COMPUTERIZED DATA SYSTEMS IN THE U.S. DEPARTMENT OF AGRICULTURE

- Four computerized data systems used in USDA are described, each in one part of a four-part article. The four systems are (1) an official estimates data base for agricultural statistics, (2) a remote access system for firm enterprise data, (3) the Land and Water Resources and Economic Modeling System (LAWREMS), and (4) the Comprehensive Resource Inventory and Evaluation System (CRIES).

- Keywords: Computerized data systems, agricultural statistics, resource inventory, firm enterprise data.

AN OFFICIAL ESTIMATES DATA BASE FOR AGRICULTURAL STATISTICS

By Mervyn R. Stuckey*

Agricultural statistics for the U.S. Department of Agriculture are collected, processed, summarized, and disseminated by the Statistics unit (formerly the Statistical Reporting Service) of the Economics, Statistics, and Cooperatives Service (ESCS) (see organization chart, p. 38). The Crop Reporting Board, part of ESCS-Statistics, was formed by a USDA regulation in 1905 to:

> prepare and issue . . . the official State and National estimates of the Department relating to crop production, livestock and livestock products, . . . Among these reports shall be the Monthly Crop Report . . . (USDA Administrative Regulation Title 1, Chapter 6, Paragraph 32E).

Members of the Crop Reporting Board make national crop and livestock estimates after carefully reviewing all State indications and recommendations submitted by the 44 State Statistical Offices from the 50 States. Thus, ESCS-Statistics is responsible for all official crop and livestock estimates and it must have them readily available for users.

During the late sixties, more and more users wanted these estimates in a machine-readable form. Probability sampling was being used in more surveys, which required computer processing. As a result, the former SRS set up an automatic data processing (ADP) intern training program. Analysts evolved a concept to make all crops and livestock official estimates readily available to users. As the ADP profession began to evaluate this concept, it became clear that a data base approach would best fit this need. Such an approach was to offer several advantages:

- Estimates would only need to be stored in one place, thus be easier to maintain (for example U.S. winter wheat production estimates for July 1977 would only be stored one place in the data base, not in several files);
- Many clerical and typing errors would be eliminated;
- Machine-readable data files could be developed upon request.

In many cases, data would not need to be rekeyed into a machine-readable form; thus, data users would save time and money.

Some disadvantages emerged as the Official Estimates Data Base was being developed. New skills were needed for data base design, different security risks would be encountered and, finally, operational procedures were changed, which required a higher skill level than had been needed with a manual system.

However, the advantages appeared to outweigh the problems. Design efforts were begun around 1971, and continued in 1973. The Official Estimates Data Base (OEDB) was to be one of several data bases which would include a data base each for crops, livestock, and prices. They were to be written in System 2000, a data base management system (DBMS) software package developed by a private corporation. This system is moderately complex and experienced ADP personnel must query and update it.

The crops data base was designed first, developed and tested in 1974. It became operational in a limited way early in 1975. Eventually, the crops data base would include official estimates for field crops (acreage, yield, production and season averages prices), noncitrus and

*The author is Head, Data Administration Section, Data Services Branch, Survey Division, ESCS.

citrus fruits, plus seed crops. Currently, it covers the previous 3 years of estimates and the current year's estimates.

In late 1975, design and development of the Crop Report System began, and indications and recommendations from the State Statistical Offices began to be incorporated every month into the crops data base. Camera-ready copy began to be produced for the monthly *Crop Production* report. Field offices of the former SRS were being connected to a telecommunications network (INFONET), which provided a definite advantage over mailing the indications and recommendations from the field offices and mailing copies of the report to them.

As a result, some field offices had more time to analyze the data, and the *Crop Production* report could be released about a day earlier. Field offices were getting camera-ready copy the same day it was released in Washington, D.C., which eliminated retyping some statistical tables. Field offices place their indications and recommendations into their own libraries; D.C. operations personnel access these libraries and bring all materials together for the Crop Report System.

A vegetables data base is also operational. A separate data base for vegetables had been considered essential because of the number of estimates involved (over 100 vegetable commodities); the need to maintain a "small" crops data base which would keep the response time within reasonable limits; and the low correlation between requests for data on vegetables and on crops.

Two data bases are now "on-line"; that is, they can be readily accessed by specific users: the crops data base and the vegetables data base. Each contains estimates from January 1975 through March 1978. These range in size from 2 million bytes (characters) for the 1975-78 crops data base to 200,000 bytes for vegetables. For the latter, a data base for 1969-74 is stored off-line on magnetic tape. Also stored this way are two crops data bases for field crops estimates (1954-62); field crops and nuts (1963-74); and citrus crops (1972-74).

A data element dictionary is maintained for the crops and vegetables data bases. Currently, it is a manual system. A separate page maintained for each data element contains such information as data element name, field length, relationships to other data elements if any, and a brief definition.

All data bases are on the INFONET network and tied into its Chicago computer center. Any technical maintenance is done by the Data Base Unit, Data Administration Section, Survey Division (see organization chart, p. 38). Unit members also develop new systems which access these data bases. Members of the Crops and Prices Data Unit in the Data Administration Section handle all production runs, making updates and queries as needed to fulfill data user requests. Access to the data bases is limited primarily to members of these two units, to whom requests are routed to avoid disclosure of confidential information. Machine-readable data files can be provided on magnetic tape, either in the Systems 2000 data base format or in a fixed-field, sequential-file format.

Additional systems are planned which will access the crops data base, in such areas as grain stocks, disposition and income. Refinements will also be made to the Crop Report System.

Further information on how to get a machine-readable data file from these data bases can be obtained from the Secretary of the Crop Reporting Board, Room 0233 South Building, U.S. Department of Agriculture, Washington, D.C. 20250.

A REMOTE ACCESS SYSTEM FOR FIRM ENTERPRISE DATA: AN OVERVIEW

By Paul Fuglestad*

Working with analysts at Oklahoma State University, members of the Agricultural Policy Analysis group in the Commodity Economics Division (CED) of the Economics, Statistics, and Cooperatives Service have modified the OSU budget generator system for USDA purposes. This revision known as the Firm Enterprise Data system (FEDS), provides "enterprise cost data that would be available for research purposes including interregional competition types of supply analysis and for analysis of policy questions."[1] Beginning in 1973, CED personnel have provided clientele with agricultural cost of production estimates, by area, for major crop and livestock types. Printed budgets, updated annually, are available upon request (from CED). They have proven very useful to program evaluators and researchers.

The FEDS budget generator is a versatile set of computer routines which "provides a completely computerized system for the development, modification, updating, and comparing of budgets."[2] Previously, access to the budget generator has been difficult for those wish-

*The author is an agricultural economist in the Natural Resource Economics Division, ESCS, stationed at Iowa State University.

[1] Krenz, Ronald D. "Current Efforts at Estimation of Costs of Production in ERS." *Am. J. Agr. Econ.*, Vol. 57, Dec. 1975, p. 931.

[2] *Ibid.*, p. 932.

ing to build new budgets or to modify existing budgets, except for experienced users of the system. The system, which is cumbersome, requires a considerable amount of storage space, and it is difficult for the uninitiated to master.[3] The initiation fees are considerable in time needed by the researcher or other staff member to install and use the system. In addition, many potential users may not have the required computer facilities at their disposal. Thus, means were needed to enable access to the FEDS system from remote locations.

The Natural Resource Economics Division (NRED) has developed a "Remote Access System." It will allow entry to the FEDS computer files at Iowa State University from a remote telephone-access type terminal. The budget input data can be revised in any manner suited to the user and the revised budget can be listed on the terminal. This capability can help researchers and other staff who want to know the impact on production costs and returns for situations which differ from those assumed when the FEDS budgets were developed.

For instance, a researcher may not agree with an existing FEDS budget's product or input prices, yields, input mix, or schedule of operations. Through the FEDS Remote Access System, the researcher can tailor a budget.

This remote system was created partly in response to NRED's Pesticide Impact Analysis Group's need to evaluate the effects of proposed Environmental Protection Agency (EPA) pesticide bans on production costs and returns.[4] It is an outgrowth of a system I designed for Iowa State University Extension economists to allow field access to the OSU budget generator, also installed at ISU. For an understanding of the budget generator system and its use, see the OSU user's manual[5] and the companion FEDS users' manual.[6]

Running the system basically involves four separate phases: accessing a stored budget, revising it, submitting it for processing, and listing the results. The user responds to computer prompting to enter the requisite data. The FEDS "program" is automatically written and submitted to the operating system for processing. Upon completion, the results are retrieved and the budget is listed on the user's terminal.

In its current form, the Remote Access System is limited. For example, revised input data, monthly input summary, and labor and machinery requirements cannot be listed. Totally new budgets cannot be prepared and altered budgets cannot be stored. These additional procedures can be built into the access system—and may be if demand warrants. However, a point can be reached in which the expense and effort spent in learning and using the Remote Access System exceeds that in learning and using the FEDS system, making the former redundant. The Remote Access System offers ease of use through remote computer facilities.

THE LAND AND WATER RESOURCES AND ECONOMIC MODELING SYSTEM (LAWREMS)

By Howard Hogg and Paul Dyke*

A geographically dispersed, computer-linked data base and modeling system has resulted from efforts of a USDA task force acting on a request by the Senate Committee on Agriculture, Nutrition, and Forestry. In December 1976, the committee asked for USDA assistance in evaluation of the Department's land and water conservation programs. A digest of relevant information was transmitted to the Senate in December 1977. The task force identified four additional elements for study:

- Provide an overview of current physical, social, and institutional setting for these programs.
- Determine the effectiveness and efficiency of installed conservation practices in meeting intended objectives.
- Initiate evaluation of selected USDA programs and develop a plan for continuing evaluation.
- Coordinate, through interagency modeling team, development and maintenance of a Land and Water Resources and Economic Modeling System (LAWREMS).

The interagency LAWREMS team, formed as a result of the task force report, had three goals: (1) Promote cooperation among agencies by explaining the uses, limitations, and linkages among existing models and data systems; (2) Recommend needs for model development or modifications, along with data requirements, to permit land and water program evaluation and impact

*The authors are economists with the Natural Resource Economics Division, ESCS.

[3] The FEDS budgets for 1975, input data, and supporting software routines require 14 million bytes of storage.

[4] A second component of this activity involved development of a remote access version of the ISU-NRED linear programming system to support the pesticide program. See Fuglestad, Paul. "Using the ISUMODEL Remote Access System." Econ., Statis., Coop. Serv., U.S. Dept. Agr., unpubl.

[5] Kletke, Darrel D. *Operations Manual for the Oklahoma State University Enterprise Budget Generator.* Okla. State Univ., Agr. Expt. Sta., Res. Rpt. P-719, June 1975.

[6] U.S. Department of Agriculture. *User's Manual, Firm Enterprise Data System.* Econ. Res. Serv., July 1974.

analyses; and (3) Coordinate an effort to improve USDA and Congressional access to existing models, data systems, and proposed systems. The team did not assume responsibility for developing or operating models. Basic maintenance and control of these remains with the developer.

Several models and data systems with high potential for use in evaluation are already accessible through LAWREMS. These systems range from single-equation, site-specific models (such as those for estimating soil loss and flood damages) to the large-scale linear programming model used in the National Water Assessment. Appendix F of the *Initial Report on the Land and Conservation Program* to the U.S. Department of Agriculture contains both a preliminary inventory and classification of those models and also the related data sets which can be used in evaluating conservation practices.

The LAWREMS team expanded this inventory by compiling additional documentation and detail of these and other potential models and data sets, and they developed documentation on the accessibility of the existing models. This latter achievement was a crucial one because the systems involved are geographically scattered and serviced by many different hardware configurations. The team outlined a comprehensive system that will help fulfill the long-range analytical needs of USDA and the Congress for land and water program evaluation. Useful linkages among the inventoried models and data sets were designed and built into this system, and additional model development and data requirements were recommended.

Because the data and models incorporated in LAWREMS were not moved from their parent agency and location, an "intelligent" terminal was programmed to access various computers throughout the United States. This terminal has storage and programming capabilities which facilitate the transfer of data and model results, and it can send commands to the host computer and give instructions to the user. Run time is reduced as the user instructions are not transmitted over telephone lines. A users manual has been prepared for models and data now in the system. If current testing of LAWREMS is successful, ESCS will make it available to other agencies. This technique should provide a method of quickly consolidating results from several models into a single report.

THE COMPREHENSIVE RESOURCE INVENTORY AND EVALUATION SYSTEM

By John Putman, Larry Harrington, and T. B. Johnson*

The capacity to analyze the impacts of policies and programs on agricultural resources and their use is invaluable to developing countries, not only for resource policy analyses but also for a broad spectrum of economic analyses. Such capacity requires the development both of an inventory of agricultural resources and also of an appropriate analytical model.

Data deficiencies and limited local experience in model building hinder such development, however. Data are scarce and often of undetermined validity while trained personnel and funds for the collection of data are usually in even shorter supply in these countries. Limited experience in modeling leads these countries to import the required expertise. All too often, however, they are left with an imperfectly understood "black box" and with no capability for refinement and updating.

The Agricultural Production Potential Study Group (APPS) of the Natural Resource Economics Division (see organizational chart, p. 38) is currently constructing a Comprehensive Resource Inventory and Evaluation System (CRIES) for developing countries. Work is underway in the Dominican Republic, Nicaragua, and Costa Rica. The project is jointly funded by the U.S. Agency for International Development (AID) and USDA. Michigan State University analysts also have a major role through a cooperative agreement with ESCS.

The objectives of the study represent a merging of the interests of the two agencies:

- To provide in-country assistance to developing countries in analyzing the extent, quality, and use options of agricultural resources and in estimating the impact of alternative policies and programs (of major interest to AID);
- To develop a coordinated data concept, data system, and analytical framework adaptable to many countries and able to accumulate and transfer consistent information among countries to improve forecasts of regional and worldwide food production potential (of major interest to USDA).

The approach to resource modeling in developing countries has three main points:

- Generation and maintenance of a consistent and accurate agricultural sector data base for individual countries,
- Institutionalization of the formation system and analytical model in local decisionmaking processes,
- Development of an inventory and classification process which may collect and aggregate resource information from many countries in a conceptually consistent manner.

The basic analytical unit used is the Resource Produc-

*The authors are agricultural economists in the Natural Resource Economics Division, ESCS, stationed at Michigan State University.

tion Unit (RPU). RPU's are defined as units of land sufficiently homogeneous in agroecological factors of soil, climate, and water resources to be reliable planning units for national/regional planning. RPU's are constructed by combining soil maps based upon soil taxonomy used by USDA's Soil Conservation Service and plant life zones used by USDA's Science and Education Administration.

The RPU map is the base, but just one component, however, of the CRIES system. In this system, one can cross-tabulate resource information with land tenure, cropping pattern, political boundary, and other information.

The development of CRIES has been marked by difficulties not usually encountered in U.S. resource analysis. Examples will be given for the Dominican Republic, the first country in which we have worked. The first major problem involved estimating national crop production and harvested area. There were enough data but inconsistency existed among the wide array of estimates available. The final selection of normalized national totals was made jointly with Dominican counterparts through a consensus process based on their view of the relative reliability of the data sources.

Information on major land use and cropping patterns was sparse. Although the Dominican Republic has what may be considered a good data set for a developing country, there is no comparison with the U.S. Conservation Needs Inventory; a consistent set of series on area planted, total production, and other variables; and an agricultural census. Cropping patterns were painfully reconstructed through use of a wide variety of secondary data sources, sample information on crop location, crop-specific maps, and expert opinion.

A major obstacle in estimating national area and production totals and current land use in the Dominican Republic is the abundance of multiple-cropping and intercropping. While there is some seasonality in rainfall, the Dominican Republic does not have even a short winter; thus, a given land unit can be harvested up to three times each year. Statistical reporting of crop area is in terms of "harvested area," which must be converted to "physical area" or a "land occupation" unit for inclusion in CRIES. If 3 crops were harvested from a given acre in 1 year, "harvested area" measurements would count 3 acres of land used. "Physical area" measurements would reduce this amount back to a single acre, with a note that each crop cycle only uses one-third of an acre annually. The ideal solution would require detailed, disaggregated knowledge of multiple-cropping and intercropping patterns, knowledge not available. An interim solution was the use of regional multiple-cropping coefficients to convert aggregate annual cropland to "physical area" terms.

Further work is underway in the Dominican Republic to develop improved estimates of land use, and national area and production totals. Satellite imagery is being used to identify major land use patterns. Data newly available from a national land ownership survey may be tapped for information on cropping patterns in general, and multiple-cropping and inter-cropping in particular. Preliminary results indicate that both satellite imagery and the cadastral survey may be surprisingly helpful. In the near future, however, the Dominican area sampling frame, developed with the help of USDA statisticians, will continue to provide most of the usable data on the agricultural sector.

The analytical model used in CRIES resembles that used in the U.S. river basin planning assistance and national water assessment. The CRIES model is a least-cost linear programming model constrained by the availability of agricultural resources, especially land, and exogenously determined levels and mixes of crops.

The activities of the linear programming model used in the Dominican Republic are identified by crop, RPU, and production technique. For each activity so defined, a yield and associated cost of production is estimated. Problems arise because "average" yield estimates have no budget data directly associated with them. Further, budgets expressed in physical inputs are rare and factor prices tend to be inconsistent. Resolution of the problem of correlating yield data with appropriate budget data is being given high priority.

The "first-generation" analytical model has been installed on a Dominican computer. Analytical model refinement should follow, not precede, development of in-country capability and refinement of the data base. The choice of a relatively simple model facilitates its institutionalization by minimizing the requirements for skilled personnel, for data, and for computer capacity.

The CRIES system has been used by Dominicans to examine the characteristics and alternative uses of state sugarcane land. The real benefits of a system such as CRIES to developing countries must be measured in increased quality of agricultural sector decisionmaking through applications such as this one.