13077

8

UNITED NATIONS
ECONOMIC AND SOCIAL COUNCIL

Distr.
LIMITED

E/CN.14/CAP/18
21 August 1967

Original: ENGLISH

ECONOMIC COMMISSION FOR AFRICA
Conference of African Planners
Second session
Addis Ababa, 4-15 December 1967

# DATA COLLECTION FOR PROJECT IDENTIFICATION

M67-978

This study was prepared by Mr. Rogers S. Cannel, Vice-President, URS Corporation (San Mateo, California, USA) and Consultant to the second session, Conference of African Planners, Addis Ababa, 4 - 15 December 1967.

The views expressed are those of the author and do not necessarily reflect those of the secretariat of the Economic Commission for Africa.

DATA-COLLECTION FOR PROJECT IDENTIFICATION

Leonardo DaVinci invented an airplane, but it would not fly. He had the concept. He observed birds, and the form of his flying machine was sound. He did not have the materials, the energy source, or the means of computing the design factors.

In another field today, economic planners are passing through a comparable stage. Plans for developing countries have been designed by observing the activities in developed countries. These plans have a concept and form, but they do not "fly." They do not fly because they are more a model of another set of circumstances than a valid reflection of what can be built with the available human and material resources. They rely too much on aspirations and too little on computation. Planning from the top can provide useful guidance, but unless supported by a detailed knowledge of what is possible, the plans will lack the resources and energy "to get off the ground."

This analogy is intended to emphasize the importance of deriving plans specifically for the area, country, or region under consideration. The derivative approach requires techniques for inventorying the resources with which to build; and models for describing the alternative ways of developing resources, the infrastructure required to support the development and industries supporting infrastructural and resource development. At each step the cost of including development schemes or support projects must be weighed in economic terms against markets, competition for markets, and alternative internal investments. The Resource Allocation Method described in separate papers was prepared to show the cause and effect relationships involved in this derivative process. It is usually accepted that plans can theoretically be derived with this procedure. It is, however, frequently asserted that the data needed to use the system are unavailable and too expensive to gather. This may be true if one takes the point of view that a resource inventory is a task that must be completed on a national or regional scale before planning work on projects begins.

If, however, one considers data collection, with a common set of standards as a phase in each project, then the cost of the data base is the cost required to insure viable development investments. The cost of one major investment mistake due to inadequate resource information would amply pay for a good resource information system, and nearly every developing country has dramatic examples of project failures due to inadequate planning information. These value judgments, however, do not help the planner to know:

- How do you start a data system?
- What do data cost?
- How can you determine how much to spend on it?
- What data do you need?
- How long does it take to obtain data?
- Where can you find data?
- How do you keep the system operating?

This paper is prepared to answer these questions and suggest one way of going about the business of setting up a Resource Information System.

Building a Resource Inventory

The mental obstacle to improvement of the information available for planning is "the need for totality." The planner thinks in terms of a completed system. The scope of his immediate task appears too limited in terms of his overall objective, so he doesn't start at all. He knows that the government he works with and the external institutions that support development are activists: they want to invest in tangibles, and will spend neither the time nor the money to build a complete information bank as a first step towards project identification and planning.

If, on the other hand, the planner can view a resource information system as an evolutionary body of knowledge expanding and improving over time according to a predetermined set of standards, the system will start to build. Each task, each project will gradually add to the body of resource information.

The evolutionary approach of continued updating and expanding coverage is a part of most government operating departments. Political data affecting franchisement or allocation of services and production, or import data affecting taxation, are important to nearly every government. Comparable data for planning purposes, by contrast, is seldom collected and organized over time. This may be because a good standard system for collecting planning data is not available; it may be because planners enter a developing country, plan, and leave, bearing no long term responsibility or identification with their plan; or it may be because economic planners adept at guiding going economies with economic indicators have not recognized the need of measuring what tools and resources are available in an economy yet to be built.

Whatever the reasons for inadequacy of information required for planning, the first step in building a resource data system is to understand that a few data are better than none and that in each succeeding study, new data can be added to the resource information system. Occasionally planning studies go beyond general observation or repeating information from previous studies and undertake development of new information. In the Tanzania/Zambia Highway study literally dozens of small studies by individuals about specific resource data were found. These usually existed in one or two copies and varied widely in quality, format detail, units of measure, and filing system. The studies were largely uncataloged and known to few people. Much information of this sort was found in connection with farming operations. Raw data were virtually useless for project identification on a regional basis since they were not comparable, or pieces of information needed for interpretation were missing.

The second step in developing a resource inventory system is to develop a standard set of rules for collection and organizing data. Many types of systems could be used. The exact system is not critical, but it must be uniform and all existing data must be organized into this system and all new data collected according to it. Characteristics that are desirable in a resource system used for project identification over a large area include:

- Data should be collected to common units of area so that the data can be sorted to present all information available about specific units of area.

- Data should be coded so that all data for a given category can be collected.

- The system for filing information or identifying detail in terms of geography should permit a logical breakdown from the general to the specific as more information becomes available.

- The system should permit the design of statistical sampling approaches. This is essential to providing an early overview of project alternative at a minimum cost. The sampling procedure can guide the early data collection to the most fruitful areas and later collection to progressively less rewarding areas. For some areas, in which the sampling and visual observation indicate little potential there may never be data in the system, although the system has made provision for it. Data should not be included in the system for the sake of completeness, only area and resource information that is likely to be used.

- The system should provide an evaluation of the information included in terms of recency, accuracy, source, etc.

- The system should permit easy adaptation to a computer. This includes not only the coding of data for storing and access, but also for use in the types of computations necessary for identifying and evaluating potential projects.

- The system should be simple to use and easy to understand. It should relate in a clear way to what it represents. The rules for use should be virtually self-evident.

- The system should be easy to update and permit logical aggregation or greater detail.

If a system has these features and is made available to all potential sources and users of information, it will grow merely because it is the easiest way to work. In Zambia, for example, where the RAM system of data inventory was provided as a part of the road study, RAM is being used and extended within the ministry of agriculture; not by edict, but by professionals who find it the easiest way to work.

The final step in establishing a resource inventory system is to formally establish it as a standard and create a statistical group to be responsible for maintaining the files and information. While this step could theoretically be the first, it is not likely to happen until there is general acceptance through the use of an established system. Funds will not flow until after system usefulness is demonstrated, and a group starting a system will have a difficult time agreeing as to what the system will be. Worse, such a group is likely to be drawn from existing government statistical groups, who will want to include what is easy to count and not what is needed for planning, and who will be negative about building the data base.

## Planning Factors

Resource inventories are conducted for three general purposes:

1. Project identification (feasibility studies)
2. Project planning
3. Crop models

The data detail required and the professional training of the team conducting the inventory vary according to these purposes.

The objective of an inventory for project identification is to provide sufficient information about a region from which to evaluate alternatives within the region, determine how the areas should be developed, and conduct benefit/cost studies to provide a basis for a financial evaluation.

This type of inventory must be led by a professional in agriculture, timber, or minerals. These talents seldom exist in the same individual, so three separate inventories usually are required, of which agriculture generally is of most immediate concern. The potential for agriculture is more diverse, harder to observe without measurements, and requires more careful planning at the national level to ensure proper exploitation. Timber tends to occur in easily observable areas where the specialists can concentrate his efforts. Mineral inventories are usually conducted on a continuing basis by private companies, and government investment is not required. Thus the cost of overall regional survey is essentially the cost of collecting all information related to the production of crops. The information for all resource potential--crops, timber, and minerals; however, should be collected according to the same file systems so that the combined value of the area canbe determined if more than one potential exists--for example, the value of timber plus the subsequent value of agriculture if the timber is cut.

The kinds of information required for an assessment of agriculture potential have been identified in the "Inventory Procedures Paper." This paper describes how to estimate the cost and undertake data collection.

First, the file system must be prepared. The region is divided into 100 kilometer squares, and interesting squares are further divided into 10 kilometer squares. Data sheets required for the analysis are filed behind the maps, one sheet for each square. Preparation of the file system takes a professional about two days and a clerk needs about a week to assemble a blank data book. The size of the area does not affect the time or cost of this step very much, since most of the time is involved in assembling the maps and finding the starting point. If maps covering the area are of several scales, the time and cost bringing the maps to a common scale is a little higher since photo enlarging equipment is required. The printing cost of the data books is in direct proportion to the number of pages and is likely to equal or exceed the labor of putting the work sheets together. U.S. Costs for the map pages run about $5.00 a page for up to fifty copies whether it is a 10 kilometer or 100 kilometer square.

Data sheets can run up to $2.00 a page for fifty copies. These, of course, have to be adjusted according to the equipment available and local costs.

The author has assembled pre-field survey data books for four countries at the following costs (in U.S. dollars):

| | Grid Size | Portion of Country Covered * | Map Sheets | Data Sheets | Man-Days Preparation | Printing Costs |
|---|---|---|---|---|---|---|
| Zambia | 100 | 20/75 | 20 | 320 | 6 | $ 700 |
| Tanzania | 100 | 29/120 | 29 | 345 | 10 | 825 |
| Eastern Region, Nigeria | 10 | 67/910 | 67 | 275 | 8 | 800 |
| Ethiopia | 10 | | | | 7 | |
| Average per grid cell | | | | | .3 | $ 22 |

In each case, the file system (grid codification) was laid out for the entire country, but maps and data were included in survey books only for the areas under study. These costs vary, of course, according to organizations performing the task, and serve only as an indication of the magnitude of expenses involved.

The second step is to assemble the data already published relating to the area under study. This may be accomplished by preparing an information matrix for use as a check list and then visit each government department, professional organization, educational institution, and commercial organization that might have data.

The matrix is arranged with classes of required data appearing as column headings across the top, and each area for which the data are required listed along one side. The information requirements are derived from an analysis of the decision to be made, the study required to provide a basis for the decision, and finally the data required to make the analysis is determined.

* Ratio of cells included to the total.

The procedure is to go through available documents searching for data. Each time information is found it is recorded on the appropriate data sheet and checked off the matrix. In the Tan/Zam Highway study several hundred documents were collected. These documents provided about one-third of the information required. It took two highly skilled professionals about a week to abstract the data from the published sources and convert it to the grid cell file system. The information matrix acted then as a guide to what new information was required. From this the field work was planned and scheduled in detail.

The third step is to conduct the field survey. In the Tan/Zam Highway study, this survey included soil bearing characteristics for highway design, soil characteristics for agriculture, present land use along the routes under study, dietary habits of the people, farming practices and crop productivity.

The team for conducting this part of the work comprised two professionals. Their judgment was required to select the areas for taking the samples to ensure they were representative and that appropriate observations were made about other factors that might affect crop productivity. It is not felt that the sampling for project identification can be conducted by subprofessionals since much judgment is required.

The areas under study had passable roads. The information described above could be gathered for about 100 miles of road per day, including about three hours travel, one hour breakfast, lunch and rest breaks, and six hours collecting data. No attempt was made to collect the data to a uniform density On the contrary, much of the area was eliminated by inspection and the collection effort concentrated in areas of greater initial promise. The collection of soil samples at one location and filling out the data form to accompany it required about 15 minutes; it involved the taking of three separate borings (samples), marking the samples, taking identifying photographs, and marking the sample spot on the map.

On an average day, samples were taken at approximately 20 locations. If the farmer was interviewed another 15 minutes might be required.

The farmer would be asked questions about how much he grew, how many crops a year he got, how much he sold or traded, when and how much it rained, etc. His answers would later be cross checked with water marks in drainage channels, type of natural vegetation, published data for similar appearing areas, etc. His diet and production estimates would be checked at local markets.

In Nigeria, added checks were made using aerial photographs for comparisons of land use identifications with population counts. In principle, the data were checked much as one works a crossword puzzle. Each interrelated piece of information was cross checked to be sure the separate answers were in balance.

In total, 350 samples were taken over approximately 1,600 kilometers of road. Except in sections of road of obvious uniformity or lack of fertility, one sample was taken for each 10 kilometers. This allowed a little more than half of the sampling to be concentrated in specific areas.

The cost of chemical analyses of the samples in the United States was $15 per sample. In summary, the field survey would include the following costs, producing about 20 samples per day (in U.S. dollars).

| Item | Cost per Day |
|---|---|
| Land rover and fuel | $ 10 |
| Two professionals | 200 |
| Subsistence | 24 |
| | $ 234 |

For a broad area project identification, an average of seven samples per square are required, so on the average it costs about $80 to inventory a 100 kilometer square, plus about $100 for chemical analysis or a total of $180. Note that at the rate of 20 samples per day, ample time is provided to gather general information through interview.

The fourth step in the resource inventory is to build soil maps of the area under study and take off the number of acres of land in each soil type.

The soil maps are built by analyzing all sources of information available. In the Tan/Zam and Nigeria studies, geologic survey maps and aerial photographs showing vegetation were coupled with the field survey. This approach has been verified as generally correct by follow-on studies in some of the areas by a U.S. Bureau of Reclamation team. This team took field samples over a portion of the Usangu Plain in Tanzania where the soil characteristics had been projected by comparisons and extrapolations. The resulting field measurements were quite close to what had been predicted by the sampling approach. It should be pointed out, however, that the purpose of this approach is to generally define areas and their production capability for purposes of selecting between alternatives. If two alternatives appear similar, in terms of relative advantages and disadvantages, one might want to extend the sample if it is important to demonstrate that one alternative is better than the other or it may be desirable to make the choice on other than technical grounds. In any event, the area selected would require a detailed inventory before the detailed planning and implementation could be undertaken. Since this latter survey would use the same file system, it would merely amount to taking more tests at smaller intervals.

The primary difference between the general survey for project identification and the detailed survey for project planning is in the location and concentration of samples. Where the first general survey amounted to a probing of areas to discriminate between alternatives, the detailed survey is conducted along a fixed pattern. To draw an analogy: The first survey indicates where the mountains are, the second draws the contours. For this latter purpose, the measurements should be taken in a fixed pattern at regular intervals. The points for the measurements can be designated on a map. These points might be as close as every kilometer in either direction.

For this sampling, the team can be semiprofessional (about the equivalent of a surveyor), with the major skill of map reading required. Since the points are much closer together, less time is required in travel and no time is required for interview.

It should be possible to take at least four samples per hour over an eight-hour working day or approximately 30 samples per day. The costs then would be:

| Item | Cost per Day |
|---|---|
| Land rover and fuel | $ 8 |
| Two semiskilled workers | 100 |
| Subsistence | 24 |
| | $ 132 |

Thirty samples per day at $132 would be $44 per sample, and chemical analysis per sample would be $15. At 100 samples per 10 kilometer square, the cost per square (25,000 acres) would be about $6,000.

To put this in proportion, a development scheme in East Africa to grow wheat or some other grain product over this land area could easily involve an investment of over $4,000,000. Thus the detailed data collection is less than 0.15 percent of the total cost. Clearly this is a very small investment to avoid an improper expenditure or, on the other hand, to demonstrate to a bank the wisdom of the investment.

The use of the resource inventory to build crop models is an alternative to agricultural experiment station data. It is not as good, since there is neither the same control not is a given variable always available over a wide enough range. It is, however, a quick way to determine the sensitivity of production to the various input factors. The process is described in the crop model book and will not be further treated here.

The final step of establishing a permanent resource data bank is to assemble the information into books so that it can be used by others. The books can be either permanently bound or gathered loose leaf into a binder. If the data is to serve as a record and part of the evidence to support a proposed program, it should be page numbered and bound to insure against tampering. If, on the other hand, the data is to be used to support future studies and updating or adding information is anticipated, the data should be filed in binders. The only new effort required here is to prepare the soil maps to replace the initial grid maps and summary tables for the various classes of information within the square.

A professional can prepare the soil maps at the rate of one grid cell per day; summaries, at the rate of about three cells per day. The printing costs for the soil maps could run as high as $20 per page since they are usually printed in at least two colors. Actual field data sheets may be reproduced for the book, therefore, these require no additional work. Photographs describing the area can be added four to a page at the rate of about $10 per printed page.

The system is designed to be coded for filing data in a computer. However, this step is not necessary until the system has reached a stage of high activity; containing thousands of pieces of information to which new data is added regularly, and from which current information is requested. This set of conditions is not likely to prevail in any East African country for some years to come.

## Summary

In summary, a resource inventory for project identification should be viewed not so much in terms of its absolute costs, but rather in terms of the number of viable projects it turns up, and the credibility it adds to the soundness of the projects. Projects have to be selected in terms of alternative expenditures in order to know where to put initial efforts, and they have to be evaluated in terms of their expected rate of return to attract international investors.

In the countries of East Africa, there are numerous hidden opportunities for resource development. Some are even generally known. What is not known is which are best, what investment is required, how they should be developed, and what rate of return should be expected. Based on limited experience in East Africa, it is possible to outline the return from a project identification program in £ (based on an area of twenty 100 kilometer squares).

| Item | Thousands of £ |
|---|---|
| Project indentification survey . . . . . . . | 25 |
| Value of projects identified* . . . . . . . . | 50,000 |
| Detail planning cost . . . . . . . . . . . . . | 250 |

Clearly the speculative portion (project identification survey) of building a resource inventory is a very small portion of the ultimate value. Viewed in benefit/cost terms, no step in planning has a greater return than the conduct of a resource inventory.

A systematic resource inventory will not find projects if they are not there, but it will identify many very good projects that otherwise would go unnoticed; it also will prevent investments in poor projects that otherwise might look attractive.

* Value is indicated in terms of investment required for projects with a positive benefit/cost ratio at an 8 per cent interest rate and 15-year payoff.