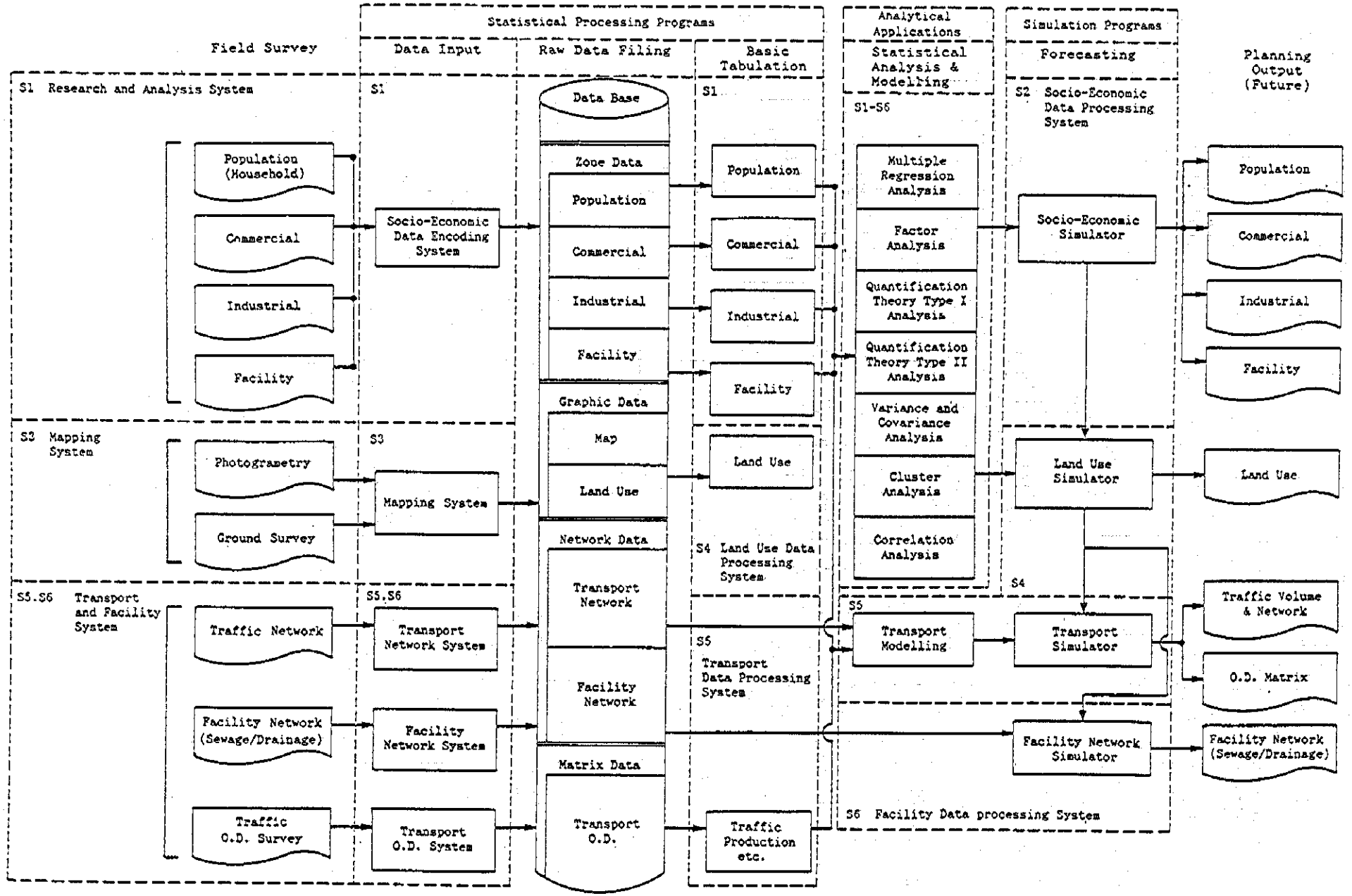


Fig. 4.1.1 BASIC ORGANIZATION OF SOFTWARE SYSTEM FOR THE PROPOSED DBMS

4.2 Definition of System Functions

Depending on the concept of DBMS and the outline of system designed in response to city planning techniques, system functions can be defined as in Fig. 4.2.1. However, these system functions are not only standardized by data items as shown in a later section, but also integrated into a consistent system of various functions. DBMS must fulfill its functions in a whole system of integrated applications. Each functions should be defined as follows:

Fig. 4.2.1 STRUCTURE OF DATA BASE MANAGEMENT SYSTEM (CITY PLANING)



4.2.1 Mapping System

In this function, the current map is presented as final maps of different types and at different scales.

Mapping system function is divided into 11 parts as follows:

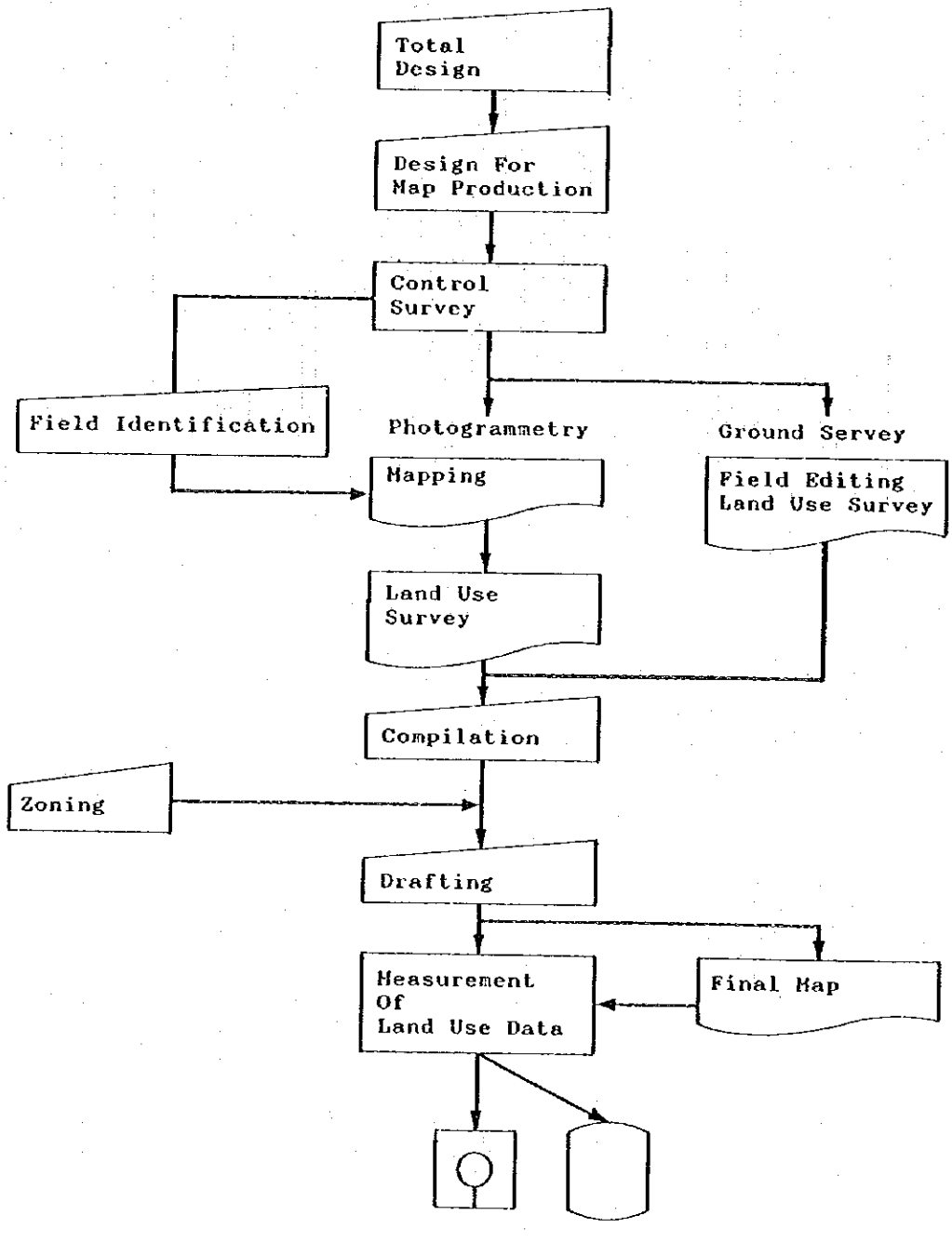


Fig. 4.2.2 Mapping System

4.2.2 Basic Tabulation Programmes

In this function, the data is processed into the form of tables, and the totals for a particular groups of data are calculated. Basic tabulation function is divided into 6 parts as follows:

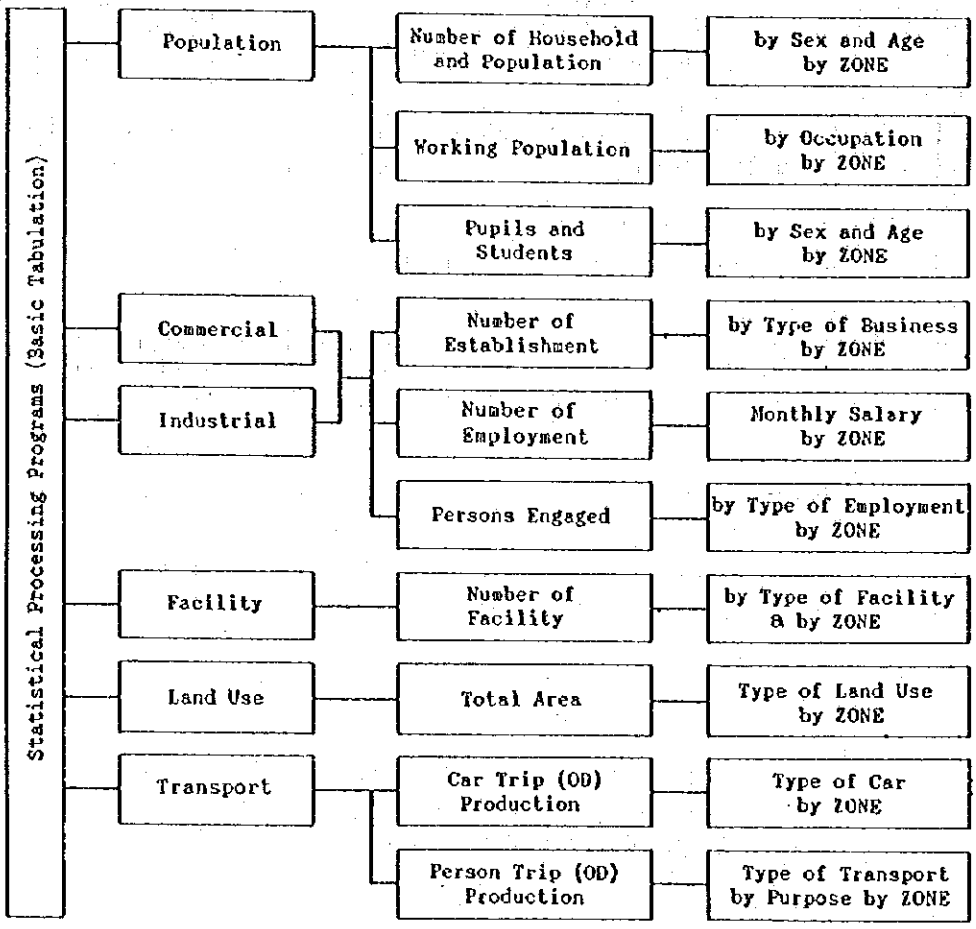


Fig. 4.2.3 System of Basic Tabulation Programmes

4.2.3 Analytical Applications

DBMS should have statistical analysis programmes. Representative techniques are shown in Fig. 4.2.4.

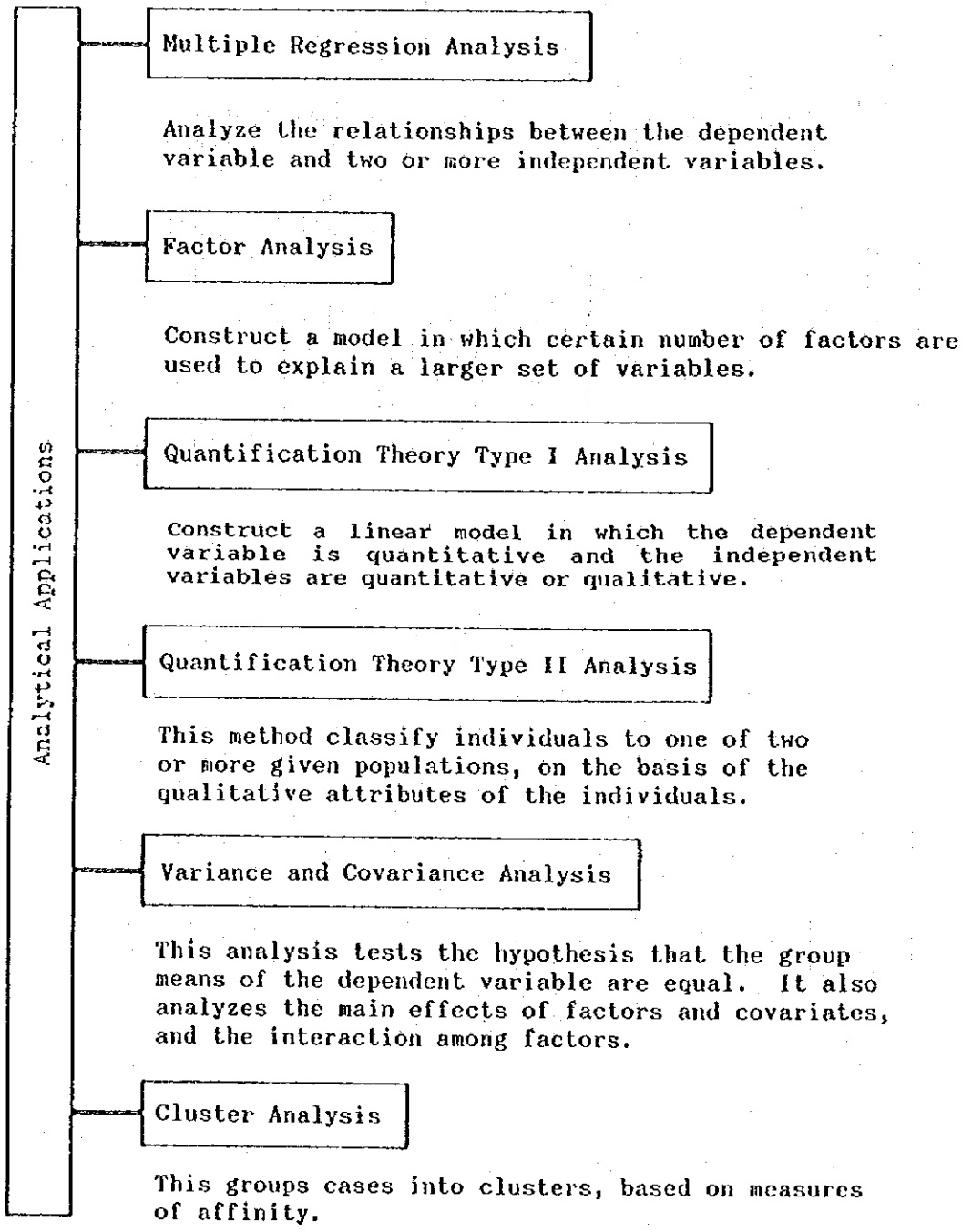
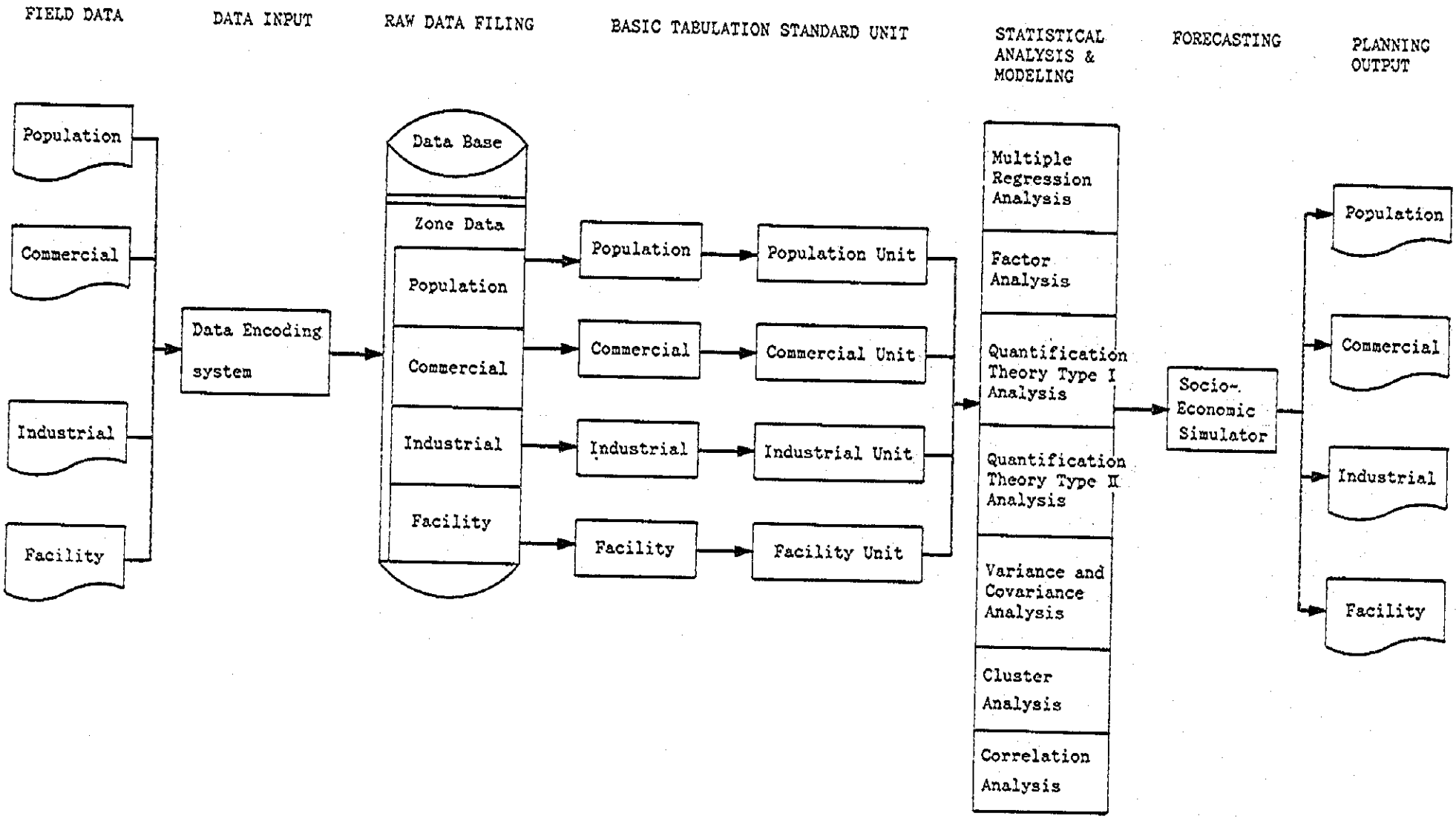


Fig. 4.2.4 Example System of Analytical Applications

4.2.4 Socio-Economic Simulator

Socio-Economic Simulation will be carried out after basic tabulation. Then they will be selected and arranged for analysis by using statistical analysis and modelling as shown in Fig. 4.2.5.

Fig. 4.2.5 STRUCTURE OF SOCIO-ECONOMIC SYSTEM



4.2.5 Land Use Simulator

These steps will appropriately indicate zonal population of residential, commercial, industrial, and agricultural area projections and also urban facility potential area and forecasting model.

Fig. 4.2.6 shows each functions as stated above.

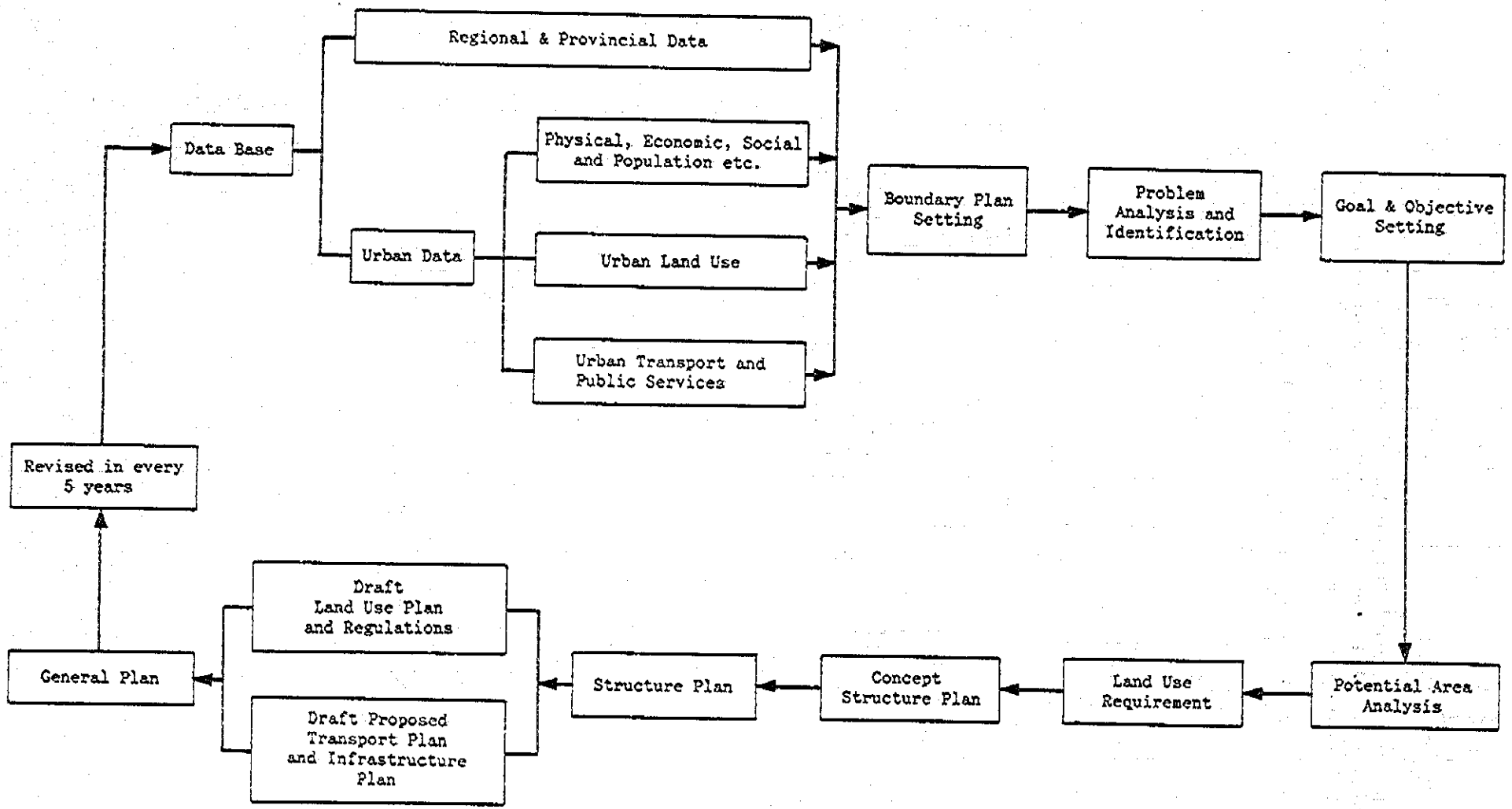


Fig. 4.2.6 LAND USE SIMULATOR

4.2.6 Transportation Simulation

Transportation Simulator should be based on Transportation O-D system as shown in Fig. 4.2.8. This function will be divided into 4 parts such as Traffic Data, Traffic Movement, Transport System, Related Data.

Fig. 4.2.7 shows each functions as stated above.

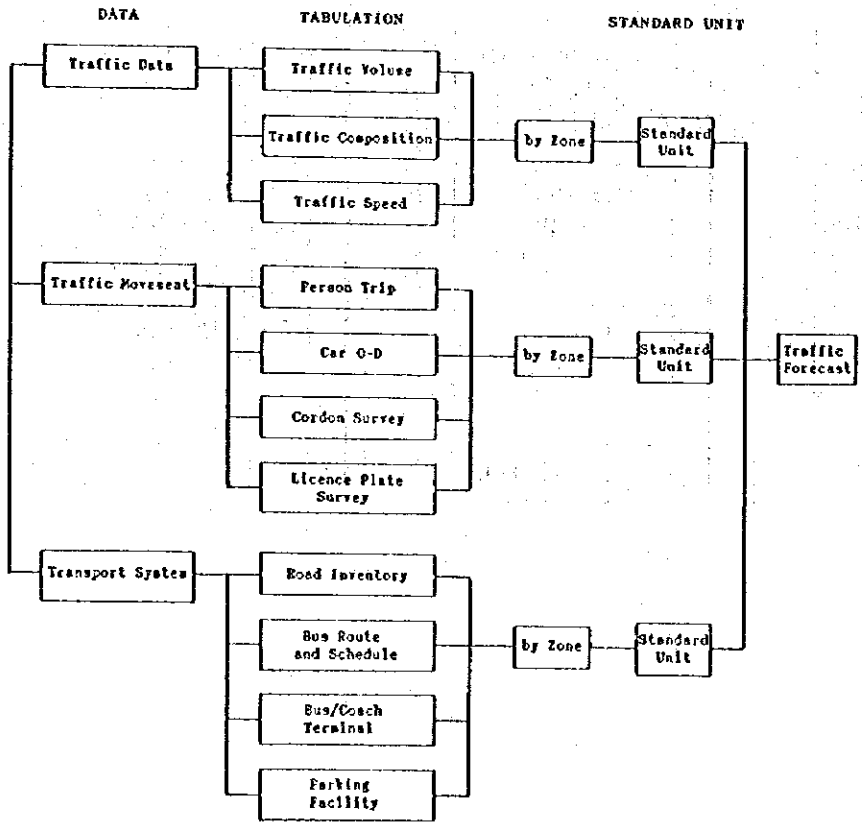


Fig. 4.2.7 Transportation O-D System

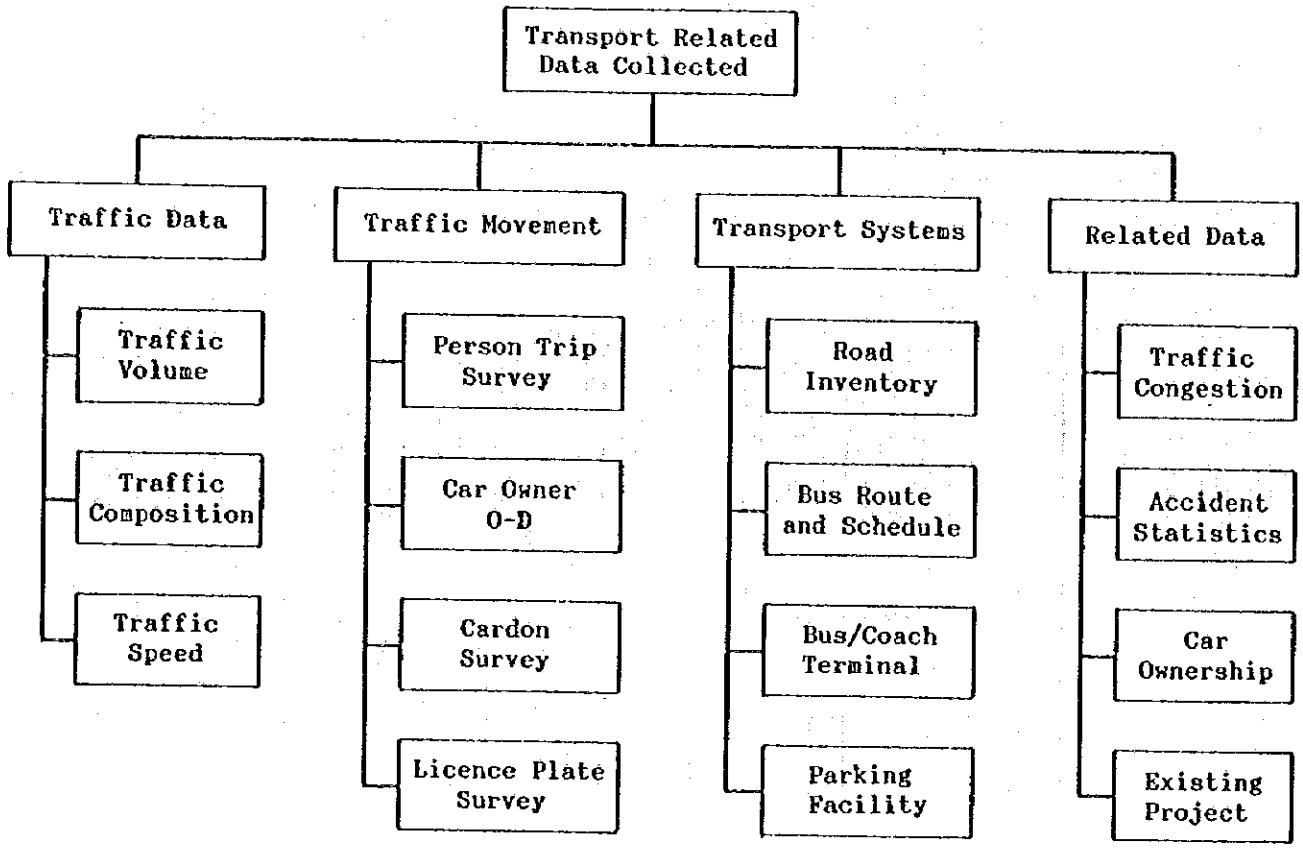


Fig. 4.2.8 Transportation Simulation

4.2.7 Facility Simulator

With regarding to the drainage and sewerage system, this function should be defined as shown in Fig. 4.2.9.

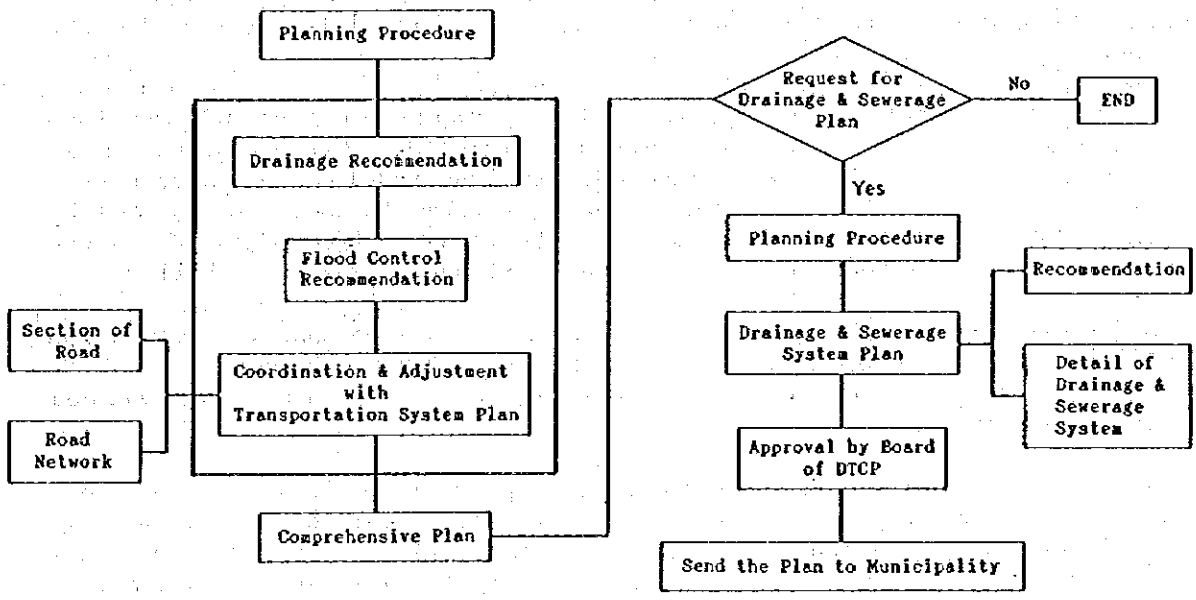


Fig. 4.2.9 Flow of Drainage & Sewerage System

4.3 Input/Output Data Items

4.3.1 Data Processing Procedures (Fig. 4.3.1)

The structure of the data processing chart shows the work procedures for plan formulation in any city planning situation. It is a framework for further investigating and developing a more practical and specific set of procedures for DTCP. An explanation of these procedures at each stage will be given as follows:

The whole process can be grouped into 4 main subjects such as Mapping, Socio-Economic, Land Use and Transportation Systems. Mapping System is a basic system from which a base map is produced. The work starts with a designated planning area on which a method of map producing is selected. The methods include used of areal photographs or by carrying out a ground survey only.

From this system, a land use map or other maps can be produced. The land use map can be used for zone division or the measurement of land use areas which are important values for the later steps in the process.

In the Socio-Economic System, a survey on population, household, land use, and economic activities are of prime concern. These data, including the land use areas measured, are inputted into computer system for basic tabulation, e.g. population tabulation by sex and age and by zone. This basic tabulation, however, is of considerable value in the process of simulation. Each simulation will be explained as follows:

In Socio-Economic simulator, future projected estimates of population and employment are of concern. To complete this work, a standard unit for each matter has to be considered, e.g. sex ratio or school attendance ratio.

The Socio-Economic data should be incorporated with land use areal units, and density or certain standard units can be obtained e.g. Population density, Industrial land standard unit. This standard unit will be used to estimate land requirements in the land use simulator.

The final outputs from the land use simulator are areas required for each activity, e.g. commercial use, industrial use, or residential use, for a determined future time.

These areas will later be allocated in the plan to their appropriate locations having considered all kinds of factors. A draft land use plan will eventually be formulated. In transportation simulator, the 4 steps of transport modeling i.e. trip generation, trip distribution, modal split, and assignment will be carried out. A number of transportation plans can be simulated and evaluated along with the respective land use plan. Then, the final draft general plan can be finally decided.

In addition, if more detailed and complex analysis is to be made for a transportation plan, it can be done through analysis of survey on transportation O.D. system. Although, this system gives more reliable information, it is rather costly to conduct.

4.3.2 Specification of Data Item

Outline of the software system as stated above may be formulated for a standardized/applicable data processing for city planning, but it cannot be realized without specification of data items. Because, each function depends on DBMS, they must be related to each other and to the entire system by processing data. Therefore, data items should be specified taking the following points into consideration:

- Review of current data item and availability of existing data;
- Identification of applications (program detail) in each function; and
- Adjustment of each function by processing data items.

In addition, data items must be specified in response to the requirements of city planning techniques. By considering the review of existing data items and data processing by conducting case studies, data items should be specified first, as shown in Fig. 4.3.2.

However, they cannot be determined without definition of zoning boundaries and a coding system, which should be improved in parallel with the advancement of city planning techniques.

4.4 Code Design [1]

4.4.1 Reasons for Using Codes

Codes are used extensively in DP systems to make processing easier.

When codes are used:

- data is simplified and standardized, and the number of mistakes is reduced;
- classifying, verifying, and totaling operations are easier;
- data is easily identified; and
- processing is more efficient.

4.4.2 Choosing Defective Codes

(1) Classifiability

Codes make it easier to classify (group) the data.

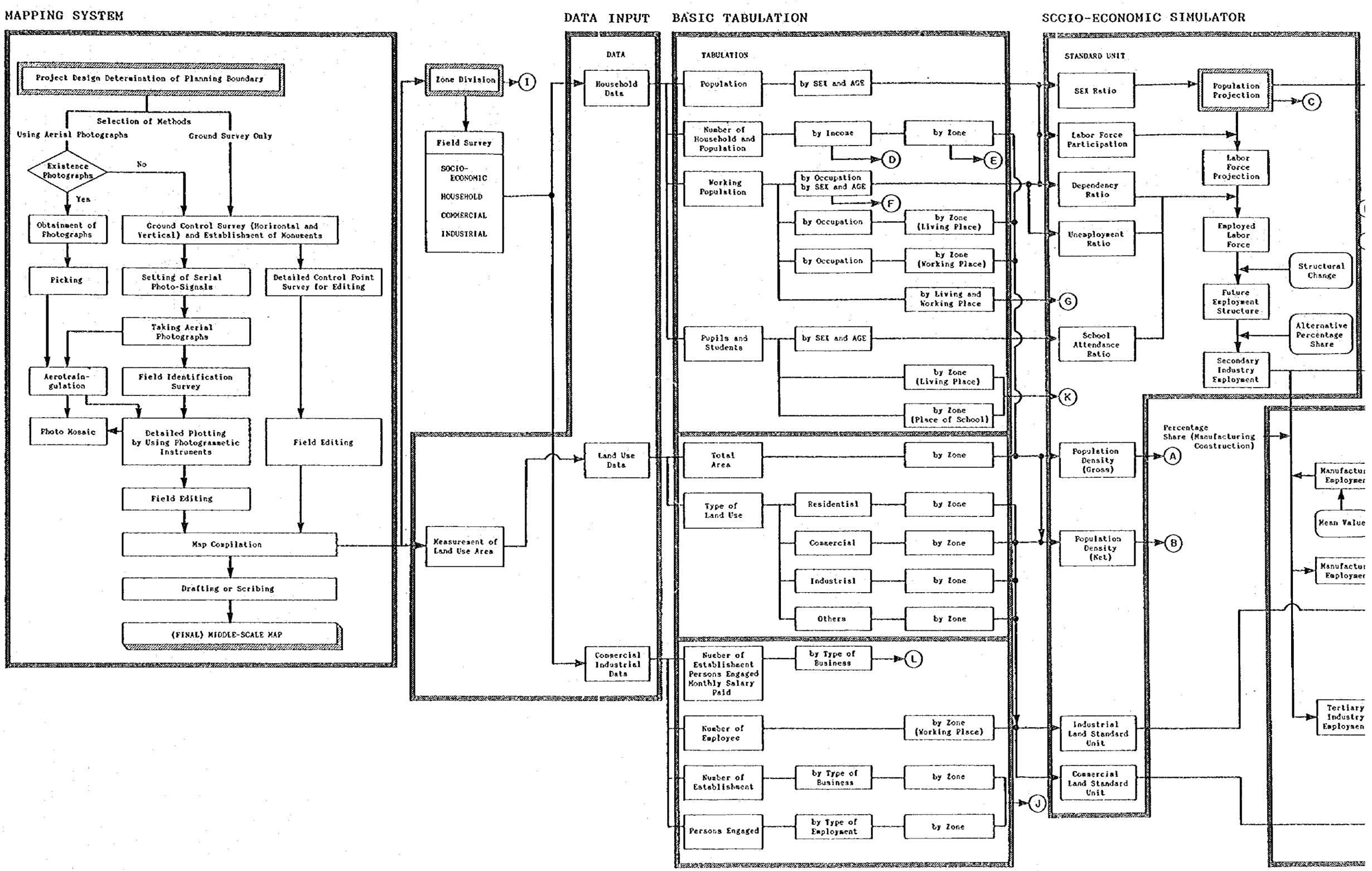
(2) Identifiability

The meaning of the codes should be easily deduced from the codes.

(3) Commonality

Codes should be defined so that they can be used in many different applications. All departments of an organization should be able to use the same codes for the same things.

Fig. 4.3.1 STRUCTURE OF DATA PI



3.1 STRUCTURE OF DATA PROCESSING

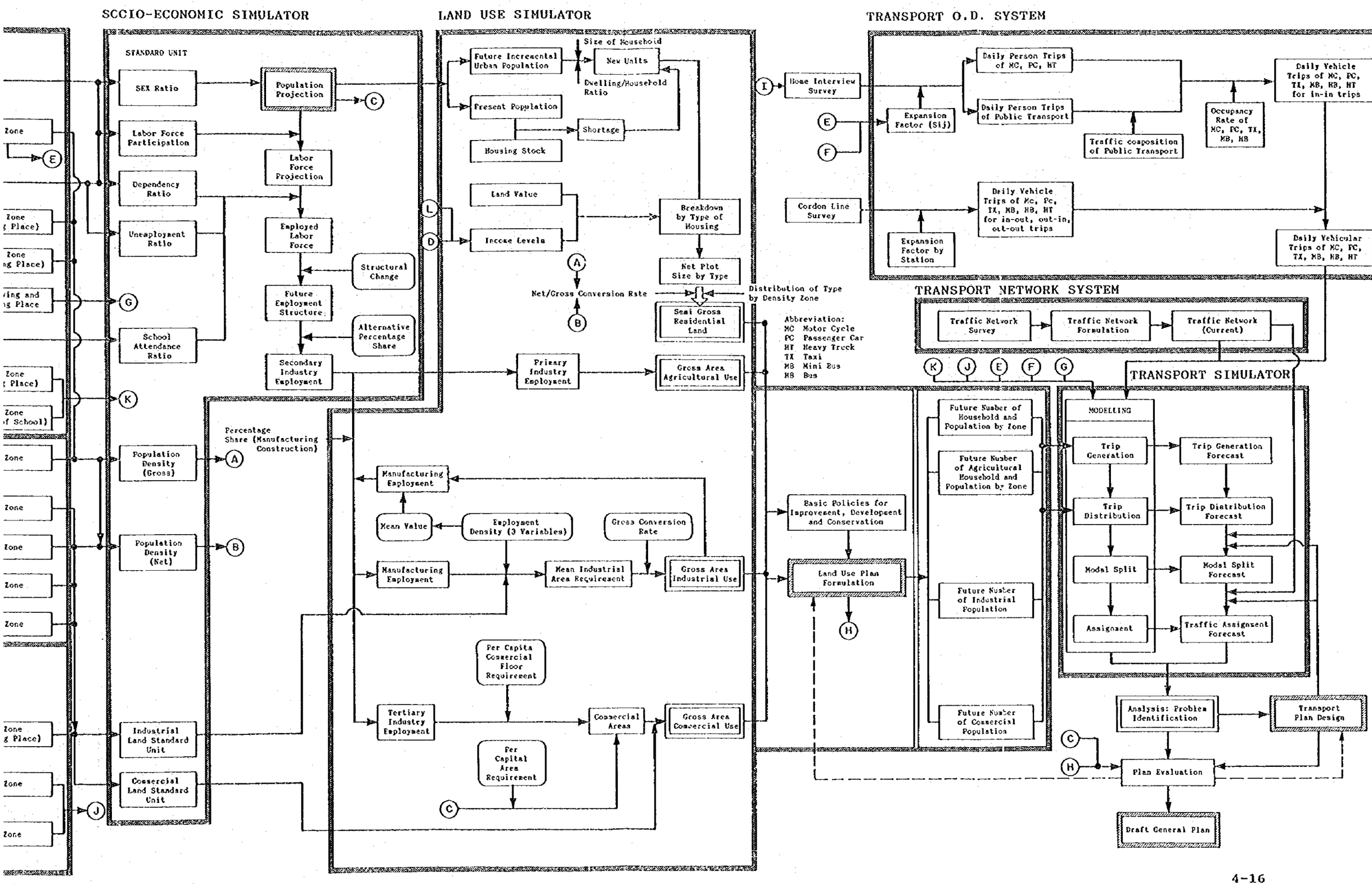


Fig. 4.3.2 ITEM SPECIFICATION

	POPULATION		COMMERCIAL	INDUSTRIAL	LAND USE	
	R & A	ENG	R & A	R & A	MAP	LAND USE
Zone Code	Area Data No. Location L Code	Address	Area Data No. Location L Code	Area Data No. Location L Code		
Item	Type of Household		Duration of Business	Duration of Business	Residential Area	Preservation Area for Residential Purposes
		Number of People	Type of Commercial	Type of Industry		Low Density Residential Area
		Number of Children Aged under 5 Years Old	Major Products	Major Products		Medium Density Residential Area
		Number of Vehicles Owned	No. of Person: Family Male	No. of Person: Family Male		High Density Residential Area
		Average Monthly Income of Family	Family Female	Family Female	Commercial Area	Commercial and High Density Residential Area
	Sex	Sex	Employee Male	Employee Male	Industrial Area	Industrial and Warehouse Area
	Age	Age	Employee Female	Employee Female		Specific Industrial Area
	Race		Age Group	Age Group	Religious Institution	Religious Institution Area
	Nationality				Educational Institution	Educational Institution Area
	Place of Birth		Place of Birth	Place of Birth	Recreation Area	Open Space for Recreation and Conservation of Environmental Quality
	Literacy				Warehouse Area	Warehouse Area
	School Grade Attending		Educational Level	Educational Level	Agriculture Forest, Vacant Livestocks	Rural and Agricultural Area
	Occupation	Occupation				Preservation Area for Rural and Agricultural Purposes
	Place of School	Place of Work or School				Agricultural Land Reformed Area
	Place of Work					Conservation of the Thai Identity, Art and Culture Area
	Monthly Salary		Total Salary Paid per Month	Total Salary Paid per Month	River, Canal	Open Space for Conservation of Environmental Quality and Fishery
	Move In				Government Area Institution Utility	Governmental Institution, Public Utility and Facility Area
	Move Out					
			Residence	Residence	Roads, Soys	
			Area	Area		

(4) Expandability

Code structures should be designed to be able to grow and change as needs grow and change.

For example, if you only have three product categories now, but think that you may have 150 in two years, it would be wise to start with a three-digit numeric category code. A two-digit code would have to be changed when the 100th category was introduced.

. CODE ITEM SELECTION
. CODE ITEM SURVEY
. CODE DESIGN

Fig. 4.4.1 Item Survey Form

ITEM SURVEY FORM

ITEM	SUBFIELD	NUMBER OF VALUES	REMARKS	STABILITY	EXISTING VALUES
ZONE #	CATEGORY SUBCATEGORY	6 200	LOCATION CODE	Static Static	1 6 A113 K774

Fig. 4.4.2 Code Structure Definition

ITEM	LENGTH	DATA TYPE	REMARKS	DATA NAME
ZONE #	10	C	X XXXX XXXXX CATEGORY SUBCATEGORY DATA NO.	ZONE

[1] This section is based on the JICA training material "System Design: Training for Information Processing" OKINAWA International Center, Japan International Cooperation Agency, 1985.

4.4.3 Selecting Items for Coding

When selecting items for coding, it is important to consider the whole organization. If a code has already been designed for the item in another application, the same code should be used in the new system. Any data item may be a candidate for coding, whether it is used for input, output, or in processing logic.

Selected items should be analyzed, taking into consideration the possibility of combining several of them into one code. For example, a book code may consist of the category, subject, and author codes.

4.4.4 Combining Codes

Consider an example where land use class, land use name, and density are selected as being suitable for coding.

Consider an example where class, product name, and capacity are selected as being suitable for coding.

If each were encoded separately, there would be three codes:

- Land use class code
- Land use name code
- Density code

Codes for the same theme should be combined and simplified as much as possible.

In the example shown on Fig. 4.3.3, the three codes could be combined into one product code, with the three items defined as subfields.

The order of subfields within a combined code should be considered carefully.

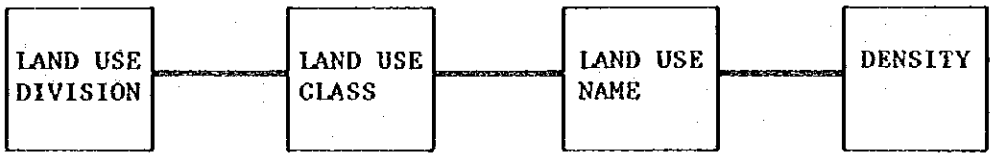


Fig. 4.3.3 Combining Codes

4.4.5 Analyzing Items for Coding

When analyzing items for coding, prepare a draft code diagram by considering the current use of the item and future use in this and other applications.

It is helpful to prepare a code analysis table (as shown on Fig. 4.4.4) by:

- combining any items which are closely related
- listing documents in which the items appear
- listing those subfields of the code which appear in each of the documents
- considering whether the code will be used in other applications
- arranging subfields of a combined code into some meaningful hierarchy from higher to lower level

Code Item	Related Documents	Subfields in Document
Land Use	Land Use Record Socio-Economic Record	Land Use class, name Land Use class, density
Where Used	Land Use Simulator Transportation Simulator	Land Use class, density "
Where Produced	Land Use Record	Produced in Mapping Division

Fig. 4.4.4 Analyzing Items for Coding

4.4.6 Code Diagram Preparation

A code system must be based on accurate and comprehensive information.

(1) Code description

Examine output documents and process flows, and prepare a code description for each application.

Clarify the order of subfields and define the meaning of each.

(2) Code diagram

Assess the growth potential of each subfield and prepare a code diagram from the above, taking into account:

- the possibility of additional subfields or changes in existing subfields
- an evaluation of possible changes in data quantities in each subfield

A change may affect representation of the code (group classification code, sequence code, and so on) and the number of digits in the code.

(3) Product code diagram

Consider expanding the code so that it will be useful for future applications.

(4) Final examination

Make a final evaluation of the code in the light of the following:

- Is it suitable for output?

  Can items be classified and arranged in an appropriate manner? For example, the code should be organized in a suitable way for sorting so that output can be listed in some meaningful order.

- Are there any problems with the process flow?

  Is the code adequate for identification?

  For example, it may be easy for a program to tell that the code value is wrong, but is it possible to tell which part of the code is wrong such as land use name, or density?

- Is there any problem with future use ?

  Consider the possible need for additional data, more detailed fields, and greater amounts of data than estimated. if any of these needs is likely to arise, redundant or dummy digits should be included in the code.

- Is it suitable for other applications?

4.4.7 Preparing Code Establishment Standards

Prepare a document to describe the code in detail. It should be include:

- code name
- subfields
- code length
- format (Such as decimal or character)
- any special values (for example, '9' means produced inside the company)

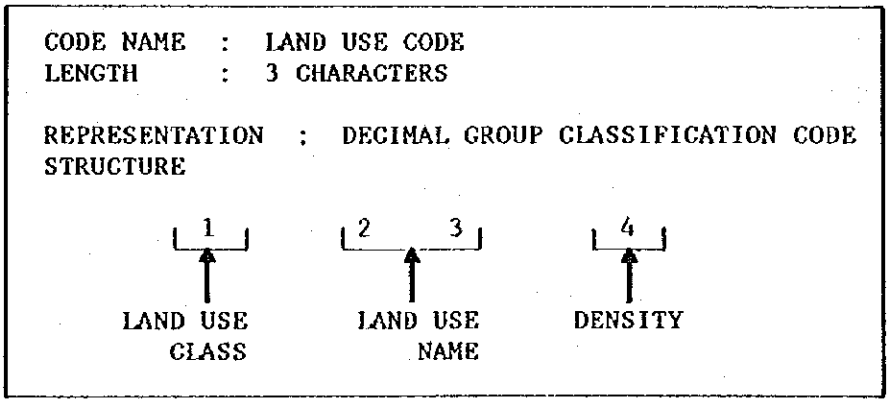


Fig. 4.4.5 Code Design

4.4.8 Definition of Code Design

In response to current city planning procedures, Code Design should be defined as follows:

Table 4.4.1 CODE TABLE FOR HOUSEHOLD SURVEY

CODE	ITEM
Identification:	Type of Household
1	Residential household
2	Commercial household
3	Industrial household
	Number of Family Number
1. 2, 3...n	No.
	Sex
1	Male
2	Female
	Age
1, 2, 3...n	real age
	Race
1	Thai
2	Chinese
3	Indian
4	Others
	Nationality
1	Thai
2	Chinese
3	Indian
4	Other
	Place of Birth
1	Native born
2	Born outside general plan or other Changwat
	Literacy
1	Literacy = (Yes)
2	Illiteracy = (No)
	Status of Education
1	Learning
2	Finish
3	No education

CODE	ITEM
	Level of Education
1	Kindergarten and pre-primary
2	Elementary (Primary Level)
3	Secondary Level
4	Vocational, Cert in ed. and Dip in ed.
5	Bachelor's degree and higher
6	Other education
7	None education
	Place of school
	Code No. of grid cell
	Occupation
1	Professional, Technical and Related Workers
2	Administrative, Executive, Managerial Workers and Government Officials
3	Clerical and Related Workers
4	Sales Workers
5	Agricultural, Animal Husbandry and Forest Workers, Fisherman and Hunters
6	Miners, Quarrymen, Well Drillers and Related Workers
7	Transport Equipment Operators and Related Workers
8	Craftsmen, Production Workers and Laborers
9	Service Workers
10	Workers not classifiable by Occupation
11	Unemployed Persons
12	Housewife
13	Student
14	Unable to work
15	Unknown
16	No education
	Place of Work
	Code No. of grid cell
	Monthly salary
(1, 2, 3...n)	real income
	Reasons for Migration (Move in/Move out)
1	To study
2	To change marital status
3	To work, to look for work
4	Following persons in household
5	Others

Table 4.4.2 CODE TABLE FOR IDENTIFICATION (LOCATION AND SURVEY AREA)

Code	Item
Identification	
	Survey Area
1	General Plan of Tha-Rua
	Data Number
1, 2, 3....n	Data No.
	Location for Area
1	In Tambon Tha-Rua Municipality
2	In Tha-Luang Sanitary District
3	In Tha-Lan Sanitary District
4	Rural in Amphoe Tha-Rua, changwat Phra Nakhpon Si Ayuthaya
5	Rural in Amphoe Tha-Luang, changwat Saraburi
6	Outside General Plan of Tha-Rua
	Location for Grid Cell
A113 ------ A774 B111 ------ B774 C111 ------ C774 D111 ------ D774 E111 ------ E774 F111 ------ F774 G111 ------ G774 H111 ------ H774 I111 ------ I774 J111 ------ J774 K111 ------ K774	
	Number of family number
1, 2, 3....n	No.
	Code of Changwat
1	Krabi
2	Kanchanaburi
3	Kalasin
4	Kamphang Phet
5	Khon Kaen
6	Chanthaburi

CODE	ITEM
7	Chachoengasao
8	Chonburi
9	Chainarth
10	Chaiyaphum
11	Chumphon
12	Chiang Rai
13	Chiang Mai
14	Trang
15	Trad
16	Tak
17	Nakhon Nayok
18	Nakhon Pathom
19	Nakhon Panom
20	Nakhon Ratchasima
21	Nakhon Srithammarat
22	Nakhon Sawan
23	Nonthaburi
24	Narathiwat
25	Nan
26	Buri Rum
27	Pathum Thani
28	Prachaup Kirikhan
29	Prachin Buri
30	Pattani
31	Phra Nakhon Si Ayuthaya
32	Payoa
33	Pang Nga
34	Pattaluang
35	Pichit
36	Pitsanulok
37	Petchaburi
38	Petchabun
39	Prae
40	Phuket
41	Mahasarakham
42	Maehongson
43	Yasothon
44	Yala
45	Roi-Et
46	Ranong
47	Rayong
48	Ratchaburi
49	Lop Buri
50	Lampang
51	Lampoon
52	Loei
53	Sri Sa Ket
54	Sakon Nakhon
55	Songkhla
56	Satun

CODE	ITEM
57	Samut Prakarn
58	Samut Sonkhram
59	Samut Sakhon
60	Saraburi
61	Singburi
62	Sukhothai
63	Surat Thani
64	Surat Thani
65	Surin
66	Nhongkai
67	Ang Thong
68	Udon Thani
69	Uthai Thani
70	Uttaradit
71	Ubon Ratchathani
72	Bangkok Metropolis
73	Mukdahan

Table 4.4.3 CODE TABLE FOR COMMERCIAL SURVEY

Table Code	Table for Commercial Survey
Identification:	
	Duration of Business
1, 2, 3....n	No. of years
	Religion
1	Buddhism
2	Christianity
3	Islam
4	Others
	Type of business
1	Retail trade
2	Wholesale trade
3	Services
4	Retail and Wholesale trade
5	Retail trad and services
6	Wholesale trade and services
7	Retail, wholesale and services
	Kind of goods and services
1	Other food and beverages stores
2	Grocery stores
3	Miscellaneous retail stores
4	Finished cloth stores
5	Cloth making shops
6	Women's hairdressing and beauty, Men's barber shops
7	Wearing apparel and accessory stores
8	Car, motorcycle, bicycle and accessory stores
9	Electric lighting equipment stores
10	Pharmacy stores
11	Plumbing, sanitary equipment and other construction materials stores
12	Furniture and fixture stores
13	Jewelry stores (Gold ornament stores)
14	Book, periodical and newspaper, stationery stores
15	Photographic studios
16	Agricultural machinery and equipment
17	Gasoline service stations
18	Turkish bath and massage parlors
19	Souvenirs and Thai handicraft shops
20	Hotels and other lodging places
21	Theatre
22	Bank
23	Transportation services

CODE	ITEM
24	Other personal services not elsewhere classified
25	Pawn shops
26	Watches, clocks, glasses stores and repair
27	Offerings
28	Advertising place
29	Laundries, laundry services and cleaning
30	Movie and V.D.O. for rent
31	Others
	<u>Persons Engaged in the business</u>
1, 2, 3....n	Family member - male
1, 2, 3....n	Family member - female
1, 2, 3....n	Employee - male
1, 2, 3....n	Employee - female

Table .4.4.4 CODE TABLE FOR INDUSTRIAL SURVEY

Table Code	Table for Industrial Survey
Identification:	
	Duration of Business
1, 2, 3....n	
	Characteristic of Building
1	Isolated
2	Share space
	Type of Industry
1	Service industry
2	Manufacturing industry
3	Handicraft industry
	Major Product
1	Manufacture of foods, drinks and products
2	Manufacture of made-up textile goods wearing apparel and leather products
3	Manufacture of wooden
4	Manufacture of pulp and paper, print and printing services
5	Manufacture of basic industrial chemical rubber, plastic
6	Manufacture of non-metallic mineral products
7	Manufacture of metal product
8	Manufacture, repairing of machinery, tools electric lamps and transportation equipment
9	Other industries
	Persons engaged in industry
1, 2, 3....n	Family member - male
1, 2, 3....n	Family member - female
1, 2, 3....n	Employee - male
1, 2, 3....n	Employee - female
Observation:	1. Lacking of time for preparation before starting the Tharua field survey, we used the old survey sheets which were difficult for coding
	2. The statistical officers have to spend a lot of time for coding because they have never been Trained to do this process before.

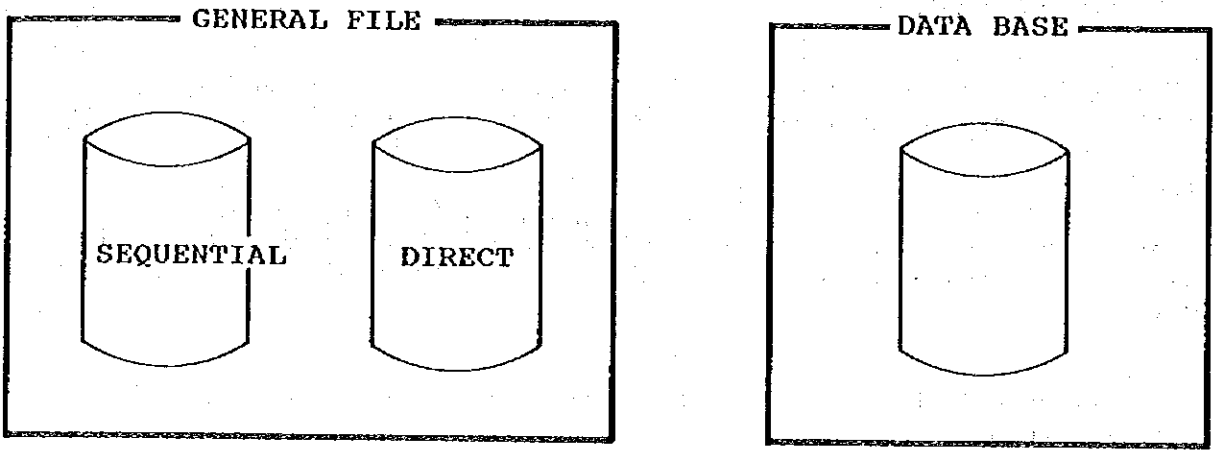
4.5 File Design

4.5.1 File Design

A DP system collects, edits, stores, and retrieves data. The way in which this data is managed can greatly influence the efficiency of the system.

During system design, decisions must be made about the types of files to be used.

We will use the term "file" to mean general files which have a linear structure, and the term "data base" for files which have a hierarchical or relational structure.



Design
- Data items to be recorded
- File organization

Fig. 4.5.1 File Design

4.5.2 File Types

Files are used in several different ways:

(1) Master File

A master file contains information which serves as the basis for a particular application. Transaction data adds, chages, or deletes master file information. Examples are a customer master file, product master file, and stock master file.

(2) Transaction File

Transaction files contain information used to update or access the data in a master file.

---

[1] This section is based on the JICA training material "System Design: Training for Information Processing" OKINAWA International Center, Japan International Cooperation Agency, 1985.

(3) Intermediate File

Intermediate files provide temporary storage for processing results. For example, a transaction file may be edited by a programme to remove obvious data errors. The resulting intermediate file would then be used to update the master file.

(4) Other Files

These include the special-purpose files such as history files and summary files.

4.5.3 File Design

Decide how the file will be used and determine its file organization by considering the following:

Data Items
Select only those data items which are necessary. For example, is it really necessary to store the name of customer's grandmother ?

File Layout
Group data items which are processed together and arrange them in order of priority. For example, street number, street name, suburb, town, and postal code should be stored together as they are all items in the address.

Volume
Calculate record sizes. Determine the minimum, maximum, and most likely number of records.

4.5.4 Organizing the Data

Once the programmes have been designed and flowcharts created, the common files and tables can be designed. The following points should be considered:

Key Items
Select appropriate fields to be used as a key to access the records in the file. If many programmes access the file, try to decide on a key field which will suit them all. Example are customer code or customer surname.

Splitting or Combining Files
If a file contains many large records, it may be more efficient to split the file into two files with smaller records. If a file contains records with only one or two items, it may be better to combine it with another file.

Interface between Programmes
Make sure that data to be passed from one programme to another is in a format that suits them both. For example, dates should have the same format in all files.

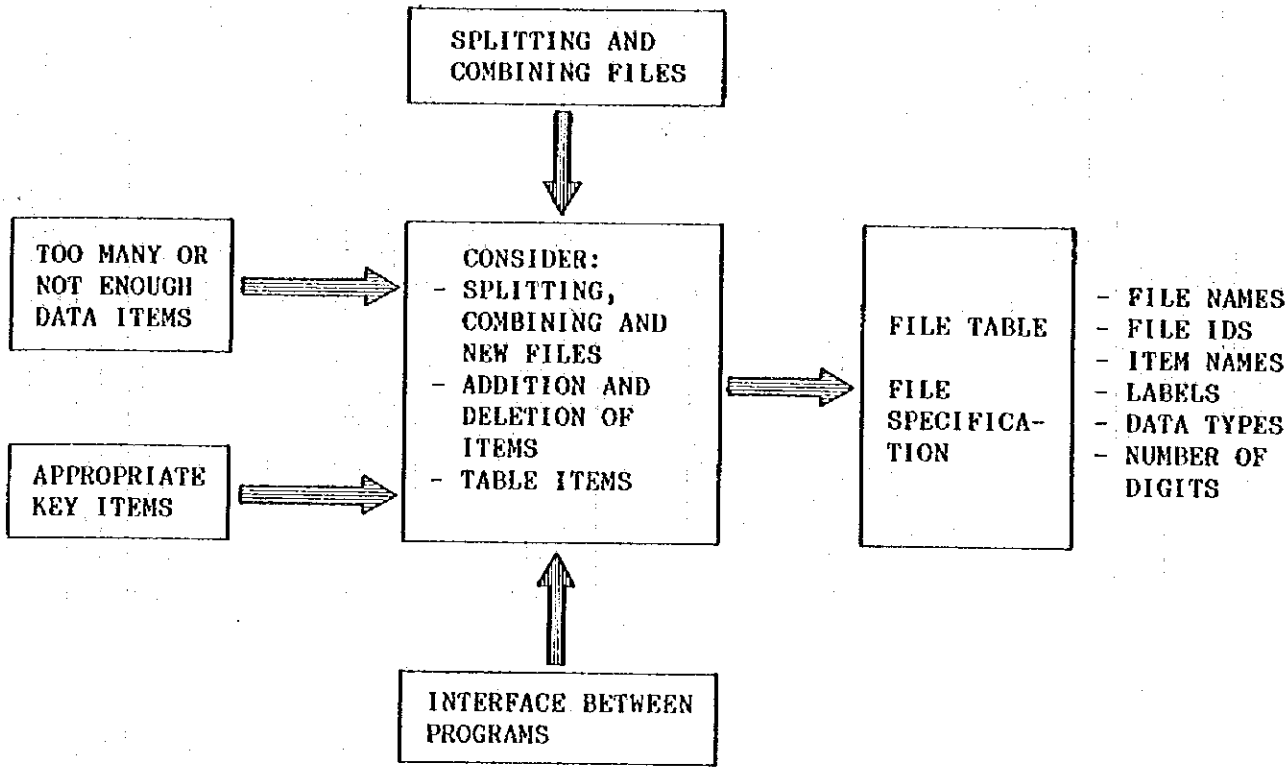


Fig. 4.5.2 Organizing the Data

4.5.5 Deciding File Format

For each file, decide and document the following:

- File Name
- Data Item Names
- Data Item Types
- Data Item Lengths
- Data Value Meanings and Limits
- Provision for Future Growth

Fig. 4.5.3 HOUSEHOLD FILE

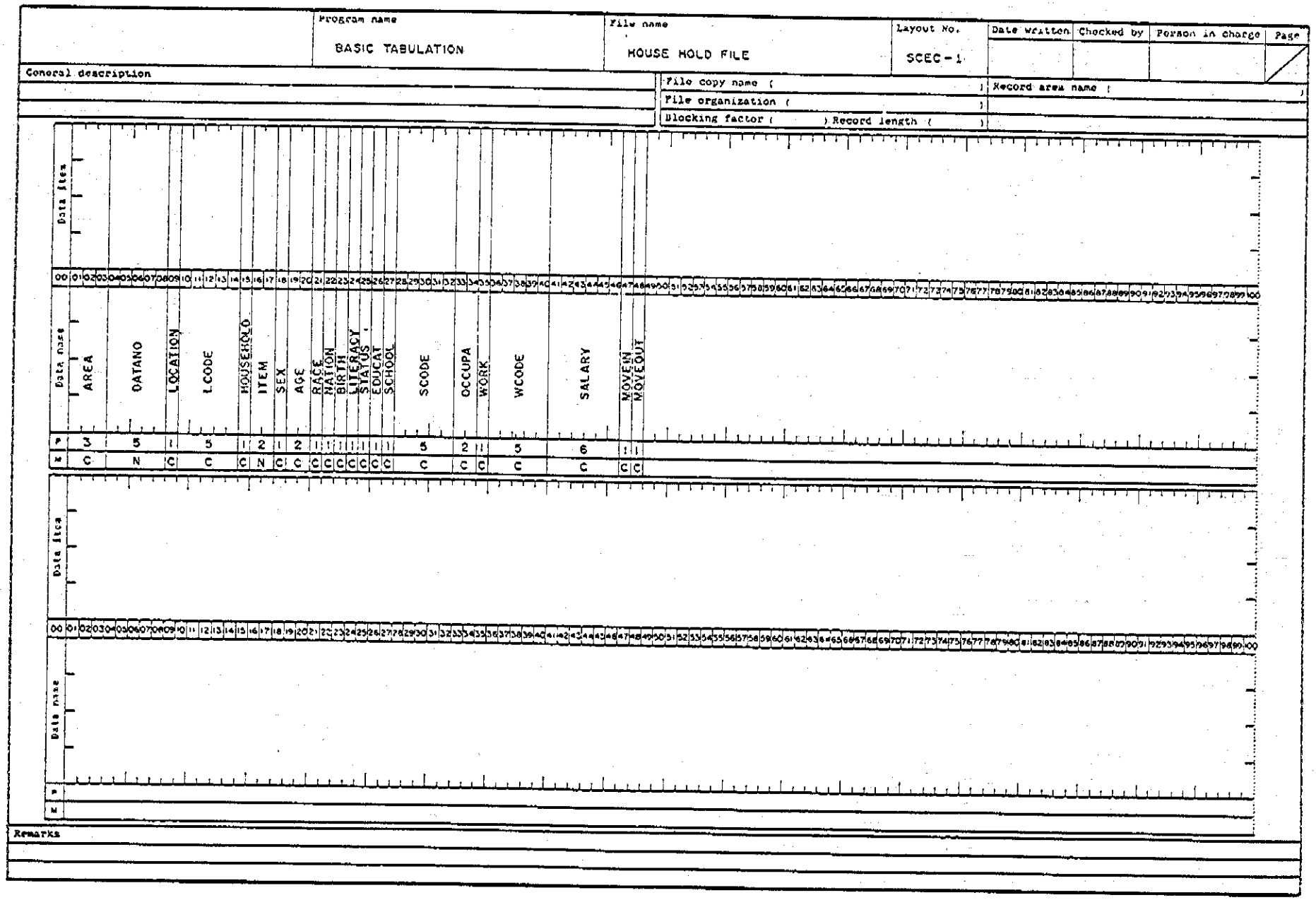


Fig. 4.5.4 COMMERCIAL FILE

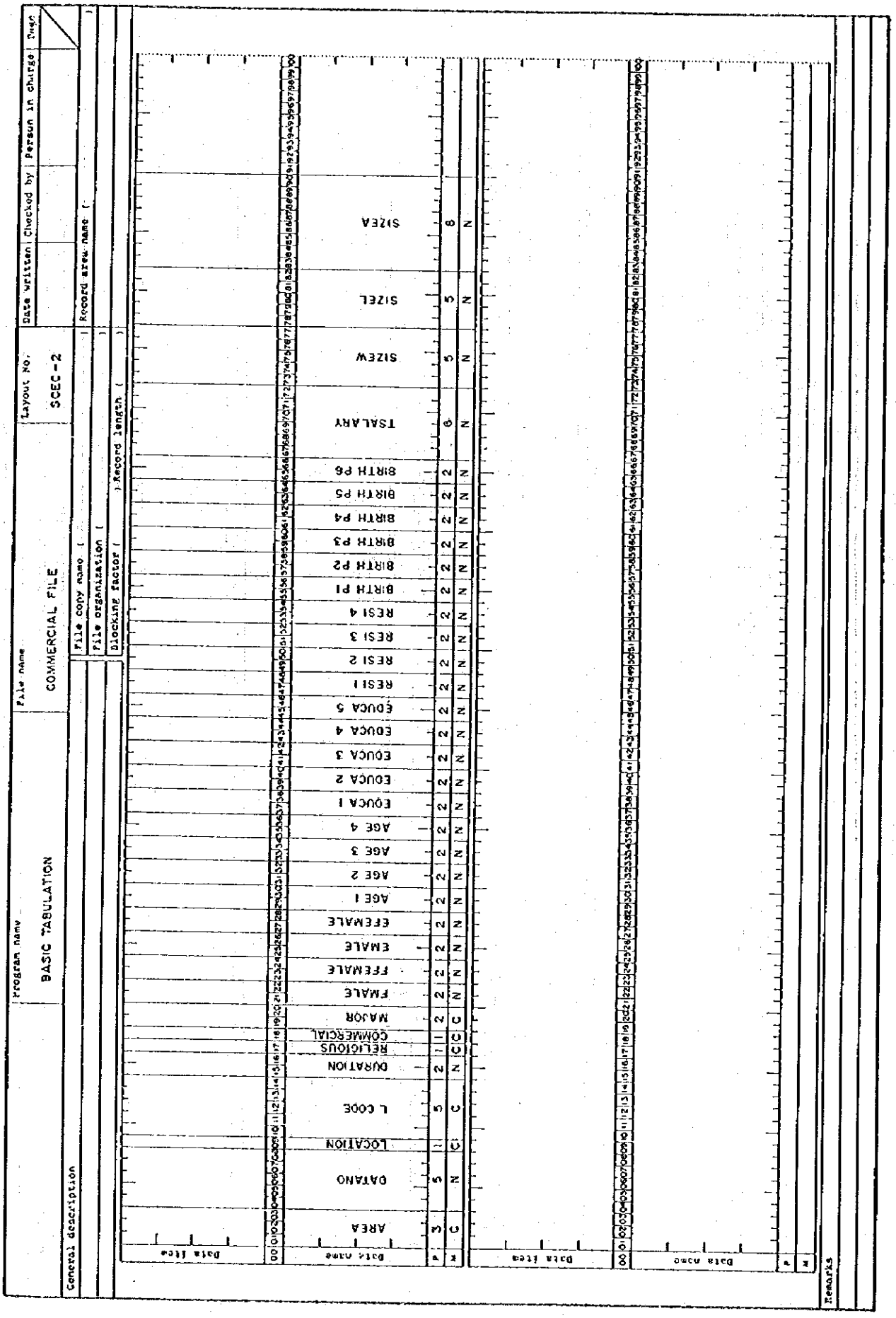
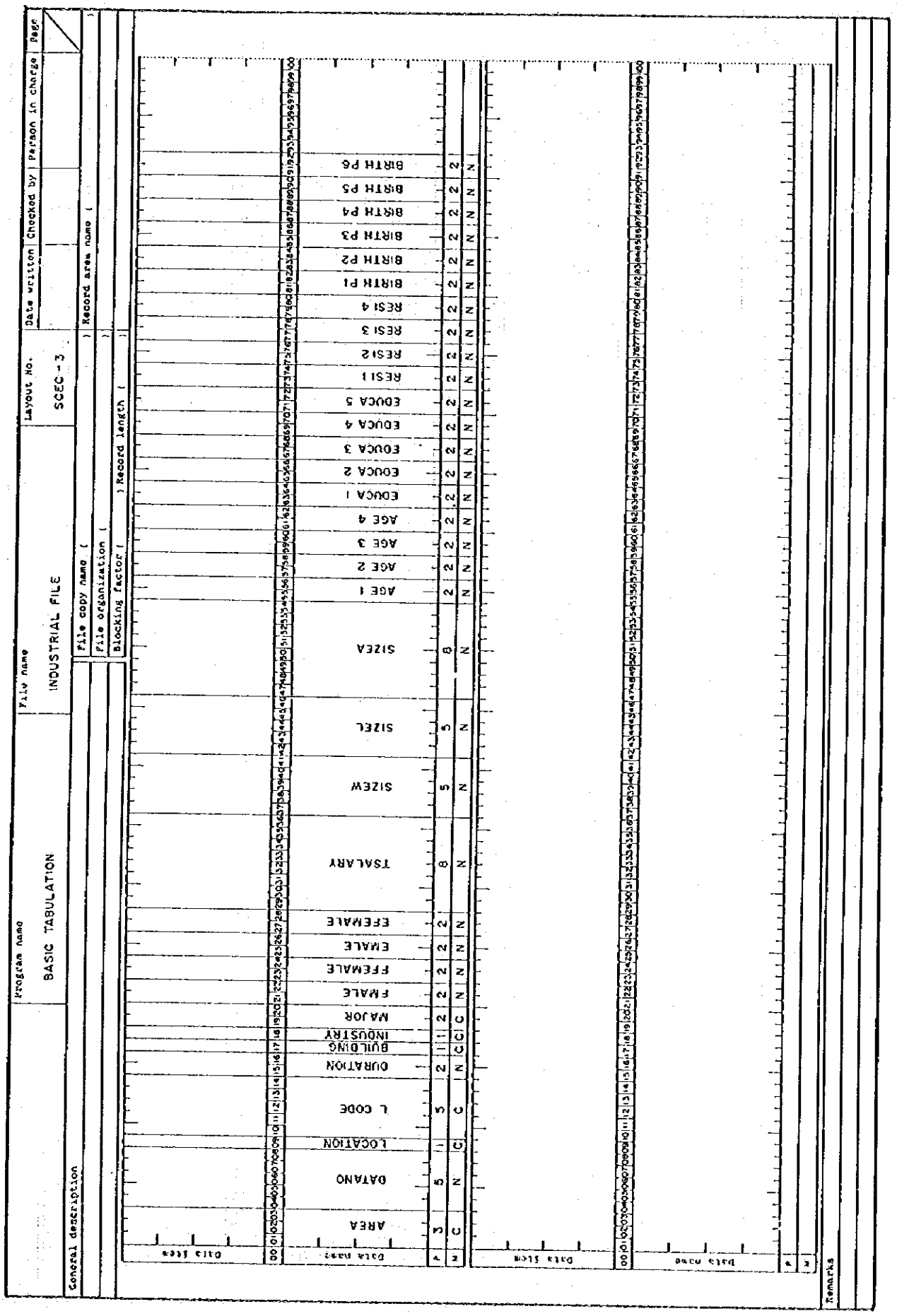


Fig. 4.5.5 INDUSTRIAL FILE



4.6 Data Base Management Package [1]

4.6.1 Data Base Management on Micro-Computers (dBASE III Plus)

(1) Introduction of dBASE III Plus

dBASE III Plus is a commercially distributed software package. It is used for data base management on micro-computers.

dBASE III Plus provides a powerful interactive environment and supports a sophiscated language for managing data.

In the interactive mode, data may be manipulated, and results displayed or printed by entering simple English-like commands.

Assume that there are two data base files on the disk. They are SCEC.DBF and LANU.DBF.

Two commands have to be entered to list the contents of the SCEC file:

The First Command Is:

. USE SCEC Name of a data base file

This command is used for bringing the SCEC.DBF file into use. Once the file is open, all data manipulation commands will operate on this file only.

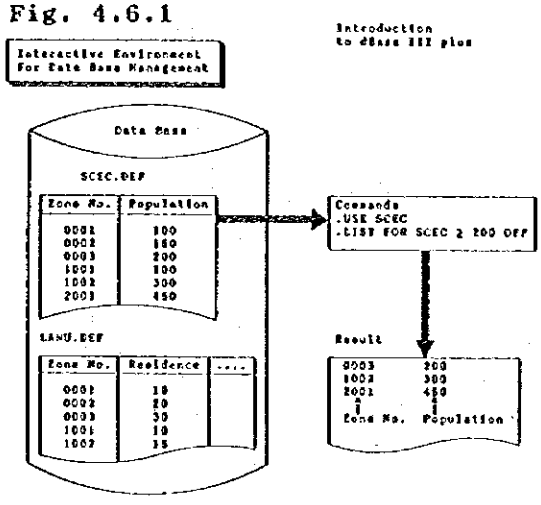
The Second Command Is:

. LIST FOR < search - condition >

This command is used for listing on the console the contents of the SCEC file.

It has also been given a search condition. In the diagram, the search condition specified is population ≥ 200.

Fig. 4.6.1



[1] This section is based on the JICA training material "Database Introduction: Training for Information Processing" OKINAWA International Center, Japan International Cooperation Agency, 1987.

(2) Tabular Format

As shown in the diagram, data items are stored in the data base in a tabular format. Each row in the table is known as a record, and each column is known as a field. Every field contains a data item. In the table shown, there are five fields within a record, and there are seven records in the file.

Each field is identified by a unique name. In this case, the field names are NO, PRODUCT, IN, OUT, and STOCK. Each record is identified by a record number. The record number indicates the physical location of the record in the file. The record at the top is record number 1.

The structure (tabular format) of a file has to be specified when it is first created. To do this, the following information is required:

Field Name

This is the unique name by which a field can be identified. The name can have a maximum length of ten characters and may be alphanumeric. It should not have embedded blanks but may contain colons (:).

Data Type

The three data types are indirected by C, N, and L.

-- C: Indicated character data. These may contain alphanumeric characters including special characters, but excluding control characters.

-- N: Indicates numeric data. These may contain numbers, with or without minus sign. They may be used in calculations.

-- L: Indicates logical data. These indicate true or false values. They are denoted by T, t, Y, or y for true (or yes) and F, f, N, or n for false (or no).

Field Width

This indicates the maximum width of data to be stored in a field. A logical field has a width of one character.

Decimals

This indicates the number of decimal places in a numeric field.

Here are the limitations and constraints:

- Records in one data base file: 65,535 (maximum)
- Characters in one record: 1,000 (maximum)
- Fields in one reocrd: 32 (maximum)
- Characters in one field: 254 (maximum)
- Accuracy of numeric fields: 10 digits

4.6.2 Data Base Management on Online System

(1) Development of Data Base

In conventional processing, each application has files for exclusive use.

Data is stored in these exclusive-use files so that it can be efficiently used for each application.

In a data base system, exclusive-use-file data for each application is collected and managed collectively as a data base.

Each application does not use exclusive-use files any more. Instead, all applications share the same data base. Data bases were developed independently of online systems.

However, as the scale of online processing increased, data bases were introduced into online systems because data in the system must be collectively managed and always kept current.

(2) Introduction of Online Data Base System

Online Data Base should be developed by combining Online System with Data Base System. The concept of Online Data Base Management System is shown in Fig. 4.6.3.

Online DBMS is enclosed in bold lines. Online DBMS operates together with the Operating System and uses the job management, task management, and data management functions of the operating system.

When data is to be sent or received in online mode, a communication control programme is used. This system consists of four parts called software components: DCMS, ISMS, DBMS and DDMS.

Moreover, this system has an exclusive-use file to store information required for operations. This file is called DD/D.

Fig. 4.6.2

Data is stored in a tabular format

No.	Zone No.	Socio-Economic Transportation		Land Use		
		Population	Trip Production	Residence	Commercial	Industrial
1	0001	100	150	10	5	0
2	0002	150	250	20	5	10
3	0003	200	300	30	10	0
4	1001	100	120	10	2	20
5	1002	300	450	15	5	0
6	2001	450	600	40	4	0

Field Field Field Field Field Field Field

Record

Data Base Structure

FIELD	NAME	TYPE	WIDTH
Field 1	SAMNO	C	005
" 2	ZONNO	N	006
" 3	POPUL	N	006
" 4	TRIPP	N	007
" 5	LURES	N	003
" 6	LUCOM	N	003
" 7	LUIND	N	003

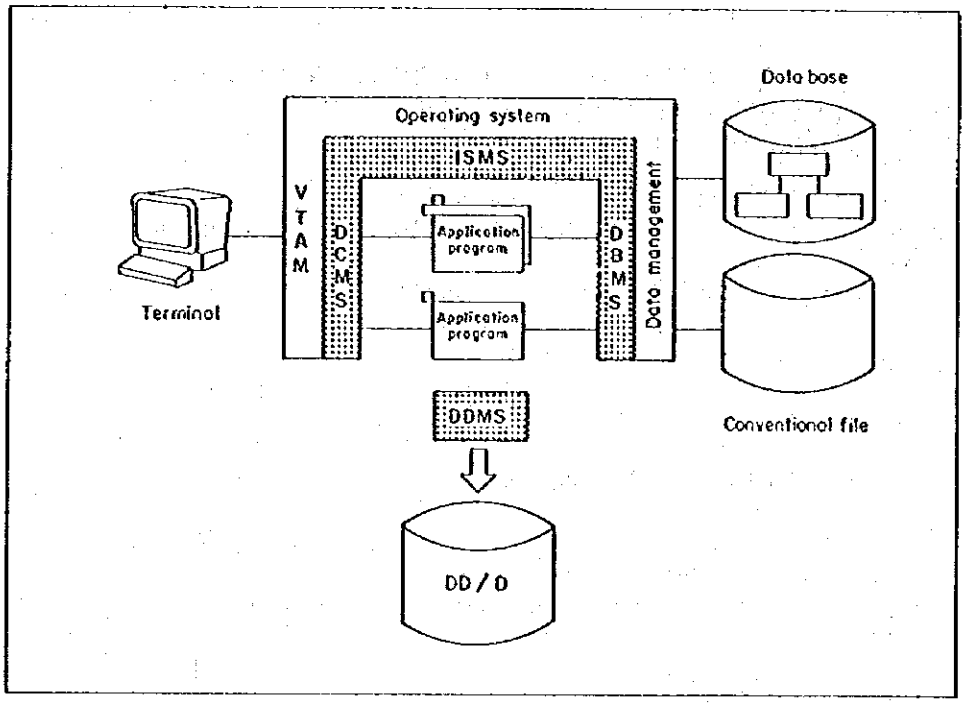


Fig. 4.6.3 Concept of On-Line - Data Base

Reference

VTAM: Virtual Telecommunication Access Method
ISMS: Integrity and Schedule Management Sub-System
DCMS: Data Communication Management Sub-System
DBMS: Data Base Management Sub-System
DDMS: Dictionary and Directory Management Sub-System
DD/D: Data Dictionary/Directory

4.7 Programme Specification

Define a processing outline for each module. This becomes the basis of the detailed programme specification which may be prepared later. The outline should include:

- a description of the input data items
- a description of the processing flow, including editing rules and error handling
- a description of the output data items and files

4.7.1 Mapping System

(1) Design for Map Production

a. Definition of boundaries of project area
   - sheet size
   - sheet lay-out
   - existing geodetic control
   - geographical coordinates
   - etc.

b. Type of product
   - line map
   - digital data concerning points or lines in the form of lists, floppy disk, magnetic tapes

c. Map scale

d. Sheet size, sheet lay-out, projection, grid

e. Type of semantic information to be presented and symbols to be used (legend)

f. Type of relief presentation
   - by spot height
   - by contours
   - by form-lines
   - by other means

g. Marginal information
   - Name or code of map-sheet
   - Legend
   - Method of compilation
   - Magnetic declination
   - Name of executing organization
   - Date of executing of different phases
   - Accuracy
   - Index

h. Type of reproduction
   For all types of maps detailed specifications concerning the method, means and materials to be used for reproduction have to be specified.

(2) Control Survey

- Existing survey framework
- Field reconnaissance
- Planimetric control (e.g. Traversing + EDM equipment)
- Height control (e.g. Spirit or direct levelling)

(3) Field Identification (before restitution)

Identification in the field of those natural and cultural details that cannot be readily identified as such in the photography, and subsequent annotation of these details on the photographs on behalf of the later restitution e.g., deepwells, culverts, power line, vegetations, etc.

(4) Mapping by Means of Photogrammetry (Stereo restitution)

- Orientations (inner, relative and absolute orientations)
- Plotting (planimetry and height plotting)

(5) Mapping by Means of Ground Survey

a. Field Editing
   - line measurement (track, laterite or asphaltic or concrete road, etc.)
   - building measurement (wooden house, concrete house)

b. Land Use Survey (existing land use)
   - Residential area
   - Commercial area
   - Industrial area
   - Warehouse area
   - Religious institution
   - Educational institution
   - Institutional, utility and facility area
   - Livestocks
   - Open space for recreation, play ground and park

c. Place names, etc.

d. Boundary and property of government ownership determination

e. Accurate measurement of short distances for cadastral or engineering purposes

(6) Land Use Survey and Field Completion (after restitution)

- Road and railway classification
- Drainage system classification
- Vegetation classification (rice field, garden, horticulture, etc.)
- Boundary and property of government ownership determination
- Place names, etc.
- Eaves measurements if bases of buildings must be determined for large scale mapping
- Detail survey to obtain information not available from the photograph such as details obscured by trees or clouds
- Accurate measurement of short distances for cadastral or engineering purposes

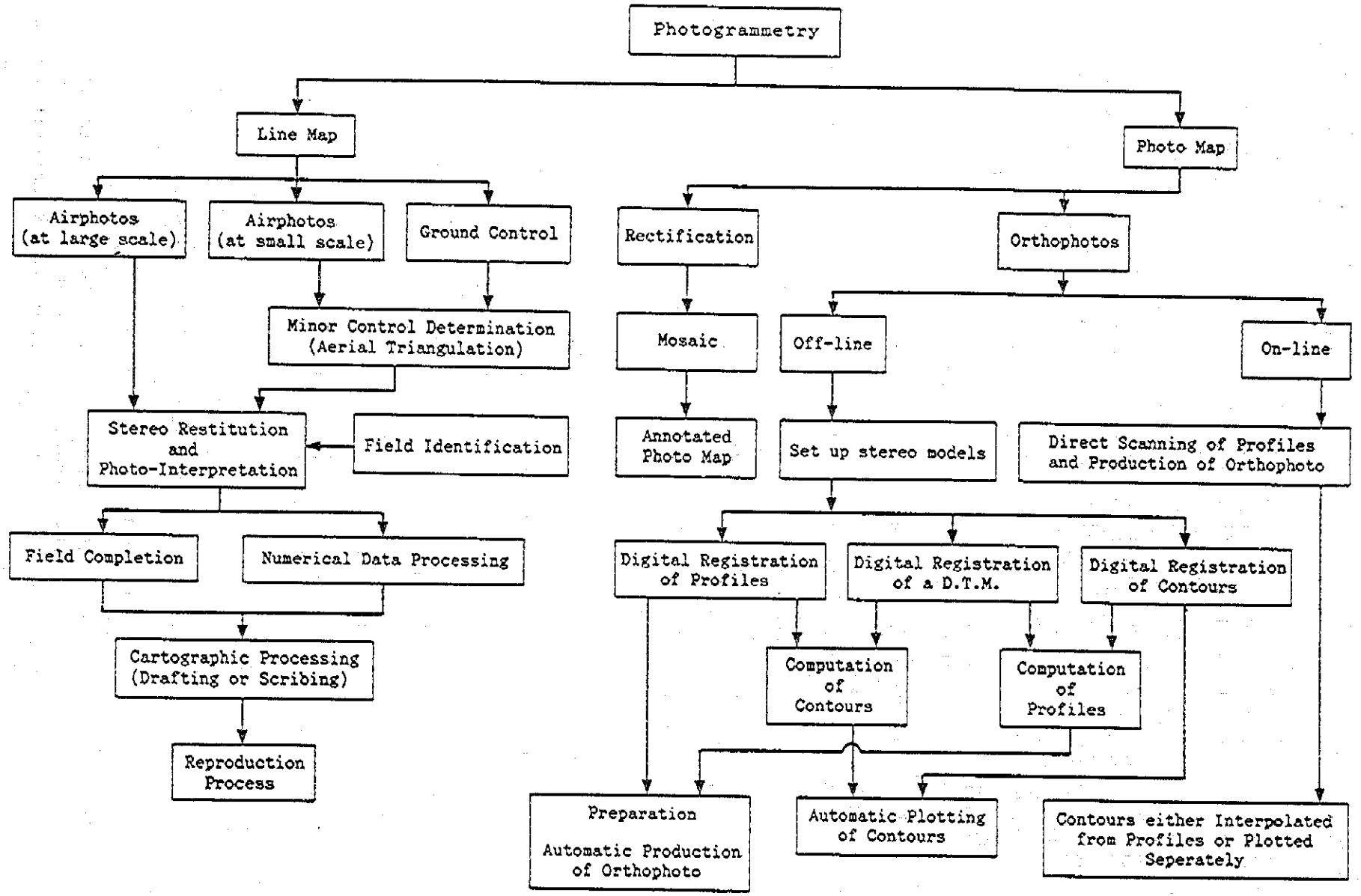


Fig. 4.7.1 MAPPING SYSTEM

(7) Compilation

All data from the field completion must be transferred to manuscripts or map sheets.

Separately plotted models must be compiled together on the map sheets.

(8) Drafting

All details must be scribed or inked for reproduction or final presentation.

(9) Final Map

Line map must be presented as final maps at different types and scales.

- Detailed Map 1/4,000
- Building Use Map 1/4,000
- Land Use Map 1/10,000
- Government Land Map 1/10,000
- Administration Boundary and Place Name Map 1/10,000

(10) Measurement of Land Use Data

Each type of land use data is collected in tabular form by means of planimeter or digital planimeter or digitizer and shown as a list and stored in a floppy disk or magnetic tape.

(11) Digital Mapping

a. Data Acquisition

Wild Tachymat TC1 (Electronic Recording Tacheometer) + GRE1 (Recording Unit) + GLE1 (Cassette Reader)

Wild DI4 Distomat + GRE2 (Recording Unit)

Digitizing Table
The optional digitizer is useful whenever information has to be taken from existing maps and plans. It can be adjusted in angle and height.

b. Data Processing

Tektronix 4054

This computer with CRT screen is the operator's work station from which he controls data input, processing and output. A feature is the interactive communication between the operator and the computer. The screen is unusually large, has excellent resolution and provides very fast interactive graphics.

Floppy-disk unit

The computer, the floppy-disk unit and the line printer from the basic equipment for data processing.
Programmes and data are stored on floppy disks. There are 2 disk drives with one controller.

Menu Tablet

The Tablet is an option. It facilitates work when the same commands have to be carried out repeatedly.

c. Data Output

Wild Aviotab TA and TA 2 tables

The Aviotab TA is a top-quality electronic plotting table with a high output rate but the TA 2 is a larger, faster table for plotting of the highest accuracy.

Line Printer (Model LA 120 DEC)

The line printer forms part of the basic equipment.

Hard-Copy Unit

The copier, which is an option, can often useful to be able to make a hard copy of the contents of the screen-graphics and alphanumerics.

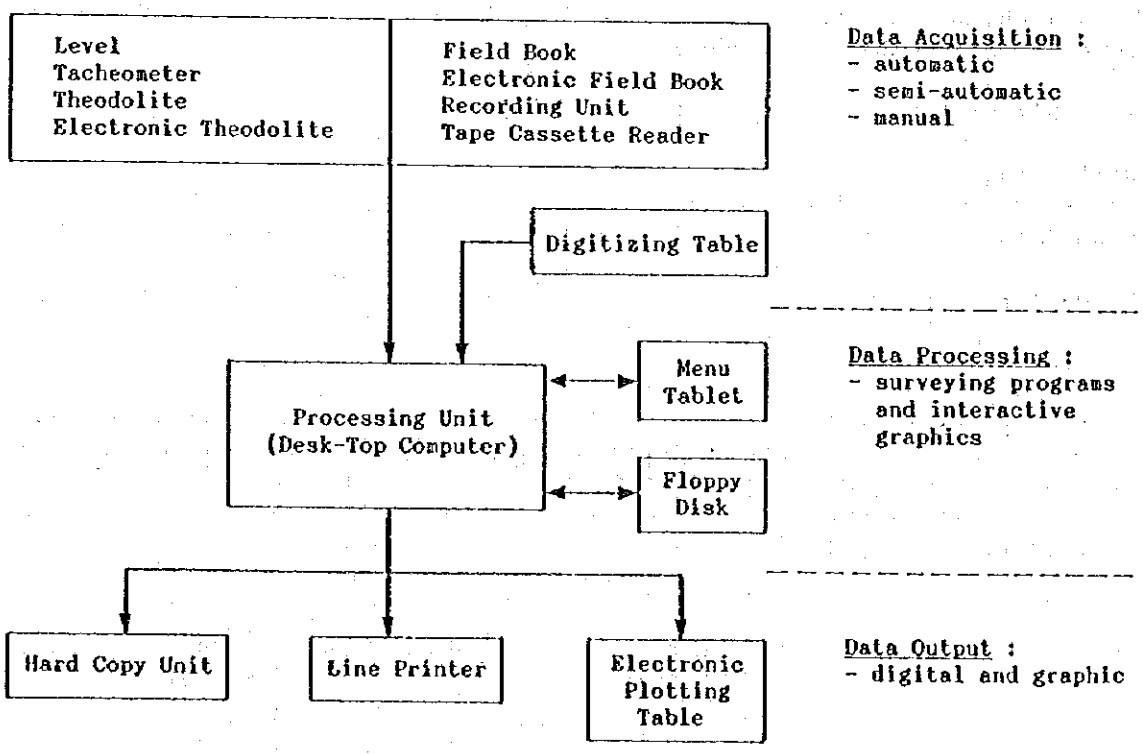
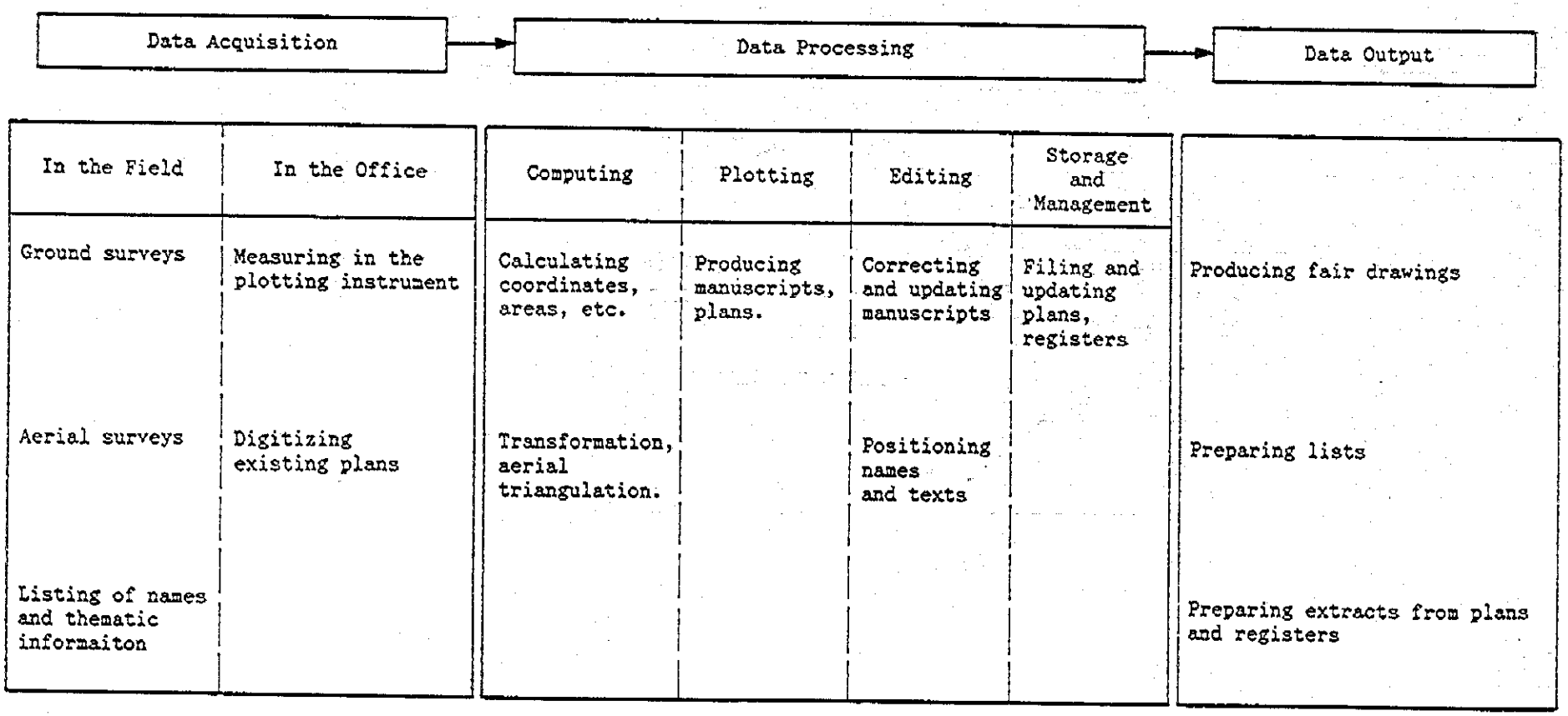


Fig. 4.7.2 Digital Mapping

Fig. 4.7.2.1 DIGITAL MAPPING



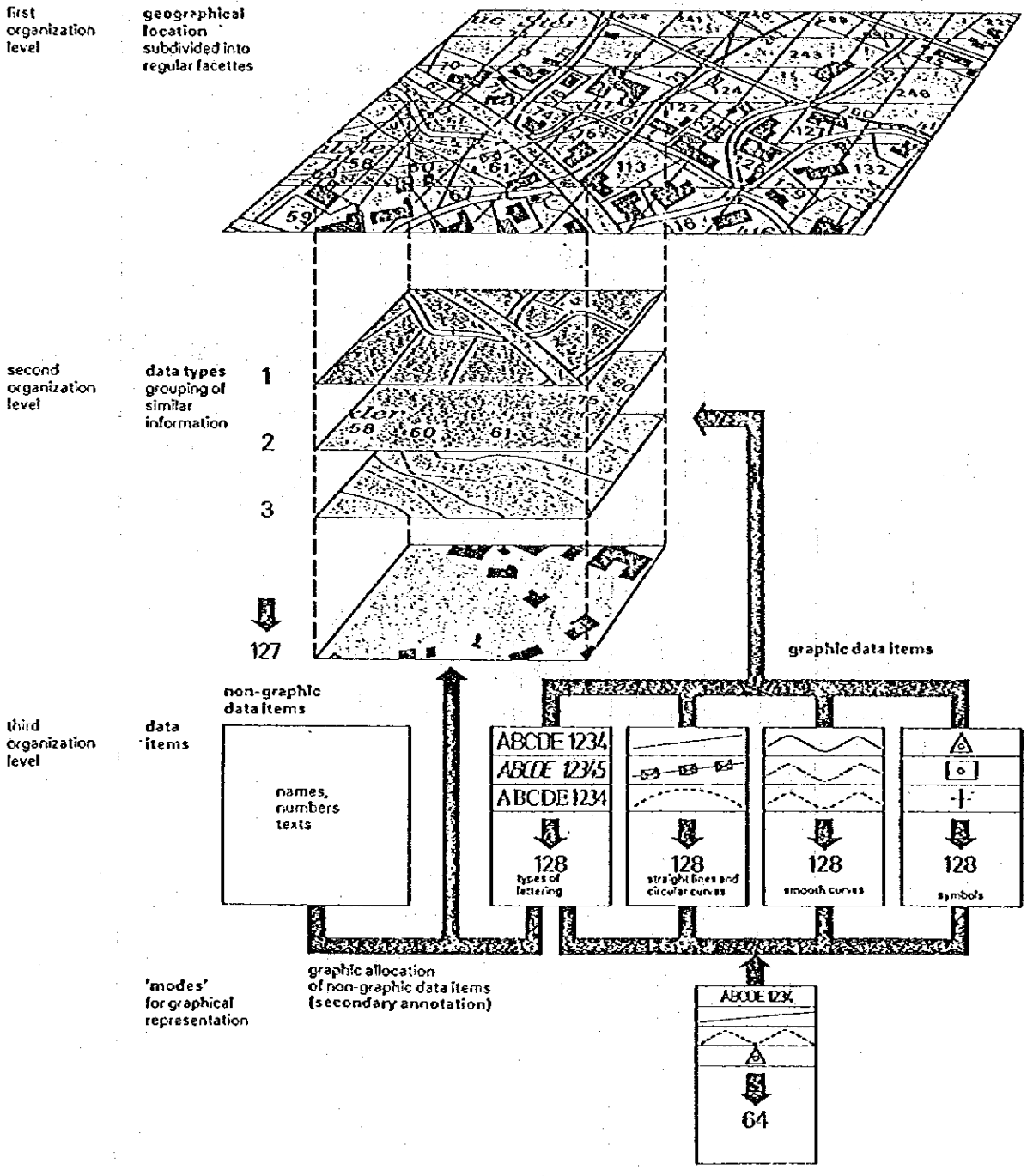


Fig. 4.7.3 MAPPING SYSTEM

4.7.2 Basic Tabulation Programs

(1) Field data

Site surveying is divided into 4 parts : population, commercial, industrial and facilities. Each part can be summarized as follows:

Population
sex, age, migration, religion, race, nationality, literacy, status of education, level of education, place of school, educational institution

Commercial
occupation, place of work, monthly salary, duration of business, type of business, persons engaged in the business, kind of goods and services.

Industrial
place of work, monthly salary, duration of industry, type of industry, persons engaged in industry, characteristics of building, major products.

Facility
hospitals, health centers, religious institutions, governmental institutions, parks and recreation areas, markets, hotels, bus stations, post office, police station, tourist attraction place, library services, service centers, telephone services, water supply, electricity services, waste disposal, slaugther house, fire protection.

(2) Data input

All the surveyed information including information from the central government agencies will be interpreted into coding.

(3) Raw data filing

The detail of each zonal information will be recorded on tape.

(4) Basic tabulation

Basic tabulation will be provided as follow:

Population

Table 1-1	showing number of population by age group and sex
Table 1-2	showing number of household and population by zone
Table 1-3	showing number of working population by occupation by sex and age
Table 1-4	showing number of working population by occupation by zone (living place)
Table 1-5	showing number of working population by occupation by zone (working place)
Table 1-6	showing number of working population by living and working place
Table 1-7	showing number of pupils and students by sex and age
Table 1-8	showing number of pupils and students by zone (living place)
Table 1-9	showing number of pupils and students by zone (place of school)
Table 1-10	showing number of population in each level (municipal, sanitary, district, province) by 20 years period.

Fig. 4.7.4 INPUT INFORMATION PROCEDURE OUTPUT INFORMATION

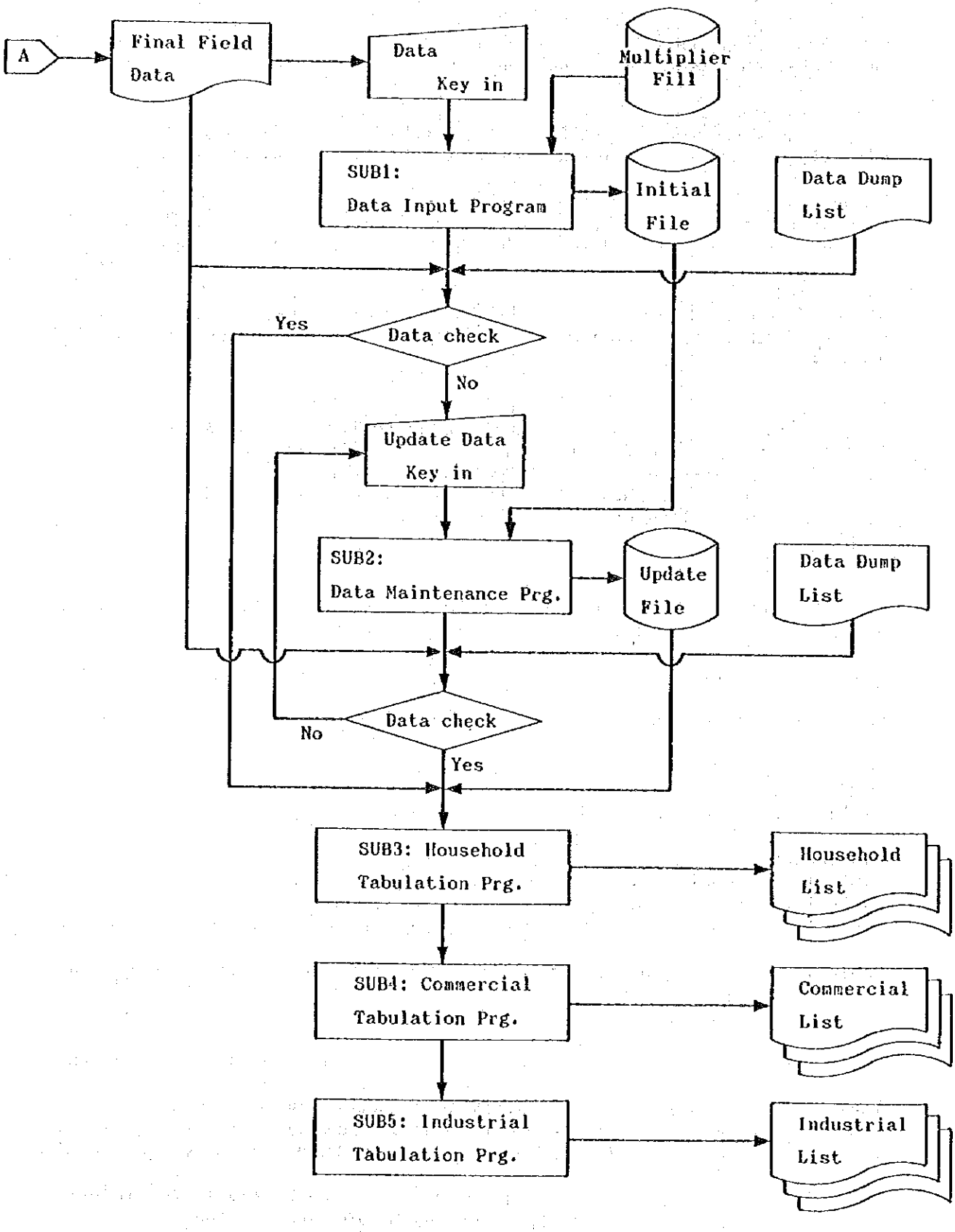
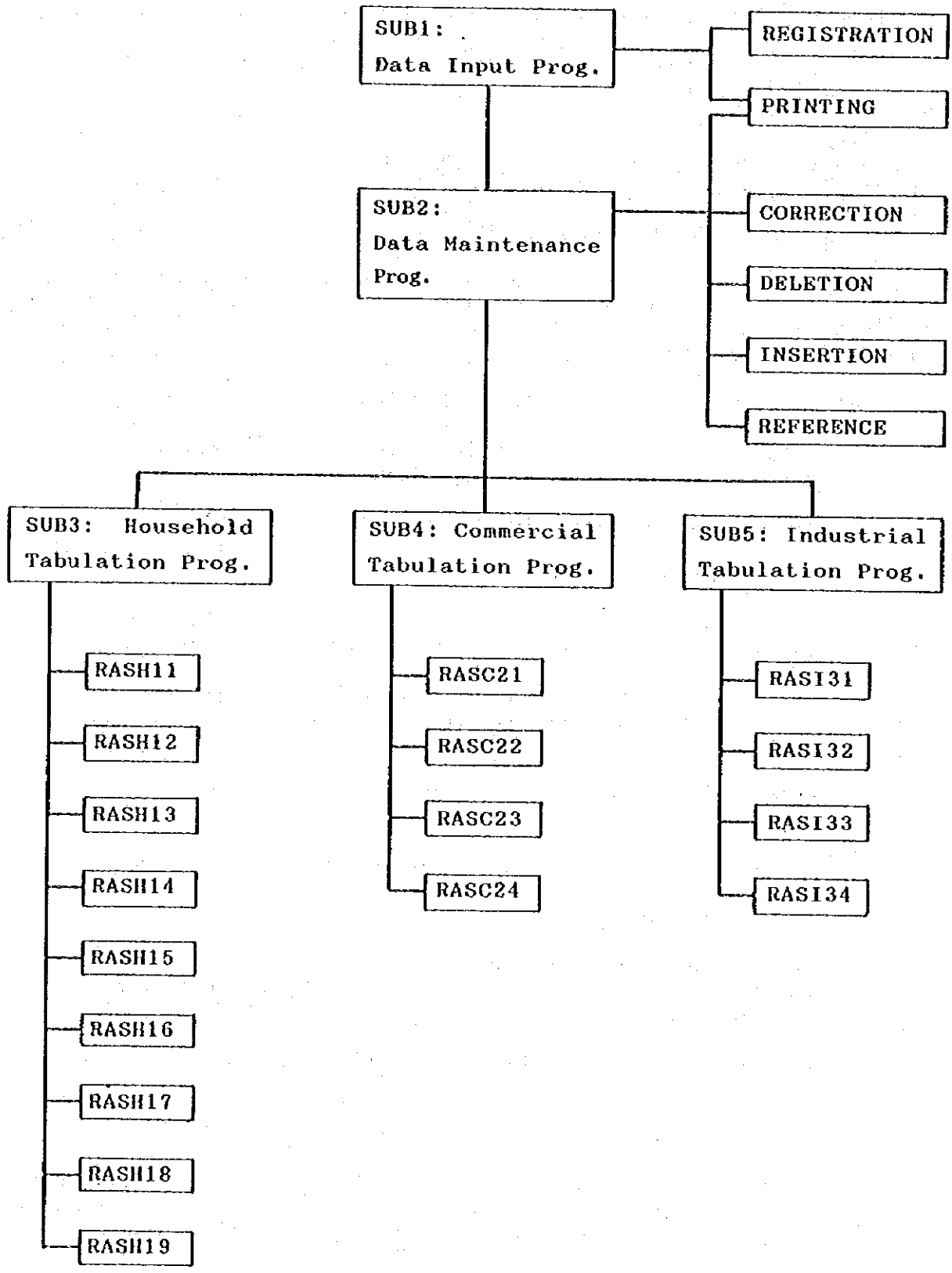


Fig. 4.7.5 STRUCTURE OF R & A DATA PROCESSING SYSTEM



Commercial and Industrial

Table 2-1 showing number of establishments by persons engaged by monthly salary paid by type of business
Table 2-2 showing number of employees by zone (living place)
Table 2-3 showing number of establishments by type of business and by zone
Table 2-4 showing number of persons engaged by type of employment and by zone

Facility

Table 3-1 showing number of schools by zone
Table 3-2 showing number of hospitals by zone

4.7.3 Analytical Applications[1]

The Data Base Management System should have statistical analysis programmes to be performed by the analyst.

The main topics of this chapter are:

a. Multiple regression analysis (REGRES command)

This analyzes the relationship between one dependent variable and two or more independent variables.

REGRES dependent-variable = independent-variable...

b. Factor analysis (FACTOR command)

This constructs a model in which a small number of factors are used to explain a larger set of variables.

c. Quantification of a linear model in which the dependent variable is qualitative and the independent variables are qualitative or qualitative.

QUANT1 dependent-variable = independent-variable....

category variable — number or category variables

d. Quantification theory type II analysis (QUANT2 command)

This allots an individual to one of two or more given populations, on the basis of the qualitative properties of the individual.

QUANT2 dependent-variable = independent-variable....

category variable — category variables

[1] This section is based on the JICA training material "Statistical Analysis: Training for Information Processing" OKINAWA International Center, Japan International Cooperation Agency,

e. Variance and covariance analysis (ANOVA)

This tests the hypothesis that the group means of the dependent variable are all equal.

This command also analyzes the main effects of factors and covariates, and the interaction among factors.

ANOVA dependent-variable = independent-variable....

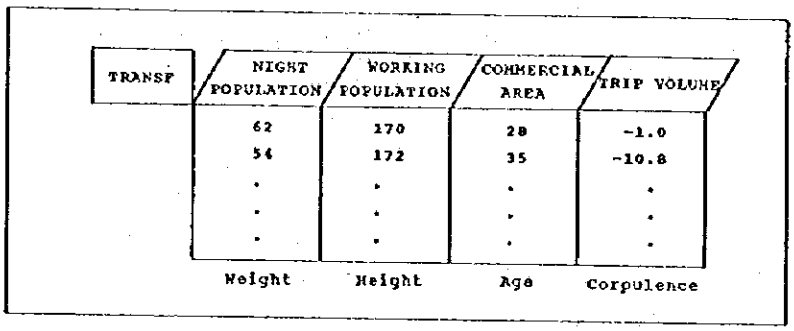
category variable — numeric or category variables

f. Cluster analysis (CLUSTER)

This aggregates cases into clusters, based on measure of affinity.

CLUSTER variable-to-be-analyzed...

(1) Multiple Regression Analysis (REGRES)




```
ANALYST_GET LEISURE
ANALYST_REGRES CORPLNCE=HEIGHT AGE
REGRES_
```

The example above retrieves the data of member TRANSF from the data bank and does a multiple regression analysis on the data.

We will examine the relationship of night population commercial area to trip volume.

The model which we will use is:

Trip Volume = (a * NIGHT POPULATION) + (a * Commercial Area) + c

In multiple regression analysis, a linear model showing the relationship between a dependent variable and several independent *variables (explanatory variables) is estimated.

The variables which are to be used in a multiple regression analysis are ordinarily specified in advance. Alternatively, the variables for the model may be selected from a range.of available variables.

If a variable selection method is used in the construction of a linear regression model, statistically relevant variables are selected from the available independent variables by a chosen selection method. There are four methods for selecting variables: forward inclusion, backward exclusion, forward stepping, and backward stepping.

Variables are specified in the following format:

```
REGRES dependent-variable = independent-variable....
```

An OPTION subcommand may be used to control the selection of analytical procedures and the range of summary statistics to be produced. If this subcommand is omitted, then standard procedures are used and the standard results are produced.

```
* * * * * * * * * * * * * * * * * INFORMATION OF COEFFICIENT * * * * * * * * * * * * * * * *
VARIABLE          STANDARD     COEFFICIENT   STD. ERROR OF   T-VALUE   PROB.     PARTIAL
                  COEFFICIENT                COEFFICIENT               (2-TAIL)  CORRE.
COMMERCIAL AREA     -0.23429     -0.26420       0.09464     -2.7917   0.0061    -0.2441
NIGHT POPULATION     0.27259      0.51209       0.15767      3.2479   0.0015     0.2811
CONSTANT                         29.04770      16.91629      1.7171

* * * * * * * * * * * * * * * * * INFORMATION OF GOODNESS OF FIT * * * * * * * * * * * * * * * *
RESIDUAL SUM OF SQUARES                      4139.68359
STANDARD ERROR OF RESIDUAL                      5.80138
AKAIKE'S INFORMATION CRITERION                444.00366

MULTIPLE R                  0.37526   :   ADJUSTED R                0.35616
R SQUARE                    0.14082   :   ADJUSTED R SQUARE         0.12685
* * *

* * * * * * * * * * * * * * * * * * * * ANALYSIS OF VARIANCE * * * * * * * * * * * * * * * * * * * *
                  SUM OF SQUARES     DF     MEAN SQUARE     F-VALUE     P(TAIL)
REGRESSION            678.50586       2       339.25293    10.08002     0.00009
RESIDUAL             4139.68359     123        33.65596
TOTAL                4818.19141     125        38.54552
* * *
```

Some of the results of a multiple regression analysis are shown above.

The figures given in the "COEFFICIENT" column of the table are the required estimates of a , a , and c.

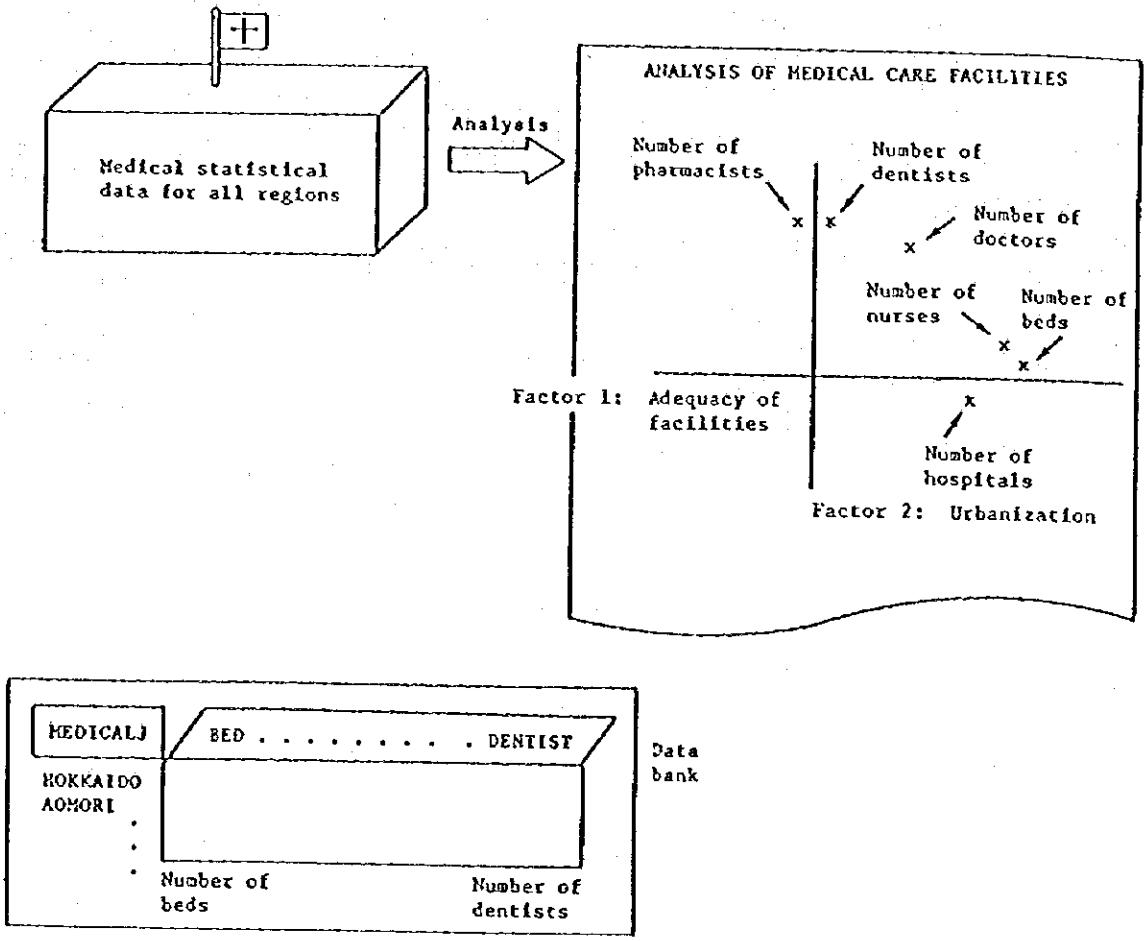
Thus, the estimated model equation is:

```
Trip Volume = (-0.26420 * Night Population)
            + (0.51209 * Commercial Area) + 29.04770
```

The equation indicates that corpulence will increase with age and decrease with height.

The value of the multiple correlation coefficient is low, and so the adequacy of the model is doubtful.

(2) Factor Analysis (FACTOR)



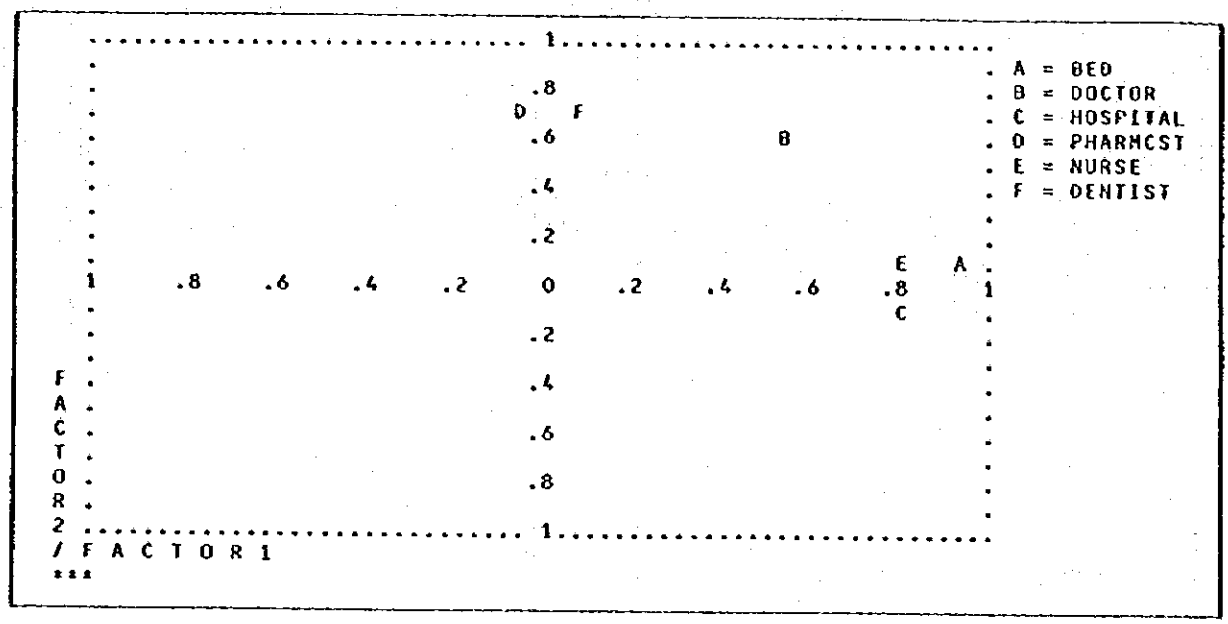

```
ANALYST_GET MEDICAL
ANALYST_FACTOR BED:DENTIST
FACTOR_OPTION METHOD (PFA) PLOT
FACTOR_
```

We will now perform a factor analysis using the data of member MEDICAL in the data bank.

The data are retrieved from the data bank, and then two factors are extracted from six medical variables. The FACTOR command is used to perform factor analysis. It is a method of constructing a model in which a small number of factors are used to explain a large set of variables. The FACTOR command may use either the principal factor method or the maximum likelihood method. Factor rotation may be carried out by either orthogonal or oblique rotation methods. The FACTOR command can also be used to compute factor scores.

The variables which are to be analyzed are specified after the command name. These variables must all be numeric variables. If no variables are specified, then the analysis will be carried out on all available numeric (quantitative) variables.

An OPTION subcommand may be used to control the selection of analytical procedures and the range of summary statistics to be produced.



Some of the results of a factor analysis are shown above.

The diagram is a plot of variables in rotated factor space.

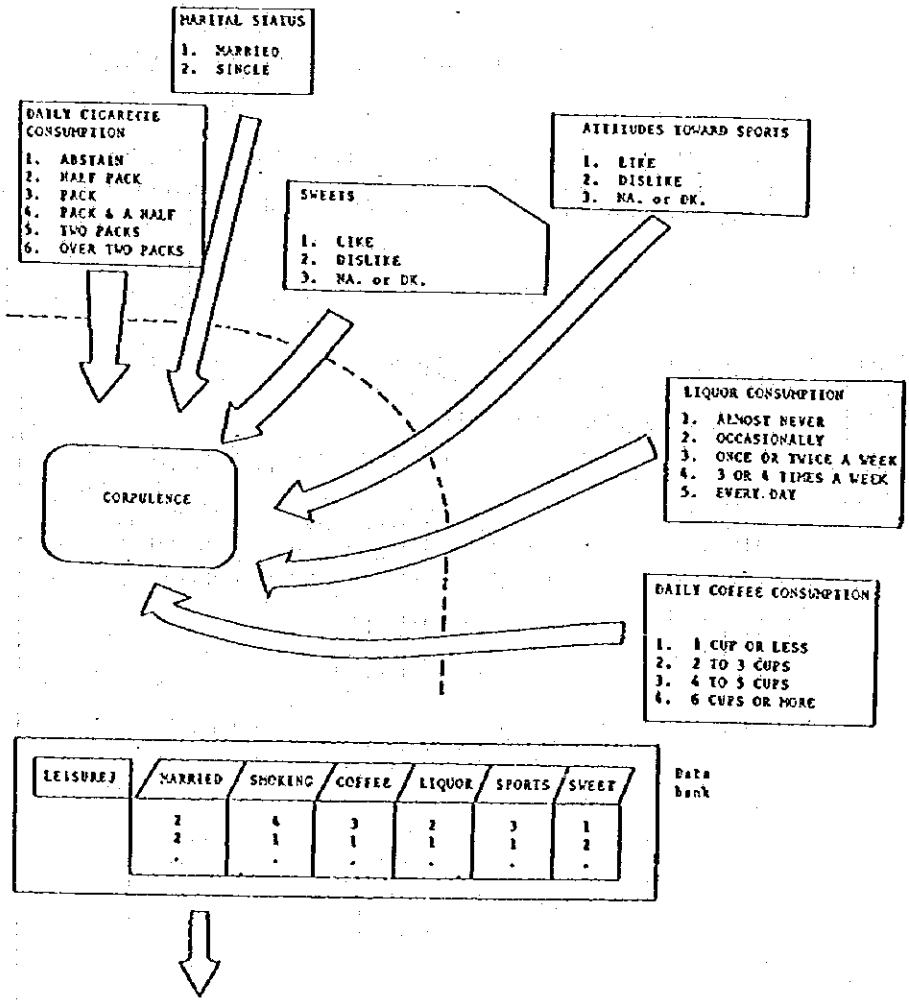
The number of beds, doctors, hospitals, pharmacists, nurses, and dentists are represented by the letters A to F respectively.

The numbers of beds, hospitals, and nurses have high positive values along the factor 1 axis. This indicates the adequacy of the facilities.

The numbers of pharmacists and dentists have high positive values on the factor 2 axis. This indicates the degree of urbanization.

Factor analysis is a method of constructing a small number of factors from a larger set of variables. The user must then interpret the meaning significance of the factors.

(3) Quantification Theory Type I Analysis (QUANT 1)




```
ANALYST_GET LEISURE
ANALYST_QUANT   CORPLNCE=MARRIED   SMOKING COFFEE LIQUOR SPORTS
SWEET
QUANT1_
```

We will now retrieve the data of member LEISURE from the data bank and analyze it using quantification theory type I analysis.

Here we want to construct a linear model which relates to marital status, daily cigarette consumption, daily coffee consumption, liquor consumption, attitudes toward sports, and attitude to sweet food.

The QUANT1 command is used to carry out the analysis. Quantification theory type I is concerned with constructing a linear model in which the dependent variable is a qualitative category) variable and the independent variables are quantitative (numeric) or qualitative (category) variables.

Variables are specified in the following format:

```
QUANT1 dependent-variable  =  independent-variable....
       /                      |
category variable             numeric or category variables
```

An OPTION subcommand may be used to control the selection of summary statistics to be produced. If this subcommand is omitted, then the standard statistics are produced.

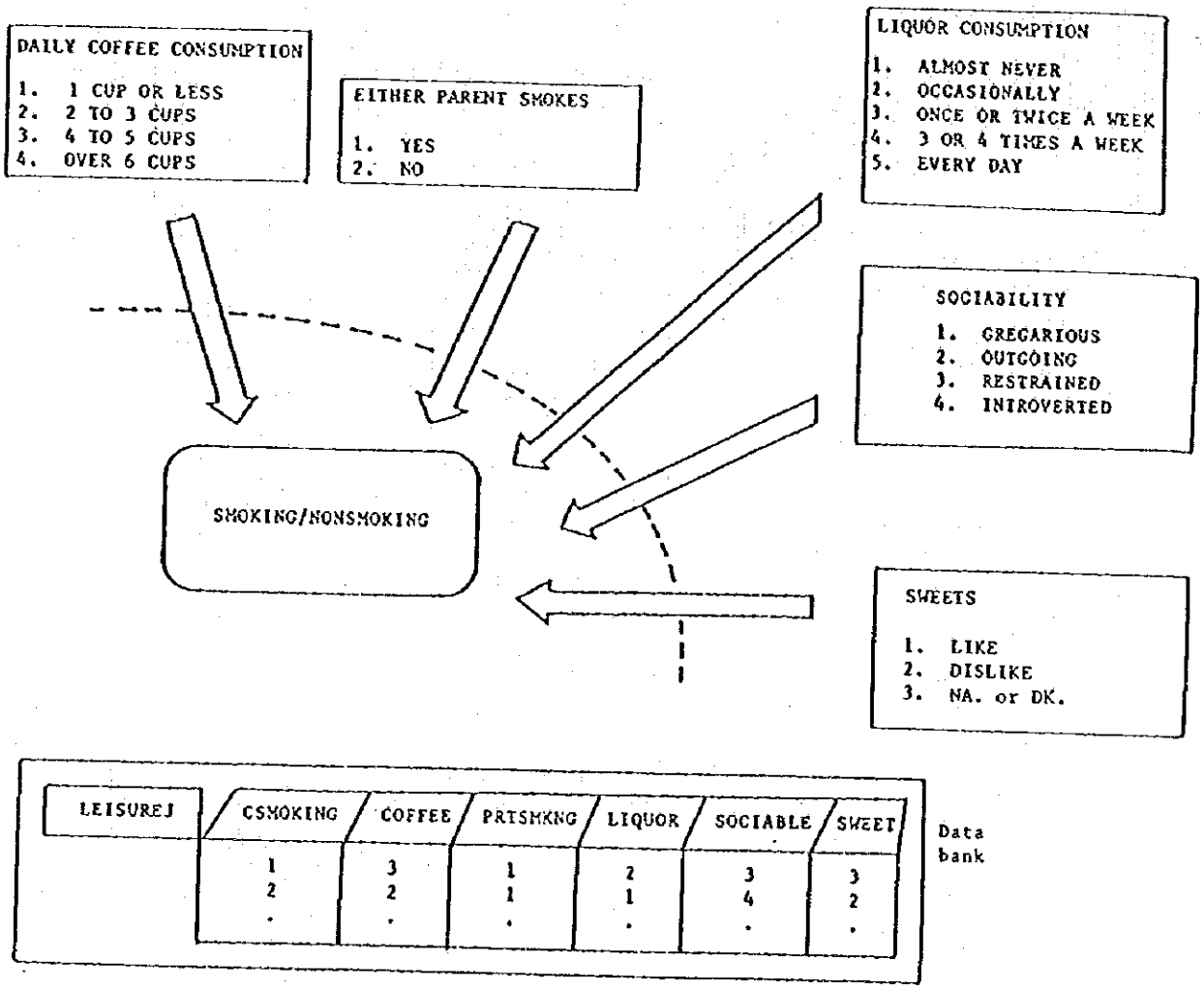
```
CRITERION = CORPLNCE * * * * * NORMALIZED CATEGORY WEIGHT * * * * * STEP = 1
            DEGREE OF OVERWEIGHT

CONSTANT ..........        -2.5444361

VARIABLE  CODE                                  COUNT      WEIGHT      RANGE
                                                                  GRAPH OF WEIGHT

MARRIED   MARRIED OR SINGLE                                        2.4087970
            1  MARRIED                           40    1.8440995              |*****
            2  SINGLE                            86   -0.7846975           **|

SMOKING   NUMBER OF CIGARETTES SMOKED IN ONE DAY                   3.7374530
            1  ABSTAIN                           53    0.0190127              |
            2  HALF PACK                          8    2.1694073              |******
            3  PACK                              24   -1.3326227         ****|
            4  PACK & A HALF                     24   -0.6013309           **|
            5  TWO PACKS                         12    2.4046303              |*******
            6  OVER TWO PACKS                     5   -0.1612012              |

COFFEE    QUANTITY OF COFFEE CONSUMED IN ONE DAY                   8.7516247
            1  UNDER 1 CUP                       37   -0.1713403              |
            2  2 TO 3 CUPS                       58   -1.1437619          ***|
            3  4 TO 5 CUPS                       29    2.9090656              |*******
            4  OVER 6 CUPS                       2   -5.8425592 ***************|

LIQUOR    DRINKING FREQUENCY                                       2.1948177
            1  ALMOST NEVER                      13   -1.4864724         ****|
            2  OCCASIONALLY                      36    0.4662993              |*
            3  ONCE OR TWICE A WEEK              33    0.3031009              |*
            4  3 OR 4 TIMES A WEEK               15    0.7083453              |**
            5  EVERYDAY                          31   -0.5534692           **|

SPORTS    ATTITUDES TOWARD SPORTS                                  0.8603361
            1  LIKE                             102   -0.0325378              |
            2  DISLIKE                            7   -0.4711190            *|
            3  NA.DK.                            17    0.3892171              |*

SWEET     SWEET TOOTH                                              1.9008607
            1  LIKE                              64    0.7826568              |**
            2  DISLIKE                           25   -1.1182059          ***|
            3  NA.DK.                            37   -0.5982367           **|
```

The results of a quantification theory type I analysis are shown above.

This report shows normalized category weights, that is, the contribution of each category of each independent variable to the dependent variable. Contributions may be positive or negative. Bar charts of normalized category weights are shown at the right hand side of the table.

It can be concluded from this report that married people tend to be more corpulent than single people, and that people who like sweets tend to be more corpulent that people who do not.

(4) Quantification Theory Type II Analysis (QUANT 2)




```
ANALYST_QUANT2 CSMOKING=COFFEE KPRTSMKNG LIQUOR SOCIABLE SWEET
QUANT2_
```

We will now do a quantification theory type II analysis using the current data in the ANALYST work area.

The analysis will be used to analyze smoking behavior (CSMOKING) in terms of parental smoking habit (PRTSMKNG), daily coffee consumption, liquor consumption, sociability, and attitude to sweets.

The QUANT2 command is used to carry out an analysis according to quantification theory type II. This theory us a generalized form of discriminant analysis. It is concerned with allotting an individual to one of two or more give populations, on the basis of the qualitative properties of the individual. These qualitative properties are the independent variables in the analysis.

Variables are specified in the following format:

```
QUANT2 dependent-variable=independent-variable ...
```

category variable — category variables

An OPTION subcommand may be used to control the selection of summary statistics to be produced. If this subcommand is omitted, then the standard statistics are produced.

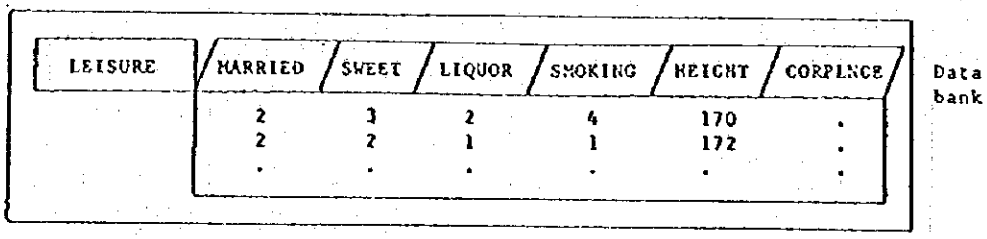
```
                DO YOU SMOKE CIGARETTES ?
            --- AXIS NO. =  1 ---

VARIABLE  CODE                          COUNT      WEIGHT        RANGE
                                                                GRAPH OF WEIGHT

COFFEE    QUANTITY OF COFFEE CONSUMED IN ONE DAY                2.1043015
             1  UNDER 1 CUP                 37   -0.9870638 **********I
             2  2 TO 3 CUPS                 58    0.0751708           I*
             3  4 TO 5 CUPS                 29    1.1172378           I**********
             4  OVER 6 CUPS                  2   -0.1192214          *I

PRTSMKNG  EITHER OF PARENTS SMOKE OR NOT                         0.4872942
             1  YES                         87    0.1508292           I**
             2  NO                          39   -0.3364651        ****I

LIQUOR    DRINKING FREQUENCY                                     1.5792309
             1  ALMOST NEVER                13   -0.9257937 **********I
             2  OCCASIONALLY                36   -0.1006042          *I
             3  ONCE OR TWICE A WEEK        33   -0.0149052           *
             4  3 OR 4 TIMES A WEEK         15    0.6534372           I*******
             5  EVERYDAY                    31    0.1982637           I**

SOCIABLE  SOCIABILITY                                            1.1414888
             1  GREGARIOUS                  35    0.5643065           I******
             2  OUTGOING                    28   -0.5771863     *******I
             3  RESTRAINED                  17    0.1236133           I*
             4  INTROVERTED                 46   -0.1237158          *I

SWEET     SWEET TOOTH                                            0.3453734
             1  LIKE                        64    0.1403622           I**
             2  DISLIKE                     25   -0.0559104          *I
             3  NA.DK.                      37   -0.2050113         **I
```

The results of a quantification theory type KK analysis are shown above.

This report shows category weights, that is, the contribution of each category of each independent variable to the dependent variable. Contributions may be positive or negative. Bar charts of category weights are shown at the right hand side of the report.

It is possible to conclude from this report that people who drink a lot of coffee or frequently consume liquor are more likely to be smokers, and that people who drink little coffee and rarely consume liquor are more likely to be nonsmokers.

(5) Variance and Covariance Analysis (ANOVA)




```
ANALYST_ANOVA CORPLNCE=MARRIED SWEET LIQUOR SMOKING HEIGHT
ANOVA_OPTION TABLE METHOD (CLASSIC COVARIATE)
ANOVA_
```

We will now carry out an analysis of variance and covariance on the current data in the ANALYST work area.

Corpulence will be analyzed in terms of marital status, attitude to sweet, liquor consumption, daily cigarette consumption, and height.

The ANOVA command is used for the analysis of variance and covariance. Analysis of variance tests the hypothesis that the group means of the dependent variable are equal.

The groups are defined by one or more category variables, which are referred to as "factors". Numeric dependent variables are called "covariates".

The ANOVA command may be used to study the effects of covariates, the main effects of factors, and interaction effects. Appropriate F tests are also carried out.

Where the dependent variables are all factors, the procedure is referred to as analysis of variance. Whereas, if even one of the dependent variables is a covariate, then the procedure is referred to as analysis of covariance.
Variables are specified in the following format:

```
ANOVA dependent-variable=independent-variable ...
```

numeric variable — category or numeric variables

An OPTION subcommand may be used to control the selection of summary statistics to be produced. If this subcommand is omitted, then the standard statistics are produced.

```
* * * * * * * * ANALYSIS OF VARIANCE (ANOVA) TABLE * * * * * * * * (SUMMARY)
                        CORPLNCE  DEGREE OF OVERWEIGHT

                                  SUM OF      D.F.    MEAN       F-VALUE P(TAIL)
                                  SQUARES             SQUARE

WHOLE EXPLAINED VARIATION         3691.9072    73     50.5741    2.3350 0.0008
RESIDUAL VARIATION                1126.2848    52     21.6593

TOTAL VARIATION                   4818.1920   125     38.5455

* * * * * * * * ANALYSIS OF VARIANCE (ANOVA) TABLE * * * * * * * * (DETAIL)
                        CORPLNCE  DEGREE OF OVERWEIGHT

                                  SUM OF      D.F.    MEAN       F-VALUE P(TAIL)
                                  SQUARES             SQUARE

COVARIATES                         323.4650     1     323.4650   14.9362 0.0003
 HEIGHT                            323.4650     1     323.4650   14.9362 0.0003

MAIN EFFECTS                       418.3381    12      34.8615    1.6095 0.1178
 MARRIED                           161.4962     1     161.4962    7.4562 0.0086
 SWEET                              90.1412     2      45.0708    2.0809 0.1351
 LIQUOR                             34.9105     4       8.7277    0.4030 0.8056
 SMOKING                           113.0192     5      22.6038    1.0436 0.4022

TWO WAY INTERACTION EFFECTS       2286.6743    44      51.9244    2.3973 0.0013
 MARRIED       :SWEET              109.6804     2      54.8402    2.5319 0.0893
 MARRIED       :LIQUOR             356.2730     4      89.0682    4.1122 0.0057
 MARRIED       :SMOKING            416.2758     5      83.2552    3.8438 0.0069
 SWEET         :LIQUOR             522.0300     8      65.2537    3.0127 0.0074
 SWEET         :SMOKING            348.0090     9      38.6677    1.7853 0.0935
 LIQUOR        :SMOKING            910.2665    16      56.8917    2.6267 0.0045

THREE WAY INTERACTION EFFECTS      665.4317    16      41.5895    1.9202 0.0395
 MARRIED :SWEET     :LIQUOR        288.3796     4      72.0949    3.3286 0.0188
 MARRIED :SWEET     :SMOKING       128.9224     3      42.9741    1.9841 0.1278
 MARRIED :LIQUOR    :SMOKING       286.8233     2     143.4116    6.6212 0.0027
 SWEET   :LIQUOR    :SMOKING       213.2218     7      30.4603    1.4063 0.2228

FOUR WAY INTERACTION EFFECTS      UNCOMPUTED
 MARRIED:SWEET  :LIQUOR:SMOKIN  . UNCOMPUTED
```

Some of the results of a covariance analysis are shown above.

This part of the output is referred to as an "analysis of variance table" and shows the results of F tests.

An examination of the right most column in the table above shows that the following are all significant at the 1% level: the covariant "height", the main effect of the factor "marital status", the secondary interactions "marital status - liquor consumption", and "marital status - daily cigarette consumption", and the tertiary interaction "marital status - liquor consumption - daily cigarette consumption". That is, the probability value for these is less than 0.01. The main effects of factors other than marital status are not significant even at the 5% level.

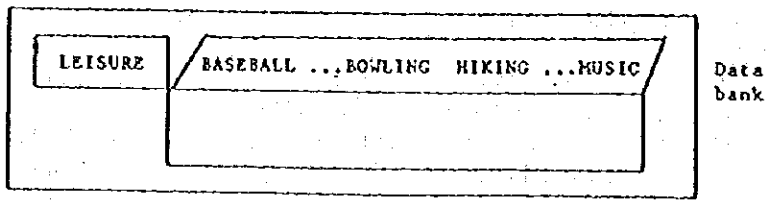
[Reference] - Significance levels

The significance level of a statistical test is the probability of rejecting a true null hypothesis.

Examples of the commonly used significance levels are 0.1%, 1%, 5% and 10%.

In this example, we have been computing the relationship between corpulence and a number of dependent variables, and establishing the probability that the assessment are incorrect.

(6) Cluster Analysis (CLUSTER)




```
ANALYST_TRANS COUNT=0

ANALYST_REPEAT A=BASEBALL:BOWLING
                 HIKING:MUSIC

ANALYST_TRANS IF(A=1) COUNT=COUNT+1


ANALYST_TERMINATE

ANALYST_QUANT3 BASEBALL:BOWLING
               HIKING:MUSIC

QUANT3_OPTION DUMMY AXES(4) BANKOUT

QUANT3_CONDITION COUNT>3

QUANT3_

ANALYST_GET QUANT3

ANALYST_CLUSTER Q3AXIS1 Q3AXIS2
                Q3AXIS3 Q3AXIS4

CLUSTER_OPTION ALL NOSTANDARD EUCLID NEAREST
               STATISTICS(3)

CLUSTER_
```

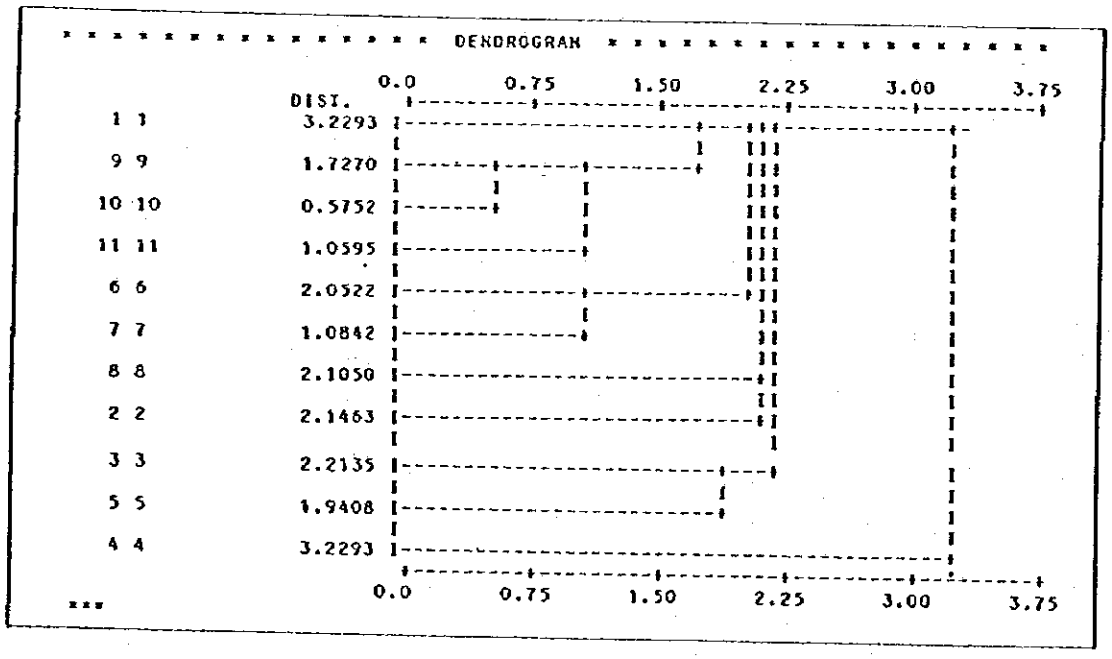
We will now do a cluster analysis on the current data in the ANALYST work area.

The QUANT3 command will be used to carry out a classification of hobbies, and the results will be subjected to cluster analysis.

The CLUSTER command is used for cluster analysis. Cluster analysis is a method of constructing clusters of cases on the basis of measure of proximity. Case classification and matrix classification are also possible with the CLUSTER command. The CLUSTER command may be used with both cases and matrix materials, and measures of affinity may be computed. Seven methods of cluster linkage are available: single linkage or nearest neighbor, complete linkage or furthest neighbor, average linkage between groups, and Ward's method. The results may be displayed on a dendrogram or icicle plot.

The names of the variables to be analyzed are specified after the command name. All of these variables must be numeric variables. If no variables are specified, then the analysis will be carried out on all available numeric (quantitative) variables.

An OPTION subcommand may be used to control the selection of analytical procedures and the summary statistics to be produced. If this subcommand is omitted, then the single linkage method is used and Euclidean distances are computed.



Some of the results of a cluster analysis are shown above.

This diagram is called a dendrogram (tree diagram). It shows affinities among hobbies selected by respondents and is based on analysis of preferences. Eleven choices were given: baseball, tennis, swimming, golf, skiing, skating, bowling, hiking, movies, reading, and listening to music.

The numbers 1 to 11 at the left of the diagram correspond to the eleven hobbies: baseball, tennis, swimming, golf, skiing, skating, bowling, hiking, movies, reading, and listening to music.

Distance values can be obtained by extending the vertical lines in the diagram (joining two options) to the top or bottom axis. The smaller this distance value is, the closer are the distribution patterns of these two options. Note that options with similar patterns have been positioned near each other.

For example, the diagram shows that the respondents' selection of options 9, 10, and 11 (movies, reading, and listening to music) are closely related, as are the choices of options 6 and 7 (skating and bowling).

4.7.4 Socio-Economic Simulator

(1) Standard Unit

Standard unit in each part will be calculated after basic tabulation made which is shown as following:

Population
sex ratio, age structure, dependency ratio, school attendance ratio.

Commercial and Industrial
- Employment structure,
- Employment ratio,
- Employment density,
- Commercial and industrial establishment structure.

Facility
- Pupils / teacher ratio
- Pupils /classroom ratio
- Population / patient bed ratio
- Population / doctor ratio

Examples of Standard Units

- Population

Sex ratio : The number of males per 100 females

Labor force participation rate : The ratio of economically active population of ages 11 years and above to the population of the same ages (%)

Unemployment rate : The ratio of unemployed population to the economically active population (%)

Dependency ratio	: The ratio of population in the age groups 0-14 and 65 and over to the working age population of 15-64

- Population x Land

Population density (Gross)	: The ratio of zonal population to the total area of the zone (persons per hectare)
Population density (Net)	: The ratio of zonal population to the total site area of the zone (persons per hectare)
Commercial land standard unit (by type of business)	: The ratio of commercial site to the persons engaged in commercial activity (sq.meters per person)
Industrial land standard unit (by type of industry)	: The ratio of industrial site to the persons engaged in industrial acitivity (sq.meters per person)
Recreation land standard unit	: The ratio of recreation land to population (sq.meters per person)

- Population x Income

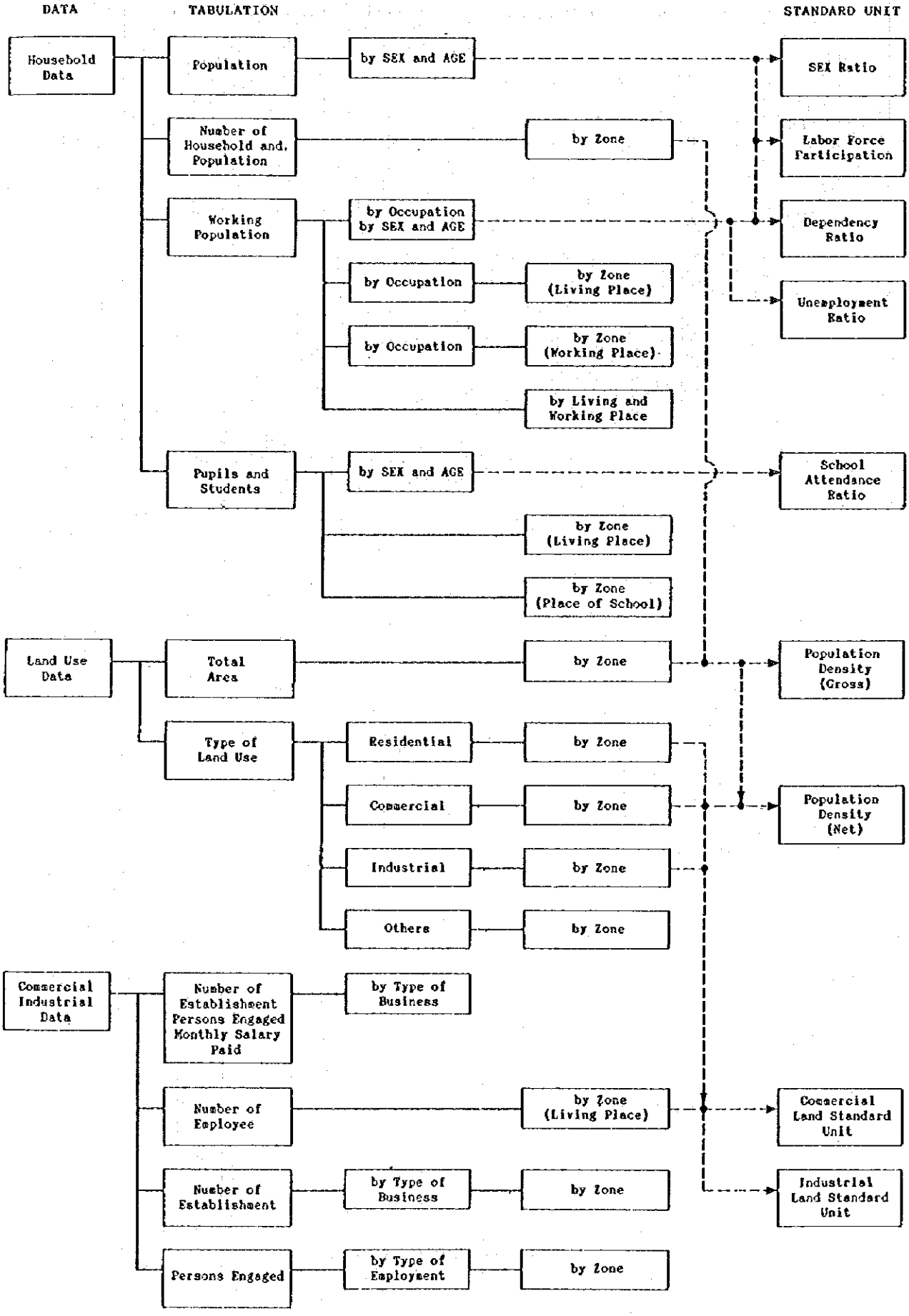
Average household income	: The ratio of household income to the number of household (baht per household)
Average income by occupation	: The ratio of income to the working population by occupation (baht per person)
Average paid salary for commercial employee	: The ratio of paid salary to the commercial employee (baht per person)
Average paid salary for industrial employee	: The ratio paid salary to the industrial employee (baht per person)

(2) Statistical Analysis and Modeling

All the information will be analyzed by use of the following statistical models:

- Multiple regression analysis
- Factor analysis
- Quantification theory Type I analysis
- Quantification theory Type II analysis
- Variance and covariance analysis
- Cluster analysis
- Correlation analysis

Fig. 4.7.6 STANDARD UNIT



(3) Forecasting and Planning Output

Forecasting for the following items of information will be carried out:

Population
population forecast by total and by age structure

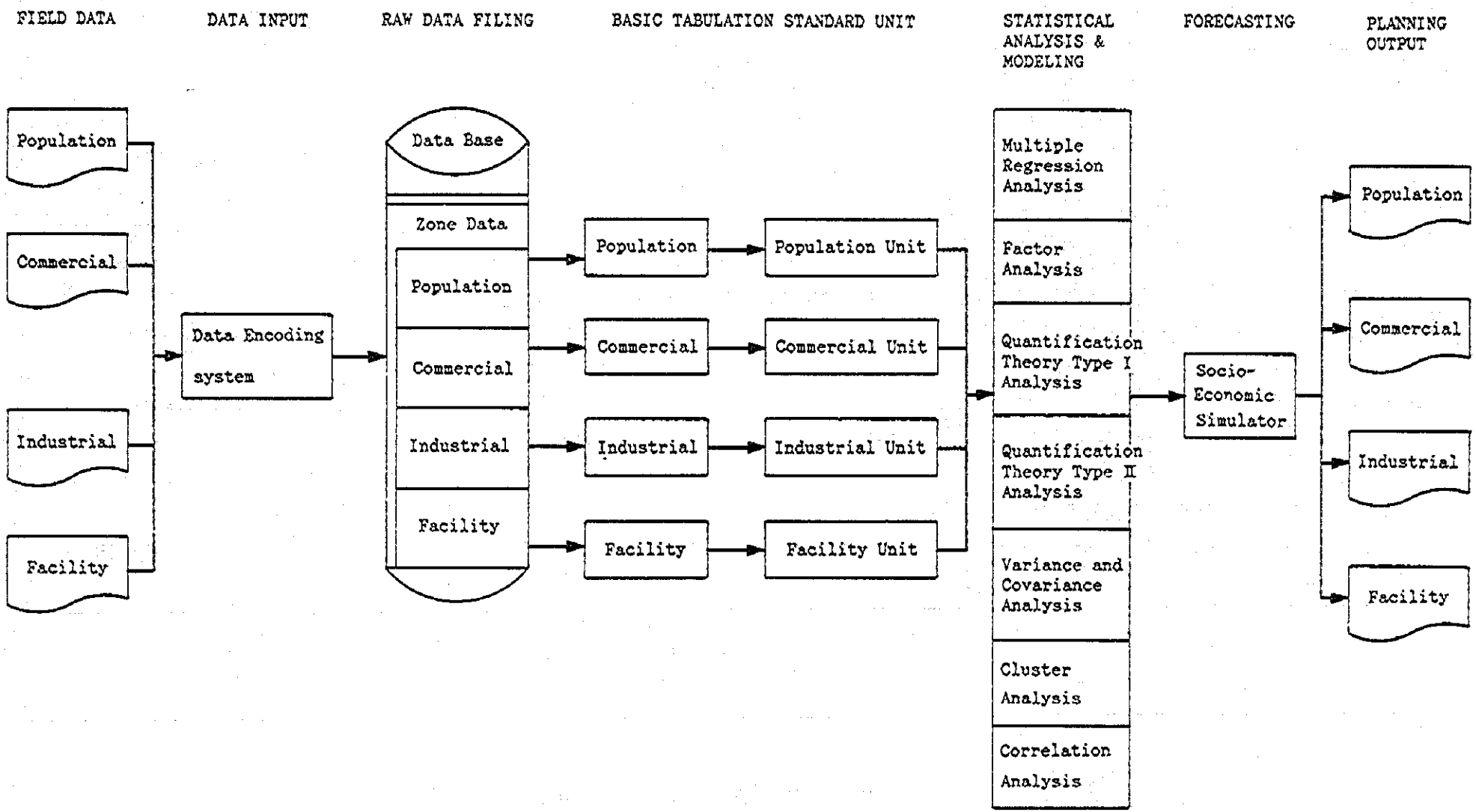
Commercial and industrial
employment forecast by total and by sector

Facility
student forecast

Table 4.7.1 SCOPE OF ANALYSIS

Population	Economic	Urban Facility
- Population growth trend. - Population distribution/density. - Age distribution. - Sex distribution. - Future projection by using the following analytical techniques: Arithmetic growth model. Geometric growth model. Ratio method. Cohort survival method.	- Age and sex distribution of employed population - Sectorial employment distribution. - Commercial and industrial activities. - Person engaged in commercial and industrial establishments. - Employment projection by using Ratio method. - School enrollment forecasting by using ratio method.	- Existing condition of urban services and facilities. - Comparing existing conditions. to accepted standard in order to determine the adequacy ex; class room : student teacher : student doctor : population patient bed: population - Pupils and students distribution and their characteristics. - School enrollment forecasting by using ratio method.

Fig. 4.7.7 STRUCTURE OF SOCIO-ECONOMIC SYSTEM



4.7.5 Land Use Simulator

(1) Input Information

Mapping

Maps prepared by Mapping Division will be submitted to the Land Use Planning team in order to define the planning boundary and study the existing conditions such as physical geography, road network, urbanization area, land use, etc.

Socio-Economic Analysis

Draft Socio-Economic analysis will be submitted to Land Use Planning team by Research and Analysis Division. This analysis will guide any further analysis by the planner if he feels this is necessary.

(2) Procedure

Boundary Plan Setting

Planners will define the boundary of the planning area. Existing maps are generally analyzed thoroughly, including site surveying by planners themselves.

Goal and Objective Setting

Setting up the goals and objectives by studying the characteristic of each area will help to guide the plan formulation process.

Socio-Economic Analysis

This will give the image of the planning area in terms of Socio-Economics.

Land Use Analysis

All the land use information will be collected and analyzed by using any appropriately technique to forecast the future land use trend.

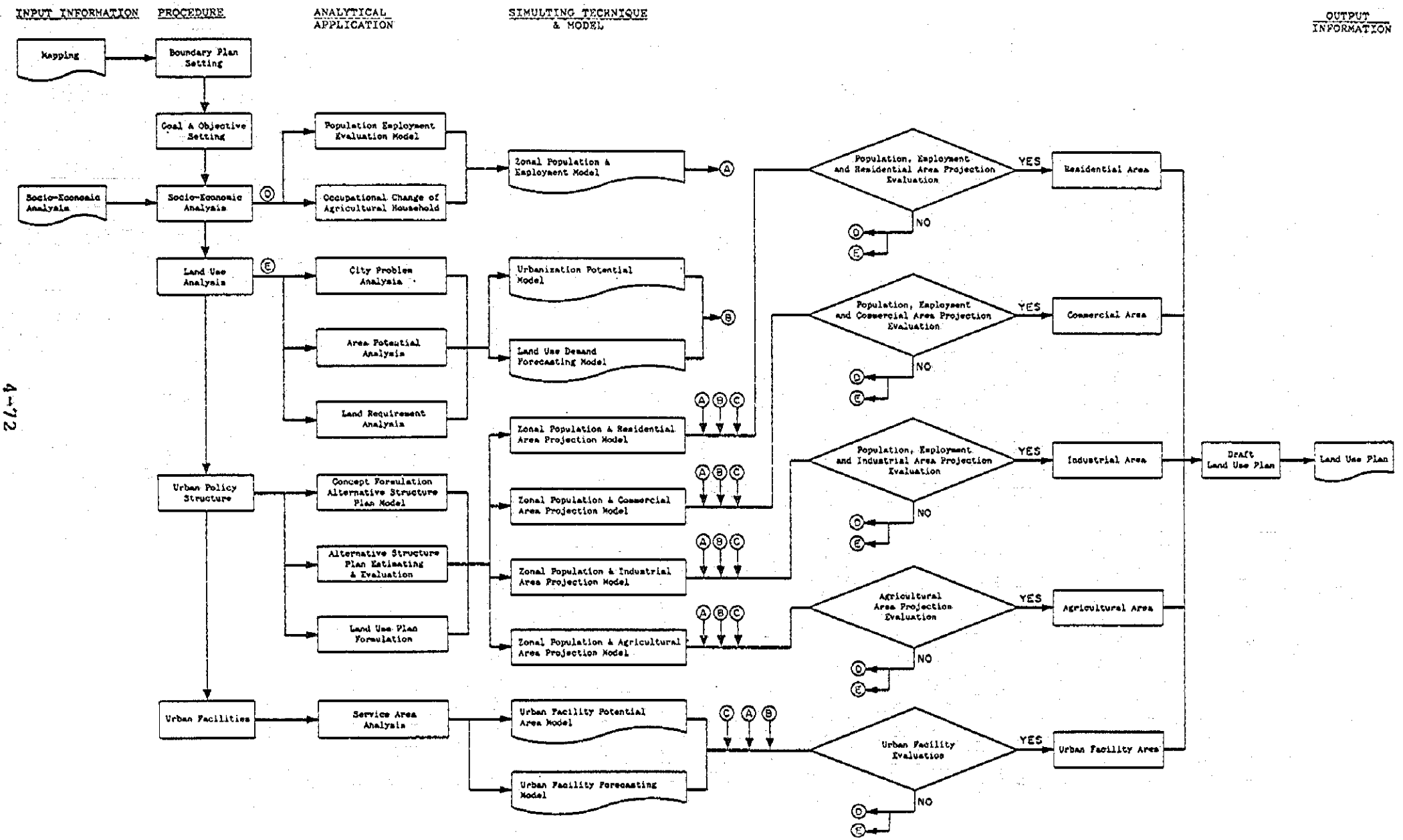
Urban Policy Structure

Planning structure formulated by appropriately selected concept is set up in this stage.

Urban Facility

The facilities such as parks, infrastructure, etc, supporting each land use planning area will be designed in order to complete the planning structure.

Fig. 4.7.8 LAND USE SIMULATOR



(4) Simulating Technique Model

Zonal Population and Employment Model (A) and Urbanization Potential and Land Use Demand Forecasting Model (B)

These steps will appropriately indicate zonal population of residential, commercial, industrial, agricultural area projection and also urban facility potential area and forecasting model.

Urban Facility Potential Area Model and Urban Facility Forecasting Model (C)

These models will contribute to the factors which indicate zonal population of each land use projection and they will be used to evaluate urban facility.

After this step, each land use area will be defined and a draft land use plan will be set up, then the final output the Land Use Plan will be produced.

4.7.6 Transportation Simulator

Transportation planning process, in general, consists of the following five steps; which roughly describes the structure of this manual.

1. Data collection (PART 3);
2. Analysis (PART 4);
3. Forecast (PART 5 and 6);
4. Plan design (PART 7); and
5. Plan evaluation (PART 8).

At the data collection stage, necessary data for planning, for example, socio-economic indices, topographic data, existing condition of transport systems and land use, among many others, are collected through field survey. Certain data are collected from local authorities and other governmental organizations.

The collected data are, then, used to analyze the existing situation of city, and the problems which need to be dealt with in the course of plan designing are identified. The planning strategies (policies) are formulated based on the result of these analysis.

Physical Planning usually looks ahead for twenty years. Target population and expected levels of activities are provided as a framework to the transport systems planning. The future transport demands are forecasted by using this framework. The methods of transport demand forecast should preferably be based on quantitative approaches but simplified methods may be applied in certain cases.

At the plan design process, the transport planner considers a variety of factors to prescribe transport system so that the transport problems identified at the earlier stage may be overcome and the expected future transport demand may be accommodated. The coordination with the land use planners is of particular importance at this stage in order to balance transport supply and demand, and to provide the transport system to guide the urban development into preferable directions.

The plan design is made in parallel with plan evaluation. A set of evaluation indices are determined and the adequacy of the generated plan is tested against these indices. The best plan is, then, put forward as a draft general plan for the transport system.

The transport planning process in the formulation of comprehensive land use/transport plan is shown in Figure 2-1. Appendix 2-1 shows the detailed steps of transport planning tasks in the preparation of general plan as a statutory planning process.

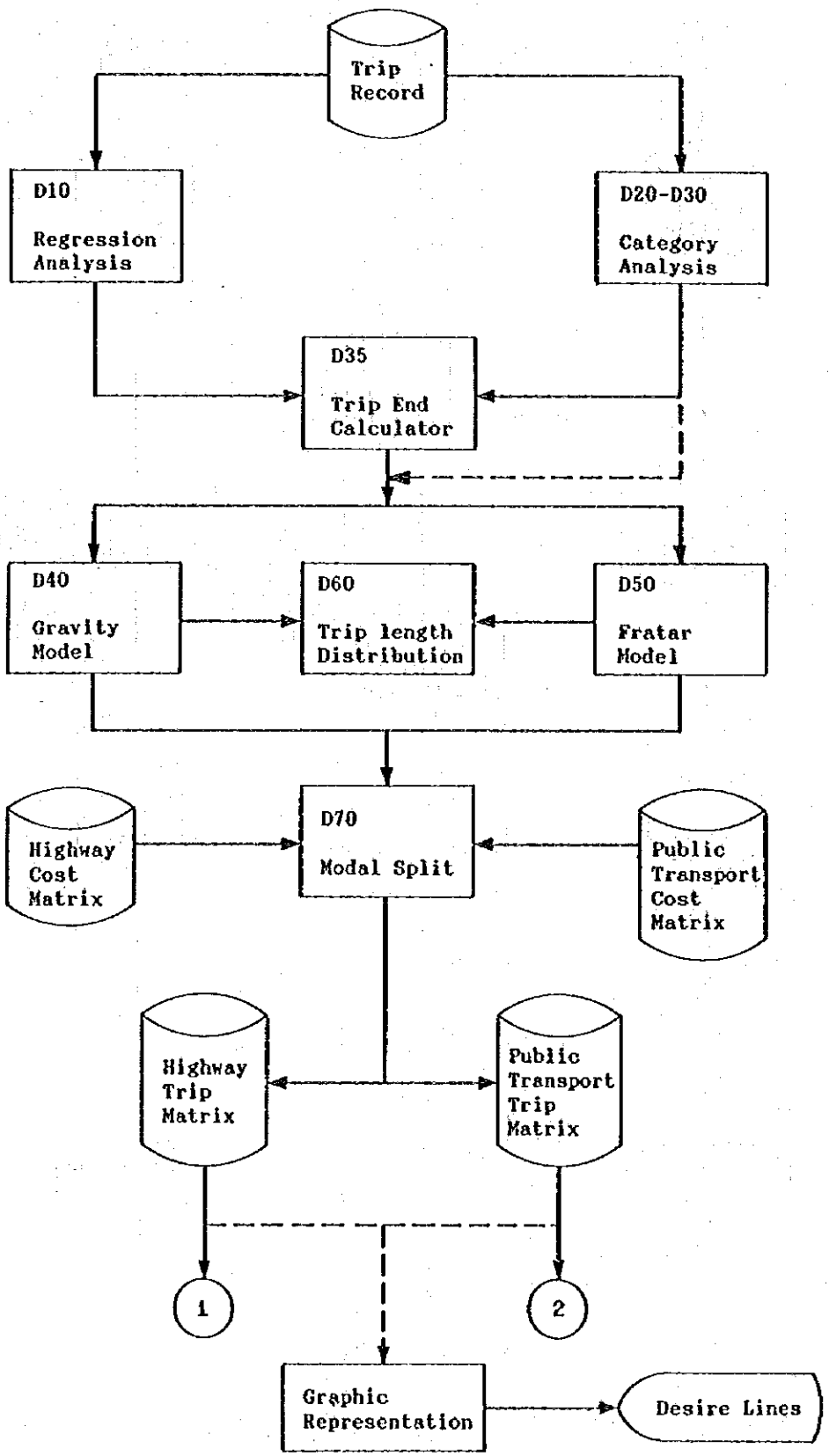
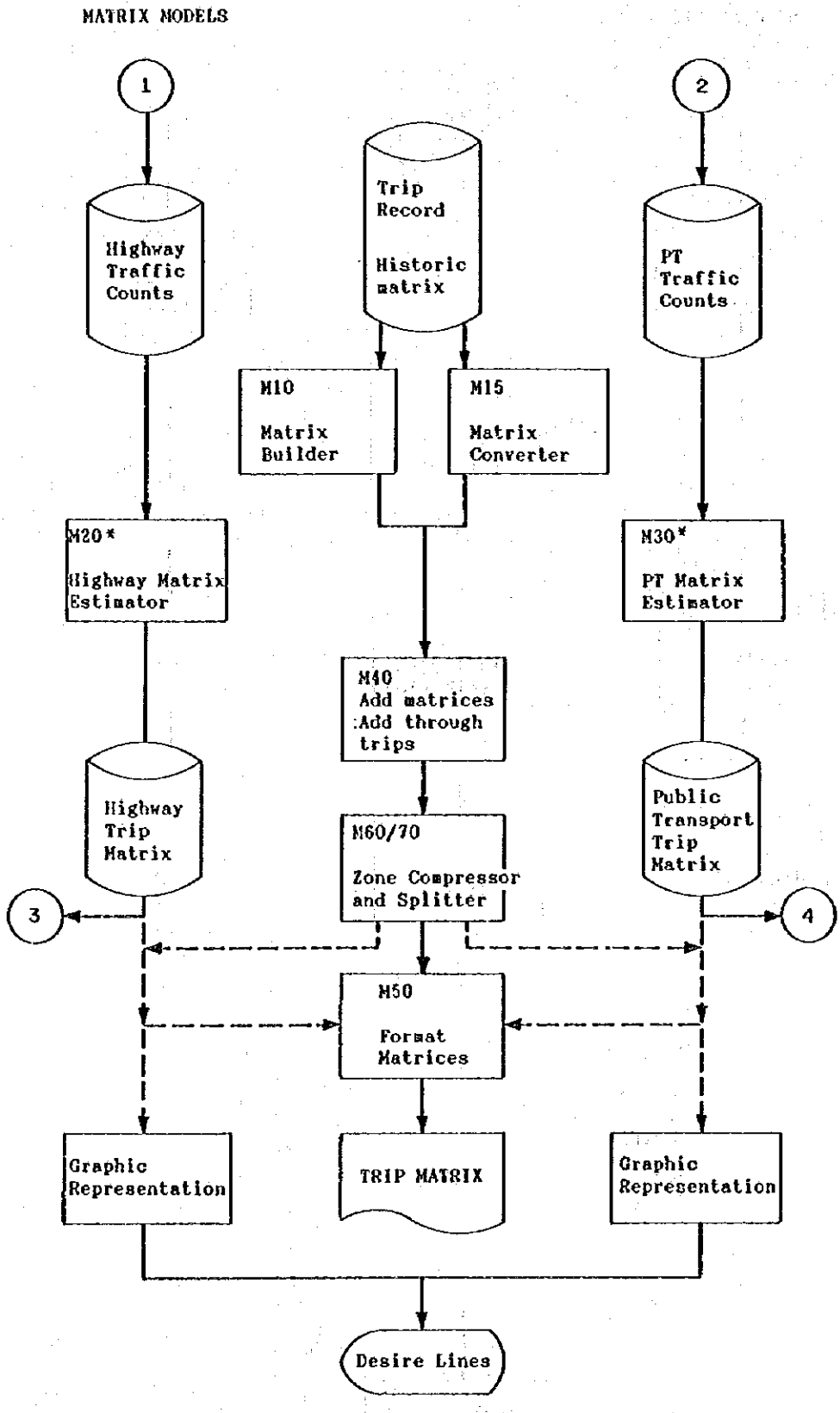


Fig. 4.7.9 DEMAND MODELS



* these programs also use network data as given from the P modules

Fig. 4.7.10 MATRIX MODELS

PRIVATE TRANSPORT MODELS

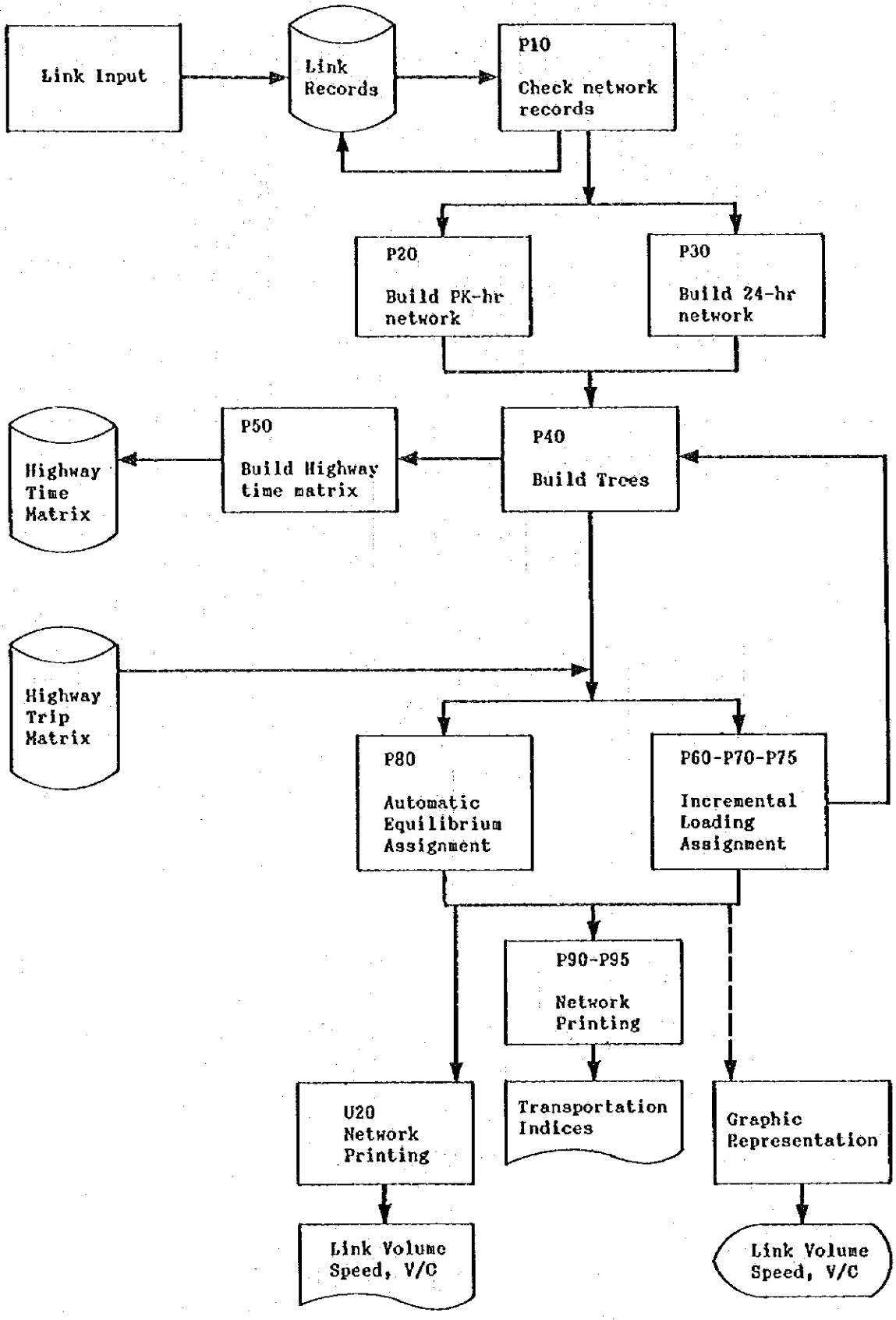


Fig. 4.7.11 PRIVATE TRANSPORT MODELS

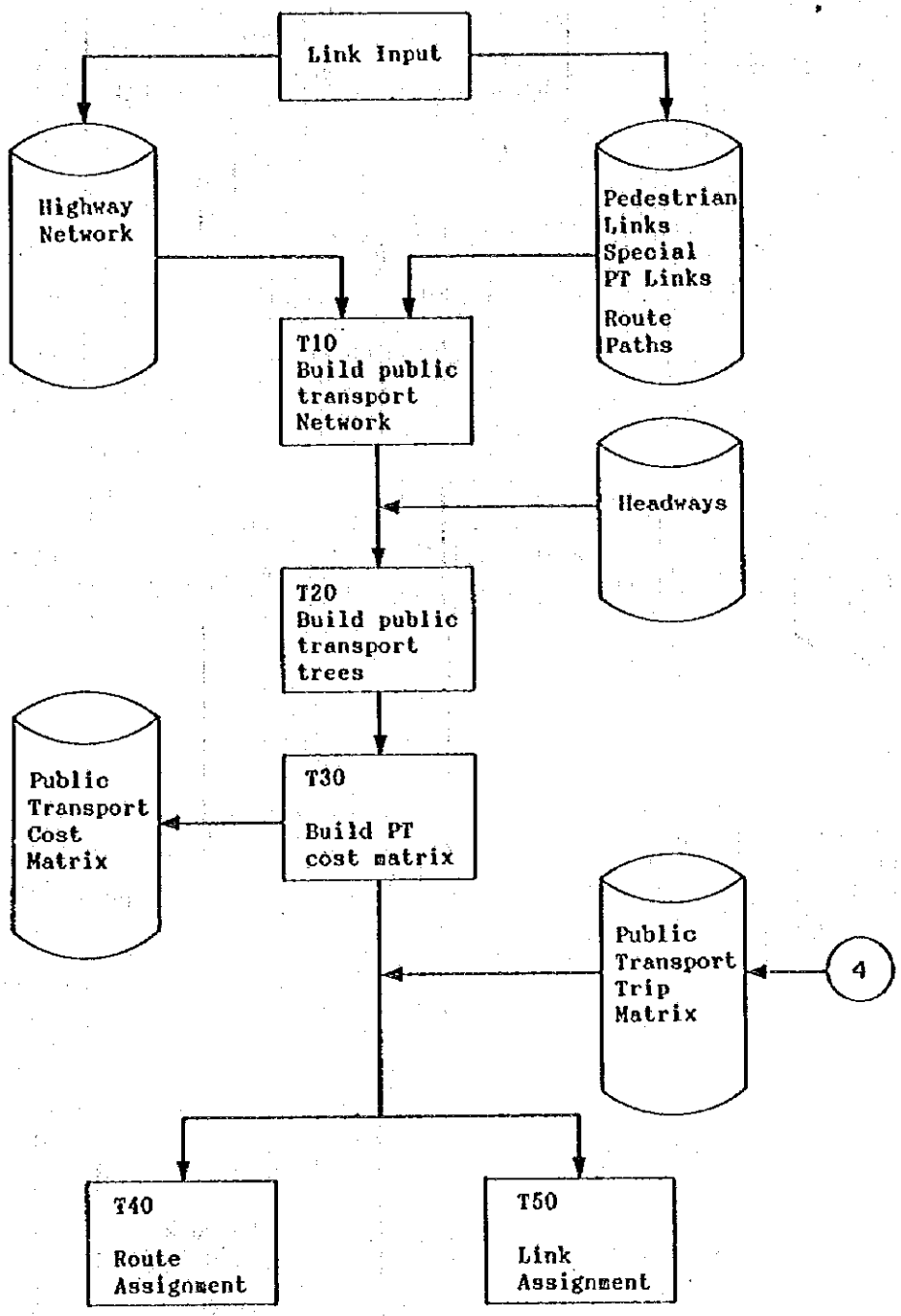


Fig. 4.7.12 PUBLIC TRANSPORT MODELS

CHAPTER 5
COMPUTER SYSTEM DEVELOPMENT

CHAPTER 5
COMPUTER SYSTEM DEVELOPMENT

5.1 Outline of Computer System Development

(1) Online and Off-line Data Processing

The computerized system should be developed from the inception stage, at which JICA Study Team proposed a system that is based on the existing situation of DTCP, to the innovative stage and full scale development stage.

Depending on the forms of information processing, the above stated stages can be classified into off-line processing and on-line processing, according to the difference in the means of information transmission. It can also be classified into batch processing and real-time processing by the difference in the response performance of the processed results. The relation between the system development stage and the processing form is shown in Fig. 5.1.1.

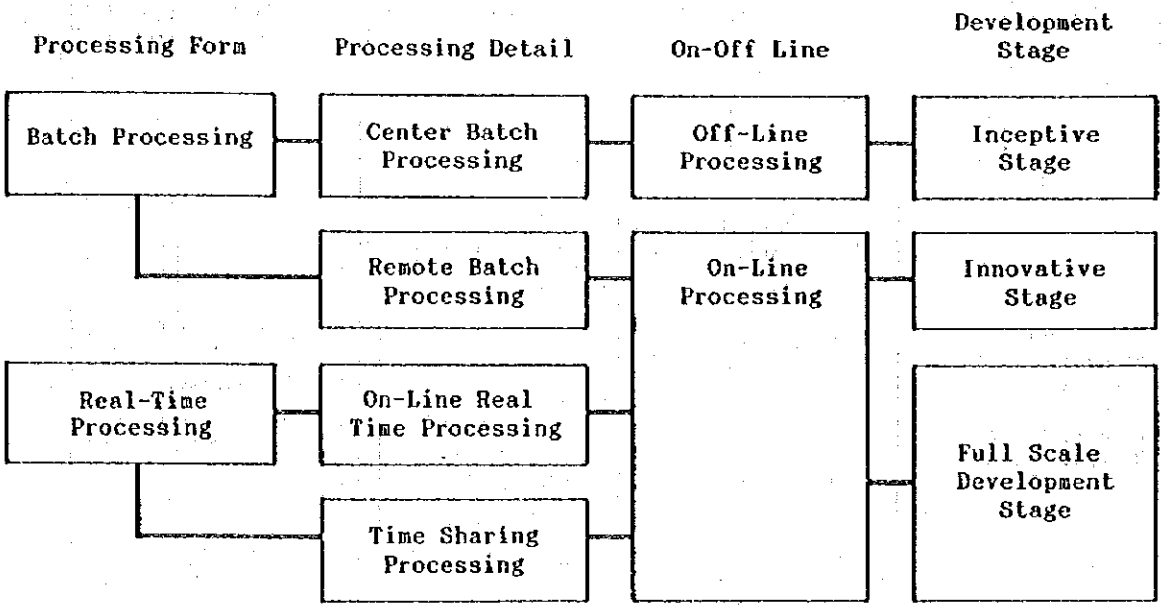
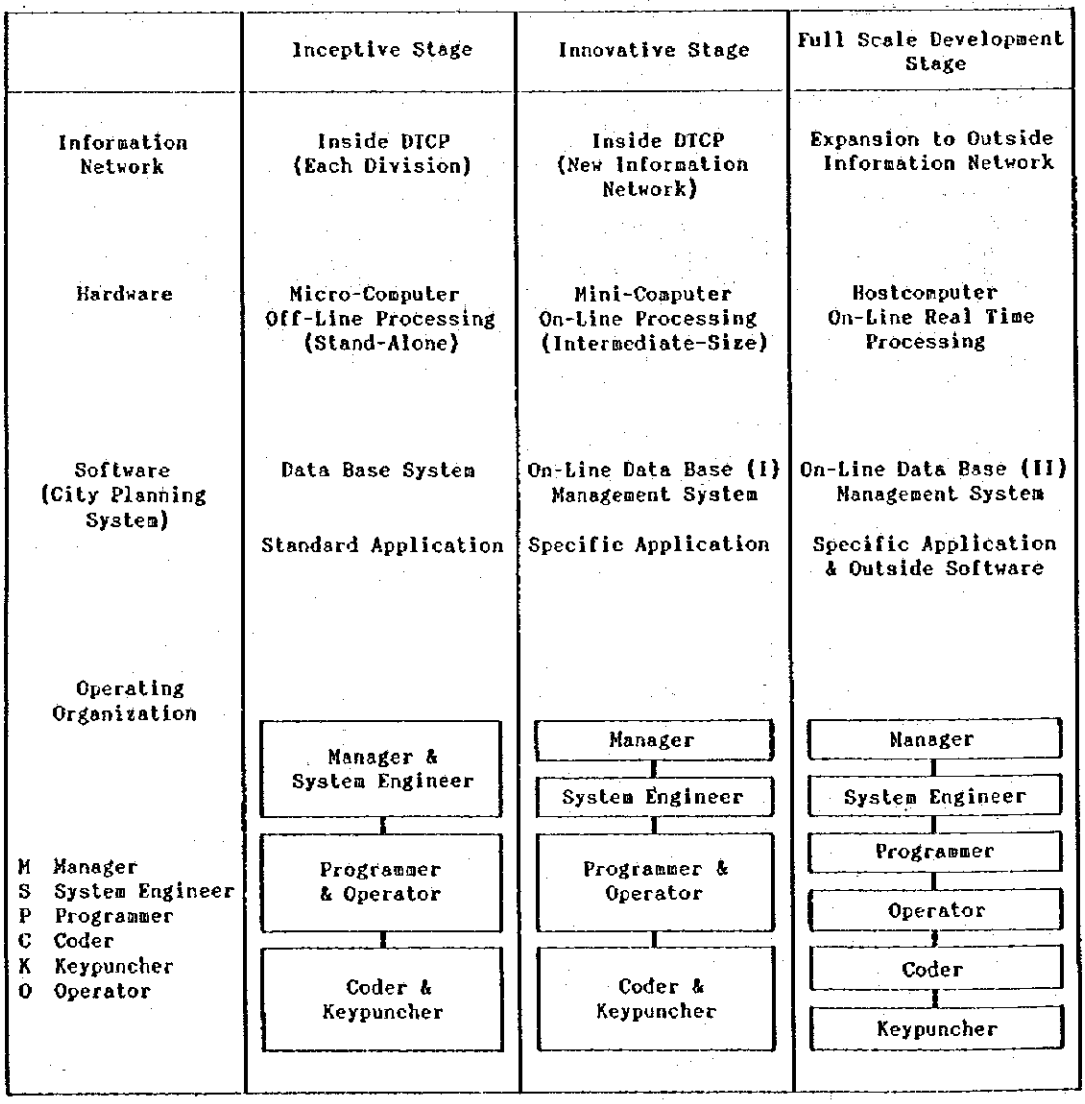


Fig. 5.1.1 Processing Forms

(2) Advancement of Data Base Management System

Regarding to the advancement of the Data Base Management System, DTCP must firmly establish the different stages of computer systems to cope with their progressive advancement of city planning as shown in Fig. 5.1.2.

Fig. 5.1.2 ADVANCEMENT OF DATA BASE MANAGEMENT SYSTEM



5.2 Inception Stage

Computers were first used for batch processing, and recently, micro-computer containing a micro-processor has been used.

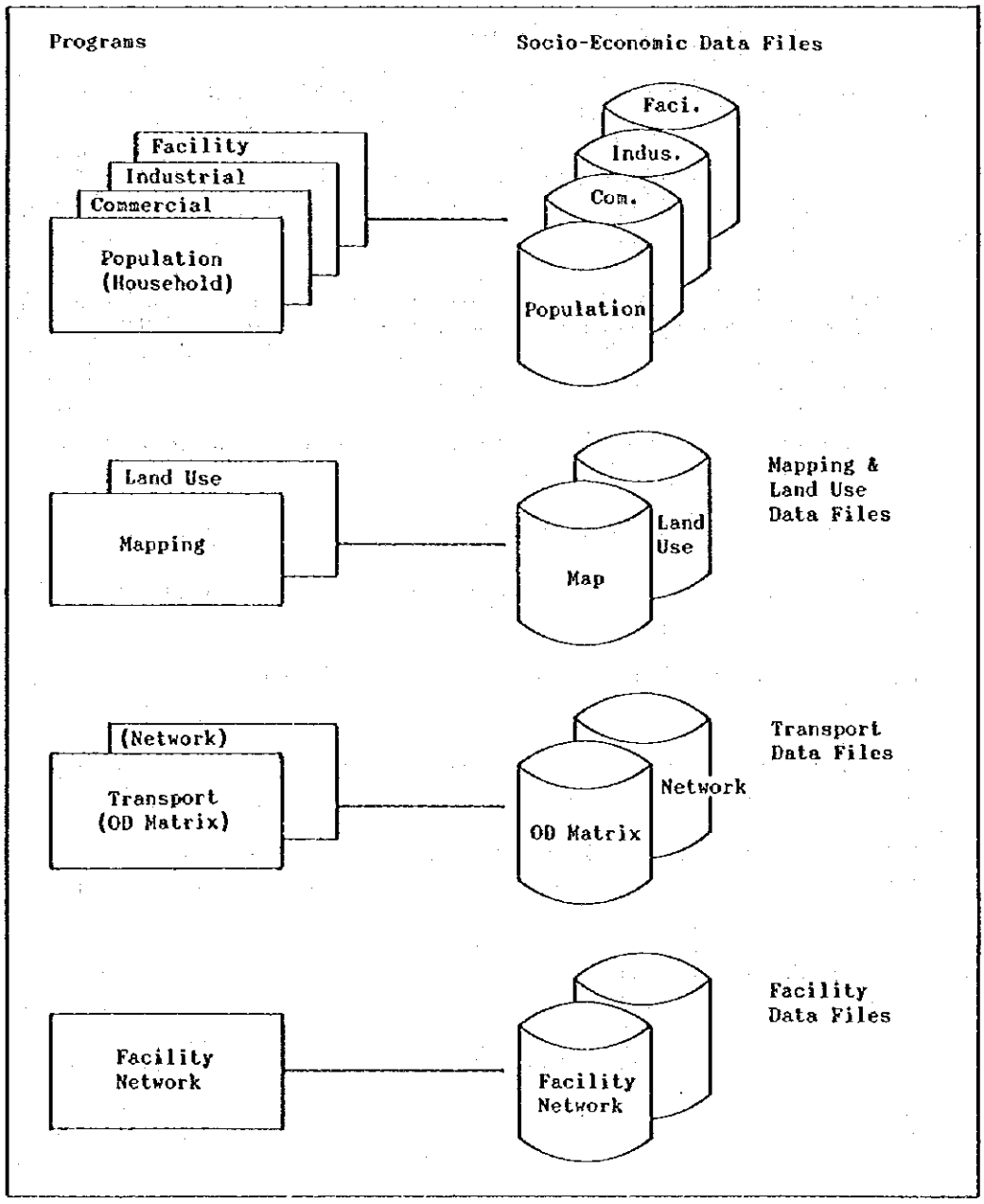


Fig. 5.2.1 File-Oriented Programming before Data Base (Inception Stage)

Batch processing is defined as follows:

- the processing of an individual programme, that does not permit interaction between the programme and user once the programme has been read by computer.

- the efficient processing of a collection of related programmes grouped together as a single run.

Computers were first used for batch processing and recently.

Micro-computers which are small, low-cost computer containing a micro-processor were used.

Batch processing of micro-computer is stand-alone which is designating an operation performed by a device, programme, or system independently of another.

In this stage, each programme is connected directly to disk files it uses. Fig. 5.2.1 shows the pre-data-base method of programming.

Batch processing requires the following work steps:

1. Data generated at various locations is collected at the computer center.

2. The collected data is collectively input to a computer.

3. In the computer center, data is processed in units of one day, one week, or one month.

4. The processing results are collectively output at the computer center.

5. The processing results are distributed to those people who require them.

   Since people must collect the data and distribute the output results, much time is required for the entire processing from data collection to distribution of the output results.

Taking the existing situation of DTCP into account, the computerized system should be started from batch processing by micro-computer. At this stage, each programme is connected directly to the disk files it uses. Fig. 5.2.1 shows this initial concept of programme structure.

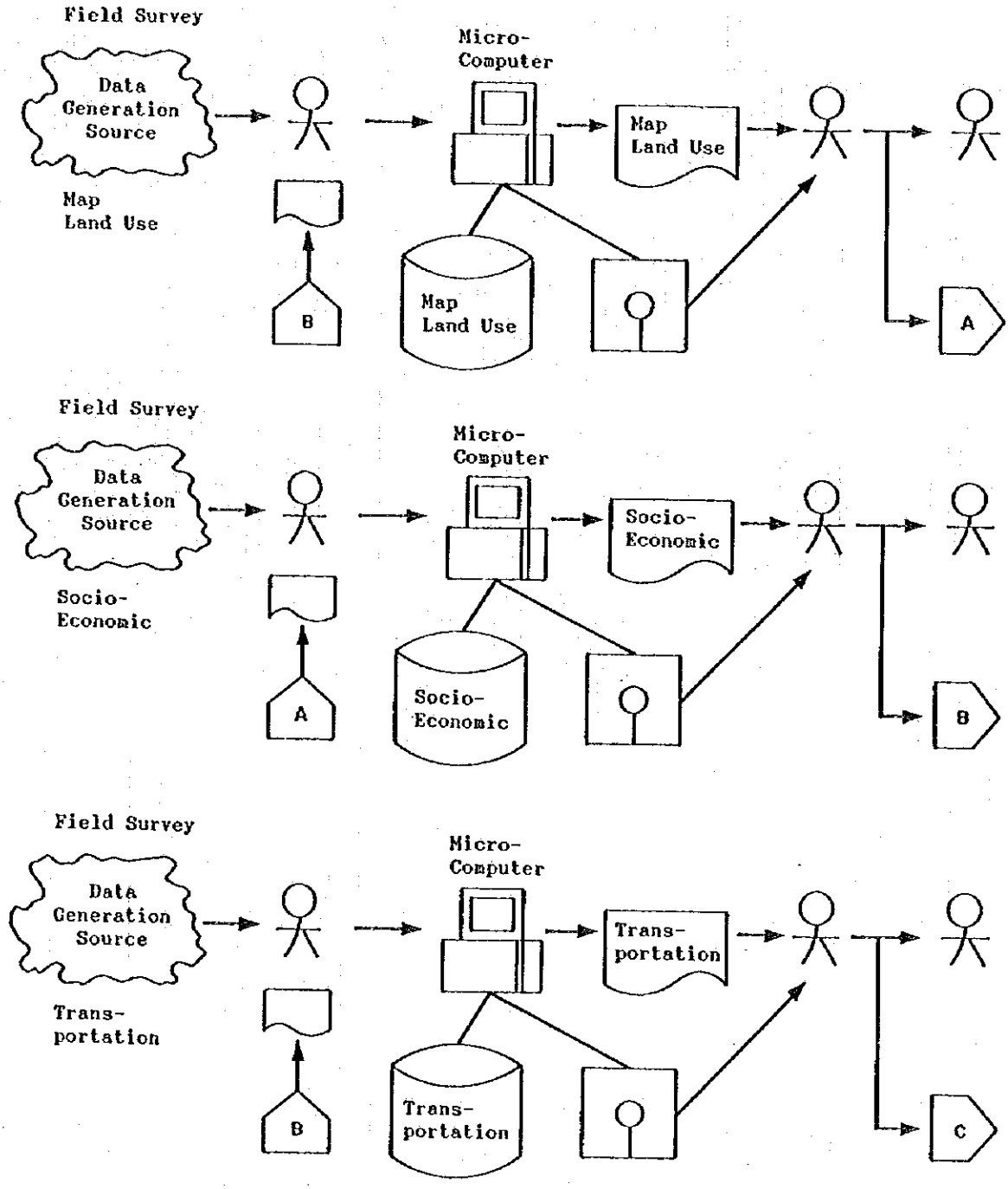


Fig. 5.2.2 BATCH PROCESSING BY MICRO COMPUTER (INCEPTIVE STAGE)

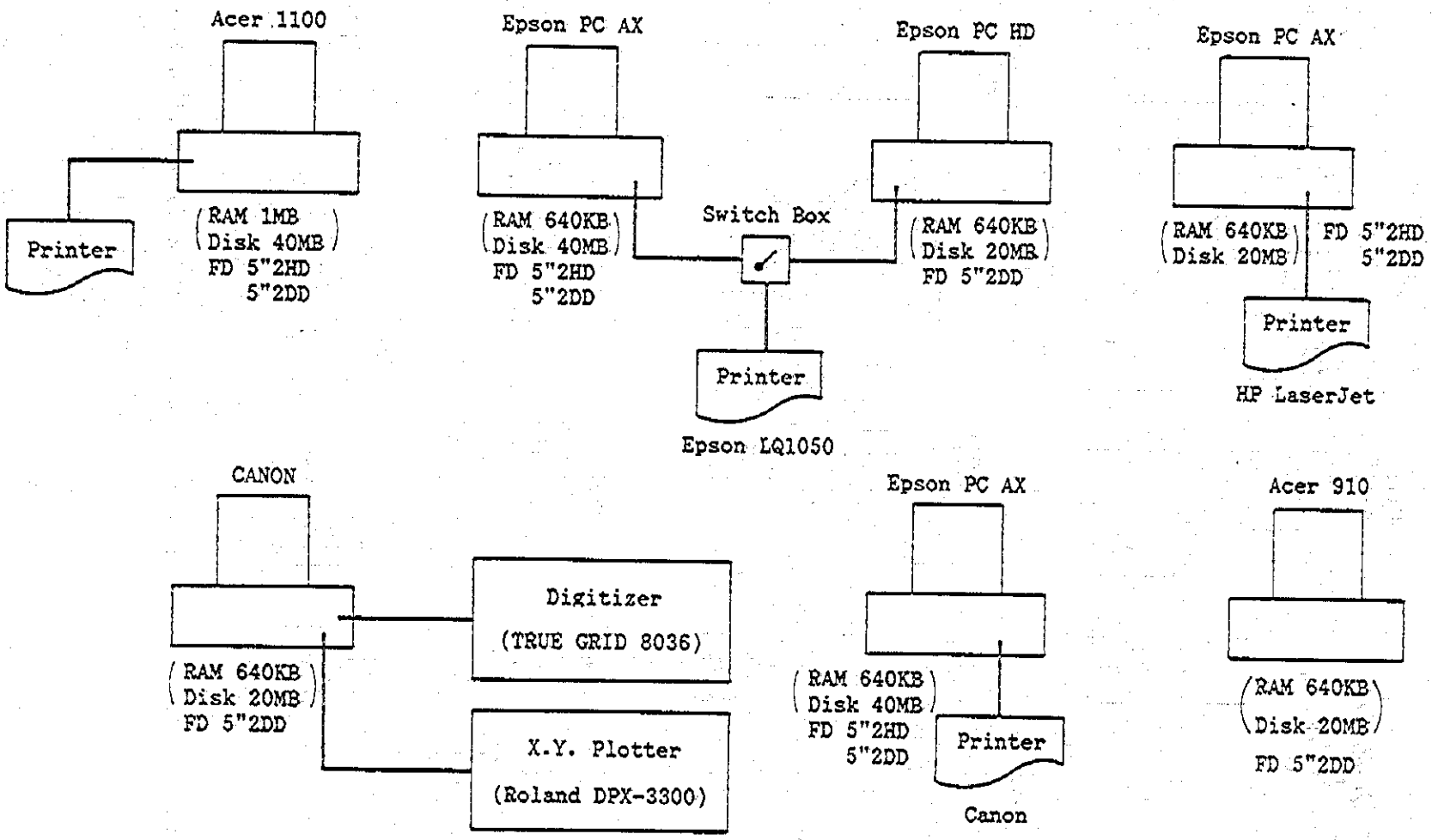


Fig. 5.2.3 STAND ALONG SYSTEM

5.3 Proposal of Innovative Stage (On-Line Processing)

In response to the technical advancement of city planning, computer system should be advanced in terms of data processing and the degree of subordination assigned to the particular data within a hierarchical structure. This new system is developed to reduce the time needed to collect the data and distribute the output results. Moreover, it could standardize the data file and application programmes.

Fig. 5.3.1 shows the data base image in which DBMS controls all access to an integrated data set.

- In the first case, programmes are forced to look at data stored in files, and there is no provision for queue processing.
- In the second case, the programmes have a logical view of the data that is independent of the way it is laid out on the disk.

The data on the disk can be changed or supplemented without requiring programme changes, and it can be also be viewed in different ways.

Based on the completion of "Inception Stage System", DTCP should establish an internal information network towards the "Innovative Stage". Innovative Stage requires mini-computer, and the software system should be developed in the key areas of statistical processing techniques (large amounts of data), analytical models, simulation models and so on.

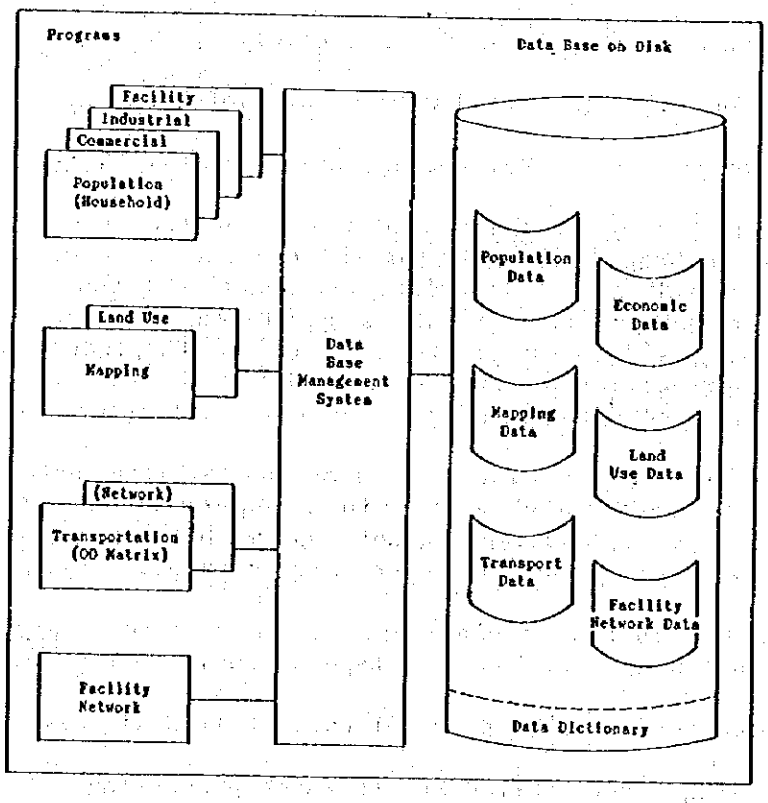


Fig. 5.3.1 The Data Base Management System (Innovative Stage)

Remote Batch Processing is defined as follows:

In the next stage, Fig. 5.3.2 shows the Data Base approach in which the DBMS controls all access to a single collection of Data.

In the first case, Programmes are forced to look at data stored in files, and there is no provision for queries.

In the second case, the programmes have a logical view of the data that is independent of the way the data is physically laid out on the disk.

The Data on disk can be changed or supplemented without requiring programme changes, and it can be viewed in different ways.

This new system is developed to reduce the time needed collect the data and distribute the output results and to standardize the Data File and applications.

This system is called on-line system. (Remote Batch Processing) In on-line processing, data is collected and the output results are distributed through communication lines.

Therefore, even if data is generated in a remote place, it can be sent to the computer center in a very short time.

Thus, the time needed to collect data and distribute the output results is reduced, but data processing in this example is the same as batch processing.

That is, collect data is processed collectively.

Such on-line processing is called remote batch processing.

5.4 Proposal of Full Scale Development Stage (Future System)

Remote Batch processing has been used in the on-line processing system of the Innovation Stage. However, as the on-line computer applications are processed, new demands will arise. It means that the data should be processed and output as soon as it is generated. At this stage, any request for using a data or programmes must be responded immediately. Such immediate processing of data is called a real-time processing. At present, on-line processing usually means real-time processing.

At the Innovative Stage, DTCP will have a complete data processing system based on DBMS which can guide, supervise and assist city planning. This system, however, cannot cope with the requirement due to the DTCP's efforts to extend the coverage of city planning.

In terms computer system development, it means a new computer center based on the full scale development stage as shown in Fig. 5.4.1.

Real Time Processing is defined as follows: (Full Scale Development)

On-line processing first started Remote Batch processing.

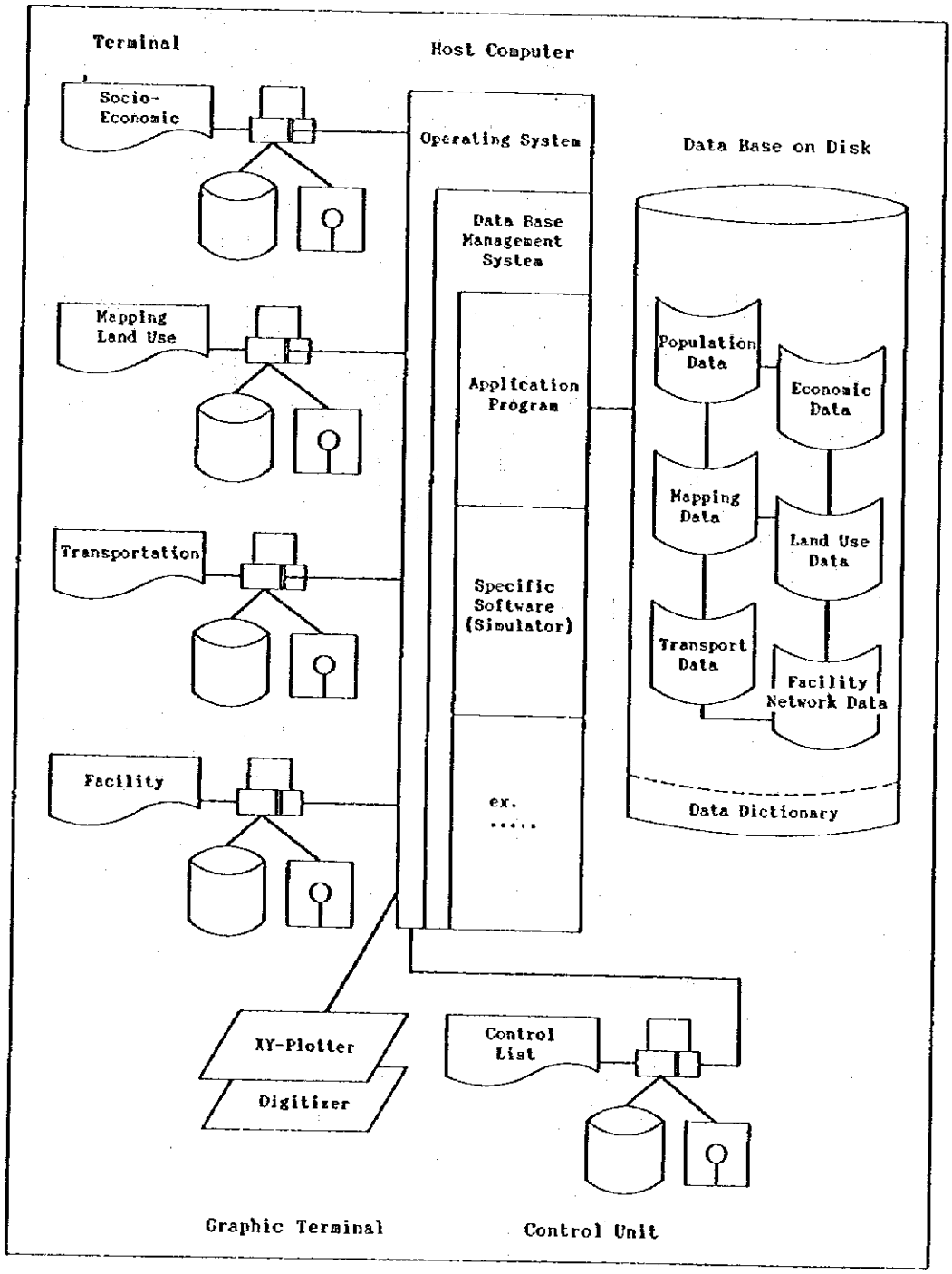


Fig. 5.3.2 REMOTE BATCH PROCESSING (INNOVATIVE STAGE)

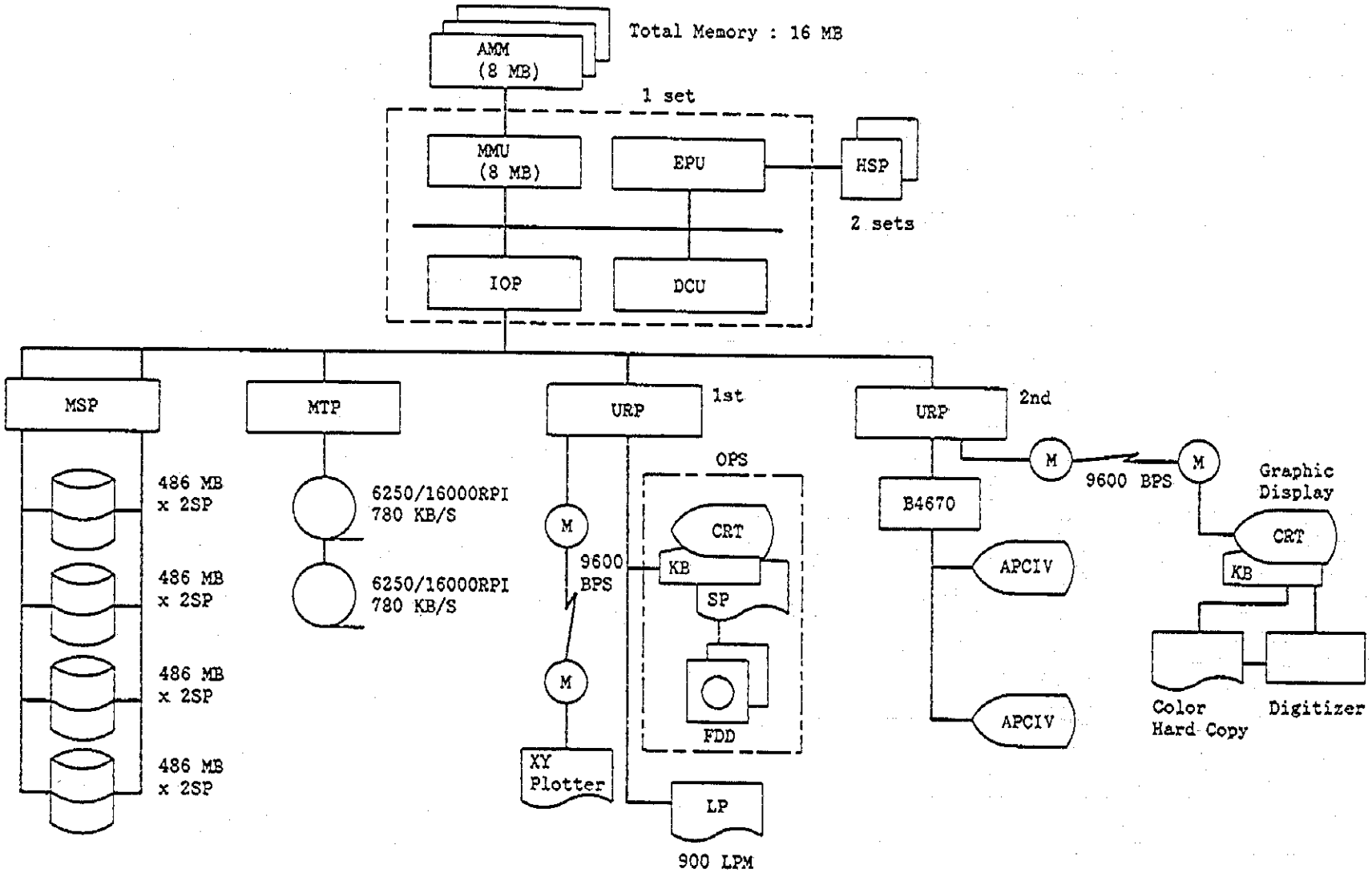


Fig. 5.3.3 An Example of Hardware System Configuration

However, as computers were applied to various on-line jobs, there arose a new demand where "data should be processed and output as soon as it is generated".

Any request for using a data or programmes must be given in responded to as soon as possible.

Such immediate processing of data is called real time processing.

At present, on-line processing usually means real time processing.

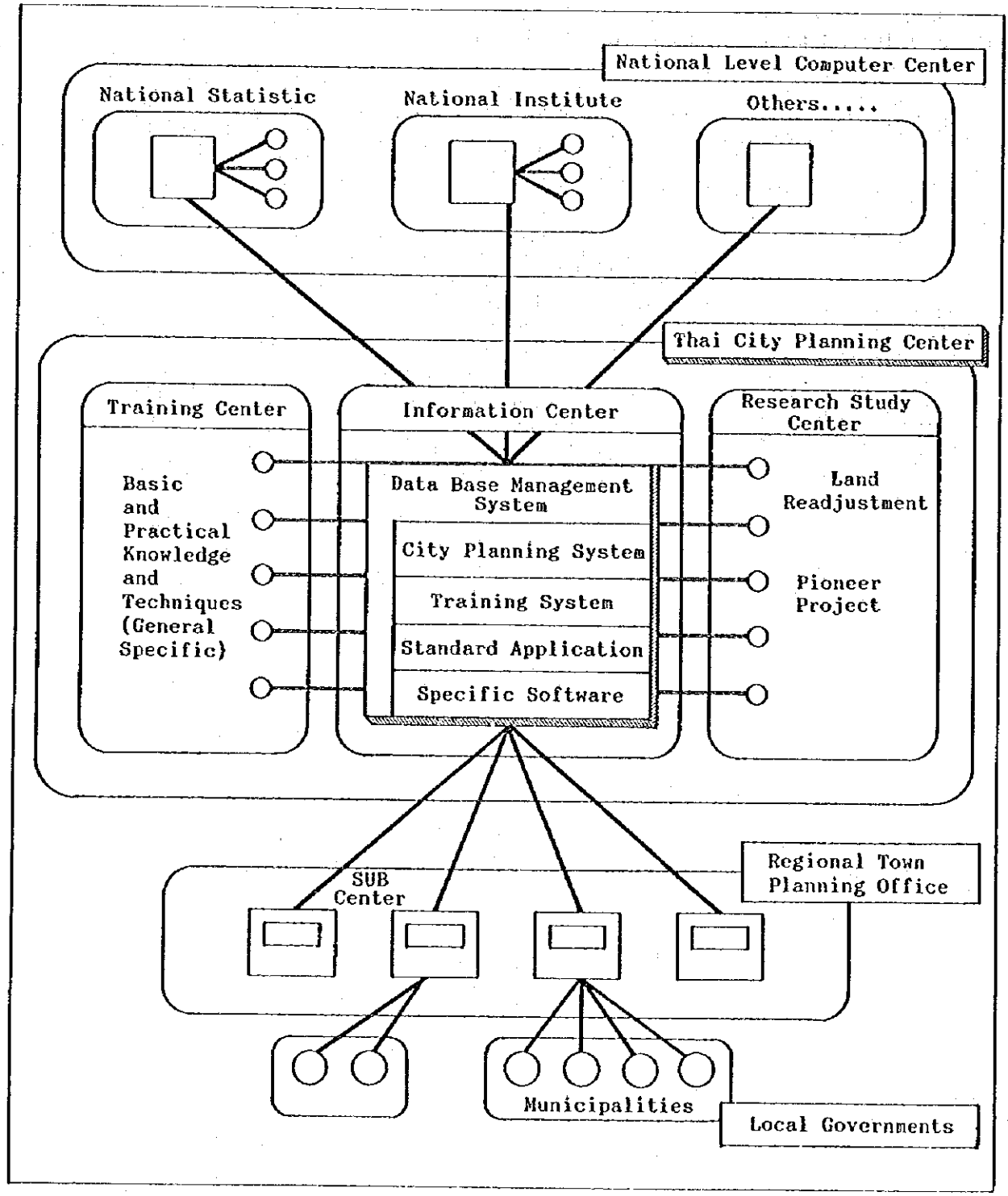


Fig. 5.4.1 FUTURE SYSTEM (FULL SCALE DEVELOPMENT STAGE)