OCCASION

This publication has been made available to the public on the occasion of the 50th anniversary of the United Nations Industrial Development Organisation.

DISCLAIMER

This document has been produced without formal United Nations editing. The designations employed and the presentation of the material in this document do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations Industrial Development Organization (UNIDO) concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries, or its economic system or degree of development. Designations such as "developed", "industrialized" and "developing" are intended for statistical convenience and do not necessarily express a judgment about the stage reached by a particular country or area in the development process. Mention of firm names or commercial products does not constitute an endorsement by UNIDO.

FAIR USE POLICY

Any part of this publication may be quoted and referenced for educational and research purposes without additional permission from UNIDO. However, those who make use of quoting and referencing this publication are requested to follow the Fair Use Policy of giving due credit to UNIDO.

CONTACT

Please contact publications@unido.org for further information concerning UNIDO publications.

For more information about UNIDO, please visit us at www.unido.org

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
Vienna International Centre, P.O. Box 300, 1400 Vienna, Austria
Tel: (+43-1) 26026-0 · www.unido.org · unido@unido.org

DO 1880

United Nations Industrial Development Organization

Distr.
LIMITED

ID/WG.84/2
22 October 1970

ORIGINAL: ENGLISH

Expert Group Meeting on the Industry Files System
and other Data Bank Techniques for
Industrial Programming

Vienna, Austria, 16 - 20 November 1970

MAIN STEPS FOR BUILDING A DATA BANK FOR PLANNING PURPOSES 1/

prepared by

Jean V. Salmona
L'Observatoire Economique Méditerranéen
(I.N.S.E.E.), Marseille, France

1/ The views and opinions expressed in this paper are those of the author and do not necessarily reflect the views of the secretariat of UNIDO. This document has been reproduced without formal editing.

id.70-5893

We regret that some of the pages in the microfiche copy of this report may not be up to the proper legibility standards, even though the best possible copy was used for preparing the master fiche.

Table of contents

PREFACE

This report has been prepared in accordance with the request of the United Nations Industrial Development Organization in support of the East African Working Party on Industrial Programming Data, which met at Nairobi (Kenya) in November 1969. The purpose of the report is to present, in relation to the experience of developed countries, the main elements of a data bank and the main steps of its implementation and to try to define the way in which a similar problem should be handled in a developing country. A preliminary summary was submitted and discussed at the East African Working Party. The following report was in fact written after the Working Party, and takes into account the discussions held during the Party's sessions.

When the manuscript of this report was about to be completed, the author received the report of the East African Working Party, which had been prepared by the UNIDO Secretariat rather independently of this one.[2/] The author gladly acknowledges the similarity of the theme between this report and the UNIDO report. The two reports share the same principles and guidelines in drawing up a data bank programme suitable for the conditions prevailing in developing countries in general. Apart from the style of writing, the main difference is that the description of the critical steps for data bank building in this paper is not explicitly geared to the particular version referred to as the Industry Files System, but is rather consciously based on the author's own experience with the I.N.S.E.E. Observatoire Economique Méditerranéen (Marseille, France). On that score, it is hoped that this paper touches upon at least some of the elements which the UNIDO report on the Industry Files System ought to have covered in more orderly or clearer terms than it does.

As in the case of the UNIDO report, the author wishes to emphasise that the data bank does not involve necessarily sophisticated systems and big computers. But the very purpose will be met if it can reap what could

---

2/ THE INDUSTRY FILE SYSTEM: Draft Report on the East African Working Party on Industrial Programming Data

become available from existing data sources and provide it to those who really need it.

During the past decade, many data bank experiments were born in developed countries, and especially in Sweden, Norway, USA, Great Britain and France: sectoral data banks for management, and general data banks for national, regional and urban planning, for government or for private users. Some of these experiments failed, others have succeeded. From these failures and successes, it is possible to draw conclusions about what to do and what not to do.

In writing the following, an effort was made to analyse, in relation to these experiments and their conclusions, the general principles and elements to be considered for data bank building, and to draw up a step-by-step guide for definitive action towards a data bank in a developing country. The ideas expressed in this paper are, like those in the UNIDO report, still in very general terms and need to be adapted for the conditions and needs of any particular country for actual implementation.

It must also be noted that this report deals only with data banks for planning purposes and investment decision making. Thus the problem of a management information system is not explicitly considered here. One of the main characteristics of information systems for planning decisions, as distinguished from those for management decisions, is that information for planning typically covers a number of different sectors while a management information system concentrates on one sector only. For instance, the information for the management of a transportation network consists of data on the existing public and private networks, on daily traffic, vehicles used, workers employed, etc. Information for transportation planning includes data on the future population and employment per district and city, daily migration, etc.

The various sets of information necessary for different planning or decision-making units will have, of course innumerable intersections among them. Now, in a given country, it may happen:

- that a sector of the economy or an institution wishes to build its own information system for its planning decisions; or
- that the State decides to create an information system (or an information system network) for national, regional or urban planning; or
- that many groups of economic institutions try to build a common information system for planning.

Whichever may happen to be the case, one can hardly over-emphasise the merits of having access to a number of different sources of information and sharing a common information system with a number of sectors or different decision-making authorities. This paper primarily deals with an information system for planning opened to any users: that is, a data bank as a completely open information system. As a strategy for maintaining input-output efficiency, a partially closed system may be favoured under certain circumstances. Although emphasis is laid on "users' needs" throughout this paper, no specific model for the grouping of users and suppliers is pre-supposed. In this sense, too, the problem of a management information system has been left out of the scope of this paper.

I. INTRODUCTION: DATA BANK AND INFORMATION SYSTEM

A. Definitions

An information system is a logical way of organizing information to fit given needs.

An information system is comprised of:

- a set of information elements: data related to a person, a building, a block, a firm, a document, etc. on the one hand, and
- means allowing integration, retrieval, access, processing, edition of these data, on the other hand.

An information system may be opened or closed:

- closed if it is intended for the use of only the organizations which are the builders of the system (these systems are generally intended for specific uses: budget management, social security, taxes, etc.);
- opened if any user has access to the system.

A data bank is, by itself, an opened information system. The sets of data of the data bank are called files.[3/] A data bank is constituted of the following elements:

1. The basic files, called the data base: each file contains data on elements of the same category (persons, establishments, etc.);
2. A file management system liable to up-dating the files and to merging the data extracted from various files: this involves the coding of individual elements (e.g. industrial establishments for an industry file system), with the same code for a given unit (e.g. establishment) in all the basic files (same identification number);
3. A files processing system, making it possible for the users to draw out of these files the information fitting their needs, in the shape suitable to the models of their decisions.

We shall see below that these conditions normally lead to the use of computers, but that they do not necessarily, from the very beginning, lead to complicated and sophisticated systems.

3/ These definitions are drawn out of "Banques de Données et Systemes d'Information Régionaux et Urbains aux U.S.A.", by Ferragu, Salmona, Timmel and Robequain, 1969.

B. Sources of data

There may be two sources of data:

- data arising from administrative action (see Fig. 1);
- sample surveys and census data.

The results of all experiments to date show that the data base must be supplied generally as by-products of administrative action.

As a matter of fact, administrative action is of a repetitive and periodical nature. The data it provides as a by-product can therefore be automatically and regularly up-dated, which is obviously not the case for a survey or a census.

Besides, experience shows that the cost involved in making available information from administrative action is generally low, and very much lower than the cost of a census or of a survey. Also, administrative action linked to the law (taxes, social security, registrations, etc.) ensures the desired degree of stability and regularity in terms of record coverage.

On the whole, ad hoc surveys are to be made only when no information is available from regular administrative records.

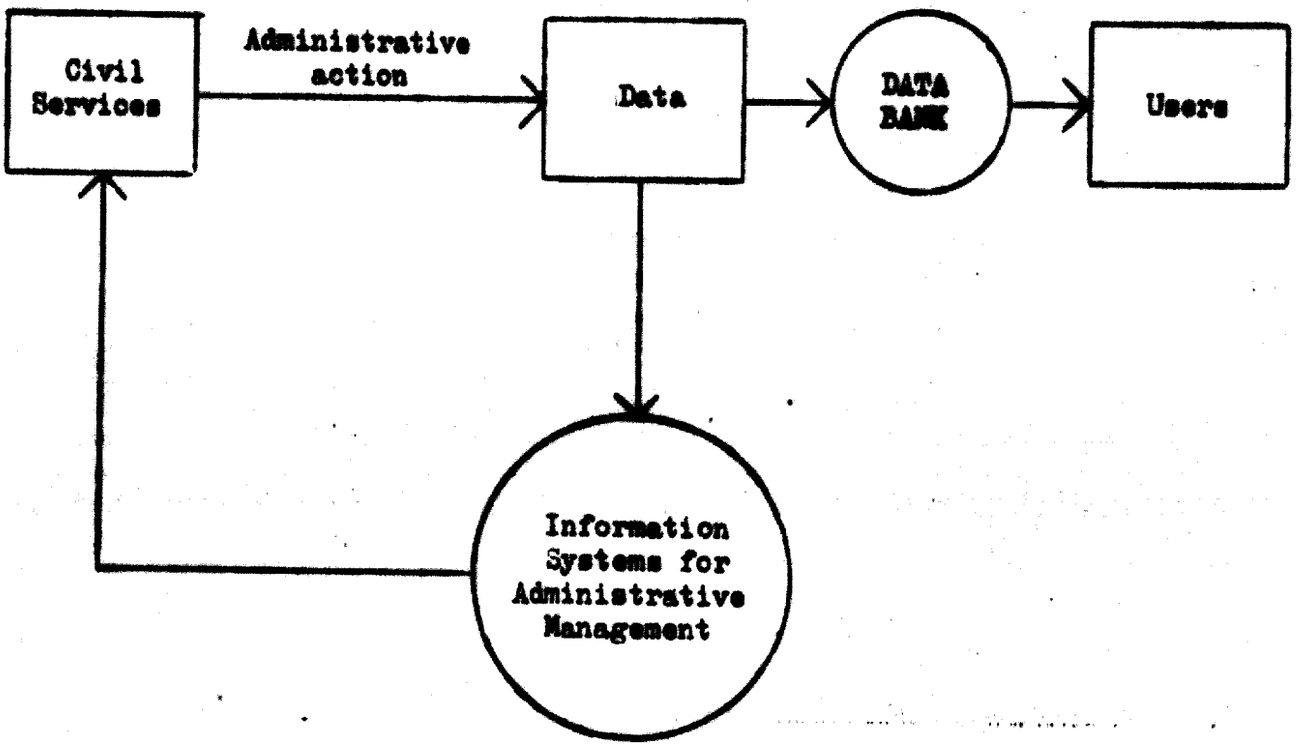
C. Sectoral versus general data banks

A data bank may be related to one given sector (for instance, a data bank on industrial establishments) or to many sectors.

As it has been mentioned above, an information system for management may be set up at the sectoral level. But the data required for planning must cover many sectors.

For industrial planning purposes (state planning as well as corporate planning), data on the existing industry and commerce, on manpower, on inland market, on foreign trade, etc. are necessary. In fact, it is not one but many information systems, related to different fields, which are necessary for planning, even if the planning is restricted to one field only.

Fig. 1 : A data bank for planning purposes



D. Data banks and their users

One must never forget that a data bank is not a purpose in itself. It is a tool, intended for certain users who are going to use it for certain purposes. Therefore, a data bank must be user-oriented. This is true for the types of data to be stored as well as for the systems themselves.

Most of the failures which occurred in various countries are to be imputed to a wrong orientation at the beginning: the data bank did not fit the users' needs and so the users did not use it.

II. THE MAIN STEPS FOR BUILDING A DATA BANK

This part of the paper is devoted to a technical analysis of the various steps to be taken in building a data bank. Problems of particular relevance to an implementation programme in a developing country will be studied in the third part of the paper.

A. The users and their decisions

The data bank must be organized in relation to the users and the uses for which it is intended. Before building a data bank, the first task is therefore to take an inventory of the users and of their decisions, to analyze these decisions and to determine the data and design the systems which fit the structure of these decisions.

A.1 Inventory of potential users

Taking duly into consideration the given economic and political structure of the country, one has to decide who, among the following various types of decision - making entity, are to be the users of the bank:

- central government policy makers and planners;
- regional and local authorities (policy makers and planners);
- parliament;
- public corporations;
- private firms, domestic and foreign;
- political parties;
- trade unions of workers;
- research centres;
- universities.

For each of these groups must be drawn up an inventory of the kind of decisions which the use of a data bank may help optimise. Such decisions may be divided into categories such as:

- location of an establishment (plant, supermarket, etc.);
- programming and locating public investments;
- marketing a new product;
- definition of a national education policy;
- general orientation of the economic policy;
- etc.

It is not necessary, at this stage, to get involved in the sophisticated problem of geography, although it is essential to be able to distinguish different geographical levels: nation level, region level, city level, district level, block level, etc. For example the decision concerned with the establishment of a new school takes place in many stages:

- First, definition by the government, at the national level, of relative priorities of various budget sectors and, particularly, of various investment sectors;
- Then, choice in aggregate volume of investments on the regional level: for instance, school investments relative to the population of school age, etc.
- Within a region, decision concerning the best location of schools taking into account the school age population, transportation networks, eto. in different cities and towns;
- Finally, within a city or town, choice of premises for location.

In dealing with a hierarchy of decisions of the above kind in relation to the "demand" study for a data bank, the following weighting method might be employed:

To each operation O_i (which may involve one or many decision makers: e.g. the location of a plant involves generally, at least, the decision of a private firm, of the central government, and of a regional authority) a weight is given. The weight can be given either to the whole operation O_i or to the decision D_{ij} of each decision maker A_j involved in operation O_i. In the latter case, a weight must be given, too, to each decision maker A_j.

These weights can be given by organizing a meeting with various national authorities interested in the data bank programme and asking each member of the group to give a subjective weight from 0 to 10. A kind of Delfi method may be employed to seek a convergence of views on relative weights.

If weights are given separately to decision makers A_j (weight a_j) and to operation O_i (weight o_i), weight d_{ij} of decision D_{ij} is the product of weights $o_i \times a_j$.

So, if possible, a weight will be given to each decision of each decision maker (for instance: the decision of a private firm of the industrial sector to locate a new plant). If this is not possible, at least a weight must be given to each operation (for instance: location of a new plant, including decision of private firms, government, etc.).

A.2 Analysis of decisions and data hierarchy

Then it is necessary to analyze every decision and to determine what data are necessary to optimize that decision.

The best method for this purpose is the case study method. Taking, for each decision, a concrete and real case, related to a decision to be made, or already made, analyze the way the decision is to be made (or has been made) and determine the information which can optimize (or could have optimized) the decision.

The case study will be realized for each decision with the participation of the involved decision maker(s), who must be representative of the group to which they belong. These case studies will be then used in group seminars (see D.2 of this chapter).

For each analyzed decision (or operation) will be thus established a list of the data necessary to its optimization. To separate the essential data from the non-essential (data not having the same importance in decision making: for instance, to locate a school in a given region, the knowledge of the school age population is much more important than any other information), each data category will be given a weight. Thus, for a given decision D_{ij}, each information I_k will be given a weight w_{ijk} which measures the importance of I_k in D_{ij} optimization. The product of $d_{ij} \times w_{ijk}$ gives a measure of the weight of data I_k related to decision D_{ij}.

Of course, when given information is useful to optimize many decisions D_{ij} the weights are to be added. Sum $b_k = \sum_{ij} d_{ij} w_{ijk}$ measures the absolute weight of information I_k. The final result of the analysis is a list of the

main data with a hierarchy of weights, measuring their utility for all the potential users concerned.

B. Data to be collected and stored in the bank

The operation is divided into two stages:

- Inventory of the existing data and of their availability;
- Choice and programming, in relation to the priorities calculated as above, of the data to be put in the bank and of the operations to make these data usable.

B.1 Inventory of existing data

B.1.1 Sources of data

Elementary, summarized and aggregated data, as available, are all considered in this inventory. These data can be divided into two types:

- Periodically and automatically up-dated data: these data are by-products of administrative periodical actions;
- Data which are not up-dated or which concern fields where no up-dated data are available: these are structural data, which will be put in the bank if their utility does not decrease too quickly when they get older (population census, transportation surveys, etc.).

The inventory will be realized in each administration source of data as follows, for instance:

i. tax administration
   - household income, balance sheet and accounts of firms and establishments
   - land-use data (property files);

ii. social insurance service;

iii. customs and foreign trade administration;

iv. public works and building administration;

v. schools and universities administration;

vi. inland trade administration;

vii. agriculture and industry administration;

viii. health administration;

ix. local authorities (registrations);

x. censuses and surveys.

Of course, if there is a Central Bureau of Statistics or an equivalent administration, the inventory must begin with this Bureau.

B.1.2 Form of available data

For each of the data existing in these sources, the form in which it is available should be examined: magnetic tape, punched cards, paper (files), books and reports, microfilm, etc. Attention should also be paid to potential data, that is, data to be generated by existing administrative actions which do not yet provide data in any explicit form.

B.1.2.1 Computerised data (files)

For each file, indicate:

- specimen of basic documents (if any);
- format and record lay-out;
- definition of unit (range);
- codes (composition and definition);
- geocoding: post address, block number, co-ordinates, etc.;
- geographic levels;
- frequency of up-dating;
- accuracy: reasons for under-estimation or over-estimation;
- existing ways of processing these data (tabulations and listings) and frequency of the processing.

B.1.2.2 Administrative forms

- specimen of the form;
- range;
- geocoding;
- geographic levels;
- frequency;
- accuracy.

B.1.2.3 Aggregated data

- statistical tables: design, codes, definitions, frequency of production, etc.;
- time series data (for instance: production indexes, etc.);
- maps and graphs;
- reports and other documents.

B.1.2.4 Inventory of recorded administrative facts, which can provide data.

B.2 Operations necessary to provide the required data

Now, the choice has been made of the data to be stored in the bank (see A.2) with weights measuring their utility. Then an inventory of existing data and their characteristics has been compiled. From the comparison between these two lists one can determine the operations necessary to obtain the required data in the required form.

In fact, the inventory of such operations may already have been taken at the previous stage. But now the costs of these various operations have to be considered. One cannot compare the cost of making available computerized administrative data for users to the cost of obtaining new data which involve the creation of a form and of a new administrative organization.

It is therefore necessary to compare the efficiency of the various data (measured by their weight) and the costs of making them available to users, in order to decide definitely which data shall be made available first.

B.2.1 First preliminary operations: registers and geographic files

For decision-making purposes, it is usually necessary to merge data from different sources. Thus, it will be necessary to merge various data related to given firms coming from tax files, customs and trade files, and other files (like electricity companies' files, etc.). In these cases, the merging must be done at the basic unit, that is, at the elementary level, even if one wants only aggregated data from these elementary data. It is essential, therefore, that the basic units be common to all the files considered. For instance, if the basic unit is a firm and if a firm includes all its establishments in the first

file, while only inland establishments are included in the second one, no merging is possible.

All essential units, such as establishments, firms, persons, dwellings, equipment, etc., must be identified by the same identifying numbers in any file.

The preliminary operation in making elementary files data available is to assign to a given unit the same identification number in all files. These numbers are assigned at the time of the birth (or of the change) of the unit and recorded in a register, that is a file where the invariable characteristics of each unit are stored together with its identification number. For example:

- register of persons:
  - identification number
  - date and place of birth
  - name and identification number of the parents;
- register of establishments:
  - identification number (which automatically indicates the firm to which it belongs)
  - location (post address, for instance)
  - kind of establishment (plant, offices, etc.);
- registers of the blocks (in a city):
  - identification number
  - streets which delimit the block and numbers of the houses along each street in the block
  - co-ordinates of the block nodes.

There are two kinds of files, to be distinguished according to the kind of unit involved:

- economic units (persons, establishments, dwellings);
- geographic units (cities, districts, blocks).

The files with geographic units are called geographic files. The problems of correspondence between economic files and geographic files, that is, the problem of geographic location of economic units, are studied below.

Once the problem of identification number is out of the way, decisions have to be made as to the techniques of data storage:

- Elementary computerized data:
  - identifying with an identification number
  - change in the format
- Elementary not computerized data - choice among the various techniques of computerizing:
  - punched cards machines
  - data recorder on tape
  - optical scanner
- Aggregated data - choice among various storage techniques:
  - microfilm for the data which are not to be processed
  - computerizing for the data to be processed, such as data concerning all the units in an individual region, city or block, which is to be up-dated and/or compared among different regions, cities, etc.

B.2.2 Second preliminary operations: administrative and political problems

Operations on information involve actions on the sources of the information. It is not always easy to get data from an administration for another administration. The confidentiality problems provide an often pursuasive plea for not giving the required data. To effect a desired change of the characteristics of data and even a change in the administration procedures to generate the data is still more difficult.

This problem is studied later, along with other political problems of a data bank (see III.C).

The operations at this stage involve:

- modification of existing administrative forms to obtain new data required by planning decisions of the users of the data bank: e.g. modification of income tax forms to get data on household equipment;
- recording of administrative actions by means of new forms which did not exist before (e.g. making it obligatory to fill in a form for every person leaving the country);
- creation of a new administrative action to record required data (e.g. recording of the sales of land and dwellings).

B.2.3 Generation of supplementary data: censuses and surveys

Censuses and surveys are to be avoided as far as possible, because of the high cost and of the lack of automatic up-dating. However, it is necessary to resort to a census or a survey in the following two cases:

- Important field in which no data are available (e.g. population and employment if there is no population census);
- Elaboration of the data base in a field where administrative files provide periodical up-dating, but where the base does not exist (e.g. if administrative actions provide every month a recording of newly-constructed buildings, it is necessary to make a census of existing buildings at one given moment to get an up-dated building file).

If some of the required information belongs to one of these two cases, a census or a survey will be planned and their cost evaluated:

- a census if it is necessary to elaborate a complete base for a file system (building files, population files, etc.);
- a survey if it is sufficient to get mean characteristics (survey on the consumption behaviour of households).4/

Sometimes, a desired survey may be conducted by partial processing of the existing census data which contains needed information. For example, if the questionnaire of the population census includes questions on household equipment, the processing of a random sample will be sufficient to get satisfactory mean data on the pattern of equipment holding in various classes of household.

C. Information systems of the data bank

The data bank is made, as seen above, of information systems including:

- the data
- the means of up-dating, retrieving and processing the data.

The systems must be able to change, that is, as the mass of data and the budget increase, they must be developed progressively and become more efficient. Of course, the starting point must not be complex and sophisticated, but a simple system making the immediately available data quickly usable.

4/ For instance: a survey provides data on the household consumption of each product in relation to the socio-professional category of the households. The population census gives the socio-professional structure of the population of any area. These two sets of data are sufficient to provide an acceptable estimate of domestic consumption in any area.

The systems described below are the most general systems of any data bank. The different stages which lead to a complete data bank, and which may be spread over many years, will be considered in Chapter III.

The data to be stored in a data bank may be divided, as seen above, into two categories:

- Mass data, to be up-dated and processed frequently. The storage of these data must allow easy processing: magnetic tape or discs to be processed with a computer, possibly punched cards to begin with;
- Other data, such as statistical tables, economic accounts, models, for which the most important problem is not processing but retrieval (for instance, retrieval of any information related to a given field).

These data are to be stored, if possible, in microfilm, and must be described in a documentation system.

Thus, a data bank is made of many information systems:

- mass information systems from among which must be separated:
  - elementary files systems;
  - summary files systems;
- aggregated data systems, including a documentation retrieval system.

C.1 Mass information systems

The mass data to be stored in a data bank intended for all decision makers are the following:

- data relating to persons: from these data can be derived data on working persons, employment, disposable incomes, etc. The data relating to households are summarized data which belong to this category.
- data relating to establishments: the data on the establishments of a given sector (sectoral data) belong to this category, as well as data relating to firms (that is, all the establishments of each firm), which are summarised data;
- data relating to buildings and public works: e.g. data on dwellings, on roads, on networks, etc.

C.1.1 Elementary files

An elementary file contains data on the indivisible element of a given category (a person, an establishment, a dwelling, etc.). The merit of an elementary file system lies in the possibility of merging data from different sources related to the same element.

As noted in the preceding section, the various data related to the same element (e.g. an establishment) can be merged only if this element is identified by the same identifying number in the various files. This is why it is necessary to build a register including the invariable characteristics of each element and its identification number.

The various elementary files and the means of up-dating, retrieving and processing them constitute the elementary data systems.

N.B.1 Some of these systems may be partial: for instance, students information system, industrial establishments information system, dwellings information system.

N.B.2 There may be two kinds of processing of these files:
- retrieval and extracting of elementary data (data relating to one person, one establishment, etc.);
- tabulation, providing statistical tables.

As a general rule, even if the laws of the country forbid the disclosure of elementary data, it is necessary to have such files within the data bank, in order to provide users with the tabulations which fit their needs exactly, and also because the up-dating is generally at the element level. Under the confidentiality rule, the only data to be given to decision makers will be statistical data. However, the problem of confidentiality must be solved technically by installing appropriate keys for controlling access to the elementary data. For example, certain administrations, such as police and tax administration, must not have access to the elementary files.

C.1.2 Summary files

The elementary data can be aggregated into summary data, that is, data relating to grouped units such as households file derived out of persons file, firms file out of establishments file, etc. Many

administrative files in fact relate to aggregate units: tax files on firms, registration files on households, etc.).

C.1.3 Geographic location, correspondence files and geographic summary files

Geographic location is a characteristic of the units of a file. This characteristic plays a particular part, especially as regards regional and urban planning and location problems.

This is why it must be emphasized that:

- every non-movable unit (establishment, dwelling, etc.) in an elementary file must have a location identifier (address, block number, etc.);
- it is essential to build correspondence files giving the connexion between various location identifiers (post address, block number in the city, co-ordinates, etc.).

Summary files at a given geographic level (for instance, block level, city level) are statistical files, very useful for decision making in location problems.

It must be noticed that the up-dating of geographic files and of correspondence files is very difficult, especially in a country where the pattern of land use is changing quickly - and that is often the case in developing countries. The administrative procedures to be set for the up-dating of locational records must be conceived very carefully.

C.2 Aggregated data, annex file and documentation

C.2.1 Aggregated data

The aggregated data are:

- statistics tables
- time-series data
- other data like national income accounts, input-output tables, etc.

This information is to be stored in microfilm. Each information is identified by an index number. The indexing of aggregated data according to an automatic documentation processing system is very advisable (see C.2.3).

C.2.2 Annex file

It is essential to know the exact definition, accuracy, and limits of application for each data file. All such "information on information" is included in a general annex file.

This file is made of various sub-files relating to various aggregated data. Each sub-file of annex information on given aggregated data is identified, in the general annex file, by the same index number that is used to identify the aggregated data itself. The annex file includes, also, information on the elementary and summary files which bear upon given aggregated data.

C.2.3 Documentation

Reports, maps, graphics, etc. which may be useful to the users of the data bank may be stored, too, in the bank. They may be stored in microfilm and their description included in a document retrieval system, which will constitute the automatic documentation processing system.

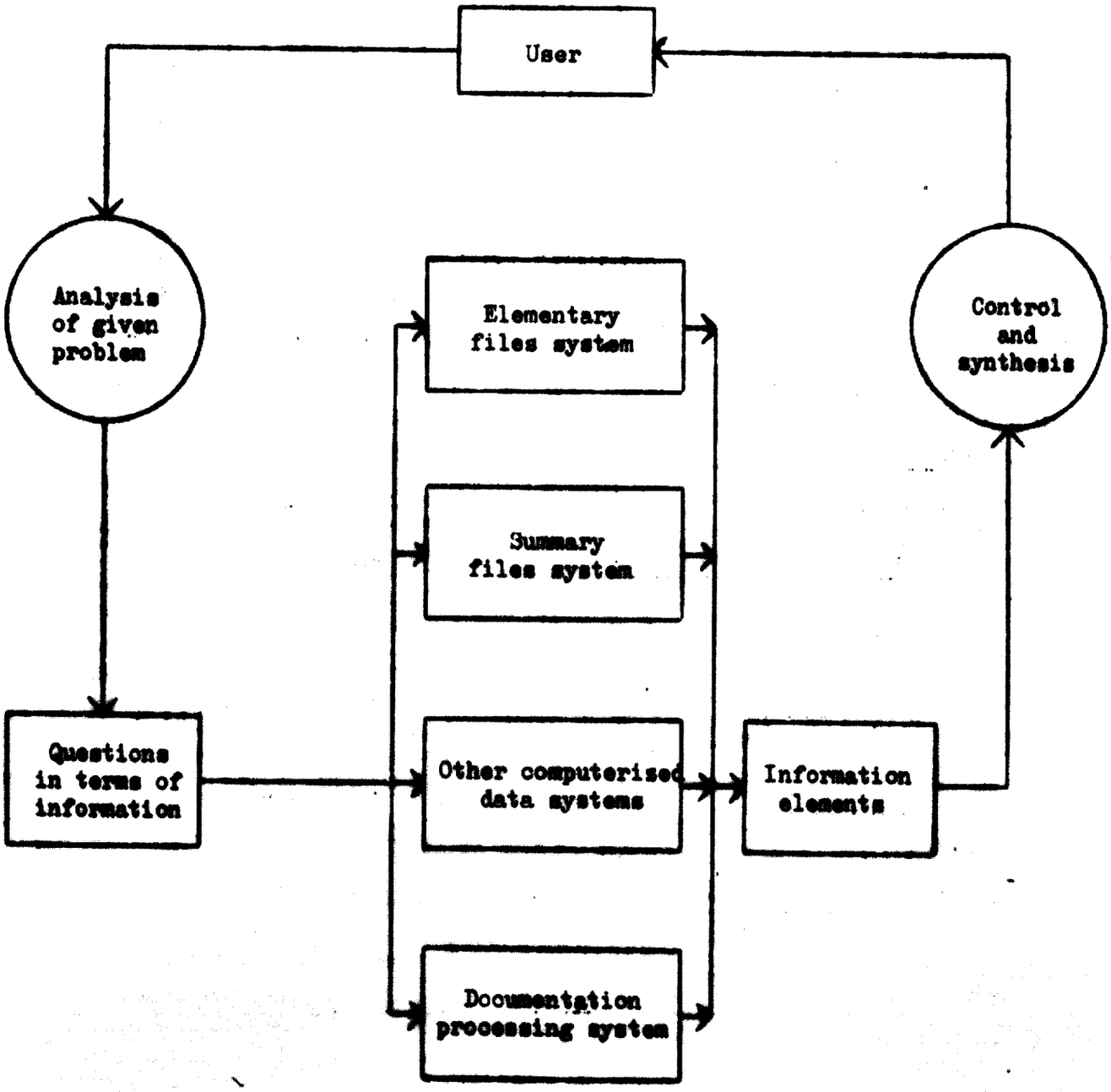
C.3 Information retrieval

To identify and assemble all the data in the bank that are to be used to deal with a given problem, a documentation processing system is set up. This system is made of:

- a general description file including a description of each of the following items of the bank:
  - elementary files
  - summary files
  - aggregated data
  - documents
- a software for automatic research of the information items which may provide answers to a given question.

However, this documentation processing system is not an essential part of the data bank. It does not have to be set up in a developing country in the early phases of data bank building.

Fig. 2 : Organization of the data bank



C.4 Entries in the system

The elementary files and summary files are up-dated by means of administrative files. The best way is for each administration to send its files to the bank periodically. Two cases arise:

- The administrative files are readily convertible into the systems' files (with a change in the format if necessary);
- The data from administration files are used to up-date the systems' files. This up-dating involves the use of the registers and the correspondence files.

When the files of the bank are computerized and the administration files are not, the computerization of the data for up-dating the bank's files is to be done by the bank. This calls for designing of appropriate entry routines (punched cards, data recording on tape, optical scanner, etc.).

C.5. Software

The software of the bank is made of the following various modules:

C.5.1 Up-dating of elementary files out of administrative files, registers and correspondence files (and up-dating of registers and correspondence files themselves).

C.5.2 Making of merged files out of two or more elementary files on the one hand, and registers or correspondence files on the other.

C.5.3 Making of summary files, and especially of geographic summary files.

C.5.4 Making of derived files (for instance, a file of scholars out of a persons file).

C.5.5 Interrogation of elementary and summary files and processing of the data of these files.

Many kinds of interrogations must be possible:

(a) retrieval of one unit in an elementary or summary file (e.g. characteristics of a given establishment) and calculations combining these characteristics (e.g. ratios);

(b) statistical tabulation out of an elementary file;

(c) retrieval, in an elementary or summary file, of the units meeting certain conditions;

(d) aggregation, in a summary file, of the data related to a group of units (e.g. providing data relating to a district out of the data relating to the blocks which constitute the district);

(e) various calculations and processing such as statistical analysis, factorial analysis, taxonomy, estimation of the parameters of a model, etc.

The size and number of files, the number of users and the necessity of satisfying quickly very different users, makes it impossible to pre-fix a special programme for every processing. This is why it is essential to call for new techniques whereby the computer writes by itself the required programme out of a few instructions.

The best solution is that the user writes the instructions by himself, without the help of a data processing engineer. This implies that the software should be based on a user-oriented language.

C.5.6 Edition of the results in a form fitting users' needs

- listings
- computer mapping and graphics
- magnetic tape, etc.

C.5.7 Finally, automatic retrieval of documentation.

The problems connected with data bank softwares can be very similar in different countries. Thus, it may be possible for one country to use a software written in another country. In the field of software, there is great scope for international co-operation and co-ordination, and especially developing countries will be able to benefit from such co-operation.

C.6 Hardware

Hardware must obviously be chosen in relation to the systems. For instance, the volume of the files determines the type of memories. Two main factors that affect the choice of hardware are:

- that the bank must be able to use administrative files; and
- that the use of a data bank involves tele-processing and installation of terminals, to facilitate direct use of the bank by those who are not located near the computer. Time-sharing will provide, at a sophisticated stage of data bank, a reduction of the cost in the management of terminals.

N.B.1 It is not absolutely necessary to have real time access to the files, even to geographic summary files, at least in the first stages.

N.B.2 The softwares provided by the computer companies seldom fit the data bank's needs. They are too general and too heavy. Therefore, the software available in the computer companies needs not be taken into account for the choice of hardware.

D. Continual interactions with the users

The users of the data bank, in the public or private sector, are seldom able to use directly the information systems of the bank and this for a simple reason: they are not fully aware of what the bank has and can do. It is always necessary that a team of specialists can analyze the problems set by users and translate them in terms of questions to the systems of the data bank. They should be able to help each decision maker in his choice of models for decision making - even very simple ones (indicators, standards) - and determine what data can fit the model. Therefore, these specialists must be aware of the information stored in the bank and of the possibilities open to the systems of the bank. They must be able to use these systems by themselves, without an intervention of computer engineers. Finally, they must be capable of making a choice among the data relevant to a given problem and a synthesis of these information elements for the decision maker.

It must be noted that no machine can do this work and that this team is the most important human element with which the data bank should be equipped.

The relationships with the users have two important aspects: Firstly, the systems of the data bank must fit the users' needs; and secondly, the users, who do not normally use an explicitly organised set of data in their decision making, must be incited to call upon the bank.

These two problems can be solved efficiently by organising group seminars for each specific category of users. We have already noted (Section A in this chapter) how it would be possible to conduct a case study, with a representative of each users' group, analysing the decision structure of the given user, and determine the kind of information that would improve his decisions and the effective decision models whereby the information should

be organized and processed. For this purpose, a seminar may be organized for a group of a few decision makers having similar decision problems. It is desirable to have a "case study" completed with any one representative of the group, before such a seminar is conducted.

Now, the seminar itself serves two purposes:

- First purpose: to test the case study by means of a collective psychoanalysis of the users attending the seminar, who react to the statement made by their colleague and by the team of the bank. Thus, the analysis can be amended and the case study finalized. Besides, reactions of the attendants may reveal the information needs which were not previously identified and re-examine how the decision structures can be improved when the data bank has started working.

- Second purpose: the seminar shows in a very concrete way how data has been used for a decision making purpose, and what lessons are to be learned from each such example. This second purpose is very important. In fact, when a data bank has been created, the users may still be accustomed to making their decisions without the use of data. Among all the possible marketing and advertisement actions, group seminars have proved to be the most efficient.

In organizing the seminars, decision makers are to be divided into homogeneous groups, each group representing a more or less common decision structure:

- government
- regional authorities
- local authorities
- parliament
- public and private firms
  - industry
  - distribution
  - services
  - banks
  - building and public works
- political parties
- trade unions of workers
- professional associations such as physicians, architects, etc.
- research centres and universities.

A case study should be conducted with a representative of each group, by means of a dialogue between the team of the data bank and the representative. The case study is divided into:

- analysis of main decisions
- translation in terms of data needs
- intervention of the data bank
- description of the data to be used
- decision models to be used
- the decision actually taken and the results of this decision.

Experience shows that such a seminar can be most effective if conducted with no more than ten participants, using one entire day without break (thus including perhaps a luncheon meeting).

When the first seminar covering a group has taken place, further seminars are to be organized periodically. They can later be handled by a university staff, which ought to guarantee an ever-increasing flow of professionals well acquainted with the role of the data bank in the community.

III. BUILDING A DATA BANK IN A DEVELOPING COUNTRY

A. General principles

Developing countries - especially the development authorities in these countries - need data perhaps more than those in advanced countries, since much of the development initiative should be taken on the national level and the cost of a wrong decision on such a level would be extremely high. This does not imply that every developing country should immediately build a sophisticated and expensive system, but the starting point can be very simple. All administrative actions provide data and every country has an administration (which at least collects taxes); these data may be organized and processed by means of information systems opened to various decision makers.

Thus, there is no particular threshold, in terms of level of economic development, for building a data bank in a developing country.

It must be emphasized, however, that the systems and the structure of the bank should be evolved gradually and progressively. The evolution ought to keep pace with changes of the basic economic structures (which are rather frequent in a developing country), with the development of available information, and with modifications in the decisions process of development authorities.

These changes cannot be handled in a pre-determined manner. Thus the systems should be "operational" from the outset, even at a rudimentary stage. The building of a bank must be divided into many stages, only the first of which can be really pre-programmed. At each stage, the data bank must be usable.

Both the institutions serving as sources of data on the one hand, and the users on the other hand, must be closely participating in the process of bank building from the very beginning.

As regards the sources of data, it is hardly sufficient to have simply a statutory action enforcing the communication of any administrative information to the data bank. But the technical skill required to effect the transmission of proper data makes it necessary to bring into the management of the bank the personnel of the institutions supplying data.

As regards the users, the necessity of close interdependence between the bank and its users has already been repeatedly noted in the preceding part of this report.

Finally, the data bank must not bring about any important new constraint for the participating institutions, at least during the first years. That is a condition for the bank to succeed. It is essential that the data bank does not create any opposition, for it can only function with a congruous relationship between sources and users. This principle implies:

- That if confidentiality is the rule in the country, it must be guaranteed from the very beginning; and
- That no action be started during the first years that may cause hostile reaction on the part of any important institution. New administrative forms to be filled in, new sample surveys on industrial firms, etc. ought to be avoided as much as possible at the beginning phase.

B. Point of departure - a pilot project

A small committee on the data bank should be established, including representatives of:

- Ministry of Planning
- all sources of information (governmental and non-governmental development institutions)
- Central Bureau of Statistics
- development corporations, banks, etc.,

in order to support the preliminary study to be conducted and to organise an appropriate study team. The study team should include statisticians and economists, as well as computer specialists. International experts, including a data bank expert and a planner (if the help of an international organisation is requested) may be allowed to participate.

The preliminary study is divided into three parts: Inventory of decision makers' needs, analysis of their decisions, and weighting of these decisions. This has been described already in II.A above. At this stage the study will not have to be a complete analysis of all decisions of all decision makers, but only a quick inventory of decision makers and of their main decisions.

However, the objectives of the data bank must be clearly defined at this stage. The study must include information concerning the attitude of the potential users towards the data bank concept, and also the particular users whom one can recommend to be associated with the project from the very beginning.

Inventory of data sources and analysis of available information: The main technical guideline for this work has been studied in II.B above. This study must assess the actual possibilities of co-operation between the existing sources of information and a possible data bank. This information will be taken into account in the data bank organization.

And finally, proposal for a pilot project: Indication of the types of action to be taken in order to effect a pilot project that can service those highly-weighted decisions by means of the readily-available data.

It is necessary to begin with a pilot project before setting a complete data bank. This is necessary in order to begin with something immediately operational, without entailing too much budgetary burden. The pilot project must be limited. It may have various shapes according to the nature of decision makers to be serviced, and the nature of available data to be immediately used.

Two possible categories of pilot project are:

- a sectoral data bank - for instance, a data bank on industrial establishments and investment projects (the Industry File System project, as proposed by UNIDO, is such a sectoral data bank project);
- a data bank on a limited area - for instance, a regional data bank for a pilot region, which may begin with geographic summary files (see II.C.1).

More or less readily-available data, alone, are to be considered in the pilot project.

It is possible to build a hierarchy of the needed data, according to their utility, and of the actions to make them available, in relation to their cost. But it is not possible to calculate precisely the benefit accruing to the country from the use of a data bank. The benefit will be widely spread over a variety of users and will be more palpable in long runs

than in short runs. A formal benefit-cost analysis, if well-intended, might thus produce an illusory result. Once proper weights have been established to different users, and the costs of accommodating their needs assessed, one would know what to do within a given project budget.

C. The pilot project and the preparation of a second project

The pilot project has been initiated from a quick inventory of the actually-available information, and of the main and approximatively-weighted needs of the users. It is essential that this project be quickly operational, perhaps within six months to one year after its start. The operation of this project will provide a checking of the information and decisions analysis. It will make the data bank familiar to users as well as to those who supply primary data. Gradually it will become clear if it is possible to build a complete data bank in the country and what will be the stages of this second project.

The organizational arrangements needed for operating a pilot project will be as follows:

- National Data Bank Committee as the management board

The National Data Bank Committee examines the preliminary study and, with its endorsement, a pilot project starts. An official decree may have to be issued to support a data bank even in this pilot phase. The Committee becomes the management board of the bank.

- Association of users

The users have been divided into groups for the preliminary study (analysis of decisions). These groups, which have been informed of these analyses, are instituted into Users' Commissions (one commission for each group). A council of users is created with one representative of each commission.

- Institution of the bank

Three solutions are possible:

- the data bank directly attached to a high-level national development administration (e.g. Ministry of Planning);
- the data bank as a private corporation (e.g. foundation);
- the data bank within the Central Bureau of Statistics.

For a pilot project, the third solution may prove convenient. The first solution would result in giving a privilege to the government's needs and thus lead to a "closed" data bank. A private data bank may not be able to obtain full assistance services where important sources of data are held. The Central Bureau of Statistics has specialists in information and will be readily trusted with handling problems of confidentiality. The Bureau, being traditionally an executing rather than a programming agency, might lack a creative and dynamic ability by itself and thus need a new staffing action to cope with the requirements of the data bank.

A risk to be avoided is the creation of many separate sectoral data banks. Indeed, this solution would make very difficult co-ordination in such matters as identification number. This is why the bank must be subject to an inter-administration co-ordination. The management board of the bank (formerly the National Data Bank Committee) is able to provide such co-ordination. Through this Committee, each data source will be able to control the use of the data it provides.

- Staff of the data bank

  The staff must include at least:

  - a technician from each data source, who has a good knowledge of the data provided by the institution to which he belongs and of the administrative sources of these data;
  - economists and statisticians who constitute a team which analyses the problems of the users as posed to the bank and conducts case studies as required;
  - systems analysts and data bank experts, whose work is to define the systems of the bank;
  - various technicians (computers, punched cards, etc.);
  - marketing specialists, to incite users to use the data bank;
  - an administrative staff.

As regards the use of computers, it has already been pointed out that a computer is not a necessary condition for a data bank in its beginning phase. A computer may be used only if the available data to be entered in the data bank are already computerized, and if a computer is available for the data bank's use in a civil service. In order to make the bank operational within a short period of time, it is not possible to use a new

computer for the pilot project. If there is no computer in any of the civil services, the pilot project will work only with card punching machines. If a computer is used, one should use only those softwares immediately available and already used on this computer.

When the bank has actually serviced users during six months to one year, the pilot project must be challenged. The experience may bring modifications and a second version of the pilot project may then be formulated accordingly. The second pilot project may work for one to five years.

D. Project No. 2

Project No. 2, which is the complete data bank project, will be prepared from the very beginning of the pilot project by a team which has been working on the pilot project. The preparatory study will cover the various steps as described in this report, and include a complete inventory and analysis of all the available information, as well as a complete inventory and analysis of all decisions.

Project No. 2 will be divided into stages in relation to the technical and budget possibilities.

The conclusions from the pilot project are to be taken into account in the building of project No. 2 and the latter may start by introducing the necessary alterations and amplifications in the second pilot project with due consideration of the continuity of the project.

BIBLIOGRAPHY

UNIDO - The Industry File System: Draft Report of the East African Working Party on Industrial Programming Data.

- The Register of Businesses: A first step in the development of an Industry File System (ID/WG.52/4).

UNRISD - Information systems for Regional Development Planning: Issues and Problems (by T. Hermansen).

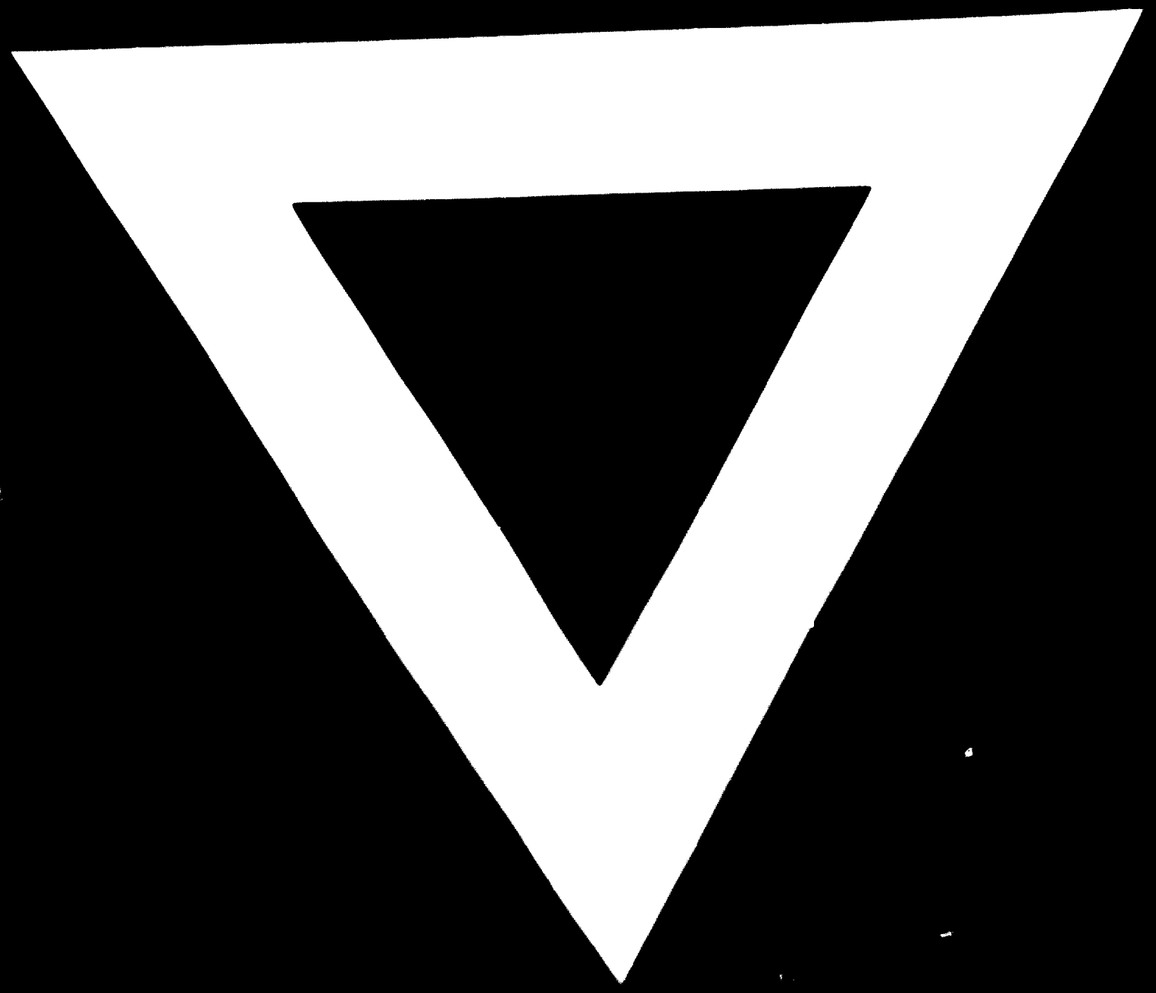
"D.B.S. Rural Geocoding - An Automated Spatial Information System for Regional Planning", by A. Terjanian, D.B.S., Ottawa, Canada.

"Banques de Données et Systemes d'Information Régionaux et Urbains aux Etats-Unis", by A. Ferragu, J. Salmona, A. Timmel (I.N.S.E.E.), D. Robequain (Ministère de l'Equipement), France.

"The New Land Data Bank in Sweden", by H. Wallner, Swedish Central Board for Real Estate Data, Stockholm, Sweden.

"Some Problems of Data Identification, Data Integration and Data Organisation in Urban and Regional Information Systems", by O. Salomonsson, National Central Bureau of Statistics, Stockholm, Sweden.

"The Role of Regional Data Banks in the Context of a System of National and Local Data Banks", by C. Leven, St. Louis, USA.



6. 10. 71