

poses to universities and research institutions. In addition, the Committee also recognized that the universities and research institutions themselves should develop a more systematic and coordinated program of data development.

This report is intended to set forth the conclusions of the Committee with respect to the problems inherent in the preservation and use of economic data collected by the Federal Government. In addition, part II will consider the problem of data development facing universities and research institutions, and make recommendations as to steps which can be taken in this area.

Part I. The Responsibilities of the Federal Government in the Preservation and Use of Economic Data

DECENTRALIZATION OF THE FEDERAL STATISTICAL SYSTEM

The statistical system of the Federal Government is highly decentralized. In contrast with many other countries, the United States does not have a central statistical office which is responsible for the recordkeeping of the Nation. Instead, each of the large number of administrative and regulatory agencies undertakes to provide much of the information which is required for its own operation. Thus, for example, the Bureau of Labor Statistics collects information on wholesale and retail prices, wage rates, employment, and a wide variety of other data relating to the role of labor in the economy. The Office of Business Economics provides data on the national income accounts, showing the progress of business activities and the functioning of the economy. The Federal Trade Commission and the Securities and Exchange Commission collect quarterly financial reports on manufacturing corporations. The Department of Health, Education, and Welfare collects the basic statistics on education and health.

In a great many instances, these statistics themselves are a byproduct of the regulatory process. Thus the Internal Revenue Service processes personal and business income tax returns and provides statistical tabulations of these returns which constitute a basic statistical source. Similarly, the Social Security Administration, in carrying out its administration of the social security program, has large bodies of information on wage and salary payments to individuals. However, one Federal agency, the Bureau of the Census, performs many of the functions normally undertaken by a central statistical office. The Census Bureau is responsible for comprehensive data on population, housing, agriculture, manufactures, retail and wholesale trade, transportation, and government bodies. This information provides other agencies with basic information about the American economy and its functioning. Thus, for example, census data provide much of the information behind the national income accounts and the detailed data about specific States and cities. To an increasing extent, the Census Bureau is undertaking special services and tabulations for other Government agencies. Certain tasks formerly undertaken by other agencies, such as the collection of foreign trade statistics and labor force surveys, have become a regular part of the census program.

RESPONSIBILITIES OF THE OFFICE OF STATISTICAL STANDARDS

The Office of Statistical Standards of the Bureau of the Budget has the function of improving, developing, and coordinating Federal statistical services. There are two specific ways in which the Office of Statistical Standards can enforce coordination and maintenance of high statistical performance. First, all survey forms which are sent out by the Federal agencies for the collection of data must have the approval of the Office of Statistical Standards. Second, since it is a part of the Bureau of the Budget, the Office of Statistical Standards participates in the review of budget requests of the various agencies for statistical activities. Both of these instruments are important, but unfortunately the task of coordination is so great that it is difficult to insure the comparability of data among various Federal agencies.

RESPONSIBILITIES OF NATIONAL ARCHIVES

The National Archives and Records Service has the responsibility for promoting improved current records, management, and disposal practices of Federal agencies, and for selecting, preserving, and making available to the Government and the public the permanently valuable noncurrent records of the Federal

Government. Before the advent of the computer, the National Archives were concerned primarily either with the basic original records or documents obtained by the Federal agencies, or with the analytic or statistical end products. The problems of intermediate worksheets and data in semiprocessed form were left largely to the discretion of the individual agencies involved. Thus, for example, with respect to the corporate tax records of the Internal Revenue Service, the National Archives has preserved in warehouses bales of tax returns filed by corporations going back to 1909. In addition, National Archives has also preserved the statistical tabulations of tax returns. With the development of machine readable data, however, it is becoming increasingly obvious that bodies of information in machine readable form which are intermediate between the original records obtained by a Federal agency and the final statistical tabulations may be more worth preserving than the original records themselves. There is a growing recognition by the National Archives of this fact. The committee was very much impressed by the active interest which the staff of the National Archives showed in this problem. However, again the problem is so vast that it may require completely new procedures and policies in the future.

IMPACT OF THE COMPUTER ON DATA PROCESSING

Data processing methods have undergone a systematic evolution which has had far-reaching implications for the Federal statistics system ever since the original punchcard equipment was introduced. Early computers were to some degree a logical extension of this punchcard equipment. Although the Univac Model I pioneered by the Bureau of the Census in the early 1950's represented a monumental step forward, it was only the modest beginning of what has turned out to be a completely new technology. Each succeeding generation of computers incorporates improvements in the size of memory, the speed of computation, and the density of data storage on tape such that the capacity and speed of operate have been increased many times over. By now the technological revolution has become so great that a reexamination of the organization of the Federal statistical system is urgently needed.

Increase in efficiency

From the outset, the computer, like other forms of automation, has reduced the amount of labor required in the processing of data. Before their introduction, a large organization of clerks and punchcard machine operators was needed to handle the huge volume of punchcards required for any substantial statistical operation. Sorting tabulating, and computing were relatively lengthy processes. Even for minimal tabulations a great many steps were requested. It is true, of course, that the computer has made necessary the development of specialists who could write programs for data processing, but once a program is written and proved out, it can be used to process large masses of information rapidly and with a small staff.

Reduction in processing time

Equally important, the time required for data processing has also been substantially reduced. Operations which formerly took 7 to 8 months to carry out now have been reduced to a matter of weeks. In the processing of the 1960 population census, the time required for certain steps was reduced from several years to several months. This shortening of time has not only meant an increase in efficiency in terms of overhead and other fixed elements in the program, but it has also resulted in making important information available more promptly. This reduction of the timelag between the collection of information and its availability greatly affects the usefulness of the information.

Improvement in data quality

The computer has also made possible new kinds of analysis which could not have been done before because of the cost and time required to carry out the necessary computations. First, it has become possible to examine and edit much more carefully than was possible heretofore. Computers can "wash" the information, and find inconsistencies which would have gone unnoticed in hand editing. Editing instructions to test the reasonableness of the basic information can be built into the processing programs. Thus, in the case of census data for manufacturing establishments, the computer can spot errors in reporting wage bills and manhours by computing average hourly earnings. Where the resulting figures are outside a reasonable range, the original information can be questioned. Other kinds of inconsistencies can be tested in a similar way, and for

each individual report the computer can make literally thousands of tests to determine which figures are out of line and which specific items should be corrected. For some Federal agencies, the ability of the computer to make such consistency checks is very important. Thus the Internal Revenue Service uses computers to check the internal consistency of items contained in each individual tax form. Such an operation is basic to one of the major administrative functions of this agency, but before the introduction of the computer it was too expensive and time consuming to be feasible. In such uses, the computer is adding a new dimension to the work and increasing the overall efficiency of the agency.

Data reduction and tabulation

Even with quite sophisticated punchcard equipment the difficulty of handling large masses of information made it imperative to reduce the information as quickly as possible to a more manageable volume. In the past this generally led to the development of a given set of tabulations, which became the final form of the data and which were all that was available to prospective users. Under these circumstances, the primary focus of attention by the producers of data was on the final published form of tabulation which was to result from the data processing. Once these tabulations were finished, there was little or no thought of utilizing the original reports for alternative analyses, since the cost and time required for additional data processing were too great. With the dramatic reduction in cost and time which the computer has yielded, however, the focus of attention is shifting to the basic information. It is now possible to use the same basic data again and again for different analytic purposes. From the point of view of analysis, the original unaggregated microinformation offer greater potential than tabulations of a more aggregative nature. Where relationships of data inherent in the basic reporting unit are important, aggregate tabulations often hide more than they illuminate.

New types of analysis

The ability of the computer to carry out detailed and complex computations on great numbers of individual cases at very high speeds has made it possible to make types of analysis which are not feasible without it. For example, prior to the introduction of the computer, aggregated tabulations of individual tax returns were used to estimate the impact which proposed changes in the tax law might have on total tax revenue and on particular classes of taxpayers. With the introduction of the computer, however, it became possible to develop a much more reliable method. A sample of 100,000 tax returns was obtained, and a computer program developed to recompute each tax return individually according to the proposed revision of the law, and thus show for the sample as a whole the exact impact of the change. This method not only provides a cumulative measurement in terms of total tax revenue, but also permits an analysis of which classes of taxpayers are affected, and by how much.

Cost of electronic data processing

Despite the very marked increase in the amount of data processing being done in Federal statistical agencies as a result of the introduction of the computer, the cost of data processing has become a smaller percentage of the total cost of obtaining information, and it represents a very small fraction of this total cost. In many cases the cost of the field survey may account for as much as 95 percent of the total cost, and processing the data less than 5 percent.

Data storage

The problem of storing basic statistical information has also been greatly reduced. A computer tape today will hold information equivalent to over 100,000 punchcards, so that a relatively small number of tapes may contain information which formerly would have occupied a great deal of space. Before the development of efficient tape storage, past data could not be kept for long periods by Government agencies, since room had to be made for the continuous inflow of new punchcards. For the first time it has now become feasible to keep the original information in machine readable form at very low cost.

Increased use of data

An important aspect following upon computer development has been the increasing use of basic computer tapes by others than the agency collecting and processing the original information. Prior to the computer, when the focus of data processing was the production of tabulations which would satisfy all users,

Federal agencies often published massive detailed tabulations which could be used by groups outside the agency for a variety of statistical purposes. With technological advances in both computers and printing, even more massive detail is being produced. As one small example, in the IRS statistics of income series the report on individuals for 1960 consisted of 165 pages, and 233 pages for 1961; on business it was 192 pages for 1960-61 and 274 pages for 1961-62. Aside from the cumbersomeness of the sheer volume of printed material, users of statistical information are now finding that the published tabulations are costly to use and often are unsuited for particular analyses which they wish to make. Even where a specific tabulation is exactly in the form desired, the user may find it necessary to put the data back into machine readable form before he can manipulate them. For these reasons, there has been an increasing tendency for Federal agencies to supply outside users with computer tapes of information to avoid the expense of recording the data. To an increasing extent, Federal agencies are considering the preservation of and accessibility to computer tapes to be a direct substitute for printed publication to make more detailed tabulations available to research users of data. Statistical programs are no longer viewed simply as projects involving the gathering, processing, and disseminating of information. Instead Federal agencies are developing the ability to tap into a source of information at one or more points in the processing stage, where data are in the form (after editing but before too much aggregation) and on the medium of recording (magnetic tape, not original schedules or printed reports) which are needed.

Interagency use of data

The ability of the computer to handle and interrelate large bodies of information has encouraged different Federal agencies to bring together information which they collect on related economic units. The recent development by the Bureau of the Census of enterprise statistics is an example. This set of information was created by linking the establishment data collected by census with corporation tax data obtained by the Internal Revenue Service. Previously, given the costs of processing and storage, only already tabulated sets of information could be brought together, and in most cases it was impossible to reconcile different sets of related data precisely. A byproduct of interagency cooperation has been an improvement in the comparability of classification systems, techniques, and methodology. In order to collate data from different sources, Federal agencies have found it necessary to use identical classification systems and to treat similar cases in a uniform manner. Information required as a basis for major legislative and executive policy decisions necessitates drawing on many kinds of data. The increasing ability of the various Federal agencies to integrate their basic data at a primary level will provide more reliable and meaningful information for policy purposes.

CURRENT PROBLEMS OF THE FEDERAL STATISTICAL SYSTEM

Although the development of the computer has solved a great many problems in the processing and handling of data, these very advances have raised problems which were not serious before, and until these problems are faced, the Federal statistical system will not reach its full potential.

Preservation of data

One of the first problems raised by the development of the computer is the preservation policy of the different Federal agencies. As already indicated, the information collected by the Federal Government represents a large investment of human and material resources both on the part of the Government in obtaining the information and on the part of the respondents in providing it. Before computers were developed the preservation of most of this information was not feasible because of the high cost of storage and the impossibility of low-cost retrieval. Now that large volumes of basic data can be kept conveniently and inexpensively in the form of computer tapes and processed at low marginal cost, the question of what should be preserved must be faced as a matter of national policy. Under the present decentralized Federal statistical system, it is extremely difficult to maintain a coherent and consistent policy with respect to the preservation of machine readable data. The various agencies are primarily responsible for day-to-day operations, and cannot give high priority to long-run considerations. There is no adequate mechanism for insuring that these agencies are following optimal policies with respect to the preservation of important

information. In view of the large number of organizations involved, it is inevitable that unless the situation is regularly reviewed by some group within the Federal Government which considers this problem to be a major responsibility, a satisfactory solution cannot be achieved. For this reason, the committee urges that the Federal Government develop procedures and mechanisms for insuring the adequate preservation of important data produced in any of the Federal agencies.

Data access

The problem of access to information is a very real one. At the suggestion of the Committee on the Preservation and Use of Data, the Bureau of the Budget and the National Archives jointly undertook a survey of machine readable data held by various Government agencies. The survey covered some 20 agencies in the Departments of Agriculture, Labor, Interior, Treasury, Commerce, and Health, Education, and Welfare, and the Board of Governors of the Federal Reserve System. Over 600 major bodies of data were listed in this preliminary survey. These data are stored on approximately 100 million punchcards and 30,000 computer tapes. The decentralized nature of the Federal statistical system makes it extremely difficult for users outside the Government, and even in other Federal agencies, to find out what data exist on various topics and how to obtain access to them. Different agencies have completely different policies with respect to access, and an outsider must know precisely whom to contact with respect to each specific kind of information. Most Federal agencies process data as an activity which is ancillary to their primary responsibilities, and therefore they find it inconvenient and costly to respond to specific requests for information which would necessarily disrupt and delay their own work. Even in the case of agencies such as the Bureau of the Census where an effort is made to respond to legitimate requests for information, it is often difficult to fit outside requests involving data processing into a work program in which the various stages of processing censuses or surveys have been carefully scheduled and timed. In other words, the present Federal statistical system is primarily geared to the production and processing of information for immediate administrative use or publication. Thus the present organization of Federal statistical operations does not lend itself to optimal use of the vast amounts of existing information, despite the fact that this use could be achieved at low data processing cost.

Development of usable data

Another major problem arising from the lack of supervision and coordination of data preservation techniques is that even where important data are involved. Federal agencies often fail to develop clean edited tapes and to provide supporting information about the data contained on the tapes. Under present circumstances, such inadequacies are quite understandable. In the processing of basic information, operating agencies are mainly concerned with achieving the results necessary for specific tabulations or given computations. It is unavoidable in this process that substantial errors will be found. Sometimes these arise from transcription or classification errors, or from errors in programming. For the purposes of the operating agency, errors can often be patched up on an ad hoc basis, but doing so leaves the original tapes with the errors in them. It is usually quite possible to correct such errors when they are known to exist. On investigation, however, the committee found that due partly to the fact that it was not necessary for immediate purposes and partly to lack of proper budget allocation agencies often neglect this task, even though its cost would not amount to more than 4 or 5 percent of the total computation cost, and in most cases would be less than 1 or 2 percent of the total budget for the project. In addition, agencies often do not provide sufficient information on the layout, classifications, and definitions of data contained in a tape. As a result, even for the agency's own purposes it becomes very difficult to go back after a few years and make use of the information, unless it happens to be in the same format and classification system employed for current data processing. The turnover of personnel within Federal agencies often make it impossible to trace back precisely what was done in the original coding of the schedules or programming. In view of these circumstances, what is needed is some system which will insure that for important data all Federal agencies will provide clean, edited data with accompanying information describing layouts, coding, and programming, so that these tapes can be served by both the agency itself and by other groups. Given the presence of day-to-day business, the shortage of funds which often occurs at the end of a project, and the priority

of other major responsibilities, Federal agencies cannot be expected to devote the required effort to the development of clean data tapes unless some specific procedures are developed to insure this result.

Disclosure

In addition to the problem of physical access to data, there is another factor which may prevent the utilization of data in their original form. A considerable portion of the information collected from individuals and businesses is obtained with the understanding that such information will be considered confidential and will not be available to other Federal agencies or anyone else. Thus, for example, in the case of the Annual Survey of Manufacturers, the data reported on the activities of manufacturing establishments constitute a confidential report to the Census Bureau and are protected by law from use by such agencies as the Internal Revenue Service for checking tax returns, or even by Congress in its investigations. It is recognized by all concerned that Federal agencies should not violate the confidentiality of their data by making them available to outside research workers or other agencies. However, it is often possible to disguise the information in such a way that specific data cannot be traced to any individual respondent. For example, the Census Bureau in the last few years has made available a sample of information on 100,000 individual households, giving considerable detail about the age, education, income ownership, occupation, etc., of the individuals in the household. In this sample the omission of detailed geographic information makes it impossible to trace the data to any specific individual. By using a similar approach the Internal Revenue Service developed a sample of 100,000 personal income tax returns. As in the case of the Census sample, data on individuals were provided without disclosing information that could be traced to any particular individual. For other types of data, the problem is somewhat more difficult. Thus, for example, merely indicating the size and industry of a manufacturing plant may be enough to identify it and so constitute disclosure, even if no additional identifying information is given. In many instances, however, there are ways in which such information can be utilized without disclosure. Since for most research purposes it is not necessary to present information on individual cases in the final results, it is often possible for researchers to provide compute programs which can be used directly upon the basic data under the auspices of the Federal agency responsible. Again, however, few Federal agencies are in a position to take the time and trouble to fill out such individual requests, even in those cases where the research would be valuable and outside financing is available.

CONCLUSION

In summary, therefore, because of the decentralized nature of the Federal statistical system and the pressure of the primary functions of the agencies, neither outside scholars nor Federal agencies are able to utilize efficiently the large amount of information which has been obtained at public expense.

PROPOSAL FOR THE ESTABLISHMENT OF A FEDERAL DATA CENTER

For the reasons outlined above, the Committee on the Preservation and Use of Economic Data urges that a Federal Data Center be established by the Federal Government to preserve and make available to both Federal agencies and non-Government users basic statistical data originating in all Federal agencies.

Need for interagency authority

The first and most basic requirement of a Federal Data Center is that it should have the authority to obtain computer tapes produced by other Federal agencies. The exact timing of the receipt by the Federal Data Center of such tapes will differ from agency to agency, and will depend on the kind of information involved. As a general rule, however, the Federal Data Center should obtain copies of the data when a clean, edited tape of the basic information first becomes available. Fortunately, because of the nature of computer processing, duplicate copies of the basic computer tapes can be produced at low cost, so that both the agency concerned and the Federal Data Center can simultaneously have the basic information available to them. In this connection the Federal Data Center should keep track of statistical projects underway in the Federal Government and make sure in advance that the budget for each project includes the proper provision for making clean, edited tapes and providing the necessary accom-

panying information on classification and programing. It should be the task of the Federal Data Center to follow statistical projects and to see that the clean, edited tapes are made available within a reasonable period.

Need for computer capability

The Federal Data Center will require substantial computer capability if it is to provide access to information by outside users and by other Federal agencies. It is important that the Federal Data Center should not only furnish basic information but also, on a reimbursable basis, it should make production runs and furnish aggregated tapes or results to scholars so as to eliminate many problems of disclosure. In a great many instances the Federal Data Center will find it advisable to develop new tapes combining information from various bodies of material produced by different Federal agencies. For example, the very considerable interest in data on specific regions or cities by State and local governments for programs such as urban redevelopment, welfare, and education, makes it desirable to combine various kinds of information pertaining to a specific area on a systematic basis. In many cases such information about communities and their characteristics does not violate any disclosure rules. These data are useful not only for purposes of public policy but also to business groups interested in market research and in planning longrun investment. It is important that the Federal Data Center be staffed with computer analysts who are subject specialists so that they can understand the nature of the data with which they work and can anticipate the analytical problems of the agencies and research organizations that want to use the data.

Need for service facilities

A Federal Data Center would provide servicing facilities, so that Federal agencies and individuals could obtain specific information directly, and it should publish descriptions of the data available. In this sense the Federal Data Center would serve somewhat the same role as the Library of Congress, inasmuch as it would be responsible for providing a systematic and comprehensive coverage of the material available in its areas of competence. It would also, of course, be serving the same function in the statistical area as Archives now does in the area of basic records and documents. It would insure that the most useful information was preserved in a usable form, and that duplicative and unwanted data did not clog the system. Finally, the Federal Data Center would provide basic information about the American economy as a primary objective rather than as a byproduct of the administration or regulatory function.

Need for new administrative arrangements

Although the functions described above for the Federal Data Center are in part covered by the activities of existing Federal agencies, no single agency is currently combining all of these necessary functions. The Office of Statistical Standards of the Budget Bureau does have the responsibility for the supervision and coordination of Government statistical activities, but it is not an operating agency. The National Archives also has interagency authority, but it has not been involved in the field of data processing and does not as currently organized have the ability or authority to undertake the task of selecting, monitoring, and controlling machine-readable data on the scale required. Finally, other statistical agencies of the Federal Government have the ability to handle, process, and combine masses of statistical data in an imaginative and productive manner, but these agencies lack interagency authority to obtain each other's records. Furthermore, although these agencies have been making an effort to provide reasonable access to their data the fact remains that they have major responsibilities for collecting and processing basic information on a continuing basis, and these responsibilities, which have first priority, make it difficult for them to devote adequate attention to individual requests.

Need for early and positive action

In view of these considerations, the committee concludes that immediate action should be taken by the Federal Government to establish a Federal Data Center and to insure the orderly preservation of important data. The Bureau of the Budget has been given the responsibility of developing programs and issuing regulations and orders for the improved gathering, compiling, analyzing, publishing, and disseminating of statistical information for any purpose by the various agencies in the executive branch of the Federal Government (see sec. 1 of Executive Order 10253, June 11, 1951). The committee therefore urges that

the Bureau of the Budget immediately take steps to establish a Federal Data Center which would have the functions described above. It should be recognized, furthermore, that the nature of such a data center is so different from anything now in existence that it may well require additional legislative authority so that its responsibilities can be well defined and recognized by all Federal agencies. It is very important that the Federal Data Center be conceived as a new and independent function, rather than an extension of present activities by any single Federal agency which has major responsibilities of another kind. In the development of the Federal Data Center it is to be expected that the Bureau of the Budget would consult with the various Federal statistical agencies involved with policymaking groups within the Federal Government such as the Council of Economic Advisers, and with congressional groups such as the Joint Economic Committee, as well as with research institutions and universities. It is to be hoped that this planning and preparatory work can begin immediately.

In addition to the early development of a Federal Data Center, the committee urges that the Bureau of the Budget place increased emphasis on the systematic preservation of important data by those agencies engaging in statistical programs. Specifically, the Bureau of the Budget should see that funds are budgeted for the development of clean tapes of important data together with the supplementary material required for their interpretation. The subsequent review by the Bureau of the Budget of ongoing statistical programs should make certain that the important data are in fact preserved in usable form. These procedures will be necessary even after the Federal Data Center is established, and they can be initiated immediately.

Finally, as an emergency stopgap measure, the Bureau of the Budget should undertake a current evaluation of the preservation policies of the various Federal agencies and together with the agencies make a joint determination of what sets of data should be preserved, and in some cases how these data can be put into a more usable form. In connection with this, it is also suggested that the Federal Government undertake to collect and publish at regular intervals an inventory of machine readable data held by the various agencies.

Part II. The Role of Research Institutions and Universities in the Preservation and Use of Economic Data

USE OF DATA IN ECONOMIC RESEARCH

Economic research has undergone striking changes during the last decade, due mainly to the advent of the computer. However, the present organization of the profession and its lack of access to major data sources impose serious obstacles in the way of optimal use of this new research development.

Research techniques and their development

Prior to the development of the computer, empirical research in economics was largely confined to the use of aggregative economic data in fairly simple models. Price indexes, production indexes, national income accounting, and industry statistics were used not only as frameworks for classifying information, but also as a means of data reduction. The limited capability of economists to process information forces them to deal with aggregations, which often obscured interrelationships among basic variables. With the development of the computer, however, low-cost data processing has been made available to economists, and as a result for many types of economic problems research technology has undergone substantial change. Economists can now specify and develop sets of data which are tailored to the research which they are undertaking. They can also process large quantities of data on a case-by-case basis, so that complex interrelationships can be studied at a microeconomic level. The use of simulation techniques on a large scale makes it possible to test the sensitivity of models to different assumptions, and to variations in specific parameters. For the first time, it has become possible to make use of the large bodies of existing information, which can be quite powerful in testing as well as suggesting theoretical hypotheses.

Research projects and individual research

There are currently in research institutions and universities many research projects each of which involves a large number of scholars. The Brookings-SSRC model of the U.S. economy, the Harvard economic research project on input-output studies, the simulation studies at the University of Wisconsin, and the

research on consumer behavior at the University of Michigan are all examples of large-scale projects using large bodies of data processed on high-speed computers. Other institutions where computer facilities exist are also carrying out research of this type. The value and productivity of this research depend in large measure on the character and quality of data available. It is not only major research projects carried out by teams of scholars that have changed, however; individual research by specialized scholars working in a particular area has also been affected. In many universities and research institutions, there is no more than one economist for a given specialty, and for this reason he must do his research as an individual scholar. It is still true that many economists engage in research on an individual basis, but where before the computer the cost of processing data and making computations was beyond the resources available to the individual scholar, today this is not as true. The existence of bodies of data and the computer is extending the horizon of such scholars and is placing in their hands powerful research tools. An increasing number of substantial and valuable research projects is being undertaken because information is available on a highly disaggregated basis in machine readable form.

ACCESS TO DATA BY ECONOMISTS

The use of the computer as a basic tool in empirical economic research does, of course, require that there exist bodies of suitable data in machine readable form. Without appropriate data, the economist with a computer would be in the same position as a biologist with a powerful microscope but no biological specimens. With limited or inferior data he will be constrained to results of limited usefulness or doubtful reliability.

Large-scale research projects

For the most part, large-scale economic research projects have a considerable advantage in obtaining the kind of information they need. However, even in these cases, the committee has found that the situation is far from satisfactory. Federal agencies are not organized to provide data, and therefore delays and administrative difficulties may make it impossible to obtain the desired information. The problem of disclosure of basic information poses additional difficulties, and Federal agencies may use these difficulties as a convenient excuse at times when they regard themselves as fully preoccupied with their own problems, although devices could be worked out to safeguard the confidentiality of the data. Where cooperation is required between two Federal agencies for the development of interrelated data, the difficulties are generally so great that research institutions hesitate to undertake the task.

Individual research

The problems facing the individual research worker are many times greater than those faced by large-scale projects. First, it is often quite difficult for an individual to find out what information exists and what form it is in. Second, making arrangements with Federal agencies often requires substantial time and effort, and usually agencies are not receptive to the individual scholar unless he is well known. The cost of having the Government prepare data in a form suitable for research purposes is very high indeed, because it must be done on a special ad hoc basis which disrupts the agency's operations. For these reasons the individual researcher is usually not in a position to obtain specially developed bodies of material. However, tapes of standard or multipurpose information specifically designed to be sold for research purposes can be developed. As one example, the 1-in-1,000 sample of the population census prepared by the Census Bureau has provided many universities and research institutions with a set of basic information which can be used in a large variety of research projects. Over the long run, the individual research scholar may have to come to depend upon such standard bodies of data much in the same way as he previously depended upon published tabulations.

DATA ACCESS FROM THE POINT OF VIEW OF THE FEDERAL GOVERNMENT

As has already been indicated, the various agencies of the Federal Government have administrative and regulatory responsibilities which constitute their major functions, and the production of statistical information and the data underlying it is usually ancillary to these major functions. Demands for data by a large number of organizations, including not only research economists but

also State and local government groups, businesses, and other Federal agencies, often place a severe and unwanted burden upon data-processing facilities and the time and energy of specialized personnel. Even when such work is done on a reimbursable basis, limitations of staff due to overall personnel and budgetary considerations and the ability to hire people making the filling of special requests a burden. Outside requests for data are often uninformed, unreasonable, and in view of the Federal agency, not worth while. Few outsiders can know enough about the data, their nature and characteristics to make sensible requests, or to have a realistic appreciation of the analytic limitations which the data impose. As already indicated, the disclosure problem is formidable and causes considerable uneasiness on the part of the responsible people in the data-producing agencies, but it also may be very useful as a shield to protect them from the nuisance of dealing with individual requests. The problem, as seen by the Federal Government as a whole when contemplating a request for data, could be reduced if research workers asking for data could get together and coordinate their requests.

ECONOMIES OF SCALE AND THE NEED FOR COORDINATION

Many of the requests for basic information on a specific subject by different research scholars are duplicative. However, since each research project will be designed in somewhat different terms and has different objectives, it is inevitable that the independent requests for information will not be identical. On the other hand, it is also quite possible that, if careful consideration were given to the matter, general master tapes might be designed in specific areas which would meet the needs of a large number of research projects. One of the difficulties with published information is that different research workers want different types of aggregations and classifications. Since it is now possible to provide data on a disaggregated basis, these differences are no longer relevant, and it becomes necessary only to specify the basic items of information to be included in the body of data.

This basic similarity in the demand for information on a given topic implies that considerable economies of scale could be achieved by coordination. If a single master tape would fully satisfy the demands of each user, designing special tapes for each user would be unnecessary. For this reason, the committee has undertaken a preliminary survey of 10 major areas of economic data to see whether or not it would be possible to construct sets of such basic tapes in these areas. On the basis of this examination it is the committee's considered conclusion that this construction not only is feasible from the point of view of economic research needs and objectives, but also would go a long way toward improving access to major bodies of data for scholars, and toward reducing the costs and alleviating the burden placed on the Federal statistical agencies.

DEVELOPMENT OF TAPES FOR SPECIFIC RESEARCH AREAS

The committee circulated to a group of research scholars working in various areas copies of the preliminary inventory of machine readable data recently collected by the Bureau of the Budget and the National Archives. In a large number of cases, these scholars prepared suggestions as to bodies of data currently in existence, which should be developed and made available to universities and research institutions on a low-cost basis. Included among the suggested data files are some which are currently available to research scholars and which certainly should remain available. For example, the Bureau of the Census has developed a program of making available for purchase large bodies of unpublished data in the form of computer tapes. The Internal Revenue Service and the Bureau of Labor Statistics have also developed specific tapes for sale. However, even where unpublished material is available on computer tape, it is often not in a form which is directly useful to the research worker. Data reduction to prevent disclosure or to select a manageable sample of data may be necessary. It would also be very useful if the research community could be better informed about what tapes exist in the various Federal agencies, and the cost of obtaining them.

A number of different data characteristics which are important for research purposes have been mentioned by research workers. (1) They point out that sets of data that are continuous over time are particularly valuable. This is especially true where information relating to a specific reporting unit is obtained at regular intervals so that changes taking place at the micro level could be ob-

served. (2) Even where continuous reporting by individual units is not available, sets of information for different periods which permit cross section analysis are very useful for research purposes since they permit examination of changes in structural characteristics and behavior. (3) It is emphasized that sets of data covering a wide range of items for a single reporting unit are more valuable for many purposes than larger sets of information which report on a smaller number of variables. In a great many cases it is the interrelationships among variables at the individual reporting unit level that are important for research purposes. (4) Sets of information which it is possible to match with other kinds of information are particularly important, even where the information contained in such sets of data may be quite narrow. Thus, for example, if a set of data tapes includes a social security number or some other identifying characteristic which would permit matching with similarly identified collateral information from another source, the tape is that much broader in its coverage. (5) Many sets of data are useful not because they are in themselves unique bodies of specialized information, but because they are already in a machine language and are capable of being manipulated at low cost, so that it is often easier and cheaper to use them than to have recourse to data already in tabulated form.

PROPOSAL FOR AN ORGANIZATION ON ECONOMIC DATA

In view of the increasing importance and usefulness of machine readable data for economic research, the committee recommends that economic research institutions and universities develop an organization to coordinate the requests by research scholars for economic data and to aid the Federal Government in the development of data for research purposes.

Usefulness of Federal economic data

The inventory of machine readable data held by the various agencies of the Federal Government and the results of examination of the inventory by research scholars in various areas has convinced the committee of the potential usefulness of such information for economic research. Out of the 600 items listed in the inventory, over 75 bodies of data can be identified as of prime importance for general research in the 10 areas listed. There is no doubt that this list would be considerably expanded if the scholars consulted had had available more detail on the exact contents of the different bodies of data. Furthermore this preliminary inventory was far from complete in its coverage of Federal agencies. Since there is such a large body of highly useful data, therefore, the committee believes that research institutions and universities should encourage the Federal Government to undertake the establishment of a Federal Data Center.

Need for coordination

There is a substantial and growing demand from scholars in research institutions and universities for bodies of machine readable data held by the Federal Government. These demands are highly duplicative in nature, but completely uncoordinated. It is quite likely that in many instances it would be possible to obtain agreement from scholars working in a given research area as to what sets of information would be most useful if developed by the Federal Government. Such sets of information would satisfy the needs of many research analysts, so that Federal agencies would not be faced with many different requests. From the point of view of the community of research scholars, there would be considerable advantages in providing a clearinghouse for information concerning economic data, since it is so difficult for the individual research scholar to discover what information exists in the different Federal agencies, who should be contacted, and how problems relating to the confidentiality of data may be solved.

Need for data development

It is not sufficient, however, merely to provide a clearinghouse and to coordinate individual demands for data. In a great many instances the research community should take an active role in advising the Federal Government how to develop and exploit a given body of economic data. The existence of a body of information can often stimulate valuable research activity. Thus, for example, the 1-in-1,000 sample of the population census was not developed as a response to specific research demand by scholars outside of the Government, but rather it was developed by the Census Bureau because they recognized the potential worth of this type of data. Those scholars who were consulted about specific research areas, furthermore, emphasized the need to integrate the different

bodies of data collected by different agencies. Although the Federal Government is continually taking steps to improve the comparability of classifications used by the different agencies, the task of integrating bodies of data is still a formidable one requiring substantial effort. For these reasons, research institutions and universities should actively participate with the Federal Government in planning the development of economic data in specific areas.

Need for a coordinating organization

In view of these specific tasks facing economic research institutions and universities, it is important to establish some continuing organization. The committee recommends that institutions outside the Government which have a sizable staff engaged in economic research involving the computer processing of large bodies of machine readable data join to form a coordinating organization on economic data. The function of such an organization would include the servicing individual research requests for economic data by providing a clearinghouse and information about the availability of data. Second, the organization should undertake an active program of data development in conjunction with the Federal Government.

Organizational structure

Although the organization might have a larger membership which would draw on its clearinghouse and information services, it would be desirable to establish an executive committee so that periodic working meetings could be held to determine matters of policy. If there is to be continuity in the organization, furthermore, there will have to be a permanent secretariat which can function on a day-to-day basis. In view of the importance of the Federal Government as a data source, it is recommended that this secretariat be located in Washington. Finally, it is also recommended that the proposed organization develop working subcommittees of scholars concerned with specific subject matter areas so they can advise the Federal Government on data development and the establishment of procedures for coordinating demands for data.

Need for early and positive action

The committee urges that at an early date the Social Science Research Council convene representatives from research institutions and universities currently engaged in research projects involving the use of empirical information, in order to develop an organization which can coordinate requests for economic data. The group which is convened should give specific consideration to (1) how the research interests of all nonprofit research organizations and universities can be facilitated; (2) what kinds of services can be provided for nonprofit research institutions and universities; (3) what kinds of coordination are considered to be desirable; (4) how the proposed organization is to be established, staffed, and financed; (5) in what way the proposed organization can assist the Federal Government in the establishment of a Federal Data Center; and, (6) in what way the proposed organization can provide the Federal Government with advice concerning the preservation and development of basic data.

The formation of a coordinating organization should not, however, be delayed until solutions are found to all of these questions. There is an urgent need for an organized group with staff support to follow through on the problems outlined in this report. Such a group would be useful to the Bureau of the Budget in carrying out the suggestions contained in part I of this report. Further delay may result in the loss of valuable data which could be saved by prompt action. Furthermore, in order to provide for the orderly flow of data in its most useful form 2 or 3 years hence, steps must be taken now to establish procedures for projects which are already in their formative stages and which, unless properly conceived, may in 2 or 3 years time present the same sort of problems which are now encountered. Finally, the very rapid growth of research needs and the large quantity of machine-readable data generated tend to produce a large number of ad hoc solutions which will make future coordination more difficult. Adequate consideration of how to meet the needs of various groups in the immediate future may forestall the development of inappropriate partial solutions.

APPENDIX

PRELIMINARY EVALUATION OF PUNCHCARD AND COMPUTER TAPES OF ECONOMIC DATA HELD BY FEDERAL AGENCIES

Although it has not been possible to compile a comprehensive listing of the comments received from research scholars, the committee has made a summary listing of some of the punchcards and tapes mentioned by research workers in specific areas. In some cases, the items discussed include tapes which can be purchased from Government agencies as well as those which are currently unavailable. In a few cases, sets of information not included in the inventory of tapes were also mentioned. The numbers in parentheses which are cited refer to the inventory numbers given in the appended listing.

1. POPULATION

The census of population data are not only basic to the study of demography, but also provide valuable information on individuals and households necessary for research on housing, employment, education, health, and consumer behavior.

Census of population samples

The Bureau of the Census has prepared 1:1,000 and 1:10,000 samples of the census of population (41-A-12 and 41-A-13) on both punchcards and computer tapes which are available for purchase. These bodies of data were warmly received by the profession and many research centers have purchased these sets of data. Many Ph. D. theses, as well as other research projects, are using this sample.

Vital statistics

The Public Health Service provides annual statistics on births, deaths, marriages, and divorces (68-12, 68-13, 68-14, 68-15, and 68-16), which are all very useful for simulation models involving population projections.

2. HOUSING AND REAL ESTATE

Data relevant to research on housing and real estate are included in the material discussed under the headings of population, consumer behavior, agriculture, banking, and taxes. In addition, however, specific housing information is collected by a number of Federal agencies. Some of these agencies, such as the Federal Housing Administration, the Federal Deposit Insurance Corporation, and the Federal Home Bank Board, were not included in the inventory of machine-readable data and so are not referred to in this evaluation.

Inventory of housing

The Bureau of the Census provides a number of different series relating to this topic. The survey of inventory change and residential financing of housing units (41-A-22), the housing vacancy survey (41-A-16), and the housing sales survey (41-A-151) are all very useful, but for many purposes some data selection might be required to reduce the number of tapes.

Building permits

A number of different census surveys are available on building permits. Building permits issued monthly and annually (41-A-148 and 41-A-152), building permits used (41-A-149), nonpermit construction starts (41-A-147), and the construction progress report (41-A-150) are all relevant and important for the analysis of the construction industry.

Prices of housing

The Bureau of Labor Statistics price data on housing include consumer price data on housing (44-B-4) and rents (44-B-5). These tapes are in addition to the information on consumer expenditures included under consumer behavior, and if available in regional detail, would provide valuable information on the relative demand and supply of housing.

Farm real estate

The Department of Agriculture provides information on farm real estate values (40-F-10). In addition there are two surveys, farm real estate market survey (40-F-12), and farm building survey (40-F-14), which would be very useful.

Other data

Financial information provided by the FHA series on insured home mortgage terminations (55-17), and data on individual parcels given by the census of governments assessed valuations (41-A-46), would be particularly valuable. Although the "County and City Data Book" is published, the computer tapes of this information which are now sold by the Census Bureau are very useful for research.

3. LABOR FORCE AND WAGES

Information on the labor force, employment, earnings, and labor unions are provided by several Federal agencies. These data are often needed in a highly disaggregated form so that they can be related at a detailed level to other regional, industrial, and demographic information.

Labor force

The basic data in this area is provided by the Census Bureau in the Current Population Survey (41-A-19), and high priority should be given to making this available. In addition, the Social Security Administration provides useful samples of employer-employee records (72-1), and continuous work histories (72-3 and 72-4). Some sample of the summary earnings record tape (72-6) would also be desirable. Finally, the Bureau of Employment Security of the Department of Labor gives data on the employment and wages of workers covered by unemployment insurance (44-A-8), labor turnover (44-B-28), and the characteristics of the insured unemployed (44-A-9). Given the current interest in the problem of unemployment data these sources are very important.

Wages and hours

The Bureau of Labor Statistics collects the basic information in this area in its survey of industry employment, payrolls, and hours (44-B-27 and 44-A-18). The data on wages and related benefits for 82 market areas (44-B-14) also are highly important. Since census data on industry and trade also contain employment and wage data it will often be found useful if these various bodies of data are available in a form that can be interrelated.

Unions and pension plans

The characteristics of labor unions (44-D-12) together with their financial data (44-D-13 and 44-D-14) provide the basic information on labor unions. Pension and welfare plans are covered by additional Bureau of Labor Statistics surveys (44-D-6, 44-D-7, 44-D-8, and 44-D-9). The growing importance of pension and welfare funds both as a source of funds in the economy and in terms of effects on the future income of the aged make this information particularly valuable.

4. EDUCATION

The increased interest in education and the magnitude of expenditures on education make it imperative that adequate data on this topic be available for research purposes. Much of the basic information is contained in the population census and other surveys where data are provided on the age, sex, and educational attainment of individuals. However, the Office of Education of the Department of Health, Education, and Welfare, provides a considerable amount of specialized information.

Primary and secondary schools

The inventory of schools for resource evaluation (51-4) provides basic data on primary and secondary schools. Additional surveys of nonpublic schools (51-7, 51-18, and 51-19) are carried out on a periodic basis. Expenditures by type per pupil (51-9) and data on various aspects of the curriculum such as science and mathematics (51-3 and 51-8) and foreign languages (51-39) furnish valuable information on the extent of educational benefits in different areas.

Higher education

There is a considerable body of information available for colleges and universities in machine readable form. Data on plant and equipment (51-16, 51-20), enrollment (51-10, 51-11, 51-15), residence and migration of students (51-12), earned degrees (51-13), faculty (51-14, 51-34), and financial statistics (51-24), are available. The survey of scientific and technical personnel (41-B-49) made by the Bureau of the Census is pertinent here. All these bodies of information are important to research projects on the role and development of higher education in the Nation.

5. HEALTH

In view of the development of both private and public health plans, economic research on health has become very important. The Public Health Service has since 1959 provided a series of continuing surveys and a number of special purpose supplements, all aimed at establishing basic and comprehensive data for research in the health field.

Health interview survey

This survey (68-1) together with the personal health expenditure survey (68-9) should be made freely available to research workers with proper measures developed to safeguard the confidentiality of the original records.

6. CONSUMER BEHAVIOR

The field of consumer behavior has been intensively studied by economists for several decades. Consumer expenditure studies, analyses of purchasing intentions, and the financial characteristics of households are all important.

Consumer expenditure

The Bureau of Labor Statistics survey of urban consumers (44-B-6) and the Department of Agriculture survey of rural consumers (40-C-2) constitute the most recent basic data in this area. The committee has already indicated that these sets of data should be available for research purposes.

Purchasing intentions

The quarterly survey of the intentions of households collected by the Census Bureau (41-A-18) constitutes a body of information which is very useful in the study of consumer behavior.

Financial characteristics of households

The 1963 survey of financial characteristics (55-1) was made by the Bureau of the Census for the Federal Reserve Board. These data are valuable for research not only on consumer behavior, but also on the role of the household sector as a source of financing in the economy.

7. AGRICULTURE

Agricultural economic research has for many decades been a major concern of many colleges and universities in the United States. It is difficult at this juncture to specify just what categories of data would be of particular interest to the various research groups in these institutions. Nevertheless it is apparent that questions of land use, conservation, productivity, farm management, and many other topics are very important.

The inventory of machine-readable data in the Department of Agriculture covers six areas: (1) Forest Service, (2) Commodity Exchange Authority, (3) Statistical Reporting Service, (4) Agricultural Stabilization and Conservation Service, (5) Agricultural Marketing Service, and (6) Economic Research Service. There are, of course, several classes of users for this information. On the one hand, there are research groups interested in the economic conditions in agriculture within specific regions of the country, and for these groups highly detailed information of a sample nature is often very useful. Other groups are more interested in the total national picture, and the functioning of agriculture as a sector in the economy. These groups want comprehensive tabulations, some of which are supplied by the Bureau of the Census.

Forest Service

The Forest Surveys are often based on a two-stage sampling scheme using aerial photographs. They are of interest primarily to those analyzing regional

forest problems. Over 20 of these surveys were listed by the research workers consulted as having considerable priority. With respect to larger bodies of data, the National Compilation of Forest Survey Statistics (40-A-59) is a very important set of data, containing information on ownership, size, forest type, species, and timber products.

Agricultural production

The basic information on farms, farm characteristics, livestock products, crops, fruit, etc., is provided by the Census of Agriculture (41-A-6 and 40-F-57). The Department of Agriculture also has tapes on the June-December enumerative survey, providing acreage reports for crops and reports on livestock (40-C-12). Both of these sets of data are important in the analysis of agricultural output.

Agricultural marketing

The data provided by the Commodity Exchange Authority showing futures transactions and trading data (40-B-3 and 40-B-5) are useful in market pricing studies. In addition, some of the data provided by the Agricultural Marketing Service on such things as fruits and vegetables (40-E-3, 40-E-5), slaughtering (40-E-8), and milk (40-E-10 and 40-E-11) provide information on specific commodities in considerable detail.

Farm management

The Economic Research Service of the Department of Agriculture provides a great deal of information about the status of farmers and farm management. Over 30 sets of data were listed as being particularly important for research on such topics as the financial condition of the farmer, transportation, housing, real estate and land use. In addition, gross income, cost of production, machinery costs, and fertilizer costs and benefits are all topics of research interest for which important sets of data exist.

8. BUSINESS AND INDUSTRY

As already indicated, reports on specific business or industrial establishments might result in disclosure of confidential information. However, highly disaggregated data for regions and industries can often be presented without disclosure. In addition, samples may be developed which would not violate confidentiality, and fuller and more detailed data could be kept in a similar form by government agencies for those research projects which require processing of the original reports.

Manufacturing and mineral industries

The census of manufactures and mineral industries for 1947, 1954, and 1958 (41-A-32, 41-A-33, and 41-A-34) and the annual survey of manufactures (41-A-38) should be made available in as disaggregated a form as the disclosure rules will permit, and specific samples of data should be integrated with the census of manufactures data and the Internal Revenue data to provide more comprehensive and complete coverage of the manufacturing and mineral industries.

Trade and services

The economic censuses of wholesale (41-A-36) and retail trade (41-A-45), transportation (41-A-37) and services (41-A-42) should be treated in a manner similar to that described for data on manufacturing and mineral industries. In addition, the monthly surveys in this area (e.g., 41-A-51 to 41-A-60) should be developed into systematic sets of samples available over time.

Banking and finance

In the preliminary inventory of machine readable data in the Federal Government, the Federal Reserve Board was the only financial institution included. It is probable that when the survey is extended to other Federal financial institutions, many important bodies of data will come to light. In the material examined in the current inventory, member bank loans to commercial and industrial borrowers (55-49 and 55-50) and small business financing experience (55-45 and 55-46) obtained by the Federal Reserve Board represent valuable research materials for analyzing business financing.

9. GOVERNMENT FINANCES AND TAXATION

A large body of information is available on the income and expenditures of Federal, State, and local governments, and Federal tax returns of individuals and business. These data are a very valuable source of research material.

Government finances

The Census of Governments financial data (41-A-47) provides information on about 91,000 government bodies. These data are valuable for comparative research on State and local governments, and when used together with other regional information provide material for analysis of regions and standard metropolitan areas.

Taxes

The Internal Revenue Service and the Brookings Institution have created in recent years tax models for individuals (48-11), for corporations (48-12), and for partnerships (48-13). All these tax models have been found to be extremely useful for research purposes. Additional tapes have also been prepared for fiduciary returns (48-9), estate taxes (48-8), and gift taxes (48-10). It is recommended that tapes be prepared on returns showing capital gains and losses, and that continuous income histories covering both individuals and corporations be developed. Finally, a considerable number of scholars consulted by the committee emphasized that it would be highly productive if the tax records could be matched with social security records, the census of population, the census of manufactures, and the financial reports of the Federal Trade Commission and the Securities and Exchange Commission.

10. FOREIGN TRADE AND PAYMENTS

Considerable progress has been made over the last decade in the development of data on foreign trade. Imports and exports on a commodity and country basis are available in considerable detail on a monthly and an annual basis.

Exports and imports

The Bureau of the Census processes the basic foreign trade data. Export and import data are available for both waterborne and airborne trade for various levels of commodity and country detail. Such data are useful for a wide variety of purposes.

Capital flows

Analysis of direct foreign investment and short-term capital flows is important for understanding the balance of payments of the United States. Some of the required data now obtained by the Treasury, the Federal Reserve Board, and the Department of Commerce are in machine-readable form. However, a great deal of other important data still are not in this form. A significant contribution would be made as a first step by putting all balance-of-payments material on tape.

11. OTHER AREAS

There are, of course, other areas of data which deserve special attention. For example, data on the natural resources of the United States are very important for studies of conservation and research on the future growth of the economy. Special topics such as research and development expenditures by government and industry, studies of pollution, highways, railways, and impacted defense areas all require and often produce special sets of data which should not be overlooked.

EXCERPTS FROM AN INVENTORY OF PUNCHCARDS AND COMPUTER TAPES HELD BY FEDERAL AGENCIES DECEMBER 1964

This listing represents an initial inventory of the machine-readable data held by a number of Federal agencies. It was initiated by the Office of Statistical Standards of the Bureau of the Budget, with the cooperation of the National Archives and Records Services. The purpose of this inventory is to provide information for development of policies relating to the preservation and use of data collected and processed by Federal agencies. This summary of the inventory was prepared by the Committee on the Preservation and Use of Economic Data of the Social Science Research Council.

DEPARTMENT OF AGRICULTURE

FOREST SERVICE

- 40-A-1 Illinois Forest Survey (decennial): Number of trees, volume, and growth in cubic feet and board feet. Reporting unit: individual tree measurements. Unrestricted; permanent; 1962; two tapes.
- 40-A-2 Missouri Forest Survey (decennial): Number of trees, volume, and growth in cubic feet and board feet. Reporting unit: individual tree measurements. Unrestricted; permanent; 1959; 80,000 cards.
- 40-A-3 Minnesota Forest Survey (decennial): Number of trees, volume, and growth in cubic feet. Reporting unit: individual tree measurements. Unrestricted; permanent; 1960-62; 130,000 cards; three tapes.
- 40-A-4 Kentucky Timber Cut (decennial): Timber products, output, and resulting timber cut reported by product in standard units of measure. Unrestricted; 10 years; 1962; 500 cards.
- 40-A-5 Pulpwood Production in Lake States Counties (annual): Pulpwood receipts by quantity, source, and species, reported by primary wood using plants (pulpmills). Confidential; 2 years; 1,500 cards.
- 40-A-6 Small Forest Ownership in Southern Michigan (one-time survey): Social and economic characteristics of woodland owners. Reporting unit: woodland owners. Confidential; 7-8 years; 1959.
- 40-A-7 Recreational Use of Huron-Manistee National Forest (one-time survey). Purpose of visit, type of area, likes and dislikes, length of stay, etc. Reporting unit: Recreational groups. Unrestricted; 3 years; 1962; 700 cards.
- 40-A-8 Boundary Waters Canoe Area Recreation Study (one-time survey): Length of stay, purpose of visit, activities, likes, and dislikes. Reporting unit: recreational party. Unrestricted; 3 years; 1960-61; 4,214 cards.
- 40-A-9 Survey of Campers in Huron-Manistee National Forests (one-time time survey): Income, age, occupation, education, family size, residence, amount of camping. Reporting unit: family campers and camping groups. Unrestricted; 3 years; 1962; 500 cards.
- 40-A-10 Fire Statistics (annual): Information on individual fires by cause, size class, fuel type, etc., man-hours of suppression action, type of action, cost of damages, etc. Reporting unit: Ranger district. Unrestricted; permanent; 1940 to present; 39,000 cards.
- 40-A-11 Timber Cut and Sold (quarterly): Volume and values of timber sold and cut by sale size class and species group. Reporting unit: National forest. Unrestricted; 5 years; 1959 to present; 9,000 cards.
- 40-A-12 Mill Scale Studies (selected intervals): Volume, quality, and value of units of output (boards and veneer) per log for a tree species. Reporting unit: processing plants. Unrestricted; permanent; 1954 to present; 275,000 cards.
- 40-A-13 Timber Inventory (decennial): Inventory of timber volumes, rate of tree growth, and tree mortality. Reporting unit: National Forest (working circle). Unrestricted; 10 years; 1954 to present; 700,000 cards.
- 40-A-14 Butte County, Fire Prevention Survey (one-time survey): The forest knowledge level (all aspects) of an individual. Reporting unit: an individual who is considered representative for a segment of the California population. Confidential; permanent; 1963; 300 cards.
- 40-A-15 Fire Weather and Fire Indices (daily): Wind speed, direction, temperature, humidity, precipitation, fuel moisture. Reporting unit: 335 stations run by DCF and FS. Unrestricted; permanent; 1951 to present; 80,000 cards; 190 tapes.
- 40-A-16 Survey of Timber Cut and Timber Products Output (one-time survey): Volume cut by species, by county of origin, by product, by ownership classes. Reporting unit: Wood using firms and operators. Confidential; 5 years; 1962; 12,000 cards.

- 40-A-17 Forest Survey, Hawaii (annual) : Location, ownership, condition of forest land, volume, kind and quality of timber trees, the net annual growth of commercial timber, mortality, losses, and annual timber cut. Reporting unit: individual forest survey plots. Unrestricted; permanent; 1958-61; 53,000 cards.
- 40-A-18 Forest Survey, California (annual) ; Location, ownership, condition of forest land, volume, kind and quality of timber trees, the net annual growth of commercial timber, mortality, losses, and annual timber cut. Reporting unit: individual forest survey plots. Unrestricted; permanent; 1958-62; 200,000 cards.
- 40-A-19 Level and Sources of Fire Prevention Knowledge of California Hunters (one-time survey) : Forest knowledge level (all aspects) of each hunter. Reporting unit: individual licensed hunter. Confidential; permanent; 1959-60; 1,400 cards.
- 40-A-20 Forest Products Marketing Research, Region 5, Timber Sales 10 (annual) : Cost and revenue of selected processing plants. Reporting unit: individual firms. Confidential; permanent; 1951-61; 5,000 cards.
- 40-A-21 Forest Products Marketing Research, Region 5, Timber Sales 01 (annual) : Characteristics of individual timber sales, date, seller, number of bidders, locations, etc. Reporting unit: individual sales. Unrestricted; permanent; 1952-61; 3,000 cards.
- 40-A-22 Information on Campground Use and Visitor Characteristics (daily) : Number of persons per group, length of stay, place of residence. Reporting unit: visitor group. Unrestricted; permanent; 1958-61; 5,000 cards.
- 40-A-23 Campground Attendance, California (daily) : Number of persons per group, length of stay, place of residence. Reporting unit: visitor group. Unrestricted; permanent; 1961-62; 500 cards.
- 40-A-24 Snow Course Water Equivalent (selected intervals) : Water equivalent, average of five or more points, course elevation, slope, aspect, forest density, and associated meteorological conditions at index station. Unrestricted; permanent; 1958-60; 2,000 cards.
- 40-A-25 Watershed Characteristics and Conditions, California (one-time survey) : Area-elevation slopes, aspects, stream lengths, geologic rock type, soil characteristics, area burned, vegetation cover types, and densities, other land uses (roads, logged areas). Reporting unit: individual watersheds. Unrestricted; permanent; 1955-64; 140,000 cards.
- 40-A-26 Individual Fire Reports (annual) : Size of fire, cause, discovery time, attack time, control time, manpower, fuel type, suppression costs, and damages for bigger fires. Reporting unit: individual fire. Unrestricted; permanent; 1940 to present; 70,000 cards.
- 40-A-27 Lumber Manufacturing Costs and Selling Values (annual) : Cost of manufacturing lumber and the selling value of it. Reporting unit: lumber mills, region 1. Confidential; permanent; 1962 to present; 50,000 cards.
- 40-A-28 Mill Scale Study, Region 1 (one-time survey) : Quantity of board recovery from trees. Reporting unit: individual trees. Unrestricted; permanent; 1961-64; 400,000 cards.
- 40-A-29 Resource Accounting, Region 4 (one-time survey) : Individual tree stands, acreage, species, site quality. Reporting unit: individual tree stands, 5 acres and larger. Unrestricted; 10 years; 1958-59; 10,000 cards.
- 40-A-30 Forest Fire Research, Region 4 (annual) : Causes of fires, size and cost of fires by classes. Reporting unit: individual fire. Unrestricted; permanent; 1942 to present; 22,000 cards.
- 40-A-31 Recreation Facilities, Region 4 (annual) : Inventory of recreation sites by classes. Reporting unit: individual campground site. Unrestricted; permanent; 6,000 cards.
- 40-A-32 Timber Inventory, Region 4 (one-time survey) : Identifies volume of timber by species as merchantable or nonmerchantable. Reporting unit; geographic area (working circle). Unrestricted; permanent; 35,000 cards.
- 40-A-33 Wilderness-Use Study, Pacific Northwest (one-time survey) : Name and address of registrant, number in party 16 years plus and minus, mode of travel. Reporting unit: recreational parties on wilderness trails. Confidential; permanent; 1961; 4,600 cards.

- 40-A-34 Wilderness-Use Study (one-time survey) : Name and address of registrant, number in party over and under 12 years, mode of travel, length of stay. Reporting unit: recreational parties on wilderness trails. Confidential; permanent; 1962; 5,000 cards.
- 40-A-35 Coast Douglas-fir Tree and Log, Lumber and Veneer Recovery (one-time survey) : Tree and log characteristics, lumber and veneer grade yields. Reporting unit: sawmills and veneer mills. Unrestricted; permanent; 1964-66; 200 cards.
- 40-A-36 Western Larch Tree and Log Lumber Recovery (one-time survey) : Tree and log characteristics, lumber grade yields. Reporting unit: individual logs. Unrestricted; permanent; 1961-62; 40 cards.
- 40-A-37 Inland Douglas-fir Tree and Log Lumber Recovery (one-time survey) : Tree and log characteristics, lumber grade yields. Reporting unit: individual logs. Unrestricted; permanent; 1961-62; 150 cards; 5 tapes.
- 40-A-38 Dwarfmistletoe Growth Impact Study (one-time survey) : Tree heights, age, d.h.h., vigor classification, disease rating, site quality, and decadal radii back to 1860 (decadal volumes were computed and are on second set of cards). Reporting unit: individual tree. Unrestricted; 10 years; 1960; 2,100 cards.
- 40-A-39 Forest Employment Data, Washington and Oregon (annual) : Number of employees in forest industries. Reporting unit: four-digit industry. Unrestricted; permanent; 1936 to present; 8,000 cards.
- 40-A-40 Forest Inventory Data and Related Inventory Studies (annual) : Forest resource statistics on area, volume, growth and drain. Reporting unit: forest inventory plots. Confidential; permanent; 1,100,000 cards.
- 40-A-41 Timber Growth and Growing Stock Projections (TRAS-2) (decennial) : Forest inventory, growth, mortality and drain. Reporting unit: data collected on forest inventory plots, Forest Survey section, for PNW Forest and Range Experiment Station. Unrestricted; 5 years; 2,000 cards.
- 40-A-42 Continuous Forest Inventory Data Cards, Region 9 (annual) : Annual survey of several local woods covering tree growth, mortality, insect and disease affliction, etc. Reporting unit: sample plots. Unrestricted; permanent; 1945 to present; 60,000 cards.
- 40-A-43 Forest Inventory and Utilization Statistics for Alaska (annual) : Furnish statistics on forest land areas, timber volumes, growth, mortality, quality, are condition, timber cut, and utilization practices. Reporting unit: trees and plots. Unrestricted; permanent; 1954 to present; 70,000 cards.
- 40-A-44 Black Cherry Lumber Grade Recovery Study, Northeastern Region (one-time survey) : Classification and measurement of lumber cut from individual logs; i.e., width, length and grade. Reporting unit: individual board. Unrestricted; 5 years; 1960; 12,000 cards.
- 40-A-45 Forest Survey of Pennsylvania (one-time survey) : Tree measurements on sample plots randomly located, diameter, height, species and grade. Reporting unit: individual trees in sample plots. Unrestricted; permanent; 1963-64; 45,000 cards.
- 40-A-46 Forest Survey of Maryland (one-time survey) : Tree measurements on sample plots randomly located: species, diameter, height and grade. Reporting unit: individual trees in sample plots. Unrestricted; permanent; 1962-63; 24,000 cards.
- 40-A-47 Forest Survey—Alabama, Arkansas, Louisiana, Mississippi, Oklahoma, Tennessee, Texas (decennial) : Forest acreage, timber volume, cut, growth, and mortality. Reporting unit: timber inventory plots. Unrestricted; 10 years; 200,000 cards.
- 40-A-48 Pulpwood production (annual) : Pulpwood procurement by county and type of wood-species, bolts, chips. Reporting unit: pulp mills. Confidential; 2 years; 4,000 cards.
- 40-A-49 Timber use other than pulpwood by State (decennial) : Timber volume removed from forests by species. Reporting unit: individual firm. Confidential; 10 years; 20,000 cards.
- 40-A-50 Stumpage Prices (selected intervals) : Characteristics of the stumpage offered and the price received. Reporting unit: individual sale. Unrestricted; permanent; 5,000 cards.

- 40-A-51 Small Private Forest Land Ownership Survey, Ohio and Missouri (one-time survey) : Study was to determine the ownership characteristics of the region, attitudes of owners toward their forest land, and if possible, reasons why owners adopt or do not adopt forestry practices and programs. Reporting unit: woodlot owners within four-county area in Ohio and Missouri. Confidential; permanent; 1958; 500 cards.
- 40-A-52 Input and Output Study of Hardwood Log Production (one-time survey) : Information on number, quality and costs of input and outputs for individual sawlogs. Reporting unit: individual boards. Confidential; 5 years; 1962; 20,000 cards.
- 40-A-53 Forage Production and Composition Survey (one-time survey) : Herbaceous and browse production and composition; overstory timber stand size, stocking class, site class, aspect position on slope, percent slope; livestock use; fire history; logging or TSI; erosion hazard index; current erosion index, soil stability rating; ground cover basal area and percent crown cover. Unrestricted; permanent; 1961; 40,000 cards.
- 40-A-54 The Market Potential for Residential Fencing, St. Louis (one-time survey) : Residential fence and home characteristics including; type of fence, volume of wood fence material, location of fence, age and type of home. Unrestricted; permanent; 1964; 600 cards.
- 40-A-55 Forest Inventory, Central States (decennial) : Volume and area information by state, county, and plot by species, dbh, ownership forest type, stand-size, site, tree class, stocking, grazing intensity, mortality, growth and quality. Reporting unit: sample plots and individual trees. Unrestricted; permanent; 150,000 cards.
- 40-A-56 Small Woodland Owners, Ohio (one-time survey) : Survey of factors such as participation in ACP woodland practices and participation in voluntary woodland practices. Reporting unit: woodland owners. Confidential; permanent; 1963; 520 cards.
- 40-A-57 Survey of Outdoor Recreation in Ohio (one-time survey) : Survey of owner's background characteristics, owner's conception of commercial outdoor recreation and owner's economic resources. Reporting unit: owners of forest recreation enterprises, such as picnic areas, camping areas, swimming beaches, riding stables, and pay lakes. Confidential; permanent; 1963; 280 cards.
- 40-A-58 Survey of Picnic Enterprises in Ohio (one-time survey) : Survey of factors such as location, size, natural or manmade attractions, services and facilities, costs of developing and operating picnic enterprises, income conversion surplus and conversion surplus ratio. Reporting unit: Forest Picnic Enterprises in Ohio. Confidential; 10 years; 1963; 72 cards.
- 40-A-59 National Compilation of Forest Survey Statistics (decennial) : By State; forest areas, inventories, annual growth, cut, and mortality. Reporting unit: Forest survey units at forest experiment stations. Confidential; permanent; 1962-63; 50,000 cards.
- 40-A-60 Wood Used by Manufacturers (selected intervals) : Amounts and species of wood used in manufacturing by industry, product, and forms of wood (lumber, bolts, veneer, plywood, hardboard, and particleboard). Reporting unit: sample survey of all manufacturing plants. Confidential; permanent; 1960; 84,000 cards.
- 40-A-61 Wood Preservation Statistics (annual) : Volume and species of wood products treated by different preservatives and fire retardants, and volume of chemicals used. Reporting unit: canvass of all wood-treating plants in the United States. Confidential; 5 years; 1959-63; 18,000 cards.
- 40-A-62 Wood Used in FHA Housing (selected intervals) : Amounts and forms of wood used by house part in FHA inspected houses. Reporting unit: sample survey of house types from a sample of FHA offices. Unrestricted; permanent; 1959 and 1962; 65,000 cards.
- 40-A-63 Working Circle Timber Inventory, Region 3 (decennial) : Data on individual sample trees, species, dbh, height, and class. Data on soil erosion, diseases, timber type, and logging information. Reporting unit: national forests. Unrestricted; 10 years; 1962; 80,000 cards; 400 tapes.

- 40-A-64 Timber Sales Inventory, Region 3 (selected intervals): Data on individual sample trees, species, dbh, height and class. Data on soil, erosion, diseases, timber type and logging information. Reporting unit: individual sale. Unrestricted; 6 to 12 years; 60,000 cards; 300 tapes.
- 40-A-65 Fire Weather Records, Intermountain (daily): Weather and fuel conditions. Reporting unit: individual station. Unrestricted; permanent; 1954 to present; 150,000 cards.
- 40-A-66 Lightning Research, Montana (selected intervals): Atmospheric electric field-lightning electrostatic field. Reporting unit: individual lightning discharges. Unrestricted; 5 years; 1960-61; 1,600 cards.
- 40-A-67 White Pine Study, Intermountain Region (one-time survey): An accounting of man-hours, vehicle miles, and costs for various phases of timber management activities. Reporting unit: individual timber management project. Confidential; 4 years; 1962-65; 2,500 cards.
- 40-A-68 Forest Survey of Northern Idaho (one-time survey): Covers area, volume, growth and mortality of trees. Reporting unit: individual sample trees. Unrestricted; 7 to 10 years; 1960-64; 30,000 cards.
- 40-A-69 White Pine Study, Intermountain Region (one-time survey): An inventory of lands capable of growing western white pine, including a description of site quality, age and species of stands presently growing on these lands, management silvicultural data. Reporting unit: sample acres. Unrestricted; 7 to 10 years; 1961-63; 1,000 cards.
- 40-A-70 Forest Survey of Western Montana (selected intervals): Covers area, volume, growth, and mortality of trees. Reporting unit: individual sample trees. Unrestricted; 7 to 10 years; 1953-58; 30,000 cards.
- 40-A-71 Forest Survey of Wyoming (selected intervals): Covers area, volume, growth, and mortality of trees. Reporting unit: individual sample trees. Unrestricted; 7 to 10 years; 1957-60; 25,000 cards.
- 40-A-72 Forest Survey of Colorado (selected intervals): First survey of forests of Colorado. Covers area, volume, growth, and mortality of trees. Reporting unit: individual trees. Unrestricted; 7 to 10 years; 1956-59; 70,000 cards.
- 40-A-73 Forest Survey, Southeast Region (decennial): Area, volume, growth, mortality, and timber cut statistics. Reporting unit: individual acres and individual trees. Unrestricted; permanent; 2 million cards.
- 40-A-74 White Pine Log Grade Study, Northeast Region (one-time survey): Classification and measurement of lumber cut from individual logs; i.e., width, length, and grade. Reporting unit: individual boards. Unrestricted; 1956-61; 37,000 cards.
- 40-A-75 Economics of Ponderosa Pine Dwarfmistletoe Control (one-time survey): Rates of return on dwarfmistletoe control investment. Reporting unit; timber stand. Unrestricted; 1 year; 10,000 cards.
- 40-A-76 Diameter Distributions for Douglas-Fir Stands (one-time survey): Number of trees by diameter class. Reporting unit: individual tree stand. Unrestricted; 1 year; 500 cards.
- 40-A-77 Financial Yields from Hardwood Stand Conversion (one-time survey): Present worths associated with conversion of stand from alder to Douglas-fir. Reporting unit: individual timber stands. Unrestricted; 1 year; one tape.

COMMODITY EXCHANGE AUTHORITY

- 40-B-1 Position Surveys—Specific Commodity Market (selected intervals): Futures holdings of each trader in the commodity market in which survey is conducted. Reporting unit: futures commission merchants, members of contract markets, foreign brokers. Confidential; 3 years; 14,000 cards.
- 40-B-2 Cash Commodity Postitions (weekly): Cash positions of respondents holding futures positions in specific commodities. Reporting unit: merchandisers, processors, or dealers in grains, cotton, eggs. Confidential; 5 years; 64,000 cards.

- 40-B-3 Futures Positions of "Special Accounts" (daily); Identification of and classification of positions of large traders. Reporting unit: futures traders whose open positions have reached reportable size. Confidential; 5 years; 210,000 cards.
- 40-B-4 Futures Positions of "Special Accounts" (daily); Identification of special accounts and their reportable futures positions. Reporting unit: futures commission merchants and foreign brokers. Confidential.
- 40-B-5 Futures Trading and Open Contracts (daily); Futures transactions, and open contracts carried on books of exchange clearing members. Reporting unit: clearing members of contract markets. Confidential; 2 months; 54,000 cards.

STATISTICAL REPORTING SERVICE

- 40-C-1 Wool Report (monthly); Pounds, proceeds, and head shorn by date of shearing and sale. Reporting unit: farmers applications to ASCS for wool incentive payments. Confidential; 3 years; 1962-64; 800,000 cards; 12 tapes.
- 40-C-2 Consumer Expenditure Survey (one-time survey); Income and expenditures for a period of 1 year. Reporting unit: rural households. Confidential; 10 years; 1961; 800,000 cards; 30 tapes.
- 40-C-3 Objective Yield Surveys (monthly); Acreage along with objective counts and measurements of plants and fruits. Reporting unit: a sample of farmers. Confidential; 1 year; 159 to present; 400,000 cards; 5 tapes.
- 40-C-4 Cold Storage Report (monthly); Storage capacity of warehouses and weights of over 80 commodities in storage. Reporting unit: all types of refrigerated warehouses. Confidential; 3 years; 1961 to present; 576,000 cards; 6 tapes.
- 40-C-5 Prices Paid Surveys—Feed (monthly); Number of reports, average price and tax, and estimated State price for 50 feeds. Reporting unit: State summary data compiled by SRS field offices. Confidential; permanent; 1958 to present; 144,000 cards; 10 tapes.
- 40-C-6 Prices Paid Surveys—Food and Clothing (monthly); Reports on farmer purchases of 80 to 100 food and clothing items. Reporting unit: individual reports for chainstores, State summary reports on independent clothing stores. Confidential; permanent; 1959-63; 160,000 cards; 24 tapes.
- 40-C-7 Slaughterhouse Survey (monthly); Number of head slaughtered, live and dressed weight, cost, and class of cattle, hogs, and sheep slaughtered. Reporting unit: a sample of federally inspected meatpacking plants. Confidential; 6 years; 1962 to present; 50,000 cards.
- 40-C-8 Nonfederally Inspected Slaughter (monthly); Number of head, average and total live weight by size groups. Totals combined with federally inspected slaughter. Reporting unit: State summary data compiled by SRS field officers. Confidential; 2 years; 1963-64; 12,000 cards.
- 40-C-9 Horticultural Specialties Survey (annual); Grower data on plants on hand, or in production, quantities sold, and value. Reporting unit: individual commercial growers in selected States. Confidential; 2 years; 1963-64; 24,000 cards; 2 tapes.
- 40-C-10 Prices Received by Farmers—Potatoes and Citrus (monthly); Revised monthly estimated prices and weights. Reporting unit: State estimates prepared by field offices and Crop Reporting Board. Confidential; 2 years; 1962-63; 8,000 cards.
- 40-C-11 Telephone and Electricity Survey (annual); Kilowatt-hours and electric bill, telephone bill, LP gas purchased and amount of bill. Reporting unit: a sample of farmers and prices they pay. Confidential; 6 years; 1961 to present; 114,000 cards; 6 tapes.
- 40-C-12 June-December Enumerative Survey (annual); Includes acreage of most crops, livestock by specie and class, farm numbers, etc. Reporting unit: farmers—a probability sample. Confidential; 5 years; 1961 to present; 1 million cards; 30 tapes.

- 40-C-13 Market Records (monthly) : Receipts of cattle, hogs, sheep, and lambs by State of origin. Reporting unit : individual livestock market and packing plants. Confidential ; 2 years ; 1963-64 ; 20,000 cards ; 2 tapes.
- 40-C-14 Beef Steer and Heifer Report (monthly) : Number of head, weight, and cost by grades for steers and heifers. Reporting unit : each of the 14 major livestock markets. Confidential ; 6 years ; 1962 to present ; 7,500 cards ; 2 tapes.
- 40-C-15 Dairy Manufacturers Survey (monthly) : Production of butter, ice cream and other frozen dairy products, cheeses, canned milk, etc. Reporting unit : plants manufacturing dairy products (16 States). Confidential ; 2 years ; 1962-63 ; 10,000 cards.
- 40-C-16 Building Values (one-time survey) : Value of property of farm, residential on farm, other buildings. Reporting unit : individual farms. Confidential ; 3 years ; 1963 ; 4,000 cards.

AGRICULTURAL STABILIZATION AND CONSERVATION SERVICE

- 40-D-1 Defense Data Program (selected intervals) : Reporting unit : individual plant locations. Confidential ; permanent ; 40,000 cards.
- 40-D-2 Storage Résumé (monthly) : Warehouse facilities by State showing location, commodity code, number of lots, units, and quantity in store by capital and appropriated fund. Unrestricted ; 2 years ; 18,000 cards.
- 40-D-3 Area Recap, Instore Processed Commodity Inventory (semimonthly) : Area recap—State total by commodity, bulk, and/or packaged with overall total. IM3R commodity class by warehouse and location giving quantity by commodity and program year. Unrestricted ; 2 years ; 18,000 cards.
- 40-D-4 Position 4—Instore—Merchandising Inventory (semimonthly) : Inventory listings showing lot number, commodity code, program code, program year, quantity, and warehouse in which stored. Reporting unit : inventory lot. Unrestricted ; 2 years ; 15,000 cards.
- 40-D-5 Position 3, Intransit, Positions 5, Ordered Not Shipped From Storage Inventory (semi-monthly) : Position 3, reference number, commodity code, program code, lot number, quantity ; position 5, reference number, commodity code, program code, program year, lot number, quantity. Unrestricted ; 2 years ; 50,000 cards.
- 40-D-6 Approved, Accepted and Reserve Warehouses (quarterly) : Name and address ; mailing address ; plant address ; type of storage, dry, cold, tank, or whey ; number of cars loaded or unloaded per day ; delivering carrier. Reporting unit : warehouses. Unrestricted ; 2 years ; 3,500 cards.
- 40-D-7 Current Warehouse Grain Activity (daily) : Grain acquisitions, dispositions, and adjustments. Reporting unit : country and terminal warehousemen. Unrestricted ; 30 days ; seven tapes.
- 40-D-8 Grain Prices and Discount Formulas (daily) : Base prices for barley, corn, flax, grain sorghums, oats, rye, wheat, and market premiums and discounts. Reporting unit : grain trade. Unrestricted ; one tape.
- 40-D-9 Grain Inventory Open File (daily) : Grains—barley, corn, flax, grain sorghums, oats, rye, wheat, and edible beans. Reporting unit : warehouse receipts. Unrestricted ; 14 tapes.
- 40-D-10 Elevator Name and Address Master File (selected intervals) : Elevator name, address, settlement markets, freight rates, storage, and loadout capacities. Reporting unit : warehouse. Unrestricted ; 1961 to present ; one tape.
- 40-D-11 CCC Commodity Loan Transactions (daily) : Loans made, repayments, collateral, acquired, loans written off, and loans outstanding. Reporting unit : individual loan. Confidential ; 60 days ; 120 tapes.
- 40-D-12 Commercial Warehouses Under UGSA—21 States (daily) : Warehouse name, address, and capacity. Reporting unit : warehouse. Confidential ; permanent ; 4,600 cards.
- 40-D-13 CCC-owned Grain Bin and Equipment Facilities (daily) : Purchases, transfers between States, counties, and bin-site locations, and dispositions. Reporting unit : individual facility. Confidential ; 97 days ; three tapes.

- 40-D-14 CCC-owned Grain Inventory in CCC Bin Sites (daily) : Grain receipts and withdrawals. Reporting unit: individual binsite. Confidential; 14 days; one tape.
- 40-D-15 CCC-Commodity Inventories (daily) : Month-end inventory; acquisitions and dispositions during month. Reporting unit: internal vouchers supporting the general ledger. Confidential; 1961 to present; 100,000 cards.
- 40-D-16 CCC Warehouse—Stored Grain Dispositions, 21 States (daily) : Commodity, trust member, quality and quantity of grain dispositions. Reporting unit: warehouse receipt. Confidential; 7 days; one tape.
- 40-D-17 CCC Warehouse—Stored Grain Acquisitions—21 States (daily) : Commodity, quality, quantity of grain acquired. Reporting unit: warehouse receipts. Confidential; 7 days; four tapes.
- 40-D-18 CCC Warehouse—Stored Grain Inventories—21 States (daily) : Description of CCC grain inventory stored in commercial warehouses. Reporting unit: warehouse receipt. Confidential; 14 days; 14 tapes.

AGRICULTURAL MARKETING SERVICE

- 40-E-1 Tobacco Stocks Report (quarterly) : Stocks of leaf tobacco owned by dealers and manufacturers. Reporting unit: tobacco dealers and manufacturers. Confidential; 5 years; 1960 to present; 50,000 cards.
- 40-E-2 Truck Shipments of Fresh Fruits and Vegetables for California and Florida (daily) : Package units of fresh fruits and vegetables. Reporting unit: carlot inspections. Unrestricted; 90 days.
- 40-E-3 Unloads of Fresh Fruits and Vegetables in 41 Cities (daily) : Carlots and carlot equivalents of fresh fruits and vegetables by commodities and origin (States or countries) unloaded in 41 principal market cities. Reporting unit: carlot inspections. Unrestricted; 30 days.
- 40-E-4 Rail Shipments of Fruits and Vegetables (monthly) : Carlot units of fresh fruits and vegetables by waybilling stations, commodities, and origin. States or countries. Reporting unit: carlot inspections. Unrestricted; 1 year.
- 40-E-5 Fruit and Vegetable Rail Shipments (daily) : Carlot units of fresh fruits and vegetables by commodities and origin (States or countries). Reporting unit: carlot inspections. Unrestricted; 90 days.
- 40-E-6 Egg Products (Liquid, Frozen, and Dried) Report (weekly) : Quantities produced. Reporting unit: plants under Federal grading. Confidential; 5 years; 1962 to present; three-tapes.
- 40-E-7 Poultry Canning Report (monthly) : Quantities of poultry used in cutting up and further processed and quantities condemned. Reporting unit: further processing plants under Poultry Products Inspection Act. Confidential; 5 years; 1960 to present; 13,000 cards; two tapes.
- 40-E-8 Slaughter and Evisceration Report (weekly) : Quantities of poultry inspected and condemned. Reporting unit: slaughter and evisceration plants under Poultry Products Inspection Act. Confidential; 5 years; 1959 to present; 100,000 cards; four tapes.
- 40-E-9 Milk Marketing Program 9110 (one-time survey) : Volume weights of milk and milk products. Reporting unit: dairy plants. Confidential; permanent; 1961-62; two tapes.
- 40-E-10 Milk Marketing Product Reports (monthly) : Sales of fluid milk products. Reporting unit: dairy plants. Confidential; permanent; 1960 to present; 16 tapes.
- 40-E-11 Milk Marketing Price Report—MO-1 (monthly) : Milk receipts, utilization, and prices. Reporting unit: dairy plants. Confidential; permanent; 1960 to present; 40 tapes.
- 40-E-12 GR-132, Volume of Grain Inspections (annual) : Volume of each kind of grain inspected at each market, by movement. Reporting unit: grain elevators. Unrestricted; 3 years; 75,000 cards.
- 40-E-13 Grain Quality (Formerly Grain Inspections) Data (annual) : A systematic sample for estimating quality of each grain crop. Reporting unit: grain elevators. Unrestricted; 3 years; 300,000 cards.

- 40-E-14 Cotton Quality Survey (annual) : Fiber and processing properties of model qualities of cotton produced in the United States. Reporting unit: cotton gins. Unrestricted; permanent; 1946 to present; 36,000 cards.

ECONOMIC RESEARCH SERVICE

- 40-F-1 Social Security Survey (annual) : Social security benefits and how employed by farmers. Reporting unit: farmers covered by social security. Confidential; 5 years; 149,000 cards.
- 40-F-2 Use of Highways in Crop Disposal (one-time survey) : Values of varying-type highways in expediting various type crops. Reporting unit: highway commission, truckers, farmers, co-ops, markets. Confidential; 5 years; 1964; 24,000 cards.
- 40-F-3 Surveys of Agricultural Finance (selected intervals) : Various lending operations, loan surveys, etc., for farmers. Reporting unit: lending institutions, Federal Reserve, individuals, etc. Confidential; 1958-61; 450,000 cards.
- 40-F-4 Great Plains Survey (one-time survey) : Landownership, water rights, mineral rights—methods of obtaining ownership. Reporting unit: individual farmers. Confidential; 5 years; 1960; 32,000 cards.
- 40-F-5 Fairfax Tax Study (one-time survey) : Comparative tax assessments on former farmer area, which is now tax-classified "suburban." Reporting unit: individual tax assessments. Confidential; 2 years; 2,000 cards.
- 40-F-6 Great Plains Pricing Survey (one-time survey) : Factors considered by sellers and buyers in pricing farm real estate. Reporting unit: individual sellers and buyers. Confidential; 5 years; 1960; 20,000 cards.
- 40-F-7 Great Plains Survey (Farmers Living Standards) (one-time survey) : Amount of land, living standards, size of family, etc. Reporting unit: individual farmers. Confidential; 5 years; 1960; 46,000 cards.
- 40-F-8 Transportation Study (one-time survey) : Transportation costs of corn between demand and supply points. Reporting unit: transportation units of various agencies. Confidential; 5 years; 1963; 10,000 cards.
- 40-F-9 Economics of Housing for Migrant Hired Farmworkers (one-time survey) : Economics of housing for migrant hired farmworkers, social security information of farmers and farmworkers. Reporting unit: farmers and farmworkers. Confidential; 3 years; 1963; 13,000 cards.
- 40-F-10 Farm Real Estate Values (one-time survey) : Values of real estate. Reporting unit: individual farmers. Confidential; 5 years; 1960; 26,000 cards.
- 40-F-11 Tax Survey (annual) : Farm real estate taxes. Reporting unit: State tax offices. Unrestricted; 5 years; 1960-63; 150,000 cards.
- 40-F-12 Farm Real Estate Market Survey (semiannual) : supply of and demand for farmland, actual sales of farmland, current prices and expected price movements, type of buyers and sellers, availability. Reporting unit: individuals. Unrestricted; permanent; 1956-64.
- 40-F-13 Voluntary Home Mortgage (one-time survey) : Insurance of home mortgage loans made by banks, home finance companies, leading agencies, insurance companies, etc. Reporting unit: individual loans. Confidential; 1 year; 1963; 115,000 cards; two tapes.
- 40-F-14 Building Value Survey (one-time survey) : Individuals asked to estimate value: their entire farm, all buildings, residence. Reporting unit: individual farmers. Confidential; 3 years; 1963; 12,000 cards; two tapes.
- 40-F-15 Credit Survey—Dairy in Wisconsin; Hog-Beef in Corn Belt (one-time survey) : Financial standing of individuals, including dairy operators in eastern Wisconsin and hog-beef feeders in the Corn Belt. Reporting unit: individuals. Confidential; 3 years; 1963; 2,000 cards.

- 40-F-16 Extent of Spraying and Dusting on Farms (one-time survey): Individual farm operations in connection with chemical treatment for insect, disease, and weed control. Reporting unit: individual farmers. Confidential; 5 years; 1958; 44,000 cards.
- 40-F-17 Machinery and Equipment Study (one-time survey): Survey of farm tractors and machinery. Reporting unit: farmers. Confidential; permanent; 1957; 10,000 cards.
- 40-F-18 Illinois Feeder Cattle Study (one-time survey): Feed conversion data for droves of cattle fed by Illinois farm cooperatives. Reporting unit: individuals. Confidential; permanent; 1938-63; 9,000 cards.
- 40-F-19 Methods Used to Distribute Fertilizer (one-time survey): Reporting unit: individual farmers. Confidential; 5 years; 1962; 48,000 cards; two tapes.
- 40-F-20 Harvesting the Hay Crop (one-time survey): Survey of farmers' operations in the harvesting of hay. Reporting unit: individual farmers. Confidential; 5 years; 1961; 90,000 cards.
- 40-F-21 Recordings of Farm Mortgages (biennial): Characteristics of farm mortgages recorded (closed): interest rates, term, size averages. Reporting unit: lenders. Confidential; 1 year; 1963; 56,000 cards; three tapes.
- 40-F-22 Liquid Petroleum Fuel Used by Farmers (one-time survey): Farmers and extent of their use of different types of fuel. Reporting unit: individual farmers. Confidential; 5 years; 1959; 64,000 cards.
- 40-F-23 Harvesting Small Grains and Field Shelling Corn (one-time survey): Extent of different harvesting methods used on small grains and extent of field shelling of corn. Reporting unit: individual farmers. Confidential; 5 years; 1960; 40,000 cards.
- 40-F-24 OECD Agricultural Exports and Imports (quarterly): SITC commodities exported and imported by country of destination and origin. Reporting unit: SITC commodities. Unrestricted; permanent; 1963; eight tapes.
- 40-F-25 State Export Equivalent Study (one-time survey): State share of production of U.S. exports and imports of selected commodities. Unrestricted; permanent; 1960-61; 10,000 cards.
- 40-F-26 SRS June Enumerative Survey, Farm Population Data (annual): Color of operator, number of persons in operator's household, number of other households on farm. Farm classification items, such as size of farm, value of products sold, etc. Reporting unit: farm operator households. Confidential; permanent; 1960-62; 18,000 cards; two tapes.
- 40-F-27 Maryland Suburbanization Study (one-time study): Characteristics of head of household and family members, information on changes in community undergoing rapid suburbanization, and attitudes of persons toward these changes. Reporting unit: households in urban fringe. Confidential; 5 years; 1960; 4,000 cards.
- 40-F-28 1/1000 Sample, Population, and Housing (one-time survey): Area and unit identification. Characteristics of persons, households, families, subfamilies, associated persons, mothers of children under 18, and housing units. Reporting unit: households and individuals within households. Unrestricted; permanent; 1960; seven tapes.
- 40-F-29 Textile Imports (monthly): Import statistics, giving commodity codes, total quantities imported, and value for cotton, wool, and the manmade fibers. Reporting unit: total imports for consumption by textile commodity reported by the Bureau of Census. Unrestricted; permanent; 1963-64; nine tapes.
- 40-F-30 Gross Farm Income (annual): Cash receipts from farm marketings; value of home consumption; value of annual change in farm inventories; index numbers of the volume of farm marketings and home consumption. Reporting unit: secondary data from governmental agencies. Unrestricted, permanent; 1949-63; 650,000 cards.
- 40-F-31 Wholesale Fruit and Vegetable Study, ME 3-30 (one-time survey): Data card. Expanded carlots, firm size and market structure. Reporting unit: wholesale firms. Confidential; 4 years; 1958-59; 114,000 cards.
- 40-F-32 Wholesale Fruit and Vegetable Study, ME 3-30 (one-time survey): Employees IRS forms 1065 and 1120. Reporting unit: IRS. Confidential; 4 years; 1959-60; 32,000 cards.

- 40-F-33 Rio Grande Tomato Study, ME 3-67 (one-time survey) : Data cards, structure of lower Rio Grande tomato market. Reporting unit: wholesale firm. Confidential; permanent; 1960-61.
- 40-F-34 Rio Grande Citrus Study, ME 3-67 (one-time survey) ; Data card, structure of Rio Grande Citrus Market. Reporting unit: wholesale firms. Confidential; permanent; 1960-61; 75,000 cards; eight tapes.
- 40-F-35 Wholesale Fruit and Vegetable Study, ME 3-30 (one-time survey) : Data cards, unexpanded carlots and firm size. Reporting unit: wholesale firms. Confidential; permanent; 1958-59; 21,000 cards.
- 40-F-36 Market for Food in Schools (one-time survey) : School characteristics, type of lunch program, and foods used. Reporting unit: public and private elementary and secondary schools. Confidential; 5 years; 1962-63; 10,000 cards.
- 40-F-37 Public Food Distribution Programs Research (selected intervals) : Food consumption and marketing information plus income and other characteristics of low income families. Reporting unit: low income households in: Baltimore, Md., Detroit, Mich., and urban and rural areas of Fayette County, Choctaw County, Okla., Escambia County, Fla., and St. Louis, Mo. Confidential; permanent; 1961-present; 150,000 cards.
- 40-F-38 Convenience Food Study (monthly) : Cost and time per serving of convenience foods and home-prepared foods. Reporting unit: supermarkets. Unrestricted; 3 years; 1959-60; 424 cards.
- 40-F-39 Low-Fat Milk Study (one-time survey) : Low-fat milk sales, composition of low-fat milk, prices for December 1962. Percent low-fat milk represents of total whole, low-fat and skim sales. Reporting unit; milk processors. Confidential; 3 years; 1962; 3,000 cards.
- 40-F-40 Marketing Horticultural and Special Crops, Promotional Practices (one-time survey) : Current marketing practices by retail florists (advertising, promotion, merchandising, pricing, etc.). Reporting unit; retail florists. Confidential; 3 years; 1964; 35,000 cards.
- 40-F-41 Dairy Promotion Study (monthly) : Pounds of milk sold per capita and prices and display size. Reporting unit: Federal Milk Order Markets and Retail Food Stores. Confidential; 3 years; 1963-65; 50,000 cards.
- 40-F-42 Expenditures for Promotion (one-time survey) : Sums spent for promotion and sources of revenue. Reporting unit: farm commodity groups. Confidential; 3 years; 1962-63; six tapes.
- 40-F-43 Food Stocks in Away-From-Home Eating Establishments (one-time survey) : Inventories of food and beverage products. Reporting unit: establishments that generally serve food for onpremise consumption. Confidential; 3 years; 1964; five tapes.
- 40-F-44 Flexibility of Dairy Manufacturing plants (annual) : Production of products by plants. Reporting unit: all plants manufacturing dairy products in United States. Confidential: 4 years; 1961.
- 40-F-45 Name and Address of Egg Assemblers (one-time survey) : Name and address of egg assemblers. Reporting unit: egg assembler and brokers. Unrestricted; 1 year; 1957-58; 26,000 cards.
- 40-F-46 McClain Cost Data for Dairy Plants (quarterly) : Sales and costs by items for 70 fluid milk plants. Reporting unit: individual plants. Confidential; permanent; 1959-63; 500,000 cards.
- 40-F-47 Egg Quality Study (one-time survey) : Gradeouts, yields of producers and price received for eggs. Reporting unit: egg packing plants. Confidential; 2 years; 1960-61; 10 tapes.
- 40-F-48 Weighted Meat Prices (monthly) : Retail prices of beef, veal, pork, and lamb. Reporting unit: cooperating chainstore. Confidential; 3 years; 1962-64; 4,000 cards.
- 40-F-49 Financial Statistics of Food Manufacturers (annual) : Advertising expenditures, total costs, total sales, total net income from IRS Source Book. Reporting unit: asset size classes. Unrestricted; permanent; 1953-60; 4,000 cards.
- 40-F-50 Interindustry Input-Output Data (one-time survey) : Outputs by producing industry and by consuming industry in producer prices. Reporting unit: industry. Unrestricted; permanent; 1947; 40,000 cards.

- 40-F-51 Farm Value, Retail Price, and Farm-Retail Spread for Food Products (monthly): Farm value, retail price, and farm-retail spread. Reporting unit: national average by commodity. Unrestricted; permanent; 1947-58; 2,000 cards.
- 40-F-52 Wool Classification Study (one-time survey): Wool quality factors and prices. Reporting unit: wool warehouse. Unrestricted; 6 years; 1957-60; 60,000 cards.
- 40-F-53 Central Market Study (cotton) (weekly): Date of sale, location, volumes, price, quality. Reporting unit: individual respondents. Unrestricted; 6 years; 1959-61; 288,000 cards.
- 40-F-54 Study of Cotton Warehouses Storage Cost (one-time survey): Cotton storage costs. Reporting unit: warehouses. Confidential; permanent; 1959-60; 2,000 cards.
- 40-F-55 Grain Storage and Handling Costs (one-time survey): Operating costs for grain elevators. Reporting unit: grain elevator operators and owners. Confidential; permanent; 1959-61; 8,000 cards.
- 40-F-56 Wool Market News Study (weekly): Wool prices and quality factors, location. Reporting unit: individual sales. Unrestricted; 6 years; 1962-63; 530 cards.
- 40-F-57 Census Data (quinquennial): Acres and production for 70 crops and 10 livestock items. Reporting unit: farmer respondents, from published census data. Unrestricted; permanent; 1959; 300,000 cards.
- 40-F-58 Normalizing Study (biennial): Acres planted, harvested, yield, production, price and value for 70 crop and 10 livestock items. Reporting unit: States. Unrestricted; permanent; 1939-62; 52 tapes.
- 40-F-59 Ohio River Basin Study (one-time survey): Estimated yields (two levels) for land capability units within land resource areas. Reporting unit: Work unit conservationists, land capability units within land resource area. Unrestricted; permanent; 1963; 33 tapes.
- 40-F-60 Conservation Needs Inventory Land Capability and Use Data (one-time survey): Land use in 1958 and estimated land use in 1975 by land capability subclass. Reporting unit: county committees, counties. Unrestricted; permanent; 1958; 7,500 cards.
- 40-F-61 Watershed Project Needs (CNI) (one-time survey): Acreages requiring project action for flood control, erosion control, drainage and irrigation. Reporting unit: county committees, watersheds less than 250,000 acres. Unrestricted; permanent; 1958; 2,000 cards.
- 40-F-62 Conservation Needs, 160-Acre Sample Plots (one-time survey): Land use, soil type, soil slope, antecedent erosion, and capability class for Oklahoma, Louisiana, New Mexico, and Arkansas. Reporting unit: 160-acre sample plots (2 percent of total area). Unrestricted; permanent; 1958; 60,000 cards.
- 40-F-63 Ownership of Farm Land in the United States (one-time survey): Characteristics of owners, acreage owned, method of acquisition and disposition. Reporting unit: landowners, random sample each county in each State. Unrestricted; permanent; 1946; 38,000 cards.
- 40-F-64 Relative Efficiency of Alternative Tenure Arrangements (one-time survey): Tenure, and farm input and output data. Reporting unit: farm operators, sample in Nebraska, Kansas, Iowa, and Missouri. Unrestricted; 10 years; 1953-57; 16,000 cards.
- 40-F-65 Land Ownership in the Southeastern States (one-time survey): Personal characteristics of landowners, amount and use of land owned, and changes in land used from 1955 to 1960. Reporting unit: Landowners, sample segments within counties of seven States. Unrestricted; permanent; 1955-60; 18,000 cards.
- 40-F-66 Land Ownership in the Great Plains States (one-time survey): Personal characteristics of landowners, amount and use of land owned. Reporting unit: landowners, sample counties of 10 Great Plains States. Unrestricted; permanent; 1957; 153,000 cards.
- 40-F-67 Study of Feed Grain Program (one-time survey): Cropland use, livestock programs, and factors relating to 1961 feed grain program. Reporting unit: farms in Iowa. Unrestricted; 5 years; 1961; 4,000 cards.

- 40-F-68 Appraisal of Soil Bank Programs in Selected Areas of Georgia (one-time survey) : Characteristics of participants and nonparticipants in soil bank program. Reporting unit: farm operators in six sample counties in Georgia (both participants and nonparticipants in the soil bank program). Confidential; 10 years; 1956-59; 12,000 cards.
- 40-F-69 Resource Use and Incomes of Farm Families in Georgia (one-time survey) : Family characteristics (age, education, occupation, income) and farm characteristics (land use, production, livestock numbers, nonfarm work, farm income). Reporting unit: dwelling units in the open country or rural households. Confidential; 10 years; 1957; 12,000 cards.
- 40-F-70 Insect Control and Related Cotton Practices Study (one-time survey) : Information on whether or not cotton insect control practices were used, and if so, the acres covered, the kind, quantity, type, and rate of application, the cost of insecticide used and method of application. Similar information on pre-emergence and post-emergence herbicides, defoliant and fertilizer, and also estimates of the 1961 and 5 year average yields of cotton. Reporting unit: individual farmers. Confidential; 10 years; 1961; 16,500 cards.
- 40-F-71 An Inventory of Land and Soil Resources in Pennsylvania (one-time survey) : Acres of land by use, soil type, slope, and degree of erosion. Two percent random sample. Unrestricted; permanent; 1957-58; one tape.
- 40-F-72 Assessment and Taxation of Farmland—Rochester, N.Y. (one-time survey) : Property and owner characteristics. Reporting unit: farmers in towns of Brighton, Henrietta, Rush, Avon, and Genesco. Confidential; permanent; 1963; 440 cards.
- 40-F-73 Market Egg Poultry Farm Adjustments (one-time survey) : Labor estimates for poultry farm operations, egg production, replacement data, buildings, and equipment requirements, costs and returns. Reporting unit: market egg producers in Connecticut. Confidential; 3 years; 1960-61; 2,000 cards.
- 40-F-74 Yield Data (selected intervals) : Acreage planted, acreage harvested and production for each crop for each county for each year. Reporting unit: all North Dakota counties. Unrestricted; permanent; 25,000 cards.
- 40-F-75 Southwest North Dakota Regrassing Study (one-time study) : Description of farm, machinery inventory, livestock inventory, many other items as reported on the survey schedule. Reporting unit: farms. Confidential; 3 years; 1957-63; 3,000 cards.
- 40-F-76 ASCS Sample of Farms (one-time survey) : Acres of cropland, non-cropland, wheat allotment, feed-grain base and normal conserving base on a 10-percent sample of farms in 32 counties in North Dakota. Reporting unit: farms. Unrestricted; 5 years; 1959-60; 3,600 cards.
- 40-F-77 Montana State-Lease Yield Data (biennial) : State-lease number, location code, seeded acres, total production, and yields for spring, winter wheat, and/or barley. Reporting unit: unpublished yearly yield data obtained from the Montana Agricultural State-lease Records. Unrestricted; permanent; 1938-62; 10,000 cards.
- 40-F-78 Yearly County Yield Data, Montana (biennial) : County code, crop code, year planted acres, harvested acres, production, yield per planted acre, yield per harvested acre. Reporting unit: yearly yield data obtained from the Montana Agricultural Statistics. Unrestricted; permanent; 1919-61; 20,000 cards.
- 40-F-79 Platte Valley Farmer-Livestock Feeder Survey (selected intervals) : Location within county, number of cattle and sheep fed, acres of irrigated cropland, acres of sugar beets. Reporting unit: farmers and livestock feeders. Unrestricted; 6 years; 1953, 1959; 9,000 cards.
- 40-F-80 Farm Adjustments on Wheat Farms (one-time survey) : Costs to produce wheat, barley, corn, sorghum, and grazing livestock for 1960 and estimated for 1970. Reporting unit: farm enterprise cost data. Confidential; 5 years; 1960, 1970; 500 cards.

- 40-F-81 Farm Machinery Costs by Size of Farm (one-time survey): Implementation, ownership costs, repair costs, use and estimated life of farm machines. Reporting unit: individual farmers on various sizes of farms. Confidential; 1960; 6,000 cards.
- 40-F-82 Great Plains Survey (one-time survey): Farm size, land tenure, land values, land use, and inventories. Reporting unit: individual farm. Unrestricted; 1957; 12,000 cards.
- 40-F-83 ASCS Survey, South Dakota (one-time survey): Farm size, land tenure, limited land use, livestock, and machinery inventory. Reporting unit: individual farm. Unrestricted; 1952; 20,000 cards.
- 40-F-84 Representative Farms in Indiana (one-time survey): Land, livestock, machinery, buildings, capital, farming practices and plans. Reporting unit: individual farm. Confidential; 5 years; 1962; 12,000 cards.
- 40-F-85 Wisconsin Farmers Home Administration Data (annual): Farm size, production, income and cost data. Reporting unit: farm and home records-individual borrowers. Confidential; permanent; 1957-62; 12,000 cards.
- 40-F-86 The Michigan Farm Credit Panel (one-time survey): Physical farm data and farm and family financial data. Reporting unit: farm families. Confidential; 2 years; 1961 3,000 cards.
- 40-F-87 NC-54 Feed Grain-Livestock Study (one-time survey): Resources, personal characteristics, and financial position. Reporting unit: Stratified random sample of farms in lower Michigan and Northern Indiana. Confidential; permanent; 1960; 3,600 cards.
- 40-F-88 Lake States Dairy Study (one-time survey): Farm resources and enterprises, personal characteristics of operator. Reporting unit: random sample farmers in lower Michigan. Confidential; permanent; 1958; 3,600 cards.
- 40-F-89 Lake States Dairy Study—Michigan Drainage Study (one-time survey): Resurvey of 1959 dairy study farms in 2 areas, resources enterprises, drainage conditions, inventory, personal characteristics of operator. Reporting unit: individual farm. Confidential; permanent; 1962; 2,500 cards.
- 40-F-90 Census Data Hay and Silage Yields (selected intervals): Total tons and total acres of each hay and silage crop by counties in the United States from U.S. Census of Agriculture for years 1949 and 1954. Reporting unit: total tons and acres of each hay and silage crop in U.S. census by counties for 1949 and 1954. Unrestricted; 4 years; 1949-54; 20,000 cards.
- 40-F-91 North Central Iowa Farm Building Survey (one-time survey): Inventory of buildings and their use, cost, repairs, crop and livestock production. Reporting unit: farmers in north-central Iowa. Confidential; 15 years; 1963; 25,000 cards.
- 40-F-92 Feed Grain Program Study (one-time survey): Cropland use, livestock, practices in farming, factor relating to participation in feed grain program. Reporting unit: farms in Pacific Northwest and Corn Belt and Texas. Unrestricted; 5 years; 1963; 30,000 cards.
- 40-F-93 Survey of Pilot Cropland Conversion Program in Iowa (one-time survey): Cropland use in 1962 and 1963, livestock programs, reasons for and effect of participation in 1963 cropland conversion program. Reporting unit: farms in Polk and Dallas Counties, Iowa. Unrestricted; 5 years; 1963; 1,500 cards.
- 40-F-94 Economic Evaluation of Alternative Systems on Corn Belt Farms (one-time survey): Costs in harvesting, hauling, drying conditioning storage, and utilization of shelled corn. General farm organization, detailed account of corn production, record of labor and equipment. Reporting unit: Illinois farmers. Confidential; 5 years; 1961-62.
- 40-F-95 An Economic Appraisal of the Use of Water for Irrigation on Illinois Farms: Confidential; 5 years.
- 40-F-96 Minimum Resources for Specified Incomes (one-time survey): Costs and returns detailed in such a manner as to permit selection of minimum resources to produce incomes of \$2,500, \$3,500, \$45,000, and \$5,500 on corn farms and hog farms with land prices established at three different levels. Reporting unit: budgeted data for corn farms and hog farms of specific income levels. Unrestricted; 5 years; 1959-63.

- 40-F-97 Field Shelling Mechanical Drying and Storing Shelled Corn in Illinois (one-time survey): Mail questionnaire on punchcards. General characteristics of farms with specific information on corn harvesting, drying, and storage equipment. Reporting unit: data from 301 farms. Unrestricted; permanent; 4,000 cards.
- 40-F-98 Crop Yields, Acreage, Pires, and Gross Income (annual): Data by counties in Illinois. Reporting unit: Illinois Cooperative Crop Reporting Service. Unrestricted; permanent; 1925-63; 6,000 cards.
- 40-F-99 Livestock and Poultry Numbers (annual): Data by counties in Illinois. Reporting unit: Illinois Cooperative Crop Reporting Service. Unrestricted; permanent; 1925-63; 6,000 cards.
- 40-F-100 Feed Grain Program in Western Ohio (one-time survey): Crop acreages, yields, animal-unit of livestock, certain crop costs, fertilizer use, etc. Reporting unit: 160 sample farms. Unrestricted; 3 years; 1962; 2,200 cards.
- 40-F-101 Farm Size and Cotton Allotment Data—Missouri Delta (one-time survey): Acres of farmland, acres of cropland, and acres of cotton allotment. Reporting unit: Agricultural Stabilization and Conservation Service farm contract units. Confidential; 5 years; 1963; 10,000 cards.
- 40-F-102 The Effect of Selected Weather Variables on Corn Yields (decennial): Corn yields, date of planting, date of tasseling, drought-day, precipitation, and average temperature. Reporting unit: Columbia and Sikeston, Mo. Unrestricted; 5 years; 1955-63; 3,000 cards.
- 40-F-103 Data From Soils Testing (one-time survey): Physical characteristic of soil, crop yields, and fertilizer treatment. Reporting unit: information sheet for soil samples reported by farmers for each soil sample. Confidential; 8-10 years; 1956-60; 4,500 cards.
- 40-F-104 Conservation Needs Inventory Data for Arkansas (one-time survey): Principal physical characteristics of the land, such as slope, soil capability, soil type, and land use and the acres associated with each characteristic. Reporting unit: randomly selected plots 40 or 160 acres in size. Unrestricted; 8-10 years; 1958; 30,000 cards.
- 40-F-105 Capital Requirements and Ownership Costs, Arkansas Rice Farms (selected intervals): Location, tenure, size unit, enterprise organization, machine inventory, replacement practices, etc. Reporting unit: individual farm operator. Confidential; 8-10 years; 1959, 1961; 1,000 cards.
- 40-F-106 Organization and Operation of Texas Rice Farms (one-time survey): Location, tenure, size unit, major land use, enterprise organizations, requirements, and returns. Reporting unit: individual farm operators. Unrestricted; 8-10 years; 1960; 1,000 cards.
- 40-F-107 Input-Output Data, Texas, Crop and Livestock Farm Enterprises: Data developed pertains to farm resource restrictions, crop and livestock production requirements, yield levels, and product prices. Reporting unit: no respondents, data worked up from publications pertaining to resource requirements, production costs, and yields of crop and livestock farm enterprises. Unrestricted; 1 year; 1957.
- 40-F-108 Oregon Wheat Study (one-time survey): Total acreage, wheat acreage, allotment, and normal yield from wheat listing sheets. Reporting unit: individual farms. Unrestricted; permanent; 1955-59; 400 cards.
- 40-F-109 ASCS Data on Farms in Southeastern Idaho (one-time survey): Total land, cropland, wheat allotment, feed grain base, etc. Reporting unit: farm firms. Unrestricted; 3 years; 1963; 2,500 cards.
- 40-F-110 Sample Survey, Farms in Southeastern Idaho (one-time survey): Land by types, crops grown, livestock, and livestock facilities. Reporting unit: farm firm. Unrestricted; 3 years; 1963; 1,200 cards.
- 40-F-111 The Farm Work Force in Kern County, Calif. (one-time survey): Characteristics of workers, employment, earnings, seasonality. Reporting unit: Farm Production Economics Division, ERS, USDA, University of California, Davis, Calif. Unrestricted; 1961; 3,800 cards.

- 40-F-112 The Farm Work Force in Stanislaus County, Calif. (one-time survey) : Characteristics of workers, employment, earnings, seasonality, migration. Reporting unit: Farm Production Economics Division, ERS, USDA, University of California, Davis, Calif. Unrestricted; 1962-63; 5,100 cards.
- 40-F-113 Irrigation Characteristics, Salinas Valley, Calif. (one-time survey) : Depth, pump lift, horsepower, discharge of each well. Reporting unit: individual well. Confidential: permanent; 1950-63; 3,000 cards.
- 40-F-114 Oregon Cattle Price Data (monthly) : Auction market location, type of cattle, average weight, price paid, date. Reporting unit: livestock auction markets. Confidential; 1964 to present; 2,500 cards.
- 40-F-115 Average Daily Gain and Feed Consumption, Cattle in Arizona (one-time survey) : Average daily gain, daily feed consumption, age, weight, and breeds of cattle; types of feed fed, etc. Reporting unit: farmers (Arizona cattle feeders). Confidential; 1 year; 1960-61; 48 cards.

DEPARTMENT OF COMMERCE

BUREAU OF CENSUS

- 41-A-1 National Location Code File (selected intervals) : Card file and tape file: place name, population total, geographic coordinates, census tract codes. Card file only: place name, population total, geographic coordinates, Universal Transverse Mercator (UTM) grid, reporting unit: standard location areas—tract and psuedo-tract. Unrestricted; permanent; 1960; 152,000 cards; two tapes.
- 41-A-2 Population Concentration (one-time survey) : Place name, geographic coordinates, population of selected points, total population within 50 miles. Reporting unit: selected cities in the United States with a total population of 25,000 or more. Unrestricted; permanent; 1960; two tapes.
- 41-A-3 County City Data Books (selected intervals) : Area, population, housing, vital statistics, manufacturers, trade, agriculture. Reporting unit: county, SMSA, urbanized area, unincorporated urban place, city of 25,000 or more. Unrestricted; permanent; 300,000 cards; three tapes.
- 41-A-4 Census of Agriculture County Summary (quinquennial) : County totals for a variety of inventory, production, expenditure, and sales items—also data about farm operator and farm. Reporting unit: farms within county; 100 and 20 percent tabulations. Confidential; 6 years; 1959; 44,000 cards.
- 41-A-5 Census of Agriculture/Sample and Specified Farm Cards (quinquennial) : Data on farms, farm characteristics, livestock and products, crops, fruits, values, etc. Reporting unit: farms. Confidential; 6 years; 1959; 8,900,000 cards.
- 41-A-6 Census of Agriculture (quinquennial) : Data on farms, farm characteristics, livestock and products, crops, fruits, values, etc. Reporting unit: farms. Confidential; permanent; 1964.
- 41-A-7 Special Census of Metropolitan Louisville (one-time survey) : Postal addresses and geographic and housing control items. Reporting unit: persons in households. Unrestricted; permanent; 1964; 300 tapes.
- 41-A-8 Census of Population and Housing (decennial) : Major characteristics of population and housing. Reporting unit: persons in households. Confidential; permanent; 1960; 732 tapes.
- 41-A-9 Census of Housing—25 percent sample (decennial) : Characteristics of occupied and vacant housing units. Reporting unit: housing unit. Confidential; permanent; 1960; 1,474 tapes.
- 41-A-10 Census of Population—Tallies—25 and 5 Percent Samples: Social and economic characteristics of persons, families, and households. Reporting unit: persons. Unrestricted; permanent; 1959-60; 4,072 tapes.
- 41-A-11 Census of Population—Basic Records for the 25 and 5 Percent Samples: Social and economic characteristics of persons, families, and households. Reporting unit: persons. Confidential; permanent; 1959-60; 7,297 tapes.

- 41-A-12 Census of Population—One in a Thousand and One in Ten Thousand Samples (decennial) : Social and economic characteristics of persons, families, and households. Reporting unit: persons. Unrestricted; permanent; 1959-60; 13 tapes.
- 41-A-13 Census of Population—One in a Thousand and One in Ten Thousand Samples (decennial) : Social and economic characteristics of persons, families, and households. Reporting unit: persons. Unrestricted; permanent; 1959-60; 198,000 cards; 7 tapes.
- 41-A-14 Censuses—Control and Identification Tapes (decennial) : Names of geographic entities for controlling and processing the censuses. Reporting unit: minor civil divisions. Unrestricted; permanent; 418 tapes.
- 41-A-15 Survey of Residential Alterations and Repairs (quarterly) : Types and costs of residential alterations and repairs. Reporting unit: housing units. Confidential; permanent; 1959-63; 360,000 cards; 10 tapes.
- 41-A-16 Housing Vacancy Survey (monthly) : Housing characteristics and vacancy status of vacant units. Reporting unit: vacant housing units. Confidential; permanent; 1959 to present; 100,000 cards; 120 tapes.
- 41-A-17 Congressional Districts (selected intervals) : Social and economic characteristics and housing. Reporting unit: congressional district. Unrestricted.
- 41-A-18 Survey of Intentions (quarterly) : Buying behavior and intentions on major household items. Reporting unit: household. Confidential; permanent; 1959 to present; 456,000 cards; 150 tapes.
- 41-A-19 Current Population Survey (monthly) : Labor force status, age, sex, veteran status, education, mobility, income, and housing. Reporting unit: persons in households. Confidential; permanent; 1959 to present; 2,720,000 cards; 350 tapes.
- 41-A-20 National and State Population Estimates and Forecasts (monthly) : Age, sex, births, deaths, and mobility for current or future years. Reporting unit: United States. Unrestricted; 2 years; 80,000 cards; 77 tapes.
- 41-A-21 Special Censuses (selected intervals) : Age, sex, and relationship. Reporting unit: persons in households. Confidential.
- 41-A-22 Housing Inventory Change and Residential Financing (one-time survey) : Components of inventory change and other housing characteristics. Reporting unit: housing units. Confidential; permanent; 1959-59; 302 tapes.
- 41-A-23 AC Summaries, Import Statistics of United States (monthly, quarterly, annual) : Quantity and dollar value at reporting level. Reporting unit: summarization of import entries in terms of commodity classification (schedule A-1959, August 1963; schedule TSUSA from September 1963 to date), country of origin (schedule C) including economic class and SITC from 1963. Unrestricted; permanent; 1960-64; 175 tapes.
- 41-A-24 No. 1 Summaries, Export and Intransit Waterborne Trade of United States (annual) : Quantity in pounds, value in dollars at reporting unit. Reporting unit: summarization of waterborne shippers export declaration in terms of type of vessel service. U.S. customs port of lading (schedule D), foreign port of unloading (schedule K), country of destination (schedule C), commodity (schedule S), flag vessel, engineer channel, trade area, U.S. coastal district. Confidential; permanent; 1958-63; 100 tapes.
- 41-A-25 No. 1 Summaries, Imports and Intransit Waterborne Trade of United States (annual) : Quantity in pounds, value in dollars at report level. Reporting unit: invoices. Confidential; permanent; 1958-63; 150 tapes.
- 41-A-26 ACD Summary—Import Statistics of United States (monthly, quarterly, annual) : Quantity of dollar value at reporting unit. Reporting unit: summarization of import entries in terms of commodity classification (schedule A 1959, August 1963, schedule TSUSA from September 1963 to date), country of origin (schedule C), customs district of entry (schedule D), traffic rate provision, type of entry (i.e., consumption or general statistics), subgroup and economic class and SITC from January 1963 to date. Unrestricted; permanent; 400 tapes.

- 41-A-27 BCD Summaries—Foreign Trade Export Air (annual) : Total quantity and value, U.S.-flag carriers, quantity and value for level stated (commodity, country, district). Reporting unit: summarization of shippers export declarations of airborne shipments in terms of commodity (schedule B), country (schedule C), and district of lading (schedule D). Unrestricted; permanent; 1962-64; 30 tapes.
- 41-A-28 CQC (Cotton Quota Control) Registers—Imports (monthly) : Separate registers for quantity and value, for commodity and country classification. Reporting unit: summarization of U.S. customs import entries in terms of commodity (schedule A, TSUSA and cotton quota control) and country of origin. Unrestricted; permanent; 1961 to present; 48 tapes.
- 41-A-29 ACD Summaries—Foreign Trade Import Air (annual) : Total quantity and value, U.S.-flag carriers—quantity and value for level stated (commodity, country, district). Reporting unit: summarization of import entries in terms of commodity (schedule A until September 1963), country of origin (schedule C), district of un-lading (schedule D). Unrestricted; permanent; 1962-64; 30 tapes.
- 41-A-30 BCD Summaries—Exports of Domestic and Foreign Merchandise (monthly) : Quantity pertaining to commodity and dollar value for levels of summary stated. Reporting unit: summarization of shippers, export declarations in terms of commodity classification (schedule B), country of destination (schedule C), district of lading (schedule D). Confidential; permanent; 1963 to present; 144 tapes.
- 41-A-31 BC Summaries—Exports of Domestic and Foreign Merchandise (monthly, quarterly, annual) : Quantity and value. Reporting unit: summarization of shippers export declarations in terms of commodity classification (schedule B), country of destination (schedule C). Confidential; permanent; 1958-63; 624 tapes.
- 41-A-32 Census of Manufactures (quinquennial) : Plant information, number of employees, payrolls, man-hours, cost of material, capital expenditures, products shipped, and materials used. Reporting unit: manufacturing establishment. Confidential; permanent; 1947; 2,500,000 cards.
- 41-A-33 Census of Manufactures and Mineral Industries (quinquennial) : Plant information, number of employees, payrolls, cost of materials, inventories, capital expenditures, products shipped, and materials used. Reporting unit: manufacturing establishments. Confidential; permanent; 1954; 5 million cards; 561 tapes.
- 41-A-34 Census of Manufactures and Mineral Industries (quinquennial) : Plant information, number of employees, payrolls, cost of materials, inventories, capital expenditures, products shipped, and materials used. Reporting unit: manufacturing establishments. Confidential; permanent; 1958; 4,500,000 cards; 752 tapes.
- 41-A-35 Economic Censuses—Retail (quinquennial) : Sales, annual payroll, weekly employment, and wages. Reporting unit: retail establishments. Confidential; permanent; 1954 and 1958; 198 tapes.
- 41-A-36 Economic Censuses—Wholesale (quinquennial) : Sales, annual payroll, weekly employment and wages, operating expenses, commodity lines. Reporting unit: wholesale trade establishments. Confidential; permanent; 1954 and 1958; 78 tapes.
- 41-A-37 Economic Censuses—Transportation (one-time survey) : Truck information: Physical characteristics and operational aspects. Reporting unit: individuals. Confidential; 2 years; 1963; 12,000 cards; 200 tapes.
- 41-A-38 Survey of Manufactures (annual) : Plant information, number of employees, payrolls, man-hours, cost of materials, inventories, capital expenditures, and products shipped. Reporting unit: manufacturing establishments. Confidential; permanent; 1949-62; 6,399,000 cards; 880 tapes.
- 41-A-39 Retail—CCBR—1 (monthly) : Sales of retail firms. Reporting unit: sample of retail trade firms. Confidential; permanent; 1962 to present; 750,000 cards; 420 tapes.

- 41-A-40 Wholesale (CTR) (monthly): Sales and inventories. Reporting unit: sample of merchant wholesale establishments. Confidential; permanent; 1962 to present; 216,000 cards, 144 tapes.
- 41-A-41 County Business Pattern (annual): Number of reporting units, first quarter payroll, and number of employees. Reporting unit: individual establishments or groups of establishments. Confidential; 3 years; 1962; 250 tapes.
- 41-A-42 Economic Censuses—Services (quinquennial): Receipts, annual payroll, weekly employment, and wages. Reporting unit: service trades establishments. Confidential; permanent; 1954 and 1958; 124 tapes.
- 41-A-43 Census of Governments—Governmental Units File (quinquennial): Population or enrollment; type of government or SMSA code. Reporting unit: State, county, city, etc., school systems. Unrestricted; permanent; 1962; 3 tapes.
- 41-A-44 Census of Governments—Local Government Directory Listing (quinquennial): Population or enrollment; type of government or SMSA code. Reporting unit: counties, municipalities, townships, and school districts. Unrestricted; permanent; 1962; 140,000 cards.
- 41-A-45 Census of Governments—Employment Data (quinquennial): Employment. Reporting unit: local governments and school systems. Unrestricted; permanent; 1962; 220,000 cards; 40 tapes.
- 41-A-46 Census of Governments—Assessed Valuations (quinquennial): Description and assessed value of approximately 1 million sample pieces of property. Reporting unit: city or county taxing jurisdiction. Unrestricted; permanent; 1962; 18 tapes.
- 41-A-47 Census of Government—Finance Data (quinquennial): Revenue, expenditure, debt outstanding, cash and investment assets. Reporting unit: cities, townships, special districts, and school systems. Unrestricted; permanent; 1962; 84 tapes.
- 41-A-48 Current Survey Directory Testing (annual): Governmental unit identification and enrollment or population. Reporting unit: sample of local governmental units. Unrestricted; permanent; 1962; 1 tape.
- 41-A-49 Sample Employment Data (annual): Full-time employees and annual rate of pay. Reporting unit: local governments and school systems. Unrestricted; permanent; 1962; 30,000 cards; 4 tapes.
- 41-A-50 Retail Trade Survey (weekly): Retail sales. Reporting unit: retail establishments. Confidential; 1962 to present; 300,000 cards.
- 41-A-51 Accounts Receivable Survey (monthly): Receivables from charge accounts and installment sales. Reporting unit: sample of retail trade establishments. Confidential; permanent; 1963 to present; 240,000 cards; 384 tapes.
- 41-A-52 Retail Inventory Survey (RIS) (monthly): Inventory. Reporting unit: retail establishments and retail multiunit firms. Confidential; 39 months; 1961 to present; 72,000 cards.
- 41-A-53 Geographic Area Sales Survey (GASS) (monthly): Sales. Reporting unit: retail establishments. Confidential; permanent; 1962 to present; 128,000 cards; 448 tapes.
- 41-A-54 Service Trade Survey (monthly): Receipts for services. Reporting unit: service establishments. Confidential; 39 months; 1962 to present; 160,000 cards; 144 tapes.
- 41-A-55 Current Trade Survey (SSA Births) (quarterly): Sales and inventory. Reporting unit: wholesale establishments. Confidential; 39 months; 1961 to present; 7,500 cards.
- 41-A-56 Canned Food Survey (selected intervals): Inventory. Reporting unit: wholesale distributors and retail multiunit organizations. Confidential; 39 months; 10,800 cards.
- 41-A-57 Retail Trade Survey—Group 1 (annual): Sales and inventory. Reporting unit: retail establishments. Confidential; 39 months; 1961-63; 150,000 cards.
- 41-A-58 Retail Trade Survey—GROUP II (annual): Sales, inventory and capital expenditures. Reporting unit: retail establishments. Confidential; 39 months; 1961-63; 8,000 cards.
- 41-A-59 Capital Expenditures Survey—Wholesale Trade (quinquennial): Capital expenditures. Reporting unit: wholesale establishments. Confidential; 39 months; 1963; 28,500 cards.

- 41-A-60 Capital Expenditures Survey—Service trade (quinquennial) : capital expenditure. Reporting unit, service establishments. Confidential; 39 months; 1963.
- 41-A-61 Iron and Steel Foundries, Blast Furnaces and Steel Ingot Producers (monthly) : Shipments and unfilled orders for castings, production of steel castings by type of furnace and production of scrap, pig iron and iron ore by type of furnace. Reporting unit: 1,200 establishments. Confidential; 3 years, 1961-63; 116,136 cards.
- 41-A-62 Nonferrous Castings—M33E (monthly) : Shipments and unfilled orders nonferrous castings by type of casting. Reporting unit: 600 establishments. Confidential; 3 years; 1961-63; 74,916 cards.
- 41-A-63 Aluminum Producers and Importers (BDSAF-122) (annual) : Total receipts and shipments; shipments on ACM and rated orders; inventories of scrap, primary ingot and secondary ingot. Reporting unit: establishments. Confidential; 3 years; 1961-63; 44,496 cards.
- 41-A-64 Farm Machines and Equipment—M35A (quarterly) : Production, shipments and inventory by type of farm equipment; value by product class. Reporting unit: 140 establishments. Confidential; 3 years; 1961-63; 9,324 cards.
- 41-A-65 Metalworking Machinery—M35W (quarterly) : Shipments and unfilled orders (domestic and export) by type of machine, including numerical control type; units value. Reporting unit: 476 establishments. Confidential; 3 years; 1961-63; 35,208 cards.
- 41-A-66 Copper Forms and Products—Brass and Bronze Foundries (BDSAF83) (quarterly) : Inventories, receipts, and shipments of copper and copper-base alloy. For brass and bronze foundries: shipments of controlled materials; also, authorized controlled materials shipments by DMS allotment number. Reporting unit: 460 establishments. Confidential; 3 years; 1961-63; 28,404 cards.
- 41-A-67 Copper Controlled Materials—Brass Mills and Copper Wire Mills (BDSAF84) (quarterly) : Copper and copper-base alloy controlled materials shipments and unfilled orders; authorized controlled materials shipments by DMS allotment number. Reporting unit: 100 establishments. Confidential; 3 years; 1961-63; 7,836 cards.
- 41-A-68 Copper Forms and Products—Copper Base Powder Mills (BDSAF-574) (quarterly) : Inventories, receipts, and shipments of copper and copper-base alloy; shipments of copper and copper-base alloy controlled materials and authorized controlled materials; total copper-base powder shipments by type. Confidential; 3 years; 1961-63; 1,884 cards.
- 41-A-69 Manmade Fiber, Woolen and Worsted Fabrics (monthly) : Productions, stocks, and unfilled orders, gray goods and finished manmade fiber fabrics and wool apparel fabrics. Reporting unit: 200 to 250 manufacturers. Confidential; 3 years; 1961-63; 77,364 cards.
- 41-A-70 Finishing Plant Report—Broad Woven Fabrics (monthly) : Gray goods inventory, finished fabrics during the period, finished goods inventory and backlog of finishing orders. Reporting unit: 130 to 150 manufacturers. Confidential; 3 years; 1961-1963; 32,436 cards.
- 41-A-71 Piece Goods Inventories and Orders (monthly) : Converters, wholesalers, jobbers, and other dealers: inventories owned by the company according to location and unfilled orders for gray and finished goods. Reporting units: 200 to 225 companies. Confidential; 3 years; 1961-63; 22,896 cards.
- 41-A-72 Tufted Textile Fabrics—M22L (semiannual) : Tufting machines; yarns and fabrics consumed in manufacture of auto and aircraft carpeting. Reporting unit: 1950 to 170 manufacturers. Confidential; 3 years; 1963; 702 cards.
- 41-A-73 Cotton and Linters in Public Storage and at Compresses—M22N (monthly) : Raw cotton and linters inventory. Reporting unit: 1,200 to 1,220 companies. Confidential; 3 years; 1961-63; 101,592 cards.
- 41-A-74 Cotton, Manmade Fiber Staple and Linters—M22P (monthly) : Consumption, stocks and spindle activity. Reporting unit : 850 to 875 companies. Confidential; 3 years; 1961-63; 135,468 cards.

- 41-A-75: Men's Apparel Cuttings and Shipments—M23B (monthly) : Cuttings, shipments and value of shipments. Reporting unit: 630 to 650 manufacturers and contractors. Confidential; 3 years; 1961-63; 87,480 cards.
- 41-A-76 Women's, Misses, and Juniors' Apparel—Cuttings and Shipments—M23H (monthly) : Cuttings, shipments and value of shipments. Reporting unit: 1,190 to 1,210 manufacturers and contractors. Confidential; 3 years; 1961-63; 151,272 cards.
- 41-A-77 Shoes and Slippers, Production and Shipments—M31A (monthly) : Shoes and slippers except those with sole vulcanized to fabric upper. Shoes and slippers with sole vulcanized to fabric upper. Value of shipments. All other products. Reporting unit: 470 to 490 manufacturers. Confidential; 3 years; 1961-63; 114,732 cards.
- 41-A-78 Broad Fabrics (Expect knit), Woven, Nonwoven and Felts—M22T (quarterly) : Loom hours, yarns consumed; looms in place and active; number of looms operating and production; and stocks of selected items. Reporting unit: 690 to 710 manufacturers. Confidential; 3 years; 1961-63; 155,184 cards.
- 41-A-79 Mattresses and Bedsprings—M25E (monthly) : Quantity and value of shipments. Reporting unit; 225 manufacturing firms. Confidential; 3 years; 1961-63; 8,900 cards.
- 41-A-80 Pulp Paper and Board—M26A (monthly) : Production—inventories and consumption. Reporting unit; 665 manufacturing establishments. Confidential; 3 years; 1961-63; 121,000 cards.
- 41-A-81 Inorganic Chemicals—M28A.1 (monthly) : Production and inventories. Reporting unit: 495 manufacturing establishments. Confidential; 3 years; 1961-63; 51,000 cards.
- 41-A-82 Industrial Gases—M28A.2 (monthly) : Production. Reporting unit: 620 manufacturing establishments. Confidential; 3 years; 1961-63; 45,000 cards.
- 41-A-83 Paint, Varnish and Lacquer—M28F (monthly) : Production and sales. Reporting unit: 225 manufacturing establishments. Confidential; 3 years; 1961-63; 3,600 cards.
- 41-A-84 Rubber—M30A (monthly) : Production receipts—shipments consumption—inventories. Reporting unit: 390 manufacturing establishments. Confidential; 3 years; 1961-63; 62,000 cards.
- 41-A-85 Rubber Consumption by Product Group—M30B (quarterly) : Consumption. Reporting unit: 39 manufacturing establishments. Confidential; 3 years; 1961-63; 62,000 cards.
- 41-A-86 Plastic Bottles—M30E (monthly) : Shipments by end use. Reporting unit: 79 manufacturing establishments. Confidential.
- 41-A-87 Oilseeds, Beans and Nuts—Report of Primary Processors—M20J (monthly) : By type—quantity of seeds or beans, crushed oil production, and inventories. Reporting unit: 323 establishments. Confidential; 3 years; 1961-63; 93,024 cards.
- 41-A-88 Animal and Vegetable Fats and Oils, Inventories—M20H (monthly) : End of month inventories—by types of oil. Reporting unit: 264 warehouses. Confidential; 3 years; 1961-63; 21,744 cards.
- 41-A-89 Fats and Oils—Report of Renderers—M20L (monthly) : Rendered production, shipments and inventories, by type. Reporting unit: 430 establishments. Confidential; 3 years; 1961-63; 61,992 cards.
- 41-A-90 Animal and Vegetable Fats and Oils—Report of Consumers—M20M (monthly) : Types and quantities of oils produced, consumed and inventories. Reporting unit: 503 establishments. Confidential; 3 years; 1961-63; 142,308 cards.
- 41-A-91 Animal and Vegetable Fats and Oils—Report of Producers and Consumers—M20N (monthly) : Production, consumption and stock on hand—by type. Reporting unit: 365 establishments. Confidential; 3 years; 1961-63; 145,332 cards.
- 41-A-92 Confectionery Survey—MA20D (annually) : Types of products produced, net shipments, and type of customer (to whom sold), and cost and type of materials consumed. Reporting unit: 244 manufacturers. Confidential; 3 years; 1961-63; 16,047 cards.
- 41-A-93 Formula Feeds for Poultry and Livestock—MA20E (annual) : Production and shipments by type. Reporting unit: 3,077 manufacturers. Confidential; 3 years; 1961-63; 16,698 cards.

- 41-A-94 Salad Dressing, Mayonnaise and Related Products—MA20F (annual): Production by type and size of container, value of shipments, number of brands and oils consumed. Reporting unit: 135 establishments. Confidential; 3 years; 1961-63; 17,475 cards.
- 41-A-95 Fats and Oils—Report of Renderers—MA20L (annual): Production, consumption, and inventories. Reporting unit: 1,170 establishments. Confidential; 3 years; 1961-63; 10,182 cards.
- 41-A-96 Animal and Vegetable Fats and Oils—Report of Producers and Consumers—MA20N (annual): Production, consumption and inventories by type. Reporting unit: 1,142 establishments. Confidential; 3 years; 1961-63; 16,062 cards.
- 41-A-97 Woolen and Worsted Machinery Activity—MA22E (annual): Spindles in place and active the last full working day of year, and worsted combs and top to tow converters in place and active on last workday of year. Reporting unit: 350 establishments. Confidential; 3 years; 1961-63; 1,518 cards.
- 41-A-98 Yarn Production—MA22F (annual): Production data for yarn spun from staple, tow and uncut top and for stretch, textured, crimped or bulked filament yarns. Reporting unit: 1,000 to 1,200 companies. Confidential; 3 years; 1961-63; 12,000 cards.
- 41-A-99 Narrow Fabrics—MA22G (annual): Production; materials consumed; narrow fabric machinery in place at end of year; value of shipments. Reporting unit: 300 establishments. Confidential; 3 years; 1962-63; 5,686 cards.
- 41-A-100 Knit Cloth for Sale—MA22K (annual): Shipments of knit cloth for sale; yarns consumed in production. Reporting unit: 300 to 400 establishments producing knit cloth for sale. Confidential; 3 years; 1961-63; 5,949 cards.
- 41-A-101 Tufted Textile Products—MA22L (annual): Tufting machinery in place at end of year; automobile and aircraft carpeting, quality and value of shipments and yarns consumed. Reporting unit: 200 manufacturers. Confidential; 3 years; 1961-63; 1,053 cards.
- 41-A-102 Stocks of Wool and Related Fibers—MA22M (annual): Stock of foreign and domestic wool and stocks of related fibers and tops and noil of January 1 of each year. Reporting unit: 500 to 600 establishments. Confidential; 3 years; 1961-63; 6,000 cards.
- 41-A-103 Cotton and Linters in Public Storage and at Compresses—MA22N (annual): Stocks on hand on July 31 of each year; material in transit on July 31 and destroyed during the year. Reporting unit: 150 to 250 establishments that do not report on the monthly M22N. Confidential; 3 years; 1961-63; 450 cards.
- 41-A-104 Cotton, Manmade Fiber Staple, and Linters—MA22P (annual): Consumption during year and stocks at end of year; spindles in place and active and spindle hours operated during the year; destroyed during season. Reporting unit: 250 to 300 establishments that do not report on the monthly M22P. Confidential; 3 years; 1961-63; 1,878 cards.
- 41-A-105 Rugs, Carpets and carpeting—MA22Q (annual): Shipments and interplant transfers; yarns and fabrics consumed in the manufacturing of specified types of carpets and rugs; machinery in place at yearend. Reporting unit: 100 to 125 establishments. Confidential; 3 years; 1962-63; 700 cards.
- 41-A-106 Cotton, Silk and Manmade Fiber Woven Goods Finished—MA22S (annual): Fabrics finished by end use, for crease resistance or wash-wear properties and against military contracts. Reporting unit: 350 to 400 finishing plants. Confidential; 3 years; 1961-63; 14,436 cards.
- 41-A-107 Broad Woven Goods, Cotton, Wool, Silk and Manmade Fiber—MA22T (annual): Production of wool woven fabrics, number of looms operating on last working day; looms in place and active by shifts; loom hours; yarns consumed; stocks. Reporting unit: 200 to 225 established not reporting on the quarterly MA22T. Confidential; 3 years; 1961-63; 3,414 cards.
- 41-A-108 Apparel Survey—MA23A (annual): Production and value of shipments, knit yarns consumed. Reporting unit: 6,000 to 7,000 manufacturers and jobbers of apparel. Confidential; 3 years; 1961-63; 167,385 cards.

- 41-A-109 Brassiers, Corsets and Allied Garments Shipments—MA23J (annual) : Quantity and value of shipments. Reporting unit: 200 to 250 producers. Confidential; 3 years; 1961-63; 6,309 cards.
- 41-A-110 Brassiers, Corsets and Allied Garments—Distribution of Sales—MA23J (annual) : Distribution of manufacturers sales by class of customer. Reporting unit: 175 to 200 companies reporting on the MA23J which had a value of shipments of over \$250,000 the previous year. Confidential; 3 years; 1961-63; 2,826 cards.
- 41-A-111 Hardwood Plywood—MA24F (annual) : Products and shipments. Reporting unit: 265 manufacturing establishments. Confidential; 3 years; 1961-63; 3,600 cards.
- 41-A-112 Softwood Plywood—MA24H (annual) : Production and consumption. Reporting unit: 153 manufacturing establishments. Confidential; 3 years; 1961-63; 2,850 cards.
- 41-A-113 Softwood Veneer—MA24K (annual) : Production and consumption. Reporting unit: 88 manufacturing establishments. Confidential; 3 years; 1961-63; 2,450 cards.
- 41-A-114 Mattresses and Bedsprings—MA25E (annual) : Quantity and value of shipments. Reporting unit: 30 to 40 manufacturing establishments. Confidential; 3 years; 1961-63.
- 41-A-115 Pulp, Paper and Board—MA26C (annual) : Receipts, production, inventories, consumption, shipments. Reporting unit: 717 manufacturing establishments. Confidential; 3 years; 1961-63; 37,500 cards.
- 41-A-116 Converted Flexible Packaging Products—MA26F (annual) : Quantity and value of shipments. Reporting unit: 125 manufacturing establishments. Confidential.
- 41-A-117 Sulfuric Acid—MA28B (annual) : Production 1-1 quantity and value of shipments. Reporting unit: 219 manufacturing establishments. Confidential; 3 years; 1961-63; 900 cards.
- 41-A-118 Shipments and Production of Inorganic Chemicals—MA28E.1 (annual) : Production, consumption, quantity, and value of shipments. Reporting unit: 995 manufacturing establishments. Confidential; 3 years; 1961-63; 8,300 cards.
- 41-A-119 Shipments and Production of Industrial Gases—MA28E.2 (annual) : Production, shipments, and consumption. Reporting unit: 696 manufacturing establishments. Confidential; 3 years; 1961-63; 5,300 cards.
- 41-A-120 Pharmaceutical Preparations, Except Biologicals—MA28G (annual) : Value of shipments. Reporting unit: 1,100 manufacturing establishments. Confidential; 3 years; 1961-63; 45,000 cards.
- 41-A-121 Plastic Products—MA30D (annual) : Quantity and value of shipments and consumption. Reporting unit: 2,846 manufacturing establishments. Confidential; 3 years; 1961-63; 65,000 cards.
- 41-A-122 Plastic Bottles—MA30E (annual) : Shipments of plastic bottles by end use. Reporting unit: 60 manufacturing establishments. Confidential.
- 41-A-123 Shoe and Slipper Production and Shipments—MA31A (annual) : Production and shipments of shoes and slippers and value of shipments. Reporting unit: 450 to 500 establishments that do not report on the monthly M31A. Confidential; 3 years; 1961-63; 18,960.
- 41-A-124 Steel Mill Products—MA33B (annual) : Quantity of receipts, production and products consumed in manufacture; quantity and value of interplant transfers, carbon steel, alloy, steel, stainless steel and conversion steel. Reporting unit: 483 establishments. Confidential; 3 years; 1961-63; 11,946 cards.
- 41-A-125 Steel Power Boilers—MA34G (annual) : Orders booked by type of boiler. Reporting unit: 100 establishments, manufacturing. Confidential; 3 years; 1961-63.
- 41-A-126 Heating and Cooking Equipment (Except Electric)—MA34N (annual) : Quantity and value of shipments and yearend inventory for gas heating stoves, gas ranges, oil burners, furnaces, stokers, and water heaters. Reporting unit: 450 establishments. Confidential; 3 years; 1961-63; 7,263 cards.

- 41-A-127 Farm Machines and Equipment—MA35A (annual): Farming machinery—production, number and value of shipments (domestic and export). Reporting unit: 1,100 establishments. Confidential; 3 years; 1961-63; 16,119 cards.
- 41-A-128 Construction Machinery—MA35D (annual): Shipments (domestic and export) in number units and value. Reporting unit: 88 establishments. Confidential; 3 years; 1961-63; 2,400 cards.
- 41-A-129 Mining Machinery—MA35F (annual): Total shipments (domestic and export)—quantity and value. Reporting unit: 190 establishments. Confidential; 3 years; 1961-63; 1,746 cards.
- 41-A-130 Internal Combustion Engines—MA35L (annual): Quantity and value by model number and engine specifications of shipments to other companies or transferred to other plants of the same company; engines converted from engines received from other plants. Reporting unit: 98 establishments. Confidential; 3 years; 1961-63; 5,868 cards.
- 41-A-131 Air-Conditioning and Refrigeration Equipment—MA35M (annual): Total shipments in number of units and value in dollars. Reporting unit: 254 establishments. Confidential; 3 years; 1961-63; 10,014 cards.
- 41-A-132 Pumps and Compressors—MA35P (annual): Pumps, compressors (shipments in number of units, and value of driven units). Reporting unit: 442 establishments. Confidential; 3 years; 1961-63; 8,298 cards.
- 41-A-133 Office, Computing and Accounting Machines—MA35R (annual): Cash registers and data processing machines; typewriters; duplicating machines; number shipped; f.o.b. plant value, and retail list price. Reporting unit: 220 establishments. Confidential; 3 years; 1961-63; 4,377 cards.
- 41-A-134 Metal Working Machinery: Metal Cutting and Metal Forming Types—MA35W (annual): Shipments and unfilled orders (domestic and export) by type of machine, including numerical control type. Reporting unit: 476 establishments. Confidential; 3 years; 1961-63; 35,208 cards.
- 41-A-135 Switchgear, Switchboard Apparatus, Relays, and Industrial Controls—MA36A (annual): Value of shipments; switchgear, circuit breakers, low-voltage panelboards and distribution boards; fuses and fuse equipment under 2,300 volts; circuit relays. Reporting unit: 573 establishments. Confidential; 3 years; 1961-63; 7,452 cards.
- 41-A-136 Electric Housewares and Fans—MA36E (annual): Total shipments, quantity and value, of electric fans (non-industrial); electric razors, and small household electric cooking and heating appliances. Reporting unit: 319 establishments. Confidential; 3 years; 1961-63; 2,616 cards.
- 41-A-137 Motors and Generators—MA36H (annual): Fractional horsepower motors; integral horsepower motors and generators other than for land transportation equipment; land transportation motors, generators, and control equipment and parts; prime mover generator sets; rotating equipment. Number and value of shipments and interplant transfers. Reporting unit: 279 establishments. Confidential; 3 years; 1961-63; 5,223 cards.
- 41-A-138 Wiring Devices and Supplies—MA36K (annual): Current carrying and noncurrent carrying devices—quantity and value. Reporting unit: 383 establishments. Confidential; 3 years; 1961-63; 3,564 cards.
- 41-A-139 Lighting Fixtures—MA36L (annual): Value of shipments—electric lighting fixtures (residential, commercial, and institutional types); vehicular lighting equipment; outdoor lighting equipment including components and parts. Reporting unit: 771 establishments. Confidential; 3 years; 1961-63; 8,295 cards.
- 41-A-140 Selected Electronic and Associated Products—MA36N (annual): Shipments of electronic component parts, equipment and systems. Reporting unit: 1977 establishments. Confidential; 3 years; 1961-63; 22,464 cards.

- 41-A-141 Selected Instruments and Related Products—MA38B (annual) : Value of shipments—test, measuring, and analyzing equipment for electronic and electrical circuits; industrial process instruments; selected analytical instruments and equipment. Reporting unit: 1,541 establishments. Confidential; 3 years; 1961-63; 5,472 cards.
- 41-A-142 Atomic Energy Products and Services—MA38Q (annual) : Nuclear reactors; reactor components and equipment—value of shipments. Reporting unit: 300 establishments. Confidential; 3 years; 1961-63; 1,497 cards.
- 41-A-143 Particle Board—BDSAF582 (annual) : Production. Reporting unit: 58 manufacturing establishments. Confidential; 3 years; 1961-63; 1,000 cards.
- 41-A-144 Lumber Survey—MA24T (annual) : Production and stocks by species. Reporting unit: 4,500 sawmills. Confidential; permanent; 1954 to present; 2,700 cards; 500 tapes.
- 41-A-145 Manufacturers' Shipments, Inventories and Orders Survey—M3 (monthly) : Value of shipments, new orders, unfilled orders, total inventory—materials, and supplies goods in process, and finished goods. Reporting unit: single unit: single unit and divisional manufacturing establishments. Confidential; permanent; 1960-64; 252,000 cards; 48 tapes.
- 41-A-146 Survey of Research and Development—RD1 and RD2 (annual) : Company funds expended resulting from research and development, company net sales, etc. Reporting unit: 3,000 companies collecting and compiling R. & D. data for the National Science Foundation. Confidential; permanent; 1957-63; 21,000 cards; 10 tapes.
- 41-A-147 Nonpermit Construction Starts (monthly) : Number of housing units started in selected nonpermit primary sampling units (PSU's) inside or outside a segment date of start and intent of building (i.e., for sale, rent, etc.). Reporting unit: 800-200 owners or contractors. Confidential; 12 years; 1959 to present; 50,000 cards.
- 41-A-148 Reports of Building Permits Issued (monthly) : Number of residential buildings, housing units, and valuation by type of structure and total number of buildings, and valuation for each non-residential item reported. Reporting unit: 4,300 building or zoning officials of selected (sample) permit issuing municipalities, towns, cities, or villages. Unrestricted; 12 years; 1955 to present; 1,040,000 cards.
- 41-A-149 Residential Building Permit Use Survey (monthly) : Number of units started in selected permit issuing places, date of start, type of structure (i.e., 1-family, etc.) and intent of building (i.e., for sale, rent or exclusive use). Reporting unit: 5,000-6,000 permit offices. Confidential; 12 years; 1960 to present; 300,000 cards.
- 41-A-150 Construction Progress Report Survey (monthly) : Type of construction, total cost (i.e., earnings, materials, etc.) and date of start. Reporting unit: 30,000 new construction owners or contractors via (F.2. Dodge data collecting agency by questionnaires. Confidential; 12 years; 1960 to present; 240,000 cards; 56 tapes.
- 41-A-151 Housing Sales Survey (monthly) : Number of units sold and unsold, stage of construction (i.e., started, completed, or not started) at time of sale, sale price, and type of finance (i.e., FHA, VA, conventional). Reporting unit: 6,000-6,5000 permit offices and camera/contractors of building sites in nonpermit places via a field enumerator. Confidential; 12 years; 1961 to present; 288,000 cards.
- 41-A-152 Report of Building permits Issued (annual) : Number of buildings, housing units, and valuation of type of structure, private and public. Reporting unit: 8,000 building or zoning officials of permit-issuing places other than those canvassed monthly. Unrestricted; 12 years; 1960-63; 175,000 cards.

BUREAU OF INTERNATIONAL COMMERCE

- 41-B-1 World Trade Director Reports File, Japan (selected intervals) : Describes foreign company, products handled, manufacturers, size, reputation, capital, annual turnover and other pertinent facts. Reporting unit: foreign posts, based on their contracts with foreign companies. Unrestricted; 5 years; 22,000 cards.

- 41-B-2 American Traders Identification File (annual): Investment interest, number of employees, annual sales import banks ref., SIC nos., address, principal officer. Reporting unit: American companies on form IA-57. Confidential; 5 years; 12,000 cards.
- 41-B-3 Dun & Bradstreet "Million Dollar Directory" Card Check (one-time survey): Number of employees, annual sales, capital, principal officer and SIC number. Reporting unit: Dun & Bradstreet based on their contacts with American companies. Confidential; 3 years; 60,000 cards.
- 41-B-4 Tariff and Trade Negotiation (Gatt) Committees (one-time survey): Tariff information and import statistics. Reporting unit: data based on government publications. Confidential; permanent; 1961-63; 250,000 cards.
- 41-B-5 Corporation Income Tax Returns (annual): Balance sheet and income statement items and asset size classes. Reporting unit: U.S. companies via Internal Revenue Service. Confidential; Permanent; 1959-60; 112 tapes.
- 41-B-6 Trade of OECD Countries and Japan (annual): Imports and exports between 20 foreign countries. Reporting unit: OECD and United Nations-data based on country reports. Unrestricted; permanent; 1961-62; 20 tapes.
- 41-B-7 U.S. Exports of Domestic and Foreign Merchandise (annual): Quantity and value by schedule B numbers for countries of destination. Reporting unit: exporters reports to collectors of custom on shipments abroad. Unrestricted; permanent; 1962; 10 tapes.
- 41-B-8 World Trade Directory Reports File, Japan (revised) (annual): Describes foreign company, products handled, manufacturer, size, reputation, capital, annual turnover and other pertinent facts. Reporting unit: foreign posts, based on their contacts with foreign companies. Unrestricted; 5 years; 75,000 cards.
- 41-B-9 U.S. Imports of Merchandise (annual): Quantity and value by schedule A by countries of origin. Reporting unit: importers report to collectors of customs on shipments arriving in the United States. Unrestricted; permanent; 1962; three tapes.

BUREAU OF MINES

- 42-1 Mine and Quarry Survey (annual): Footage of exploration and development, quality of ore mines and mining methods. Reporting unit: quantities shortages, long tons. Confidential; permanent; 1958-63; 275,000 cards; 75 tapes.
- 42-2 Clay (annual): Reports on quantity and value used and quantity and value sold, total quantity used and sold and value total used and sold. Reporting unit: producers of clay. Confidential; 1 year; 1962-63; 10,000 cards.
- 42-3 Sand and Gravel Canvass (annual): Production figures of various classes S&G by government and commercial producers and users. Reporting unit: sand and gravel producers & users (commercial and non-commercial). Confidential; permanent; 1961-63; 60,000 cards, 7 tapes.
- 42-4 Water Canvass (quinquennial): Water consumption, production, treatment, source of water recirculation. Reporting unit: various mineral producers. Unrestricted; permanent; 1962; 86,000 cards; 16 tapes.
- 42-5 Iron and Steel (scrap iron) (monthly): Shipments order for casting, production of steel castings and ingots by type of furnace, consumption of scrap, pig iron, iron ore, stocks, production, receipts and consumption by grades of scrap. Reporting unit: users of scrap iron. Unrestricted; 1 year; 1963-64; 84,000 cards.
- 42-6 Bituminous and anthracite coal (monthly): Employment, number of operators, days active, days lost, number of injuries, man-hours, man-days and production. Unrestricted; 1 year; 25,000 cards.
- 42-7 Bituminous and Anthracite Coal Canvass (annual): Annual bituminous and anthracite survey covering employment, days and hours worked, disabling injuries and production. Reporting unit: coal producers. Unrestricted; 10 years; 245,000 cards.

- 42-8 Stone Quarries Canvass (annual): Survey of the industry covering number of operations, employment, man-hours worked, days active, length of shift, disabling injuries, including fatalities. Reporting unit: quarry operators. Confidential; 8 years; 80,000 cards.
- 42-9 Petroleum Injury Experience (annual): Data from oil companies, including oil, and gas-well drilling contractors, oil- and gas-field service contractors, and natural gas companies. Reporting unit: petroleum producers. Unrestricted; 8 years; 60,000 cards.
- 42-10 Sand and Gravel Injury Experience (annual): Employment, man shifts, man-hours, injuries, days lost, degree of injuries and number of plants. Reporting unit: sand and gravel operators and users. Unrestricted; 6 years; 66,000 cards.
- 42-11 Metal Industries Including Placer (annual): Same as nonmetal industry. Reporting unit: metal operators and producers. Confidential; 8 years; 70,000 cards.
- 42-12 Nonmetal Industries (except quarries and coal) (annual): Number of men employed, days and hours worked, number of mines, injuries, days lost, degree of injury, nature and part of body. Reporting unit: nonmetal operators and producers. Confidential; 8 years; 71,000 cards.
- 42-13 Manufacturers and Coal Retailers (monthly): Tons on hand, tons received during month, tons consumed and tons on hand at end of month. Reporting unit: consumers of coal. Unrestricted; 2 years; 24,000 cards.

DEPARTMENT OF LABOR

BUREAU OF EMPLOYMENT SECURITY

- 44-A-1 Teuc (Temporary Extended Unemployment Compensation) family characteristics study (one-time survey): Personal characteristics (age, sex, number in household, etc.), economic characteristics (industry, occupation, work history), and unemployment insurance experience (weekly benefit, duration of benefits, etc.). Reporting unit: individuals-claimants for Teuc. Confidential; 3 years; 1961-62; 230,000 cards.
- 44-A-2 ES-212 nonagricultural placements, standard metropolitan areas (monthly): Major occupation group—total, veteran, under 22, 45 and over. Industry division and 2-digit manufacturing—total, veteran, 45 and over. Reporting unit: 65 selected SMSA reports. Unrestricted; permanent; 1962-63; 73,000 cards.
- 44-A-3 ES-212 nonagricultural placements, State summaries (monthly): Major occupation group—total, female, veteran, age under 22, age 45 and over. Industry division and 2-digit manufacturing—total, female, veteran, age 45 and over. Reporting unit: State summary reports (54 State agencies). Unrestricted; permanent; 1957-63; 90,900 cards.
- 44-A-4 ES-211 employer information transcript (monthly): Employment past, current, future; turnover and placements; current hours and wage rate. Reporting unit: individual firms for selected industries. Confidential; permanent; 1959-63; 32,000 cards.
- 44-A-5 ES-209 supplement—service to selected age groups (monthly): New applications, active file, initial counsel interview, nonagricultural placements—total, female, age under 22, age 45-64. Reporting unit: State summary reports (54 State agencies). Unrestricted; permanent; 1962-63; 5,400 cards.
- 44-A-6 ES-209 local office activities, standard metropolitan areas (monthly): Applicant services, employer services, nonagricultural placement activities, claims-taking activities—total, veteran. Reporting unit: 65 selected SMSA reports. Unrestricted; permanent; 1962-63; 54,100 cards.
- 44-A-7 ES-209 local office activities, State summaries (monthly): Applicant services, employer services, nonagricultural placement activities, agricultural placement activities, claims activities—total, female and veteran. Reporting unit: State summary reports (54 State agencies). Unrestricted; permanent; 1962-63; 54,100 cards.

- 44-A-8 ES-202 employment and wages of workers covered by UI and UCFE laws (quarterly): Reporting units, 3-month employment, total wages, taxable wages and contributions. Reporting unit: State summary reports (51 State agencies). Confidential; permanent; 1958-63; 1,810,000 cards; 24 tapes.
- 44-A-9 ES-203 characteristics of the insured unemployed (monthly): Age and sex of the insured employed; sex and duration by industry; industry and occupation. Reporting unit: insured unemployed individuals. Unrestricted; permanent; 1958-63; 540,000 cards.

BUREAU OF LABOR STATISTICS

- 44-B-1 Wholesale price index (monthly): Wholesale price and discounts for specified products. Reporting unit: producers of commodities for primary markets. Confidential; 5-years; 1959-present; 310,000 cards.
- 44-B-2 Comprehensive housing unit survey (decennial): Type of housing unit, occupancy, tenure, rent or value, year built, rooms, persons, equipment, income level. Reporting unit: individual housing units; selected cities. Confidential; 5 years; 1959-60-63; 500,000 cards.
- 44-B-3 Consumer price index—foods (monthly): Retail price of 120 food and household supply items. Reporting unit: grocery stores in selected urban areas. Confidential; 4 years; 1960-present; 5,100,000 cards.
- 44-B-4 Consumer price index—home ownership (monthly): Sales price, size (square feet), whether previously occupied, year built, site value. Reporting unit: data supplied by FHA; derived from insured mortgage file. Confidential; permanent; 1958-present; 1,850,000 cards; 55 tapes.
- 44-B-5 Consumer price index—rents (biennial): Monthly rental, date of rent change, change in characteristics of housing unit and rental terms. Reporting unit: individual housing units in selected urban areas. Confidential; 2 years; 1962-present; 200,000 cards.
- 44-B-6 Consumer expenditure survey (decennial): Detailed accounting of household receipts and disbursements, and family characteristics. Reporting unit: individual households. Confidential; 10 years; 1959-60-63; 5,500,000 cards; 1,200 tapes.
- 44-B-7 National survey of professional administrative and technical pay—(annual): Frequency and earnings of selected professional, administrative and technical personnel. Reporting unit: 1,700 establishments in manufacturing, public utilities, wholesale trade, retail trade, and engineering services. Confidential; permanent; 450,000 cards, 3 tapes.
- 44-B-8 Earnings distribution survey—minimum wage impact (selected intervals): Hours and straight-time earnings for individual employees, nonoccupational, e.g., all production or nonsupervisory. Reporting unit: establishments in broad industry groups—retail trade, manufacturing, wholesale trade. Confidential; 5 years; 1958-62; 790,000 cards.
- 44-B-9 Union wage scales in the building trades, printing trades, local transit and local trucking in 64 cities (annual): Minimum union wage scale and maximum hours for selected occupations. Reporting unit: union locals. Unrestricted; 3 years; 30,000 cards.
- 44-B-10 Supplementary employee remuneration surveys—manufacturing and broad industry groups (triennial): Establishment expenditure on selected supplementary fringe items. Reporting unit: establishments. Confidential; 6 years; 1959-62; 200,000 cards; 2 tapes.
- 44-B-11 Strikes and lockouts (monthly): Number of stoppages, workers and man days idle involving six or more workers in excess of one shift. Reporting unit: newspaper clipping service. Unrestricted; permanent; 200,000 cards.
- 44-B-12 Industry wage occupational studies (annual, triennial, quinquennial): Straight-time earnings for selected representative occupations and supplementary benefits. Reporting unit: establishments, primarily manufacturing, on nationwide and area basis: Confidential; 5 years; 40,000 cards.

- 44-B-13 Wages and related benefits—82 labor market areas (annual) : Hours and straight-time earnings for selected cross-industry jobs and supplementary establishment benefits. Reporting unit: establishments in manufacturing wholesale trade, retail trade, finance and insurance, and services. Confidential; 5 years; 2 million cards; 10 tapes.
- 44-B-14 Older worker performance study (one-time survey) : Production reports for selected weeks for selected individuals (mail sorters). Reporting unit: U.S. postal service. Unrestricted; 4 years; 1962; 24,000 cards.
- 44-B-15 Seasonally adjusted employment and average weekly hours (annual) : Employment and average weekly hours—50 series. Reporting unit: establishments and households. Unrestricted; permanent; 1,000 cards.
- 44-B-16 Average annual rates of change (regression coefficient) (annual) : Output per man-hour, output, man-hours, employment, and unit labor costs. Reporting unit: data from labor force and household reports. Unrestricted; permanent; 2,000 cards.
- 44-B-17 Master address tape, maritime survey (quarterly) : Name and address of employers subject to Longshoremen's and Harbor Workers' Compensation Act. Confidential; 1 tape.
- 44-B-18 Injury rates in manufacturing (quarterly) : Employees, hours, number of work injuries. Reporting unit: establishments. Confidential; 2 years; 1961-64; 650,000 cards.
- 44-B-19 Injuries and causes of accidents to seamen (selected intervals) : Description of injured, location of accident, description of injuries, causes. Reporting unit; individual accident. Confidential; 2 years; 1954-61; 7,500 cards.
- 44-B-20 Injury rates by industry (annual) : Employees, hours, injury distribution and time charges. Reporting unit: establish. Confidential; 5 years; 1958-62; 500,000 cards.
- 44-B-21 Injuries and accidents to longshoremen and harbor workers (quarterly, annually) : Deaths, other injuries, and hours worked, by port. Reporting unit: employers subject to Longshoremen's and Harbor Workers Compensation Act. Confidential; 2 years; 1962-present; 50,000 cards; 1 tape.
- 44-B-22 Master registry of work-injury reporters (annual) : Name; State, industry, reporter codes; city location. Reporting unit: establishment. Confidential; permanent; 1962; 1 tape.
- 44-B-23 Work injuries and work-injury rates in hospitals (one-time survey) : By department: average number of employees, total man-hours. For individual injured employees; kind of injury, nature of injury, part of body injured, occupation. Reporting unit: individual hospitals. Confidential; 2 years; 1963; 2 tapes.
- 44-B-24 Injuries and accident causes (one-time survey) : Occupation, extent of disability, severity, nature of injury, accident type, hazardous condition, agency of accident, unsafe act, activity of injured at time of accident. Reporting unit: individual injuries in establishments in selected industries. Confidential; 2 years; 1955-60; 10,200 cards.
- 44-B-25 Work injuries experienced by minors (one-time survey) : Industry in which employed, activity of injured, nature of injury, accident type. Reporting unit: individual reports transcribed from workman's compensation files of eight States. Confidential; 2 years; 1959; 32,000 cards.
- 44-B-26 Work injuries and work-injury rates (one-time survey) : By department: average number of employees, total man-hours. For individual injury employees: kinds of injuries, nature of injury, occupation, part of body injured. Reporting unit: establishments in selected industries. Confidential; 2 years; 1960-61; 104,000 cards.
- 44-B-27 Survey of industry employment, payroll and hours (monthly) : All employees, women, production workers, and average hourly earnings, average weekly earnings, hours, and overtime for production workers. Reporting unit: Individual establishment. Confidential; permanent; 1957-present; 100 tapes.

- 44-B-28 Survey of industry labor turnover (monthly): Total separation, quits, layoffs, total accessions, new hires. Reporting unit: individual establishment. Confidential; permanent; 1956-present; 2,500,000 cards; 3 tapes.
- 44-B-29 National survey of scientific and technical personnel in industry (annual): Employment of approximately 20 scientific and technical occupations by function (research and development, basic research, administration, etc.). Reporting unit: company and establishment. Confidential; permanent; 1959-64; 200,000 cards; 8 tapes.
- 44-B-30 Estimates of labor force characteristics from current population survey (monthly): Employment status of noninstitutional population 14 years and older by demographic characteristics. Reporting unit: household. Confidential; permanent; 1947-present; 7 tapes.

OTHER BUREAUS AND DIVISIONS

- 44-D-1 Applications for learner certificates (quarterly): Learners authorized. Reporting unit: applications. Confidential; 1 year; 2,000 cards.
- 44-D-2 Minors illegally employed in agriculture (semiannual): Minors illegally employed, age and school grade. Reporting unit: investigated farms. Confidential; 2 years; 14,000 cards.
- 44-D-3 Minors illegally employed in industries other than agriculture (semiannual): Minors illegally employed, age and hazardous order violated. Reporting unit: investigated establishments. Confidential; 2 years; 11,000 cards.
- 44-D-4 Learner investigations (quarterly): Learner in violation of certificate. Reporting unit: investigated establishments. Confidential; 2 years; 400 cards.
- 44-D-5 Reports on investigation findings (monthly): Employees underpaid, underpayments disclosed, minors illegally employed. Reporting unit: WHPC investigated establishments. Confidential; 2 years; 110,000 cards.
- 44-D-6 Register of all welfare and pension plans filed under the WPPDA (monthly): Name and address of plan; basic characteristics; types of employees; industry; benefits provided; type of administration; State location; type of funding; etc. Reporting unit: individual welfare and pension benefit plans: Unrestricted; permanent; 1959-present; 14 tapes.
- 44-D-7 Financial data of welfare and pension plans (one-time survey): Financial data; contributions, benefits paid; assets, by type; amount of insurance premiums; distributions by size, class; etc. Reporting unit: individual welfare and pension plans filing financial reports. Unrestricted; permanent; 1959; 110,000 cards.
- 44-D-8 Financial data of welfare and pension plans (annual): Financial data; contributions; benefits paid; assets by type; distributions by value of assets; amount of insurance premiums; distributions by size, class; etc. Reporting unit: individual welfare and pension plans filing financial reports. Unrestricted; permanent; 1960-63; 16 tapes.
- 44-D-9 Administrative costs of welfare and pension plans (one-time survey): Administrative cost data; total receipts; salary costs; fees, rents, interest; other administrative costs; etc. Reporting unit: sample (approximately 4,000) of individual welfare and pension plans reporting under the WPPDA. Unrestricted; permanent; 1962; 2 tapes.
- 44-D-10 Names and address of reporting unions (monthly): Name of president or financial officer of union and mailing address of union. Reporting unit: individual labor union reporting under the provisions of the Labor-Management Reporting and Disclosure Act. Unrestricted; permanent; 1959-present; 52,000 cards; 4 tapes.
- 44-D-11 Register of reporting labor unions (monthly): City and State location, union affiliation, unit designation. Reporting unit: individual unions reporting under the LMRDA. Unrestricted; permanent; 1959-64; 6 tapes.

- 44-D-12 Characteristics of labor unions (monthly) : Locations; union affiliation; unit designation; dues and fees; election rules; etc. Reporting unit: individual unions reporting under the LMRDA. Unrestricted; permanent; 1959-present; 52,000 cards; 2 tapes.
- 44-D-13 Financial data for unions (selected intervals) : Receipts, disbursements, assets by type. Reporting unit: individual unions reporting under LMRDA. Unrestricted; permanent; 1959-60; 208,000 cards.
- 44-D-14 Financial data for unions (annual) : Receipts and disbursements by type, assets by type, etc. Reporting unit: individual unions reporting under the LMRDA. Unrestricted; permanent; 1962-63; 4 tapes.

DEPARTMENT OF THE TREASURY

INTERNAL REVENUE SERVICE

- 48-1 Statistics of income—Individual income tax returns (annual) : Sources of income, adjusted gross income, itemized deductions, exemptions, taxable income, income tax, business income, and deduction items. Reporting unit: stratified sample of income tax returns of individuals (approximately 500,000). Confidential; 3 years; 1962-62; 2,503 tapes.
- 48-2 Statistics of income—Sales of capital assets, individuals (one-time survey) : Gross sales price, depreciation, cost, gain or loss, period held, and type of capital asset for each transaction; adjusted gross income and selected income items from return. Reporting unit: all returns of individuals reporting sales of capital assets included in the statistics of income sample for individual income tax returns (approximately 155,000). Confidential; 3 years; 1962; 34 tapes.
- 48-3 Statistics of income—Corporation income tax returns (annual) : Assets, liabilities, receipts, deductions, profits, income tax, and tax-related items, distributions to stockholders, and industry. Reporting unit: stratified sample of income tax returns of corporations (approximately 170,000). Confidential; 3 years; 1960-62; 1,475 cards.
- 48-4 Statistics of income—Foreign tax credit study (annual) : Foreign taxable income, foreign taxes, dividends received, foreign tax credit and related items, industry, country, and total assets. Reporting unit: all returns 1118 attached included in the statistics of income sample for corporation income tax returns (approximately 4,000). Confidential; 3 years; 1961-62; 21 tapes.
- 48-5 Statistics of income—Controlled foreign corporation study (one-time survey) : Foreign receipts, profits, taxes, total assets, dividends paid, country, and industry. Reporting unit: all returns with form 2952 attached included in the statistics of income sample for corporation income tax returns (approximately 11,000). Confidential; 3 years; 1962; six tapes.
- 48-6 Statistics of income—U.S. business tax returns, sole proprietorships (annual) : Income and deduction items and industry. Reporting unit: stratified sample of individual income tax returns—sole proprietorships (approximately 220,000). Confidential; 3 years; 1960-62; 754 tapes.
- 48-7 Statistics of income—U.S. business tax returns, partnership returns (annual) : Income and deduction items, balance sheet data for alternate years, and industry. Reporting unit: stratified sample of partnership returns of income (approximately 75,000). Confidential; 3 years; 1960-62; 232 tapes.
- 48-8 Statistics of income—Estate tax returns (biennial) : Gross estate, deductions, exemptions, estate tax, tax credits, types of property, executor's commission, attorney's fees, and funeral expenses. Reporting unit: all estate tax returns (approximately 80,000). Confidential; 3 years; 1962; four tapes.
- 48-9 Statistics of income—Fiduciary income tax returns (biennial) : Sources of income, deductions, exemptions, and tax items. Reporting unit: stratified sample of fiduciary income tax returns (approximately 51,000). Confidential; 3 years; 1962; two tapes.

- 48-10 Statistics of income—Gift tax returns (biennial): Total gifts, exclusions, deductions, specific exemption, and gift tax. Reporting unit: stratified sample of gift tax returns (approximately 34,000). Confidential; 3 years; 1962; two tapes.
- 48-11 Tax model for individual income tax returns (annual): Income, deductions, exemptions, and tax information. Reporting unit: subsample of statistics of income sample for individual tax returns. Confidential; permanent; 1960-62; 316 tapes.
- 48-12 Tax model for corporation income tax returns (annual): Income deductions, exemptions and tax information. Reporting unit: subsample of statistics of income sample for corporation income tax returns. Confidential; permanent; 1962; 10 tapes.
- 48-13 Tax model for partnership returns (annual): Income and deduction items. Reporting unit: subsample of statistics of income sample for partnerships. Confidential; permanent; 1962; five tapes.
- 48-14 Reporting characteristics of taxpayers (one-time survey): Selected data on excise taxes and extent to which taxpayer fills out return. Reporting unit: stratified sample of form 720 (approximately 58,000). Confidential; not kept; 1963.

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

OFFICE OF EDUCATION

- 51-1 School Centered Physical Fitness Programs (annual): Pupils engaged in planned program of physical activity, those passing tests of physical fitness and those who had physical exams. Reporting unit: elementary and secondary schools. Unrestricted; 2 years; 1962-63; 4,000 cards.
- 51-2 Math Teaching in Junior High Schools (quinquennial): Enrollment, facilities, teachers, contents of math programs. Reporting unit: public junior high schools. Unrestricted; not kept; 1962; 7,800 cards.
- 51-3 Science Teaching in Junior High Schools (quinquennial): Enrollment, science facilities, number of teachers. Reporting unit: public junior high schools. Unrestricted; not kept; 1962; 6,000 cards.
- 51-4 National Inventory of School Facilities and Personnel for Resource Evaluation and Damage Assessment (annual): Name, address, and location of schools, permanent buildings used for instructional purposes, number of permanent general use facilities, number of pupils and school employees. Reporting unit: public elementary and secondary school plants. Unrestricted; not kept; 1962; 12 reels.
- 51-5 Survey of Engineering Degrees (annual): By sex, engineering degrees according to curriculum and level. Reporting unit: institutions granting engineering degrees. Unrestricted; 2 years; 1961-62; 11,000 cards.
- 51-6 Survey of Engineering Enrollment (annual): For each of 26 curriculums in engineering, the number of undergraduate and graduate students by sex, year of study, and number of day and evening students. Reporting unit: institutions granting engineering degrees. Unrestricted; 2 years; 1962-63; 12,000 cards.
- 51-7 Offerings and Enrollments in NonPublic Secondary Schools (selected intervals): Enrollment, teaching staff and curriculum. Reporting unit: all nonpublic secondary schools. Unrestricted; not kept; 1961-62; 47,000 cards.
- 51-8 Offerings and Enrollments in Science and Mathematics in Public High Schools (biennial): Math and science courses and enrollment by sex. Reporting unit: public high schools. Unrestricted; 2 years; 1962-63; 9,000 cards.
- 51-9 Current Expenditures Per Pupil in Public Schools (annual): Per pupil expenditures for administration, instruction, attendance and health services, pupil transportation services, operation of plant, maintenance of plant, and fixed charges. Reporting unit: public school system. Unrestricted; 1 year; 1961-62; 3,000 cards.

- 51-10 Survey of Opening Fall Enrollment (annual) : Full time, part time, and first time fall enrollment, by sex. Reporting unit: all institutions of higher learning. Unrestricted; 2 years; 1962-63; 12,000 cards.
- 51-11 Comprehensive Report on Enrollment (biennial): Number of resident and extension students, and students taking courses by TV or radio and enrolled in short courses or individual lessons. Also enrollments by sex in the preceding summer session. Reporting unit: all institutions of higher education. Unrestricted; 2 years; 1959-61; 55,000 cards.
- 51-12 Residence and Migration of College Students (quinquennial) : Undergraduate and graduate students (full time and part time, and whether first time, new transfer, or continuing students) from each State or U.S. territory. Reporting unit: institutions with students taking creditable toward a bachelor's or higher degree. Unrestricted; 5 years; 1963; 100,000 cards.
- 51-13 Earned Degrees Granted During Year (annual) : Four-year bachelor's and first professional degrees; first professional degrees requiring 5 or more years; second level degrees; and doctorates. Reporting unit: institutions granting bachelor's or higher degrees. Unrestricted; 2 years; 1962-63; 50,000 cards.
- 51-14 Faculty and Other Professional Staff (biennial) : By sex, staff for general administration, student personnel services, resident instruction, extension instruction, organized research, elementary or secondary instruction and other facts. Reporting unit: all institutions of higher education. Unrestricted; 4 years; 1961-63; 24,000 cards.
- 51-15 Survey of Students Enrolled for Advanced Degrees (annual) : Graduate students, both full and part time, enrolled for master degrees and doctorates by field of study and year level. Reporting unit: institutions granting advanced degrees. Unrestricted; 2 years; 1962-63; 20,000 cards.
- 51-16 Higher Education Planning and Management Data (annual) : Faculty and administrative salaries, new buildings completed, and basic student charges. Reporting unit: all institutions of higher education. Unrestricted; 2 years; 1962-63; 47,000 cards.
- 51-17 College and University Library Statistics (annual) : Library collections, personnel and expenditures. Reporting unit: all institutions of higher education. Unrestricted; 2 years; 1960-62; 20,000 cards.
- 51-18 Survey of Non-Public Elementary Schools (quinquennial) : School description, enrollment by grade and number of secular and non-secular teachers. Reporting unit: nonpublic elementary schools. Unrestricted; 5 years; 1961-62; 14,500 cards.
- 51-19 Survey of Non-Public Secondary Schools (quinquennial) : School description, enrollment by grade and number of secular and non-secular teachers. Reporting unit: nonpublic secondary schools. Unrestricted; 5 years; 1960-61; 14,000 cards.
- 51-20 Special Education for Exceptional Children (quinquennial) : Enrollment and grade level of blind, partially seeing, hard of hearing, speech impaired, crippled, special health problems, socially and emotionally maladjusted, mentally retarded, and gifted students. Also, number of teachers. Reporting unit: public school systems and residential schools. Unrestricted; 5 years; 1962-63; 67,000 cards.
- 51-21 Inventory of College and University Facilities (quinquennial) : Existing physical facilities of institutions of higher education. Reporting unit: institutions of higher education. Unrestricted; 5 years; 1957; 150,000 cards.
- 51-22 Public School Library Statistics (biennial) : Library collections, number of school librarians, expenditures for books and number of school libraries. Reporting unit: public school systems. Unrestricted; 4 years; 1960-62; 75,000 cards.
- 51-23 Public Library Statistics (quinquennial) : Library collections, personnel and expenditures. Reporting unit: all public libraries. Unrestricted; 5 years; 1962; 68,000 cards.

- 51-24 Financial Statistics of Institutions of Higher Education (biennial) : Amounts and sources of income, expenditures by purpose, property utilized, endowment funds, students' loan funds, and annuity and living trust funds. Reporting unit: higher education institutions. Unrestricted; 4 years; 1959-61; 140,000 cards.
- 51-25 Organized Occupational Curriculums (biennial) : Enrollments and graduates of organized occupational curriculums. Reporting unit: institutions of higher education. Unrestricted; 4 years; 1959-61; 50,000 cards.
- 51-26 Education Directory, Part 2 (annual) : Total enrollment, superintendent's name and address. Reporting unit: public school systems having an enrollment of 300 or more. Unrestricted; 1 year; 1963; 28,000 cards.
- 51-27 Offerings and Enrollments in High School Subjects (decennial) : Enrollment by subject and grade level. Reporting unit: public secondary schools. Unrestricted; 10 years; 1960-61; 240,000 cards.
- 51-28 Statistics of College and University Libraries (biennial) : Collection, staff, expenditures and individual salary by level of personnel. Reporting unit: heads of academic libraries. Confidential; 2 years; 1962-63; 11,500 cards.
- 51-29 Public School Library Statistics (quinquennial) : Resources, staff, space, level of service, enrollment. Reporting unit: heads of school library administrations. Confidential; 2 years; 1960-61; 60,000 cards.
- 51-30 Statistics of Public School Libraries (biennially) : Resources, enrollment, staff, expenditures. Reporting unit: heads of school library administrations. Confidential; 2 years; 1962-63; 3,400 cards.
- 51-31 Statistics of Public Libraries (quinquennial) : Resources, staff, expenditure, and service area. Reporting unit: heads of public libraries. Confidential; 2 years; 1962; 79,450 cards.
- 51-32 Statistics of Local Public School Systems (one-time survey) : Instructional personnel, public, facilities, and fiscal data. Reporting unit: public school systems. Unrestricted; 2 years; 1959-60; 138,600 cards.
- 51-33 Survey of Home Economics (biennial) : Administration and organization of home economics programs, enrollment and faculty. Reporting unit: 4-year institutions which offer programs in home economic leading to at least a bachelor's degree. Unrestricted; 2 years; 1963; 3,550 cards.
- 51-34 Status and Career Orientations of College Faculty Members (One-time survey) : Personal characteristics, position, and assignment, educational background, work experience, economic status and occupational plans of college faculties. Reporting unit: higher education institutions. Unrestricted; 2 years; 1962-63; 32,000 cards.
- 51-35 Organization and Administration of Student Personnel Services (one-time survey) : Background of selected student services personnel, staffing of student services area, pattern of organization and manner in which policies are approved and implemented. Reporting unit: institutions of higher education. Unrestricted; 2 years; 7,800 cards.
- 51-36 Science Teaching in Elementary Schools (one-time survey) : Objectives, enrollment by grade, facilities, teaching aids, practices and procedures. Reporting unit: elementary schools. Unrestricted; 2 years; 1960-61; 43,500 cards.
- 51-37 Status of Industrial arts in Public Secondary Schools (one-time survey) : Objectives, classes, enrollment by sex, laboratories, teachers, methods and problems within the curriculum area of industrial arts. Reporting unit: public secondary schools. Unrestricted; 2 years; 1962-63; 9,150 cards.
- 51-38 Survey of Early Elementary Education in Public Schools (one-time survey) : Status, characteristics, practices, and policies of early elementary education; i.e., nursery, kindergarten, and primary grades. Reporting unit: medium-sized and larger public school systems. Unrestricted; 2 years; 1960-61; 1 tape.

- 51-39 Foreign Languages in Public Secondary Schools (one-time survey) : Courses taught, enrollment, materials, equipment, educational background of teachers and salaries. Reporting unit: public secondary schools. Unrestricted; 2 years; 1959; 43,000 cards.
- 51-40 Fifty-Year Programs: A survey of Policy and Practice in Classroom-Teacher Education (one-time survey) : Policies and practices in programs at the fifth-year level designed for elementary and secondary classroom-teacher education. Reporting unit: higher education institutions having fifth-year teacher training programs. Unrestricted. 2 years; 1959-60; 19,000 cards.
- BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM
- 55-1 Survey of Financial Characteristics (one-time survey) : Components of net worth. Reporting unit: Family units. Confidential; 3 months; 1962; 33 tapes.
- 55-2 Federal Funds Historical Data (daily and weekly) : Basic reserve position, Federal funds and related dealer transactions. Reporting unit: individual member banks. Confidential; permanent; 1959-present; 6 tapes.
- 55-3 Cost of Living for Latin America and Western Europe (annual) : Index of cost of living for approximately 25 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 250 cards.
- 55-4 Exports of Goods and Services, Growth Rates for Asia, Latin America, and Western Europe (annual) : Single total exports of goods and services for approximately 50 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 600 cards.
- 55-5 World Exports, Compound Annual Growth Rates (annual) : Single total of exports for approximately 112 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 1,120 cards.
- 55-6 World Imports, Compound Annual Growth Rate (annual) : Single total of imports for approximately 112 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1950-63; 1,456 cards.
- 55-7 Real Gross National Product, Growth Rates for Asia, Latin America, and Western Europe (annual) : Real GNP for approximately 50 countries. Reporting unit: U.N. Yearbook of National Accounts Statistics and official country data. Unrestricted; permanent; 1950-62; 600 cards.
- 55-8 Gold and Foreign Exchange Reserves, Growth Rates for Latin America and Western Europe (annual) : Single total of reserves for approximately 40 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 400 cards.
- 55-9 Industrial Production, Growth Rates for Latin America and Western Europe (annual) : Composite index of industrial production for approximately 15 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 150 cards.
- 55-10 Money Supply, Growth Rates for Latin America and Western Europe (annual) : Single total of money supply for approximately 25 countries. Reporting unit: International Monetary Fund. Unrestricted; permanent; 1953-62; 250 cards.
- 55-11 U.S. Exports to Asia (monthly) : Single total exports to Asian area (Pakistan east through Japan). Reporting unit: Bureau of the Census, U.S. Department of Commerce. Unrestricted; permanent; 1953-64; 72 cards.
- 55-12 Japanese Industrial Activity Index (monthly) : Composite index of output in manufacturing, mining, and utility sectors. Reporting unit: U.S. Embassy, Tokyo, Japan. Unrestricted; permanent; 1947-63; 34 cards.
- 55-13 Savings Flows to Savings and Loan Associations and Mutual Savings Banks (annual) : Levels, new deposits, withdrawals of savings capital at savings and loan associations and regular deposits at mutual savings banks. Reporting unit: Federal Savings & Loan Insurance Corp. & National Association of Mutual Savings Banks. Unrestricted; permanent; 1964; 1,500 cards.

- 55-14 Boeckh Index of Construction Costs (monthly): Residences and apartments—building costs. Reporting unit: builders. Unrestricted; permanent; 1934-59; 600 cards.
- 55-15 Wholesale Price Indexes—Construction Materials (monthly): Construction materials and components. Reporting unit: wholesalers. Unrestricted; permanent; 1947-59; 6,000 cards.
- 55-16 Mortgage Recordings (monthly): Recordings by type of holder and type of mortgage. Reporting unit: Mortgage Recording Offices and other sources available to Home Loan Bank Board. Confidential; permanent; 1939-59; 4,000 cards.
- 55-17 FHA—Insured Home-Mortgage Terminations (monthly): Date of origination, data of termination, characteristics of borrower, etc. Reporting unit: FHA-insurance offices. Confidential; permanent; 1956-62; 1,000,000 cards; 25 tapes.
- 55-18 Mortgage Debt—Net Changes Only (quarterly): One-four family mortgage debt; multifamily and commercial mortgage debt; farm mortgage debt. Reporting unit: financial institutions and other mortgage holders. Confidential; permanent; 1949-63; 500 cards.
- 55-19 Reports of Condition, All Insured Commercial Banks (quarterly): Asset and liability items—in detail. Reporting unit: individual banks. Confidential; permanent; 1959-63; 580,000 cards; 29 tapes.
- 55-20 Sample Survey of Agriculture (one-time survey): Selected data on farm debts and other characteristics of the farm and farm operators. Reporting unit: farm operators and landlords in the United States. Confidential; 5 years; 1960; 264,000 cards; 2 tapes.
- 55-21 Reports of Income and Dividends. All Insured Banks (annual): Breakdown of income, expenses, taxes, etc. Reporting unit: individual banks. Confidential; permanent; 1960-63; 240,000 cards; 8 tapes.
- 55-22 Operating Ratios—Member Banks (annual): Forty ratios showing relationship between various balance sheet items and income and expense items. Reporting unit: individual banks. Confidential; 2 years; 1962-63; 18,000 cards.
- 55-23 Electric Power Series (monthly): KWH's. Reporting unit: electric utilities and industrial self-generators. Confidential; permanent; 1957 to present.
- 55-24 Industrial Generation of Electricity (monthly, annual): KWH's—current generation capacity. Reporting unit: Federal Power Commission. Unrestricted; permanent, 1957 to present; 100,000 cards; 1 tape.
- 55-25 Federal Funds Rate (daily): Effective rate, low bid, high offer. Unrestricted; 10 years; 1955-62; 1,800 cards.
- 55-26 U.S. Government Securities Dealer Statistics (daily): Positions, borrowings and transactions. Confidential; permanent; 1963 to present; 40,000 cards; 2 tapes.
- 55-27 U.S. Government Security Yields and Prices (monthly): Yields on key Treasury bills and average yields and prices on Treasury coupon issues. Unrestricted; 10 years; 1952-63; 3,000 cards.
- 55-28 Statement Week Averages—3 Month Bill Rate and Reserves (weekly): Free reserves; bill rate; weeks high, low; spread; excess reserves; borrowed reserves; 3 weekly moving average; free reserves; 9-week moving average free reserves; spread as percent of bill rate; deviation from 3-week and 9-week average; 8- and 13-week lag in moving averages; net changes in variables. Confidential; 10 years; 1953-63; 2,500 cards.
- 55-31 Money Supply (daily): Due to and from banks, Federal Government deposits, other demand, vault cash, cash items, time deposits. Reporting unit: individual Reserve banks. Unrestricted; permanent; 1958 to present; one tape.
- 55-32 Survey of Negotiable Time Certificates of Deposit (one-time survey): Outstanding certificates of deposit with breakdowns by denominations, original maturity, and type of holder. Reporting unit: individual member banks and selected nonmember banks. Confidential; permanent; 1960-62; 600 cards.

- 55-33 Member Bank Deposits and Vault Cash (daily) : Demand deposits: due to banks, U.S. Government demand, other demand, net demand; time deposits; vault cash. Reporting unit: Federal Reserve District summaries. Confidential; permanent; 1958 to present; 58,000 cards.
- 55-34 Flow-of-Funds Accounts (quarterly) : Elements of the published accounts. Reporting unit: Flow-of-Funds Section, Division of Research and Statistics. Unrestricted; 1 year; 1946-63; one tape.
- 55-35 Spot Exchange Rates: Major Currencies Against U.S. Dollar (weekly) : Quotations on Swiss, German, I.K., Dutch, French, Italian, Canadian, Belgian and Japanese currencies. Reporting unit: market data. Unrestricted; permanent; 1959 to present; 300 cards.
- 55-36 Industrial Stock Indices (weekly) : Indexes for industrial stock prices in Canada, United States, United Kingdom, Japan, Germany and Switzerland. Reporting unit: market data. Unrestricted; permanent; 1958 to present, 352 cards.
- 55-37 Long-Term Bond Yields (weekly) : Yields on specific long-term bonds in United Kingdom, United States, Germany Canada and Switzerland. Reporting unit: market data. Unrestricted; permanent; 1958 to present; 320 cards.
- 55-38 Short-Term Interest Rates (weekly) : Yields on German, United Kingdom, United States, and Canadian Treasury bills; Swiss 3-month deposits; and Japanese bank loans and discounts. Reporting unit: market data. Unrestricted; permanent; 1959 to present; 300 cards.
- 55-39 Interest Arbitrage for German Commercial Banks (quarterly) : Yields on Euro-Dollar Deposits (London), German inter-bank loan rate, German Treasury bills, forward rates on DM, and differences. Reporting unit: market data. Unrestricted; permanent; 1960 to present; 214 cards.
- 55-40 Interest Arbitrage, New York/London (weekly) : Yields on United States and Canadian Treasury bills, forward rates on Canadian dollar, and difference. Reporting unit: market data. Unrestricted; permanent; 1959 to present; 300 cards.
- 55-41 Interest Arbitrage New York/Canada (weekly) : Yields on United States and United Kingdom Treasury bills, United Kingdom local authority deposits, forward rates on sterling, and differences. Reporting unit: market data. Unrestricted; permanent; 1959 to present; 300 cards.
- 55-42 International Money Market Yields for U.S. Dollar Investors (weekly) : Yields on Euro-\$ deposits (London); United States CD's; Canadian, United States and United Kingdom Treasury bills; and United Kingdom hire purchase, Canadian Finance Co. and U.S. Finance Co. paper. Reporting unit: market data. Unrestricted; permanent; 1963 to present; 85 cards.
- 55-43 Three-Month Forward Exchange Rates (weekly) : Forward quotations on Swiss, German and United Kingdom currencies against the United States dollar and Swiss, United States, Germany, Dutch, Belgian and French currencies against sterling. Reporting unit: market data. Unrestricted; permanent; 1961 to present; 160 cards.
- 55-47 F.R. Bulletin Profit Series (quarterly, annual) : Sales, profits before taxes, taxes, profits after taxes, dividends. Reporting unit: 180 manufacturing corporations. Unrestricted; permanent; 1955-59; 4,500 cards.
- 55-48 Statistical Data Compiled From Bank Examination Reports (quarterly) : Principal assets and liabilities; maximum interest rates on time deposits; maturity distribution of investments; amounts of criticized assets and gross losses on loans; percentage of past due loans and other ratios. Reporting unit: individual member banks (one report per year for each bank). Confidential; permanent; 1963-Present; 68,000 cards; 5 tapes.
- 55-49 Member Bank Loans to Commercial and Industrial Borrowers: One time survey. Amount and term of loan and borrower characteristics of each loan. Reporting unit: individual bank data reported by a sample of about 2,000 banks. Confidential; permanent; 1955; 4 tapes.

- 55-50 Member Bank Loans to Commercial and Industrial Borrowers: One time survey. Amount and terms of loan and borrower characteristics for each loan. Reporting unit: individual bank data reported by a sample of about 2,000 banks; confidential; permanent; 1957; 4 tapes.
- 55-51 Interest Rates Charged by Banks on Short-Term Business Loans (quarterly): Amount of loan, annual interest paid and annual rate of interest for new or renewal loans. Reporting unit: individual bank data reported by sample of banks in 19 leading cities; confidential; permanent; 1958-63; 3 tapes.
- 55-52 Interest Rates Charged by Banks on Short-Term Business Loans (quarterly): Amount of loan, annual interest paid and annual rate of interest for new or renewal loans. Reporting unit: individual bank data reported by sample of banks in 19 leading cities; confidential; permanent; 1964; 2 tapes.
- 55-53 Ownership of Demand Deposits Survey (annual): Number of accounts and amount of demand deposits held by individuals, partnerships and corporations grouped by major types of holders and account size. Reporting unit: individual bank data from a sample of about 1,800 banks; confidential; permanent; 1959-61; 6 tapes.

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

PUBLIC HEALTH SERVICE, NATIONAL CENTER FOR HEALTH STATISTICS

- 63-1 National Lung Cancer Mortality Survey (one time survey): Data include smoking history, residence history (identification of places by population size in which the deceased person resided 10 years or longer), diagnostic techniques used to establish diagnosis, and histologic type of cancer. Reporting unit: informants listed on death certificates, including family informant, attendant physician and institution in which death occurred. Data collected for a sample of approximately 3,000 lung-cancer deaths. Confidential; permanent; 1958; 4,000 cards.
- 68-2 National Mortality Survey (one time survey): Data include items analogous to those collected in the 100-percent and 25-percent stages of the census of population, including place of residence of decedent, family type and family status of deceased person, related information about decedent's family, employment status, occupation, and industry of deceased person and his spouse, if married, income of deceased, of his spouse, and other family members. Reporting unit: family informants listed on death certificates. Data collected for a sample of approximately 10,000 deaths. Confidential; permanent; 1960; 2 tapes.
- 68-3 National Mortality Sample Survey (annual): Hospital use during last year of life; diagnostic information and information on operations for each episode of medical care; varying from year to year, social, economic, and family characteristics of deceased persons, such as family income, educational attainment of deceased person and spouse, place of previous residence. Reporting unit: informants listed on death certificates; records of hospitals and resident medical care institutions. Confidential; permanent; 1961 to present; 4 tapes.
- 68-4 National Natality Sample Survey (annual): Survey topic varies from year to year. Topics covered to date include radiation exposure during pregnancy, medical care during pregnancy, family status of mother, socioeconomic characteristics of mother, and expectation of additional births. Reporting unit: Informants listed on birth certificates, including mother, hospital of birth, attendant physicians. Confidential; permanent; 1963 to present; 2 tapes.
- 68-5 Master Facility Inventory (annual): Name, address of establishment, ownership, type of service, number of beds, and number of employees. Reporting unit: hospital or resident institutions. Confidential; permanent; 1962 to present; 5 tapes.
- 68-6 Resident Places Survey I (one time survey): Admission policy, utilization statistics, and health of residents. Reporting unit: nursing and personal care homes and residents of these homes. Confidential; permanent; 1962; 2 tapes.

- 68-7 Health Interview Survey (annual) : Persons, acute and chronic conditions, hospitalization experience, other health characteristics. Reporting unit: households members in civilian, noninstitutional population. Confidential; permanent; 1957-64; 35 tapes.
- 68-8 Health Interview Survey—Hearing Ability Supplement (one-time survey) : Hearing ability, history of hearing problem; use of hearing aid, lip reading, sign language. Reporting unit: household members with hearing impairment in civilian, noninstitutional population. Confidential; permanent; 1962-63; 2 tapes.
- 68-9 Health Interview Survey—Personal Health Expenditure Supplement (one-time survey) : Expenditures by persons in household for doctors' bills, hospital bills, medicines, dentists' bills, and other medical expenses. Reporting unit: household members in civilian, noninstitutional population. Confidential; permanent; 1962; 2 tapes.
- 68-10 Health Examination Survey Data From First Cycle Program (one-time survey) : Findings of individual health examinations, and related household questionnaires and control records. Reporting unit: individuals selected to constitute probability sample of U.S. adults, ages 18-79. Confidential; 1959-62; 300,000 cards; 25 tapes.
- 68-11 Health Examination Survey Data From Second Cycle Program (one-time survey) : Findings of individual health examinations and related household questionnaires and control records. Reporting unit: individuals selected to constitute probability sample of U.S. children, ages 6-11. Confidential; 1963-65; 400,000 cards.
- 68-12 Birth and Fetal Death Statistics (annual) : County, city, sex, race, birth order, ages of parents, month, birthweight, legitimacy, and other information. Reporting unit: State, territorial, and independent city health departments. Confidential; 2 years; 1960-63; 8,000,000 cards; 87 tapes.
- 68-13 Marriage Statistics (annual) : County, race, age, previous marital status, number of marriages for each party: month. Reporting unit: State, territorial, and independent city health departments. Confidential; 1960-62; 127,000 cards, 3 tapes.
- 68-14 Death Statistics (annual) : County, city, sex, race, age, month, nativity, marital status, cause of death. Reporting units: State, territorial, and independent city health departments. Confidential; 5 years, 1950-63; 8,700,000 cards; 70 tapes.
- 68-15 Divorce Statistics (annual) : County, month, duration of marriage, legal cause, race, age, previous marriage, number of children. Reporting unit: State, territorial, and independent city health departments. Confidential; 1960-62; 54,000 cards.
- 68-16 Birth, Death, and Fetal Death Statistics (annual) : Sex, color, age, birth order, cause of death. Reporting unit: generated in data-processing operations. Unrestricted; permanent; 1947-59; 4,160,000 cards.
- 68-17 Special Death Statistics by Occupations and industry (one-time survey) : State, occupation, industry, age, color, cause of death (all males 20-64 years). Reporting unit: State, territorial, and independent city health departments. Confidential; permanent; 1950; 335,000 cards.
- 68-18 Special Death Statistics by Occupations and Industry (one-time survey) : State, sex, color, age, up to five causes of death. Reporting unit: State, territorial, and independent city health departments. Confidential; permanent; 1955; 500,000 cards.

SOCIAL SECURITY ADMINISTRATION

- 72-1 One-Percent Sample Employee-Employer Record (annual) : Year of birth, sex, race, industry and geographic codes, quarterly and annual wages. Reporting unit: social security account number holder. Confidential; 10 years; 1955-62; 90 tapes.

- 72-2 One-Percent Continuous Work History Sample (annual) : Year of birth, sex, race, earnings each year, benefit and insurance status, total earnings and coverage since 1936. Reporting unit: social security account number holder. Confidential; 3 years; 240 tapes.
- 72-3 One-tenth of One Percent Continuous Work History Sample (annual) : Year of birth, sex, race, earnings each year 1937 to date, benefit and insured status. Reporting unit: social security account number holder. Confidential; 3 years; 36 tapes.
- 72-4 Name and Address File Tape (quarterly) : Name and mailing address of employers. Reporting unit: employers reporting wages. Confidential; permanent; 70 tapes.
- 72-5 Data Employer Duplicate Check Tape (quarterly) : Wages and wage items reported by employers for 5 quarters. Reporting unit: employer. Confidential; permanent; 165 tapes.
- 72-6 Summary Earnings Record Tape (quarterly) : Annual earnings 1951 to date, total earnings 1937 to date, quarters of coverage 1951 to date, date of birth, sex. Reporting unit: social security account number holder. Confidential; permanent; 1,500 tapes.
- 72-7 Regular Transcript Master Benefit Tape (monthly) : Benefit amount, age, State and county of payee, technical and historical data on benefits. Reporting unit: person entitled to OASDI monthly benefit. Confidential; permanent; 400 tapes.
- 72-8 Employer Identification Tape (semiannual) : Geographic and standard industrial classification (SIC) codes. Reporting unit: county-industry reporting units for each employer. Confidential; permanent; 34 tapes.

APPENDIX 2.—STATISTICAL EVALUATION REPORT NO. 6—REVIEW OF
PROPOSAL FOR A NATIONAL DATA CENTER

(A Report Prepared by Edgar S. Dunn, Jr., Consultant to the Office of
Statistical Standards Bureau of the Budget)

PREFACE

This report "Review of Proposal for a National Data Center" is the sixth of a series presenting the results of a comprehensive review and evaluation of some aspects of the statistics program of the Federal Government. It was prepared by Edgar S. Dunn, Jr., Resources for the Future, Inc., as consultant to the Bureau of the Budget.

The proposal which Mr. Dunn has reviewed stems from the work of a committee established by the Social Science Research Council to study the problems of the preservation and use of economic data. In the spring of 1965 that committee made its report to the SSRC, which presented it to the Director of the Bureau of the Budget. In its report, the Committee on the Preservation and Use of Economic Data, known as the Ruggles committee, summarized its recommendations as follows: "First, * * * that the Bureau of the Budget, in view of its responsibility for the Federal statistics program, immediately take steps to establish a Federal Data Center. * * * Second, that the Office of Statistical Standards * * * place increased emphasis on the systematic preservation in usable form of important data prepared by those agencies engaging in statistical programs. * * * Third, that at an early date the Social Science Research Council convene representatives from research institutions and universities in order to develop an organization which can provide a clearinghouse and coordination of requests for data made by individual scholars from Federal agencies."

In asking Mr. Dunn to examine the proposal, and to study ways of implementing it, we were concerned primarily with the first two of these recommendations. But, while the Ruggles committee represented the interests of the academic and social science research community, we were concerned with the use of statistical data for research, policy and decisionmaking at all levels, both within and outside Government. Mr. Dunn wisely extended it to include consideration of the relationships between the collecting and compiling processes on the one hand and preservation and accessibility for further use on the other hand.

We are indebted to Mr. Dunn for this analysis and report. We are also indebted to the many persons with whom he consulted, particularly those mentioned in his letter of transmittal who assisted by their thoughtful analysis of particular aspects of the entire problem and the preparation of the important appendix material.

RAYMOND T. BOWMAN,
Assistant Director for Statistical Standards.

NOVEMBER 1, 1965.

Dr. R. T. BOWMAN,
*Assistant Director for Statistical Standards,
U.S. Bureau of the Budget,
Executive Office Building, Washington, D.C.*

DEAR RAY: Transmitted herewith is the final report containing my review of the proposal for a National Data Center.

In seeking to identify ways of implementing this proposal I undertook an intensive period of study and review covering the last 6 months. I had the benefit in this effort of the advice and counsel of numerous others. This assistance was engaged in the following way. I divided the problem into parts that could be considered simultaneously. One part consisted of an informal ad hoc committee which met with me on a number of occasions to discuss the feasibility of establishing a referral-reference function in relation to the files of the Federal

tatistical system. Joining me in this discussion were Joe Daly and Ed Goldfield of the Bureau of the Census, Bob Steffes and Rudy Mendelsohn of BLS and Ezra Glaser of the Patent Office. In a second initiative Mendelsohn undertook a more intensive study based upon the earlier survey of machine readable records conducted by OSS to try to get a better fix on what it will take to establish an archival function. He in turn was assisted by many people in the agencies. Lastly, the National Bureau of Standards was used as a vehicle to assemble a small group of knowledgeable people in an attempt to specify more clearly the essential elements of a data service center required to provide a range of facilitating services. Ezra Glaser, Marshall Wood, and Dave Rosenblatt were the principal contributors to this effort although conversations included Sam Alexander and other members of his staff. Paul Krueger and I also participated in this effort.

In addition, I have engaged in many discussions of substantive issues with a number of knowledgeable people in the Federal agencies (both statistical agencies and program agencies) and in the universities.

I particularly want to acknowledge the invaluable assistance that I have received from Paul Krueger on your staff. He has given me continuous support, assistance and encouragement. He has joined me in many of the meetings and discussions with the aforementioned and has made his own valuable contribution to the thinking process.

The form and content of the report, of course, remains my own responsibility. I believe that the general conclusions and recommendations are sound and supported in whole or in part by the informed judgment of many others beside myself, but I do not attribute the views of this report specifically to any one or all of its many contributors.

In writing the report I have incorporated material included in earlier memorandums as well as sections that contain explanation and argument that is superfluous from the point of view of the informed staff member of the Office of Statistical Standards. I did so because I assumed that this report might be used in whole or in part to communicate elements of this problem and the recommended solutions to more than one group. I attempted, therefore, to include a comprehensive discussion of the problems and opportunities.

The report makes clear that my own understanding and evaluation of this problem has modified somewhat in the course of the study. I now feel that the production standards and practices are a more important element in both problems and solution. This, as well as other considerations, leads me to be less sanguine about the possibility or the desirability of keeping the issues of organization in the background. I think that there might be some benefit in our discussing this and several related issues on an informal basis.

Let me say that I have enjoyed working with the Office of Statistical Standards on this problem. I hope that the results are constructive in serving your needs and objectives.

Sincerely yours,

EDGAR S. DUNN, JR.

SUMMARY

The Ruggles Committee report recommending the establishment of a National Data Center is only one of the more manifest expressions of concern, dissatisfaction and frustration that have been surfacing among the groups that use numerical records for research, planning or decisionmaking at all levels. The problems at issue go far beyond the forms of discontent generated by special interests or marginal interests not served by public policy. They result from major changes on both the demand and supply side of the information process since World War II. Many people in this wider circle are attached to a rather naive data bank concept of the solution that does not incorporate an adequate appreciation of the basic problems in data use and data generation.

THE PROBLEM

The central problem of data use is one of associating numerical records and the greatest deficiency of the existing Federal statistical system is its failure to provide access to data in a way that permits the association of the elements of data sets in order to identify and measure the interrelationship among interdependent or related observations. This is true at virtually all levels of use and for all purposes from academic model builders to business market researchers.

There are a number of characteristics of existing programs and procedures that stand in the way of an effective association of numerical records for purpose of analysis.

(1) Important historical records may be lost because of the absence of consistent policy and procedure for establishing and maintaining archives.

(2) The absence of appropriate standards and procedures for file maintenance and documentation lead to low-quality files that contain many technical limitations to effective association of records.

(3) Many of the most useful records are produced as a byproduct of administrative or regulatory procedures by agencies that do not recognize a general purpose statistical service function as an important part of their mission.

(4) Record association requires a good deal of intelligence about the compatibility of records in several dimensions and the circumstances that condition their availability. There exists no organized reference capability for performing the kind of reference service essential for the Federal system or even within individual agencies.

(5) Production procedures for collecting, coding and tabulating data that were appropriate when developed now lead to several types of record incompatibility that block the kinds of record association in usage that is required by current policy problems and made possible by computer technique.

(6) There are identifiable gaps in existing data records that stand in the way of bringing together records of greatest relevance for today's problems. Some of these gaps are more apparent than real and reflect the effect of the other obstacles to effective record association.

(7) The structural problems of concern to today's policymakers and the effort to bypass problems of record incompatibility force the utilization of data at levels of disaggregation that place severe strains upon regulations restricting the disclosure of information about individual respondents. Technical possibilities for using the computer to bypass these disclosure constraints have not been generally developed and made available by the agencies.

(8) There are new possibilities for more efficient management of large-scale numerical files in terms of storage and retrieval; new possibilities for rearranging files in more useful form; new possibilities for retrieving in the form of maps, graphs, charts and other media in addition to the traditional tabular forms; new possibilities for building in disclosure controls and disclosure bypasses; new possibilities for matching records to assure compatibility. These potentialities require the expenditure of time and effort on system design and software development that few agencies can justify.

THE STAKES ARE HIGH

The stakes associated with even a partial resolution of these problems of file availability and compatibility are very high. This rests in part upon an unexploited joint demand for information and an information service capability. In the Federal domain alone large amounts of money are being ineffectively spent in an effort to deal with these problems. The amount of overlapping and resource waste is substantial. The stakes are also high because the improvement in the utility of the information base could have an unmeasurable but substantial effect upon the quality of public administration.

THE SOLUTIONS ARE NOT SIMPLE

The solution to these problems do not rest, as some think, in bringing a large number of tapes into a common repository. Nor does it rest upon the fact that many different uses impact upon the same data sets making them "general purpose" records susceptible to central management. General-purpose files are always put to special-purpose uses. What makes a record a general-purpose record is for it to be constructed on the basis of standards, maintained in effective condition and serviced by institutional arrangements and a technical system capability that will allow it to be combined successfully with other records in a wide variety of ways that will meet the special requirements of a wide range of users. Thus the solution to the problems will require program modification on a broad front involving all of the agencies as well as an emergent data service center.

RECOMMENDATIONS

Accordingly, it is recommended that a National Data Service Center be established with the capability to:

- (1) Manage archival records;
 - (2) Develop referral and reference services;
 - (3) Provide explicit facilitating services for users including:
 - (a) File rearrangement, cost tabulation and extended output options;
 - (b) Tape translation and file modification;
 - (c) Record matching;
 - (d) Disclosure bypassing; and
 - (e) Standard statistical routines.
 - (4) Develop computer hardware and software systems essential to above;
 - (5) Provide staff support to work in conjunction with the Bureau of the Budget to develop and establish and monitor standards essential to the system capability; and
 - (6) Establish a research capability directed to an analytical evaluation of user requirements for the purpose of designing and developing the system components essential to perform these services.
- The National Data Service Center would perform these services for:
- (1) Archival records under direct management control of the Center;
 - (2) The current and accumulated records of administrative and regulatory agencies;
 - (3) As a system resource to be used in connection with the current records of any agency not in a position to meet the needs independently.

RESOURCE REQUIREMENTS

Figures are offered that represent estimates of program costs for several components of these functions but it is pointed out that there are joint-product, joint-cost relationships between these service activities that make these estimates questionable as a guide to overall program costs. A judgment is made that the range of services and program adjustments required (including resources for modifying agency programs to be consistent with system requirements) would call for expenditure of between \$1 and \$2 million annually during the first year or two and rising to the neighborhood of \$10 million annually over a period of 5-10 years. A serious problem will be the assembly of the kind of intellectual resources required.

ISSUES OF ORGANIZATION

All of this really raises the issue of what kind of Federal statistical system we want to develop in the next generation, and encompasses a number of issues of organization and mission that will need to be addressed in a broad context. Further progress on the whole effort must depend on some understanding of the issues at the top policy level and some preliminary policy decisions to guide the direction of further effort.

PRIORITIES

If an effort is made to undertake this kind of system development there are certain immediate requirements that need to be fulfilled.

- (a) A continuous focus of leadership needs to be established.
- (b) This focus needs to be provided with staff support providing the kind of research-analytic capacity that can evaluate use requirements as a guide to specifying program options and reduce these options to specifications, costs and a logical order of time phasing.
- (c) Begin under existing authority to:
 - (1) Develop the standards for archives and compatible statistical building blocks, and
 - (2) Begin work on the 9,000 tape nucleus archive identified in appendix B. The Bureau of the Budget should give serious consideration to requesting funds to support these efforts in the fiscal 1967 budget.

INTRODUCTION

The assignment leading to this report originated 6 months ago in the form of a request to seek out and identify ways of implementing the proposal for a National Data Center presented by the Subcommittee on the Preservation and Use of Economic Data of the Social Science Research Council (the Ruggles

committee). During the interim the proposal and the problem set to which it was addressed has been intensively reviewed. I have had the benefit in this effort of the Council and effort of numerous others as indicated in the covering letter.

In order to identify the program options that might satisfy the intent of the Ruggles committee proposal, it was necessary to try to specify more precisely the problem set at issue. This has led to a characterization of the problem and attendant proposals in a somewhat broader context than the committee report but in a way that seems consistent with the intent and the leadership council of that group.

It is important to note that the report of the Ruggles committee is only one of the more manifest expressions of concern, dissatisfaction and frustration that have been surfacing among the groups that use numerical records for research, planning, or decisionmaking at all levels. The committee report is a good representation of the interests and the concern of the academic social science research community. There are other loci of discontent. The newly emerging welfare agencies of the Federal Government (OEO, EDA, the Department of Urban Affairs, the new programs in education and health in HEW, etc.) are experiencing great difficulty in assembling the statistical data that will guide them in analyzing their missions, establishing standards for performance and formulary for guidance, and support of administrative decisions and evaluating results. Their problems in this regard, when added to the usual difficulties of new program development, threaten to delay programs and render decisions more vulnerable to attack. A wide assortment of groups associated with making policy and planning for public facilities at all levels are becoming increasingly vocal about this concern. These are the groups that plan for roads, schools, hospitals, urban and regional development, etc. As business management turns increasingly to supplement its internal sources of information with the intelligence afforded by public agencies, they, too, are becoming aware of some of the inadequacies and anomalies of the information base.

Public needs for general purpose statistical information have never been satisfied and, indeed, never will. There are serious and legitimate issues of policy about how far down the scale one goes before general purpose becomes special purpose and about the levels of support for public information services that are appropriate. However, the problems that are at issue at the present moment go beyond the forms of discontent generated by marginal interests not served by current public policy. They are the product of fundamental changes on both the demand and supply side of the information process that have come to a head since World War II.

The most dramatic and obvious change on the supply side results from the advent of the large-scale computer. The economic feasibility and technical capability of producing, managing, and utilizing large numerical files has been multiplied by factors of a thousand in some technical applications and often by factors of 10 and a hundred in the economic and engineering dimensions of program planning. These are order of magnitude changes in capability that have come with revolutionary speed. They not only represent important and discontinuous changes in scale but also changes in kind because program options become technically feasible that were unthinkable as recently as 10 years ago.

Less commonly noted has been an increasingly dramatic change in information requirements on the demand side. In part, this is a consequence of the technical capabilities offered by the computer as well. The user can now handle data matrixes of a size and complexity formerly unmanageable and can use analytical techniques of a computational dimension formerly impossible. In part, however, the changes in information requirements stem from radical changes in demand factors distinct from these responses to expanded technical capability. Public policy in recent years has turned increasingly to a concern about the problems of social structure as they relate to public welfare and public policy. The issues of poverty, education, health, area depression, urban organization, etc., all require an increase in relevant detail for sub-system components of the total economy or total culture. At the same time the analytical disciplines in the social sciences and management analysis and control have been turning increasingly to quantitative methods and procedures.

As a result of these processes the users are increasingly finding that their needs for data are not satisfied by traditional documentary formats and the producers of data are finding the need to make data available more commonly in machine readable form—often in an organization and a format unique to the

purpose. The producers and users of data find their requirements and their missions intersecting in ways unknown a short time ago. Neither can continue to live the independent life formerly customary. The computer and other dimensions of social change have performed a shotgun wedding and both parties are in the process of discovering their incompatibility.

The procedural and program difficulties that led to this report are the product of this incompatibility. It is not uncommon in cases of this type for the assessment of difficulty to be one sided. The community of users has been sensitive to and vocal about many of the limitations of the producers of general purpose data—their mission concepts and institutional forms. The producers, in turn, can perceive many inadequacies on the part of the users. This report continues to be one sided in emphasis because it addresses itself primarily to the problems of use. The charge that framed it was couched in terms of the missions of the producing agencies of the Federal statistical system. It would be very worthwhile for some group to produce a companion evaluation of the anomalies in the production and usage of statistical information that arise from the practices and concepts of the user groups.

It is sometimes true that people who have concerned themselves with this problem are content with a superficial level of diagnosis and prescription. There is considerable attachment to the notion that most of our problems can be solved by computerizing all of the data we have in the backroom. This may be characterized as the "naive data bank" notion and its widespread acceptance is a source of some concern. I should emphasize here that this is not a characteristic of the Ruggles Committee report which was produced by knowledgeable and sophisticated people. However, the tendency to see the solution to the problems in relatively cheap technical programs has led to some misinterpretation of the Ruggles Committee report on the part of both those looking for additional support for data bank schemes and those reacting to the naive data bank concept. These evaluations and solutions are not based upon an adequate understanding and appreciation of the realities of the production processes essential to data generation or the institutional forms appropriate to their purpose.

The Ruggles Committee report gave us a healthy beginning toward an evaluation of this problem in realistic terms. However, this group did not have the time or staff resources to spell out the total problem set in a way that seems essential to support a more detailed consideration of program options. There is also a tendency in this report to see the problem primarily in terms of the accessibility of existing records and the solution in terms of the extension of user services. There is much that is valid in this representation but it gives insufficient attention to the important fact that accessibility is bound up with all of the production procedures and is inseparable in a number of fundamental respects from the issues related to the quality and scope of the existing records. It seems useful, therefore, to attempt a more precise formulation of the problem set to the solution of which the data center concept is addressed.

THE PROBLEM

The central problem of data use is one of associating numerical records. No number conveys any information by itself. It acquires meaning and significance only when compared with other numbers. The greatest deficiency of the existing Federal statistical system is its failure to provide access to data in a way that permits the association of the elements of data sets in order to identify and measure the interrelationship among interdependent activities. This deficiency has been partially overcome in a few vital areas where we need to trace and analyse the performance of the economy, by the establishment of special programs to bring together data sets in the form of national accounts, special index series, etc., but remains a debilitating constraint for most uses of data for analysis and planning. This is true for virtually all levels of use and for all purposes. It is a problem that plagues the research analysts inside and outside of the National Government who, for example, are engaged in building models of the economy in the interest of analyzing and projecting the major dimensions of economic growth and stability. It has been the principal obstacle to the administration's attempt to build a postattack revaluation and recuperation model. It is just as serious a problem for the uses that do not take the form of integrating data sets into a complex and formal model structure. The organization like EDA that wishes to establish a measurable test of eligibility for its program benefits faces the same problem. The business analyst who simply

wants to identify a variety of characteristics of the firms and households that form his principal markets often faces the same problem. The State Governor who wants to evaluate elements of his own program or the problems of his State by comparing them with the problems and programs of other States faces the same problem.

While the general problem is one of associating numerical records in use, it is understood better if we can visualize it in terms of its problem subsets. There are a number of distinct obstacles to file association that need to be identified before we can talk about solutions and program options. Consider the following:

The archival problem

This is the problem that initially interested the Ruggles Committee and the door through which they entered to their concern with some of the broader issues of file management. The problem arises from the fact that the statistical agencies are oriented primarily to producing data publications and often leave their records improperly documented for further proceeding and analysis. Worse still, useful records are sometimes destroyed. These things occur because the existing system has no standards for identifying the files significant for preservation or for essential levels of file documentation. It provides no financial or organizational mechanisms for their maintenance. The decisions about the significance of archives is left to functionaries with little knowledge of their value in use and who must allocate funds for their documentation and preservation in competition with agency missions defined by previous policy in more restrictive terms and considered primary by agency personnel.

This, obviously, constitutes a major obstacle to the association for records for anyone who needs to work with data with any significant historical dimension.

Problems of file maintenance

Closely associated with the archival problem are some of the more fundamental problems of file maintenance. The utility of a file and its capacity for association with other records rests on more than the existence of a tape and a document that identifies its content sufficiently well for the data to be retrieved. Many additional problems stem from the low quality of file maintenance.

A couple of the more gross and obvious defects are associated with the fact that there are still important records that do not exist in machine readable form. Amongst the files that do exist on tape, some are in a mixed binary mode and some in a decimal mode making data association impossible without expensive and time-consuming mode translation. This is often true even between records of the same agency.

More subtle defects in file maintenance are uncovered when the need for data association requires bringing the detailed data in to accord with summary or published data. Often in the rush to meet production goals, agencies have pushed work through the processing stages of screening, reconciliation, estimation and summary in great haste without correcting prior files whenever errors or discrepancies are found. For example, corrections made "at the summary level are not carried back into the micro-detail. Indeed in some instances corrections have been made only in the published data, leaving both the summary and the detailed machine records uncorrected. Occasionally, summary data may no longer be in machine records and must be recreated by reprocessing the detail files or by keypunching and processing the new records. In some surveys, standards for screening data for creditability may have been coarser than appropriate for other uses.

The urgency to release results may also leave a disarray from the viewpoint of good file record and format management required for the files to be reused. There may be no uniform position in the records for like data and duplication of the same data can occur. For archival purposes a uniform record for the same data is essential and elimination of duplication economical. Also, the tape records may contain excess information. Certain codes and indicators used in the initial processing have no meaning in the archival context and the files must be purged of the excess information. Files may have interspersed alphabetic information useful only in the narrow survey context and which add complexity to programing efforts when used in other contexts. The absence of clear identifiers as part of the tape must be corrected to facilitate use of such tapes.

The requirements for simple access as well as the association of records are often stymied by the limitations of standards and procedures for routine file maintenance. Mr. Mendelsohn, who was loaned to OSS by BLS to conduct a

detailed review of the condition of the more important data files that might form an archival or file management center, concluded that the loss of data because of the failure to support good file management is distressing. (Appendix B.) The report of the study group formed at the National Bureau of Standards also emphasized the problem of file maintenance (appendix C) as did the Ruggles Committee report.

The reference problem

One of the serious obstacles to the fullest utilization of the information resources of the Federal Statistical System and to the effective association of its records in use, is the absence of any clearly defined reference function. The inadequate nature of this kind of service is directly traceable to the production orientation of the agencies stemming from their primary missions as data publishers. The reference function has generally been thought of as the responsibility of the documentary centers. To the extent that the agencies attempt to provide occasional referencing assistance, the task falls to an individual whose primary mission is defined in terms of the publication mission. The inadequate nature of this service is also traceable to the fragmented nature of the records of the Federal statistical system growing out of the divided responsibility for their generation and maintenance. The reference problem is made especially complex because of the decentralized character of the Federal statistical program. No agency has been in a position to perform a reference service in relation to the total file. The problem is becoming more acute as records are frequently demanded in disaggregate or special form not met by traditional documentary formats and are frequently used in ways that require extensive knowledge and understanding about the compatibility of records in several dimensions and the circumstances that condition their accessibility.

The problem of administrative records

Some of the most useful general purpose numerical records are generated as a by-product of administrative and regulatory procedures of Federal agencies. These agencies rarely interpret their missions to include the capacity to provide general purpose statistical services. Even when they might like to do so they experience serious difficulty because of the traditions, program priorities, budgetary procedures, and legislative authorities peculiar to their agency. Users who need to acquire from these agencies, and especially those who need to match these records with other files, find their task difficult. There is an important need to provide for the management and servicing of these records for general purpose statistical use.

The disclosure problem

The legal and administrative regulations on the disclosure of information supplied by individual respondents are becoming increasingly restrictive to the user. Only rarely is this because policy or research requires specific information about individual respondents per se. It is usually because of the need to associate sets of data in the interest of determining the interrelationship between two or more variables. The strain upon disclosure arises because matching several sets of data for consistency at levels of aggregation appropriate to the problem requires a retreat to elemental units in the process of constructing the necessary aggregates. It is not widely understood that the interest in micro-data and the existing pressures and constraints do not grow out of an interest in information about the specific respondent.

The fact that the strains upon the disclosure rules usually are of this form is fortunate because there are possible a number of servicing procedures based upon computer technology that can satisfy the needs of the user in most cases without violating disclosure regulations. Currently, however, the agencies of the Federal statistical system have only a very limited capability for performing the kinds of services that would lead to disclosure by-passes. The usage of the data is, thus, severely constrained, and valuable information is lost by aggregation at too early a stage in the analysis.

Problems stemming from the procedures used to generate data

It would be a mistake to conclude that the serious obstacles to effective use of the Federal statistical system under modern conditions is solely a product of its present inability to perform a series of user oriented services. Some of the most serious anomalies arise out of current practice in the production of data.

We have already seen how constructing legitimate disclosure by-passes forces one back to a manipulation of highly disaggregated components or even respondent units as building block elements. The collection and tabulation procedures of the agencies generate constraints on data use that lead in a similar direction. This arises out of some fundamental problems in dealing with the coding and classification of original source data.

As has been noted, a common form of information usage in analysis requires the matching of an attribute for two or more statistical sources or the association of two different attributes. This may require matching between different historical sets or between the files of different programs or agencies. This association of records is rendered difficult or impossible by at least three classes of record incompatibility that stem from current production practice. One difficulty in associating records from the different sources stems from the noncompatibility of classification of the data by several collection agencies and information systems. In the process of condensing and summarizing source records from initial respondents, information is tabulated on the basis of classification schemes that group items into classes. Often these classifications are inconsistent. This is sometimes the result of the failure to develop general purpose standard classification codes applicable to all programs for these intermediate aggregates. It is often a result of the fact that standard codes are applied differently by different agencies so that there is no assurance that each agency (or program) will assign the same respondent to the same cell. Either of these cases often makes a comparison of the cells meaningless or difficult for purpose of analysis.

A further difficulty grows out of the fact that the basis for classification applied by the collection agency in defining the cells may be inconsistent with the analytical or descriptive requirements of the user.

When either type of problem occurs, one solution is to return to the initial respondent unit or some other disaggregate building block and reconstruct consistent boxes of data. This yields the same class of by-pass procedure identified with the disclosure problem. At this point one may encounter another common problem in the form of the noncompatibility of the definitions of the respondent unit. This is a class of noncompatibility that not only renders questionable the comparison of seemingly similar cells for different systems (as in the other two classes of incompatibility) but may render difficult or even impossible the reconstruction of compatible cells.

The anomalies that grow out of these compatibility problems can be tackled in two ways. One often hears it proposed that general purpose standards for the classification of intermediate aggregates be considerably extended and aggressively applied to all agencies. This may not be the most fruitful line of approach. Existing standards may possibly be improved and made more general purpose by a more intensive analysis of user requirements and a concern with the issue is not unimportant. However, an attempt to force all uses into a common standards mold for intermediate aggregates has attendant disadvantages from the point of view of the user as well as the producer and the agency vested with the responsibility of formulating standards. General purpose classifications for intermediate aggregates always require some compromises in taxonomy that reduce the utility of the data for special purposes. Furthermore, a great deal of the data generated by the Government comes from programs that have a special purpose mission and restrictive legislative authorities and requirements that go with it. Forcing on these agencies a rigid application of general purpose codes for intermediate aggregates may be impossible and even undesirable because they conflict with special purpose missions. It does not appear to be a helpful possibility that all data sets can be arrayed in compatible boxes that will anticipate all uses. The attempt to deal with standards in this context will place impossible strains and burdens upon the machinery for making and enforcing policy with respect to standards.

A more fundamental way to handle this problem may be a progressive move in the direction of compatible building blocks that can be reassembled to provide compatible and relevant aggregate sets for special uses and can be used as a bypass for disclosure problems and other procedural obstacles. This suggests that the problem of standards of greatest importance in the emergent situation is the need for uniform identification, definition and coding of the respondent unit as a basic building-block unit. The absence of a uniform system of coding and classification for geographical areas is also a serious deficiency and is an important part of this same problem. This also suggests the importance of procedures for assuring that every agency puts each respondent in the same cell

and that an important criterion for evaluating existing standards for intermediate aggregates is the extent to which they can serve as useful intermediate building blocks that obviate the necessity for returning to the respondent unit for many programs.

Unless something of significance is done to modify current practice in these production procedures, the matching of data from diverse sources will remain generally impracticable and often impossible.

Associative records that do not exist

Many of the most important analytical and policy issues of today require the association of existing records with records that do not exist. There are serious gaps in the public record of social activity.

There is a particularly important class of records that is missing. It can be identified by examining the problems of some of the most important Federal programs. Missing are the records that enable policymakers and planners to understand adequately how people, households, regions, activities, enterprises, and administrative units are functionally related and how they change over time.

The importance of such a capability is readily apparent. There is a large array of new and old welfare programs involved in trying to ameliorate various forms of social pathology and transform people (e.g., poverty, education, health), and regions (e.g., EDA, rural redevelopment), and the activities that engage them (e.g., Small Business Administration and large elements of the agricultural program). There is a large array of new and old programs engaged in planning for and providing public facilities (e.g., highways, mass transportation, water resources, urban development and housing, etc.). In each of these programs considerable effort, planning, and resources are expended for program development, in establishing the formulary for program management, and in evaluating program results. Indeed, Federal legislation in these areas impose planning requirements as a condition of grants-in-aid and other forms of assistance upon State and local governments and other State, local, and regional activities in at least a dozen large programs.

To date, the problems have been formidable and the results unimpressive for one principal reason. The information base that exists and can be economically accessed tells us a great deal about the characteristics of people, households, activities, enterprises and their institutions at any one point, but tell us very little about how they are linked into functional networks or how they transform over time. These latter are the most relevant information resources for policymaking and program evaluation in these areas. What form of job training, what form of regional assistance, what kind of road networks, what modes of mass transportation, what kinds of cities are questions that need to be answered on the basis of some knowledge of functional linkages and evaluated in terms of measurement of change. The responsible planners and administrators of these programs are feeling a keen sense of frustration because of the paucity and irrelevance of much of the information available to them. Some of the records they need to associate to resolve these issues do not exist.

The information gap related to these requirements reveals two elements of significance for the present evaluation:

First, a large part of the apparent gap in the kind of information needed is a direct function of the same system anomalies outlined above. In many cases, the problem does not rest upon the fact that the relevant attributes of people, activities, or institutions are not included in existing records. It rests with the fact that these attributes cannot be associated in functional configurations or traced through a historical sequence. We see the evidence of this in the widespread current interest in what is characterized as "longitudinal" data—(the ability to trace attributes of the same respondent through time in order to identify transformations—i.e., from and to movements in relations to places, activities, occupations, institutional affiliation, welfare categories, etc.). Therefore, many of these requirements could be met if the problem of file compatibility could be resolved through an extension of the servicing capability and some modification of the production practice of the Federal statistical system. This underscores the observation made in the introduction that the problems of file accessibility cannot be successfully separated from all of the issues related to the quality and scope of the files. In a fundamental way, file accessibility is the issue of file compatibility which is inseparable from the production practices that determine the organization and quality of the file. It is important to note that steps that can be taken to improve file compatibility

and accessibility will substantially increase the effective scope and utility of present files without a change in the size of the files or the attributes of the respondents they contain though additional resources might be required.

Second, although a significant part of the gap in information is a function of file incompatibility, there are also gaps which result from missing attributes.

For example, some attributes of the population may appear only infrequently with the decennial census and be needed for intercensal periods for vital program planning and evaluation. Other attributes may not appear in any records under existing information programs. Some notable gaps appear in the fields of transportation and construction and in connection with some of the important welfare attributes of people.

The following observation is pertinent here: The system, as it currently operates, provides no authority or mechanism for the review of the statistical program and the allocation of its resources in the light of the most important changes in information requirements. The decisions about programs that determine the scope and quality and accessibility of the records are primarily made upon the basis of technical problems, cost considerations, respondent pressures, etc., that impact directly upon the production process of the individual program and agency. There is no systematic way for the requirements side of the problem to enter the decision process.

Potentialities for system development are being missed

The new technology is making feasible a number of possibilities for greatly improving the utility of existing records. There are new possibilities for more efficient management of large-scale numerical files in terms of storage and retrieval; new possibilities for rearranging files in more useful form; new possibilities for retrieving in the form of maps, graphs, charts, and other media in addition to the traditional tabular forms; new possibilities for building in disclosure controls and disclosure bypasses; new possibilities for matching records to assure compatibility, etc.

One aspect of the service potential inherent in the new technology deserves some elaboration. The association of records in analysis usually carries with it a computational burden. This may take the form, for example, of computing the ratio of two data sets or making seasonal adjustments or computing coefficients in the analysis of variance. All of these derived numbers form a latent set implicit in the original source data. The computational capacity of modern computers is such that computations of this kind can often be made as fast or faster than the tape can be passed through the machine. Once a system has been developed for providing such a service, the marginal cost of generating these numbers when the tapes are being passed for retrieval is close to zero. Adding this kind of system capability can have the effect of increasing the effective size of the files of the Federal statistical system tenfold with latent numbers involving some computation.

These potentialities require the expenditure of time and effort on system design and software accumulation that few agencies can justify in terms of their current program levels or even appreciate in terms of their existing individual program missions.

The problem of file fragmentation

This is not a problem that is conceptually distinct from the others. Instead, it intersects the entire problem set being discussed and forms a part of the explanation for some of these anomalies. Currently, files are being generated and managed by more than 20 different agencies. It is precisely this division of responsibility and fragmentation of resources that inhibits system development and generates many of the problems of file compatibility. But apart from the way this problem invades all of the others, it imposes additional constraints because of the procedural, bureaucratic, and sheer time and space restrictions upon file usage.

THE STAKES ARE HIGH

The stakes associated with even a partial resolution of these problems of file availability, accessibility, and compatibility are very high. This is apparent even if we restrict our view to the significant Federal programs mentioned above. There is manifest in these programs an impressive and unexploited joint demand for information. This jointness has two important aspects. Even where the attributes of the numerical files of importance to these agencies are disjoint, they require the same servicing capabilities in the statistical system in order to

perform the essential tasks of record association. The agencies require the same kind of system capabilities. Beyond this, a number of the programs have a joint interest in the same sorts of file extensions. The new welfare agencies, for example, have a strong joint interest in longitudinal data about the welfare attributes of people that are not currently available. Even where these agencies might have discrepancies in the attributes of specific interest, there is a good possibility that the same collection vehicle could be used in servicing their needs.

It is interesting to note that many of these agencies have had substantial resources given to them by Congress explicitly for the purpose of generating or accessing the information essential to the conduct of program. This constitutes a formal recognition by the administrative-legislative process that the established statistical programs are not meeting these needs. So far, the remedial choice has been to fund programs to meet special requirements rather than system modification. This kind of bypass, however, has proved largely ineffective for several reasons.

(1) These agencies have no effective way to apply these resources to system reform that would improve record compatibility.

(2) The funds are dissipated because, though significant in total amount and perhaps even adequate to support major improvements, they are fragmented by their attachment to specific and narrowly conceived missions with not appreciation of the overlap or jointness of interest.

Thus, while the new welfare agencies could probably finance a collection vehicle adequate to their joint requirements, no one agency can really accomplish this satisfactorily alone and there exists no coordinating authority that can identify and exploit their joint interest.

Further fragmentation occurs even within agency programs. The Corps of Engineers, for example, has for years spent enormous sums of money on information to serve water resource planning requirements. Much of this expenditure has been duplicative and wasteful because the money for this purpose has always been funded on a river basin project basis so that it was virtually impossible to take advantage of the scale economies for building the servicing capability for the entire set of projects. As a consequence, each river system has tended to be planned in functional isolation without the opportunity to define the linkage between projects or to trace economic and social costs and benefits in an appropriately general context.

The stakes in program improvement in the Federal statistical system are high because the amounts of money being ineffectively spent on statistics in these programs is very large. They are also high because the improvement in the utility of the information base could have an unmeasurable but substantial effect upon the quality of public administration. The ability to ask relevant questions and get prompt relevant answers in planning, administering, and evaluating programs is of considerable importance.

All of this is only by way of recognizing the Federal interests involved. The stakes of State and local public officials, and the business and research community are equally large. These are the decision units which, by the nature of their responsibility, require disaggregate data sets that are especially affected by the problems of file compatibility. They have a common interest in extending the capabilities of the Federal statistical system.

This common interest has an especially important new dimension. The major opportunities that exist for extending the scope of the file available for analysis with some reasonable economy of effort in the near future rest in an exploitation of the records that are (or can be) generated by the State and local public agencies. However, their utility, and the utility of the file of the Federal system, will be immeasurably enhanced if these records can be brought into reasonably compatible association.

We are witnessing a burgeoning interest on the part of the State and local groups in developing the mechanisms for setting standards for these files and for maintaining and servicing them over time. This source of information is bound to emerge in importance and size. It is particularly important, therefore, that improvements in the Federal program lay the groundwork that will permit effective integration of the Federal file with other sources as they emerge. Furthermore, these emergent efforts are going to require guidance and leadership in setting standards and designing systems in a way compatible with total requirements. Much of this leadership must be supplied by example and by cooperative effort by a Federal system that is moving in response to modern requirements and opportunities.

THE SOLUTIONS ARE COMPLEX

It seems clear from the foregoing problem characterization that the solutions to this set of problems will have to be multidimensional.

When it made its proposal for a national data center as a solution and developed its justification, the Ruggles committee revealed an understanding of many of these dimensions. Its proposal was a constructive one and intended to be interpreted with some flexibility. However, the representation of the solution in this way has had some unfortunate consequences not anticipated by the committee.

In designating the center as a national data center and placing considerable emphasis upon the collection of tapes—growing out of its concern with the archival problem—the proposal became quickly translated in the minds of many as another data bank proposal.

The data bank idea is enjoying a considerable fad at the present. Many people have grasped this as the solution to their information problems. They have been encouraged by the substantial success that some fairly restricted and specialized information systems have had. A number of businesses, for example, have enjoyed some success in pulling their management records into a compatible and useful information system. The impression is widespread that bringing machine records together into some kind of central file will be instrumental in resolving the data problems of the broad class of users who attempt to use the files of the Federal Government.

This notion is supported by a general misunderstanding about the character of the files and their use. Those supporting this view are impressed with the fact that many different users have intersecting requirements for the same sets of data produced by the Federal statistical system. These records are, accordingly, viewed as general-purpose files. The convenience and economies of scale of bringing these records together into a common repository seem obvious. The obstacles to effective use under the present system are interpreted as technical and bureaucratic limitation amenable to this kind of technical solution.

What is not often adequately appreciated is the fact that general-purpose data are always used to fill special-purpose needs. This means that, while there are many intersecting interests in the same files, the impact on the file of each use may be quite different in terms of the organization, the levels of disaggregation required, and, most importantly, in the way the file needs to be associated with other records. It is this need for record association that is paramount and the source of most of the difficulty as was represented above. What makes a record a general-purpose record is not the fact that many users have an interest in its dimensions. It rests upon the file being constructed on the basis of standards, maintained in effective condition, and serviced by institutional arrangements and a technical system capability that will allow it to be reprocessed and combined successfully with other records in a wide variety of ways that will meet the special requirements of a wide range of users.

Thus, the key to solving these problems does not reside in the assembly of the records in a center but in the capacity to provide certain forms of file management and utilization services to the user. The effective provision of these services may require the assembly of some of these records into an integrated file, but this is defined by technical system requirements and is not the central issue it is made to be by many representations. It is important to characterize such a program as a data service center. The proposal is too important and fundamental to be burdened with its association with the naive data bank concept.

The Ruggles committee explicitly formulated at least a part of this rationale in their report and, hence, were putting forward a constructive proposal worth serious consideration. However, the committee never made explicit the way in which the problems of file compatibility rest upon the collecting and tabulating procedures of the agencies. It needs to be emphasized that these are important dimensions of the problem. Extending the mission of the Federal statistical system to provide user servicing capabilities based upon the new technology can do a great deal to extend the utility of existing records. However, the logic of a flexible service capability rests upon the ability to manipulate statistical building blocks. The development of these building blocks is a production task not contemplated in the suggestions for a data center. Some modification of current production practice will be essential for success. Indeed, if this problem is not tackled on a broad front, the generation of the servicing capability will fail to provide the kind of service intended and aggravate the sources of friction and dissatisfaction vis-a-vis the producing agencies.

RECOMMENDATIONS

The concept of a national data center is an appropriate vehicle for program reform if the concept is broadened to emphasize the role of the servicing capability and if it can be given an important role in assisting the Bureau of the Budget establish standards and monitor compliance. Accordingly, the basic recommendation is for the establishment of a national data service center whose primary mission would be to provide service to users of Federal statistical data both inside and outside of the Government.

This Service Center would have to be designed from the outset to incorporate certain basic functions:

(1) Direct the file storage and management for significant archival records in machine readable form for all participating agencies.

(2) Provide a central referral and reference source for the users of Federal statistics. This would include the development and maintenance of a formal reference index and the development of statistical reference specialists.

(3) Provide explicit facilitating services for the users of Federal data. This capability would consist of the following kinds of services:

(a) File rearrangement and cross tabulation to meet special needs and provide an extended range of output options in the form of maps, graphs, charts, and other media in addition to traditional tabular forms;

(b) Tape translation and other forms of file modification to bypass some of the inconsistencies and deficiencies in file management;

(c) Record matching where file compatibility exists or can be developed by file rearrangement;

(d) Disclosure bypassing where requirements violate legislative or administrative regulations;

(e) Perform standard statistical routines that form an essential part of the strategy of record matching and disclosure bypassing and which join routine computation with retrieval in a manner that makes a whole set of computationally derived numbers a latent part of the file of the Federal statistical system.

The National Data Service Center would be prepared to perform these services for:

(a) Archival records under direct management control of the Service Center;

(b) The current and accumulated records of administrative and regulatory agencies; and

(c) As a system resource or facility available to be used in connection with the current records of any agency where the need cannot be adequately met by the agency.

(4) Develop the computer hardware and software systems essential to the file management and servicing functions.

(5) Provide the staff support to work in conjunction with the Bureau of the Budget to develop and establish standards essential to the system capability.

There are a number of areas in which new or revised standards will be essential:

(a) Standards that define the records to be preserved in archival form;

(b) Standards for documentation and file maintenance, and

(c) Standards for the classification and coding of statistical data with special attention to respondent units and other forms of statistical building blocks.

(The sooner some of the standards related to the establishment and maintenance of archival records can be established the better. The review of the program [in app. B and in the next section] indicates that the most useful archives and the most economical are those that are developed under proper control and coordination from the present forward. The necessary procedures can then be built into the routine processing of data. This suggests some urgency for making as much headway with these issues as possible. The OSS should begin right away to work on establishing these standards without waiting on any formal actions on proposals for a data service center.)

(6) A research-analytic capability will be essential to the success of these functions. This does not mean developing the capability for conducting research and analysis directed toward issues of policy and management. Such analytic functions should be centered in the Executive Office and the operating departments. Policy research and analysis should be kept separate from the supporting function of supplying and servicing information.

A research capacity directed to an analytical evaluation of user requirements for the purpose of designing and developing the system components essential to perform these services is the essential capability. The construction of the reference file, the definition of standards in every category and the design of software routines and other system techniques that perform the facilitating services would all be controlled by what could be learned through research and analysis about the systemic elements of user requirements.

Some indication of the direction this analysis will have to take can be gathered from the National Bureau of Standards report in appendix C. A modest effort made to think through the kind of knowledge about user requirements that will be essential to system design and development is described there. In sum, it will be important to identify major classes of users, to learn the extent to which their requirements intersect the same sets of data, to learn the ways in which they require record matching from similar or different sources and the acceptable levels of aggregation. Only a systematic understanding of the joint and disjoint characteristics of the major requirements can serve to design an effective reference index, design relevant standards, and guide system design.

RESOURCE REQUIREMENTS

Many of the elements of this kind of program appeared in the Ruggles committee report and in the preliminary review. The task that has occupied recent months has been the attempt to document the needs more fully and develop some notion of preliminary specifications and costs.

The problem was broken into three parts for study and discussion and assistance sought with each. First, the essential ingredients for a reference and referral service were considered. A committee of knowledgeable people was assembled on an informal basis to discuss these issues (identified with the report in app. A). Second, a more intensive study was undertaken based upon the survey of machine-readable records conducted by the Office of Statistical Standards and contained in the appendix of the Ruggles committee report. In this way, an attempt was made to specify more clearly some of the costs of the archival function. The Bureau of Labor Statistics made part of the time of Mr. Mendelsohn available to carry this out (report in app. B). Third, an attempt was made to specify more clearly the essential elements of the system that would provide the facilitating services and what it would take to provide such services. For this purpose the National Bureau of Standards was used as a vehicle to assemble several people with a considerable range of knowledge of both the uses of Federal data and the production processes that generate them (report in app. C). In addition to these organized efforts I have discussed substantive issues with a number of knowledgeable people in the Federal agencies (both statistical agencies and program agencies) and in the universities (including an interview with the professionals involved in Project MAC at MIT).

This effort has yielded a better understanding of the nature of the problem and the system requirements. However, it has been somewhat less successful in specifying in detail the components of the system and the resource requirements. Let me review first the results and then evaluate the shortfall.

The reference function

In reviewing the requirements for the reference function the committee made a rough judgment that it might take as much as 5 years and an average of \$2 million a year to provide a meaningful reference and referral service for the Federal statistical system. This appraisal is limited in two ways, however. It is not the product of the kind of staff work in program planning that would be essential to a refined estimate and, therefore, represents only an informed speculation. More important, this estimate was generated with a view of the reference function as a discrete service unit or capability. It is recognized that a reference service would be more effective as an integrated part of a total service center program because the reference problem forms only a part of the larger problem set. If the provision for reference services is combined with other user services, the professional staff (particularly in its analytical and system development capacity), the computer facilities and other components of the service system could perform many joint functions. It is believed that because of the joint product character of these services, the incremental costs of providing a reference capability as part of a larger service system would be somewhat less.

The archival function

A review of the tape file inventory was undertaken which attempted to identify the important archival records and determine the costs of bringing these files to an acceptable level of file maintenance and documentation to be incorporated into an archive. The question of which records constitute significant archives rests, of course, upon an interpretation of requirements and development of standards not yet undertaken. In the interest of getting some feel for the dimensions of the problems this question was begged by arbitrarily preparing a list of the records considered to be vital general purposes series on a judgment basis to represent a sample archive.

On the basis of the data included in this sample archive it was estimated that a more complete archive would represent about 20,000 reels of magnetic tape and require an estimated \$3 to \$3.5 million and 3-5 years to develop. Of this amount about \$800,000 would be needed to bring data not now machine accessible into usable form, about \$500,000 would be needed to transfer punched card data to magnetic tape. In addition, between \$500,000 and \$1 million would be essential for blank reels and tape copying.

One of the interesting aspects of this report is the fact that almost half of this file (9,000 reels) could be brought into a data center for about \$260,000 within a year. This indicates that the files vary widely in the quality of their maintenance and documentation. The incremental costs of the second half is about \$300 a reel as compared with \$37 a reel for the first 9,000 reels.

Obviously, the costs of bringing existing files into archives are substantial and some review and justification will be needed. This can only be done within the context of a more comprehensive review of the user requirements that must guide planning in this area. A related sidelight of the report is that the files that are best maintained and can most easily be brought into an archive are not necessarily those that are most important in a usage sense. This is largely a function of the size of the files and the frequency with which it is produced. This suggests that an archive based upon considerations of cost and convenience in assembling existing records does not assure the most useful file.

At the same time, a large part of these records can be preserved at such a small cost that there seems little question that the investment in this resource is essential and justifiable. There is another important observation. About half of the total costs estimated are for system development and will have to be incurred even if the archival objective is addressed to current and future records only. These must be thought of as a capital cost of archival development as distinct from the costs of "dusting off" existing records.

It is true that these records will still contain all of the elements of file incompatibility that are the product of the production methods and standards that governed their generation. However, they appear sufficiently vital to current and future analysis that a total program should make a serious provision for trying to salvage some of the loss in data resources that has taken place in the absence of a policy and procedure for file preservation. In addition, every step needs to be taken to place future accumulations on a sound and economical basis.

The costs of bringing these records into an archival file do not represent all archival costs, of course. There are storage costs also (less than \$10,000 annually for 20,000 reels in prime air conditioned space), but these are inconsequential when compared to the need for facilitating services resting upon hardware and software systems to allow their effective use. These latter costs, however, cannot be fairly estimated at this point because, again, these services can be provided jointly by a facility which has a broader user service mission. As in the case of a reference service, an archive that is set up as a discrete service function will cost more than one incorporated in the total service complex.

A systems capability for facilitating services

It was through the agency of the Bureau of Standards that we attempted to assemble the intellectual resources to establish the scope of the program and the costs that would be required in establishing a system capability for providing the kind of facilitating services outlined in the recommendation. It is obvious from the foregoing discussion that this is the key to the program concept and to the evaluation of costs.

This turned out to be a difficult assignment. In our early attempts we found it exceedingly difficult to estimate program dimensions and costs without a clearer specification of the requirements the system will be designed to serve. Accordingly, a 3-day study session at Camp Ritchie was planned to see if we could break the back of this problem as the report in the appendix reveals.

This session made some progress in visualizing the requirements as well as assisted in clarifying some elements of the problem to be solved. In particular, we began to formulate some notion of how an analysis of user requirements might be structured and carried out. We could not, in the time spent, get to a more precise specification of the system elements. I feel that this work has laid the groundwork so that a series of additional work sessions of the same type might lead to a formulation of more specific program options. This task was found to be large enough that it did not seem advisable to undertake this kind of intensive staff work without a more specific decision on the part of the Director of the BOB concerning the kind of effort that is going to be devoted to this problem set and under what auspices.

At the same time some informed speculative judgments about the order of magnitude of costs developed out of the discussion. A total program of the type outlined under the recommendations would probably start out with an expenditure of \$1 or \$2 million annually in the first years and grow to the neighborhood of at least \$10 million a year. In the early years the size of the program will be controlled more by the practical limitations of assembling and training the kind of staff and acquiring the kind of equipment necessary. The fund could not be spent at a rate commensurate with the need and the objectives. A more detailed specification of program objectives, their phasing and the allocation of costs will have to rest upon additional staff work and should be preceded by some tentative policy decisions that will guide the work.

Staff requirements

Just as it is not possible to detail costs, it is not possible to detail personnel requirements. However, all of the reports or the discussions leading to them emphasized one point. The kind of statistical reference specialist, user service specialist and statistical systems analyst that is required to make this kind of program work either does not exist or is in extremely short supply. This implies (1) that program development will be constrained at the outset by intellectual resources and not financial resources, and (2) a successful program in this field will have to make explicit provision for professional development and training both in its program and in its budget.

Technical requirements

This is also the place to point out that there has been nothing in the entire review to suggest that an effort of the kind outlined in this report would be technically constrained. There has been some indication that existing computer hardware has been designed with greatest concern for computational capacity and is not as economical or as flexible as possible for the management and servicing of very large scale numerical files. However, the existing state of the art contains the essential elements of a more appropriate hardware system.

Similarly, the software routines for file management and servicing will need to be developed, but there is nothing to indicate that these problems of system development are not tractable. What is indicated is that considerable work must be expended over time to create these capabilities. There is every assurance, however, that the state of the art is adequate to support fully this kind of effort.

ISSUES OF ORGANIZATION

My views on the organization issue have been strengthened by the months of study since the preliminary memorandum. First, I cannot visualize a meaningful program addressed to the interrelated set of problems discussed above without a considerable degree of centralization of function. Some form of inter-agency service center will be essential. Second, if such a center is developed with existing agency structure essentially unmodified, it cannot perform its mission without agency cooperation and without explicit accompanying modification of agency missions.

This suggests that an effort of this scope could not be implemented without seeking new legislative authority. Legislation will have to be worked out and sought that will permit the service center to receive file custody, that will relieve the agencies of their disclosure restrictions as they pertain to the release of data to the center, and, at the same time, transfer the agency's disclosure obligations to the center. No workable independent center could be developed without meeting this issue head on at the outset. Further, the kind of program coordination and control of standards that will be essential may require legislation giving some interagency program authority to the new center. In addition, a single budgetary instrument for implementing the new program would be essential.

This kind of formulation inevitably leads to speculation about the organizational forms that might serve this end. I would like to react to some of the speculation proposals known to me:

(1) A new independent agency; this is certainly the cleanest solution. It could be accomplished with a minimum of "ad-hoc-ery" and would provide maximum flexibility for planning and innovation. It would be unencumbered by many existing agency jealousies and provide the freest opportunity for developing new leadership. At the same time, it might foster a coalition of agency opposition. Given the attitudes both in the Bureau of the Budget and on the Hill about new independent agencies, it might be difficult to manage. Barring this, some existing agency would have to form the vehicle for the program. Several have been mentioned in this context.

(2) GSA-Archives. The General Services Administration and, specifically, Archives, has been pointed to as an agency with already existing interagency authority and concerned with a part of this function. I am extremely dubious about the viability of such a solution. In the first place, this is not primarily an archival problem. It is primarily a complex problem of file management and coordination and rests upon a base of production practice that must be involved in the solution. It must develop a mixed professional staff of reference specialists, statistical specialists, subject matter specialists, system design specialists and programmers and technical services staffs. These resources do not exist even in embryo in these agencies. They would be handicapped by their image in building up the quality intellectual resources the program would need to succeed. Furthermore, these functions would not represent a primary mission from the point of view of the management of these agencies. An even more compelling objection is the fact that it would take the first tentative steps toward some integration of the user services of the Nation's statistical system down an organizational path that might make more difficult the achievement of desirable emergent forms.

(3) One of the existing statistical agencies: I do not believe that any of the existing agencies offer a desirable home for this function. It is true that they have already accumulated some of the expertise and equipment and management services and experiences that a new venture of this type requires. However, their mission concepts are conservative and inhibited in this area. The leadership is lacking. Perhaps more important, old interagency jealousies, etc., would make it more difficult to develop an atmosphere of cooperation.

(4) Compromise between the existing agency and independent agency solution: Of the existing agencies Census is certainly the most logical candidate by virtue of both its mission and the caliber of its professional staff. Many of the disadvantages of assigning this function to a new agency would be offset if Census were made an independent agency itself. If this were done and the user service functions set up parallel and with equal organizational status to the Census function we might have something of the best of both worlds. Something of this type might also have the advantage of being a constructive first step toward some degree of reorganization of the Federal statistical system.

(5) The National Bureau of Standards: It has been pointed out that the Bureau of Standards performs an interagency mission and has an unusual combination of existing legislative authorities to receive funds from and distribute funds to agencies, to set up special institutes, to use visiting scholars, etc. It already has an interagency service tradition and has been explicitly given the responsibility for assisting Federal agencies in planning computer systems. They also have in existence an emerging computer utility that might serve some of the needs.

(6) National Resources Evaluation Center: This agency has been suggested because it is an independent agency in the Executive Office of the President with existing interagency authority and responsibilities that extend beyond the mission of the Office of Emergency Planning where it is housed. It already contains a very large file of integrated Federal data from the various statistical agencies and has built up a staff with more experience in integrating interagency records than any other agency. It has a large computer installation organized for large-scale file management. It has an already existing interagency committee with the major statistical agencies represented and each of the agencies have one or several professional employees assigned full time to the activity. It is a conceivable vehicle if its authorities and functions were extended and removed from OEP.

If an effort to implement such a program goes forward, a great deal of thought and discussion will have to go into reviewing the kinds of options represented

here. I would like to offer the following related observations: First, there is a major threshold that must be negotiated if we deal meaningfully with the problem and program set at issue here. The kinds of services recommended cannot be subdivided without imposing upon a more limited function, serious functional handicaps, unnecessary expense and, possibly, seriously threatening its chances of success. I am concerned that partial measures may, in the end, do more harm than good. Since such an effort cannot spring into being "full-blown," it will have to be time-phased over a considerable period. However, the program should be considered as a whole.

Second, from an organizational point of view it seems inescapable that whatever initial action is taken, the end result will be a substantial reorganization of the Federal statistical system. It is very important that the organizational vehicle used at the outset does not predetermine the future evolution of the system in a way that limits its ability to implement essential subsequent phases.

Third, it seems to me that in discussing this problem set and proposed program, we are really engaged in a discussion of what kind of Federal statistical system we want to develop in the next generation. I am much less sanguine now about whether it is possible, or even desirable, to keep these issues in the background. I am sufficiently concerned about the abortive potential of solutions that fall short of a critical threshold and organizational arrangements that might inhibit essential lines of development that I feel it important to consider the issues of organization in a straightforward way. I would go further and suggest that every one of the six interim solutions outlined above has serious limitations, although some have a more open-ended character than others. My own preference would be to handle the organizational problem at the outset as a reorganization of the Federal statistical system. There are a number of indications that this might be a favorable time to do so.

Fourth, this predilection is reinforced when I reflect upon the great importance production practice plays in the whole configuration of problems and solutions. I am also inclined to believe that a fundamental improvement in the integration of production practice can offer a constructive solution to the paperwork problem of respondents without jeopardizing important components of a general purpose information system. Another factor reinforcing this inclination is the conviction that some form of integrated leadership can go far in dealing in a creative way with the joint interest of existing Federal programs and agencies whose current large expenditures for data now constitutes a large resource waste.

PRIORITIES

The comprehensive scope of the issues presented here plus the fact preliminary staff work cannot specify and cost explicit program options suggests certain priorities. First, a continuous focus of leadership needs to be generated. The proposal has already been put forward that an interagency committee be established to provide this focus.

Second, wherever the leadership function is vested, it seems to me that the highest priority is to provide this focus with the staff support essential to identify requirements and specify the elements of the system that must be provided for. The earliest requirement is to engage the research-analytic capability identified under item 6 of the recommendations. The development of specific program options, the definition of their specifications, costing these elements, and identifying the essential order of a time-phased program will require early intensive staff support of a very special kind.

Third, a beginning can be made under existing authority upon some of these problems before new programs and organizations can be developed and funded. At least two kinds of effort could be begun right away.

(a) The standards that shape the content of archival records and determine the essential forms of file maintenance and documentation need to be worked out and made a part of ongoing programs. A beginning can also be made in formulating the kinds of standards that will produce statistical building blocks essential to file compatibility.

(b) The 9,000 tape file record identified in the Mendelsohn report constitutes a nuclear archive that can be generated quickly at a very modest cost. Funds should be made available to the agencies to begin the creation of this basic archival record.

The Bureau of the Budget should seek funds to carry out these preliminary measures under its own authority. The staff work and the extended effort applied to statistical standards should be centered in the Office of Statistical

Standards. The funds to develop the basic archive could be transferred to other agencies as a part of a controlled plan.

Several hundred thousand dollars might profitably be requested in the fiscal 1967 budget for this purpose.

APPENDIX A

APPENDIX MEMORANDUM

Subject: Report of informal committee on the reference problem.

THE COMMITTEE

An informal ad hoc committee was assembled by Edgar Dunn, acting as chairman, to discuss the problem of developing an adequate reference service for the Federal Statistical System. The committee was composed of Joe Daly and Edwin Goldfield of the Bureau of the Census, Rudolph Mendelsohn and Robert Steffes of the Bureau of Labor Statistics, Ezra Glaser of the Patent Office and Edgar Dunn, consultant to the Office of Statistical Standards. These individuals participated with the knowledge and consent of their agencies but served as individuals and informed professionals. No attempt was made to get agency clearance or establish agency points of view. It was merely an informal attempt to formulate some judgments about the nature of the problem and its solutions. The committee met on several occasions on an irregular schedule during the summer of 1965.

THE PROBLEM

It was agreed that one of the serious obstacles to the fullest utilization of the information resources of the Federal Statistical System is the absence of any clearly defined reference function. The inadequate nature of these services is traceable to the production orientation of the agencies stemming from their primary missions as data publishers. The reference function has generally been thought of as the responsibility of the documentary centers. To the extent that the agencies attempt to provide occasional reference assistance, the task falls to an individual whose mission is defined as a production responsibility. The inadequate nature of this kind of service is also traceable to the fragmented nature of the numerical records of the Federal Statistical System growing out of the divided agency responsibility for their generation and maintenance. Because of the decentralized nature of the Federal statistical program the referencing problem is made especially complex and no agency has been in a position to perform a generalized service with reference to the total file.

The problem is becoming increasingly important in recent years as important uses of the numerical files are more frequently taking the forms of records in machine readable form rather than the traditional documentary form. The problem is also fed by changes in information usage that are leading to more complicated information requirements. Records are needed more often in disaggregate or special form not met by traditional documentary formats and they are often used in combination in ways that require extensive intelligence about the compatibility of records in several dimensions.

DESIRED REFERENCE CAPABILITY

Ideally the Federal statistical system should be able to develop a reference system that has the capacity to deal with inquiries in an efficient and creative way that would facilitate access to the records and extend their utility. The clientele is conceived to cover a wide range of sophistication and types of need. However, the requirements for a reference capability stem primarily from a large and growing core of intermediate information processors that service the research aims of academia and the decision and administrative requirements of business and government at all levels.

Such a service should be able to:

1. Help the client refine his inquiry and frame it in a way acceptable to the system and, in the process, give some preliminary information about the scope and nature of materials implied in his request as an aid to further defining, sharpening, and limiting the inquiry. (Experience of the Science Information Exchange and the National Science Referral Center have pointed up the great importance of this function even in dealing with highly trained professionals.);

2. Provide, by drawing upon a reference index and other reference tools, a fairly complete documentation of formal intelligence concerning—

- (a) the number and size of relevant file sets,
- (b) their taxonomic descriptors,
- (c) their mode of preparation (census, survey, etc., questionnaire forms, etc.),
- (d) their quality characteristics including (1) quality of the data (sampling and response errors, etc.), and (2) quality of the files (state of documentation and file maintenance),
- (e) the extent to which the taxonomic and qualitative characteristics of the data will support merging and collating series for various purposes,
- (f) where the data are located and how they may be accessed including such information as the form of the file (published, machine readable, machine language and format, etc.) and access costs in both time and money;

3. To perform a switching service so that the user can access the needed records efficiently. (The uninitiated user needing to access several data sets in different divisions of several agencies can be faced with a complex switching problem difficult to handle efficiently unaided.);

4. To provide a "semiautomatic Joe finder" to facilitate access to informal and specialized intelligence concerning the records and their characteristics. This would need to be a person-oriented service that would revolve around the role of the professional reference specialist who would deal with the client by person, by mail, and by phone.

IMPORTANT CONSIDERATIONS IN IMPLEMENTATION

A general concensus emerged from the committee discussion about the principal issues or problems to be resolved in the process of implementing such a goal. There was also agreement about the general form of the solutions. The most important issues revolve around three points:

The reference file

In order to perform his role effectively the reference specialist would have to have access to a set of formal reference aids that would constitute the elements of an emerging reference system. The principal aid is visualized as a formal reference index that would probably be machine oriented. This index would attempt to bring into a reference file the kind of reference intelligence implied in the previous sections (2 a through f) that could be gathered and formulated in a formal system. Such a file would be designed to facilitate an iterative search procedure and to generate documentation to service each inquiry. After some initial period of development this file might have the capacity to generate, periodically, one or more condensed summary index documents that could serve as visual reference aids not only in the reference center itself but in documentary and service centers throughout the United States.

The construction of such an index would be a professional task of considerable magnitude and complexity. It would take time and resources to develop and maintain on a current basis. Indeed, the development of such a file would represent an ongoing task that, by its very nature, would never be completed.

The order in which the components of this reference intelligence are selected for development and the form of their organization into a file should not be random but governed by systematic priorities. Furthermore, it should be only marginally controlled by the ease with which such reference material can be organized out of existing materials. The development of the file should be controlled by a research-analytic effort on the part of the staff that would provide guides to the emergent usage. Accumulating knowledge about request incidence will be only partially helpful. Maximum effectiveness of the file will rest upon an explicit effort to identify the principal classes of users (in terms of their analytical requirements and problem orientations), the way in which their requirements are common or disjoint, the way in which they generally intersect different statistical records, and, therefore, the nature of the reference intelligence necessary to serve each class of user. Some idea of the systematic character of the requirements is essential in order to do an effective job of designing a responsive reference system.

The reference specialist

The reference file is only a tool. The key to successful data referencing is the reference specialist.

In the context we are discussing here he has three recognizable functions. First, the professional reference specialist is the essential human link in deal-

ing with the reference client. He must often assist the client in refining his inquiry to a form that will facilitate response. He will provide the interface with the formal reference tools such as the reference index and the "semi-automatic Joe finder." He will supplement these sources with his own fund of informal intelligence gleaned from experience.

Second, the reference specialist must provide the professional analytical capability to undertake system-oriented research of user requirements and develop on a time phased basis the operating characteristics of the reference system.

Third, the reference specialists must undertake the task of constructing the reference system and its component formal reference index. This will require bringing together and systematizing large amounts of technical intelligence and incorporating it into an operating system.

One of the biggest problems in developing a reference capability is that the kind of professional reference specialist that is characterized here does not exist. It is a new kind of professional capability of emerging importance. There are a few men in established agencies whose work experience fits them with attributes that come close to the functional requirements outlined above. They are very limited in number, however, and, characteristically, are key men in fulfilling agency missions.

The success of the attempt to develop a reference system will rise or fall on the strength of the kind of professional talent that will guide its development. Since the kind of experience that is necessary in this function is rare, an essential part of any program effort will be an explicit recognition of this fact and an explicit procedure for the training and development of statistical reference specialists.

Ways will have to be worked out so that new professionals could have rotating assignments that would carry them into the primary statistical agencies where they could (1) work on specific components of reference information for the purpose of implementing the reference file, (2) come under the supervision of and receive training from those people who represent the greatest fund of accumulated knowledge, and (3) receive a total system orientation that could be gained in no other way. Project financing would have to incorporate explicitly the resources that would support staff training and development.

The task is made somewhat less formidable by the fact that the three categories of reference functions outlined above are susceptible to some degree of specialization. During developmental phases the intellectual resources of the staff could also be supplemented through consulting arrangements. It would be useful, for example, for the operating agencies to assign some of their specialists to work with the reference service on a temporary or part-time basis. The analytic or system design component might be especially amenable to supplementation during the early phases.

The organizational form

The feeling was strong that a successful effort to develop a Statistical Reference Service would require some degree of centralization of function. An important part of the reference function is interagency or total set in character and cannot be handled within the context of an agency orientation. At the same time, the reference agency can perform its function without involving the primary agencies directly. In the previous section we already outlined some of the ways in which agency participation would be indispensable. The agencies will need to play a role as a breeding ground for research specialists, and, of course, will be the source of most of the reference intelligence that must be used to construct and maintain an index. It may also be necessary and desirable to establish within at least some of the larger agencies a companion reference function that would be linked with the total reference capability.

The committee did not discuss the specific institutional form of such a service center within the framework of the existing institutions.

COSTS

The committee devoted some time to a consideration of the resources that would be required to establish such a reference service. It came to the conclusion that nothing very precise could be said about costs at this stage for several reasons.

First, the kind of evaluation that can be given by a group of this type at an early speculative state is suspect. A more refined notion would require the application of more staff resources to planning and evaluation than are currently available.

Second, the question of costs is confused by the possible existence of joint costs in this program area. The general problem set of which the reference problem forms a part has several other dimensions that extend beyond the restricted problem this committee has taken for discussion. The costs of establishing a reference service of the kind discussed here would be quite different if it were established as a discrete function of, if it were developed as one component of, a more generalized user service capability. If the provision of reference services were combined with other statistical services addressed to establishing and maintaining archives, servicing administrative records, or providing tape translation, disclosure bypasses and other file management services, the professional staff and program facilities including computer facilities could perform many joint functions. The increment costs of providing a reference capability as a part of a more extensive user service capability could possibly be a great deal smaller than would be required for a separate and independent function.

Despite these obstacles to cost estimation, the committee noted that the operating budgets for two agencies performing related (though in many ways basically different) functions ranged from \$400,000 a year for the National Science Referral Center in the Library of Congress to \$2 million a year currently for the Science Information Exchange of the Smithsonian Institute. The committee speculated that it probably would take as much as \$2 million a year on the average over a 5-year period to develop a meaningful reference service capability including enough resources to involve agency programs in the way necessary.

The committee also discussed the question of the demand for reference services and whether anticipated needs were appropriate to justify such levels of expenditure. The need for the service was judged to be sufficiently great to justify a serious effort.

There is no way, at this preliminary stage, that anything more than a judgment can be offered. In the first place, nothing in present agency experience can serve as a guide to demand levels for a service that has never existed in anything like the form indicated in this prospectus. The judgment rested on several considerations. First, there is considerable evidence of frustration and inefficiency because of the absence of such a service. Second, there already exist other programs, such as the ones referred to, that have been judged essential and for which expenditures substantially exceeding the sums of money mentioned here have been undertaken. Most of these have been undertaken in the interest of improving the efficiency of documentation in the physical sciences. In the circles where public and business policy are made, and social science, management and marketing research undertaken, the most compelling needs is not for a way to handle better the traditional documentary materials generated, but for a way to acquire efficient access to relevant numerical files that constitute the main bodies of evidence and of research inputs. The need for expanding the services in this area seems equally as compelling as those in the field of physical science documentation that are already receiving extensive attention.

APPENDIX B

OCTOBER 1, 1965.

From : Rudolph C. Mendelsohn.
To : Mr. Edgar Dunn.
Subject : Report on data inventory.

DATA BANK REQUIREMENTS

An estimated \$3 to \$3.5 million and 3 to 5 years are needed to stock the proposed data center with Federal statistics now in existence. These data would probably comprise about 20,000 reels of magnetic tape. However, a Federal center could be stocked with a respectable volume and variety of data relatively quickly for about \$260,000. At the rate of about \$27 per reel, a bank of 9,000 tape files could be established in about a year. Such a course would provide a fairly representative selection of significant data including, for example, 750 reels of the census housing data; census current population data on 375 reels; the BLS Consumer Expenditure Survey on 43 reels; the BLS industry hours, earnings, and labor turnover data on 36 reels; the OBE national income accounts on 2 reels; as well as IRS tax data on 5,300 reels and BOASI social security data on 1,900 reels. An insight into the volume of data readily available is gained by the rough calculation that the equivalent of nearly 1 billion punched cards would be included (table 1).

Of the \$3 to \$3.5 million needed to supply the data center over the 5-year period, about \$700,000 would be needed to bring data not now machine accessible into usable form. About \$500,000 would be needed to transfer punched card data to magnetic tape. And about \$1.5 million would be needed to reprocess data now on magnetic tape. Looked at another way, about 100 man-years of professional aid will be required to review and correct the records, develop comments to show file contents to the serious scholar from both the substantive and machine processing points of view. Also, about 100 man-years of keypunching and a very rough estimate of 10,000 computer hours are indicated (table 2). In contrast, about 13 man-years of professional work and 1,200 hours of machine time would stock the center with half the total in a year.

The major resource requirement of the 5-year effort is for the Census Bureau where \$1¼ million is requested. Over half the amount reported for that Bureau, about \$700,000, is needed to bring the 25- and 5-percent population samples for 1960 to acceptable levels. On the other hand, the Internal Revenue Service and the Bureau of Old Age and Survivors Insurance, both among the giants of data processing, have requested relatively limited amounts \$17,400 and \$14,300, respectively.

In considering the course of stocking the center several key factors should be kept in mind. First, the cost of additional historical reels after the initial storage of 9,000 is quite high—about \$300 per reel, compared with about \$27. The cost per reel is considerably higher for some of the files. For example, the SEC Quarterly Financial data and the FRB Report of Condition of Insured Banks would cost \$20,000 and \$8,000 per reel, respectively. Obviously, a careful review and justification for high-cost files is needed before their improvement can be supported. Second, the costs indicated in this report refer only to those needed to make data accessible within the responsible agency. I am assuming that the proposed data center would defray the costs of tape copying and would supply its own blank reels. Such costs are not inconsiderable. The 9,000 reels which could be made ready in about a year would cost the center over a half million dollars for blank reels and for copying.

GENERAL COMMENTS

I have the general impression that the larger the volume of data and the higher the frequency of processing the greater is the tendency for the files to be in acceptable order. That is to say, large files like those of the IRS and BOASI and the high-frequency operations in the BLS manpower field and in the FRB are in good shape while some decennial and annual operations at the Census Bureau and the relatively small files at SEC, FTC, OBE, and the Office of Education are either not well maintained with the computer or are not well mechanized at all. In other words, the degree of accommodation to the computer seems to be a function of the work pressures to use it.

The vast majority of available information is already in machine form. The small amounts of significant data not machinable are found in OBE, the Office of Education, and in the Department of Agriculture. The files in OBE and OE are not large, would total about 200 reels, in my estimation, and are not in machine form in appreciable amounts. As you know, efforts are underway to correct this in both agencies. About 25 percent of the Agriculture data are now machinable, according to Department representatives. It was asserted that the remaining 75 percent are significant and useful data and should be available to a data bank for research in agricultural economics.

AGENCY COMMENTS

Three general comments tended to be made by agency representatives. First, and least frequent, assertions that the Budget Bureau survey and the Ruggles Committee report has brought an increased awareness of the need for more effective file maintenance and that efforts to achieve this end would be incorporated in ongoing operations where feasible. These good intentions should be supported with funds where appropriate and the dilemma in which the Census Bureau finds itself with respect to the 1960 population samples should not be allowed to be repeated. I have no doubt that the \$700,000 now required would have been far less if the job had been done as a part of the 1960 census work.

Once the records have been brought to acceptable levels through new financial support, I doubt that programs operated at high frequencies will require more

than trifling amounts to maintain this level in current operations. On the other hand, many of the recurring annual and decennial operations need explicit support for the improvement of file management practices. The 1960 population samples are examples of the latter situation. An assessment of these costs is outside the scope of this survey and, in fact, should be considered a part of the cost of the survey without regard to a proposed Federal data centers.

Then, there was the comment by the larger agencies, heavily stressed by the Census Bureau, that funds to improve the files would not necessarily generate the desired results. The argument was that only present staff members had the background and professional experience needed to do the job. Since these people were already fully occupied, could not be diverted, and new staff could not do the job, it would not be done even if money is supplied it was asserted. I do not concur with this view. The work can be done if it is properly phased; that is, if sufficient time is allowed, new workers under the supervision of experienced personnel can do the job in the long run (the 5-year span I have suggested). I believe the reaction cited above assumes a crash effort to organize the files on a high-priority basis. I agree that it cannot be done this way and advise against such an approach.

Agency representatives seemed excessively concerned with the confidentiality question. Turning data over to a Federal center would be a breach of contract with respondents who have been assured that none but agency personnel would view their reports, it was said. I tried to convey the assurance that, if a data center were established, it would assume the obligation of protecting both the agency and the respondent. Since feelings on this matter run quite deep, some steps should be taken at the outset to vitiate them or discussions beyond this narrow consideration could founder.

One constructive suggestion was made in regard to confidentiality. Mr. Robert Menke of the Securities and Exchange Commission expressed the view that corporate concern dealt mainly with current affairs. It was his feeling that, after a period of 5 to 10 years, back data could be exposed to public view without serious objection by respondents. There would be difficulty perhaps in applying such a rule retroactively but a notice to this effect on future collections of data might serve to make the problem less troublesome in the years ahead.

CONCLUSION

I have a final comment. I found the evident loss of data because of the failure to support good file management distressing. Immediate steps ought to be taken to stop this erosion of a national resource. The costs indicated above measure the deficiency of not doing it before, and they will grow as time passes. It is difficult to argue that these losses have immediate meaning. The tools, techniques, and intellectual attitudes needed for their useful exploitation are not yet reflected in our institutions. But, as you know, changes are already underway. And even though we are unable now to predict how the store of data might be used. I am convinced that actions to preserve this national resources will be appreciated by those who follow.

TABLE 1.—Data file inventory—Selected files which can be ready in 1 year, by agency and project

File name and data type	Current number of tape reels (1)	(2) Improvement through added resources				Total costs (4)
		Substantive personnel (a)	Systems analysis and programming (b)	Machine processing (c)	Number of tape reels after additional resources (d)	
Total, all agencies.....	10,340	\$11,100	\$71,540	\$126,125	8,971	\$261,475
Census Bureau	2,230	11,100	41,600	78,500	1,407	155,750
County City Data Book, 1952, 1956, 1962..... summary	3				3	
1960 Census of Housing, basic 25 percent summaries..... source	1,474		12,000	55,000	750	74,500
Monthly Survey of Intentions..... do	172		5,000	8,000	1,000	12,000
Annual Survey of Population Survey..... do	375		22,000	8,000	375	15,000
Current retail report..... summary	64	8,000	1,000	5,000	40	45,000
County business patterns..... do	32	100	100	500	32	14,500
Manufacturers shipments, inventories, orders..... do	100	3,000	1,500	4,000	25	800
Bureau of Labor Statistics	10				10	8,700
Survey of Employment, Payroll, and Hours..... source	184		23,000	29,500	79	54,700
Survey of Industry Employment, Payroll, Hours..... summary	84		23,000	25,000	30	48,400
Survey of Industry Labor Turnover..... do	2				2	400
Survey of Employment, Payroll, and Hours..... do	3				1	400
Consumer Expenditure Survey, 1960-61..... source	44			3,500	3	400
Office of Business Economics					43	4,100
National Income and Production Accounts..... summary				1,000	2	1,000
Internal Revenue Service				1,000	2	1,000
Individual Income tax returns..... source	5,303				5,303	17,400
Corporation income tax returns..... do	2,503					3,000
U.S. business tax returns..... do	1,475					4,500
U.S. partnerships returns..... do	754					3,000
Tax model for individual income tax returns..... do	232					2,050
Tax model for corporation income tax returns..... do	316					2,850
Tax model for partnerships returns..... do	20					1,000
	3				3	1,000

TABLE 1.—Data file inventory—Selected files which can be ready in 1 year, by agency and project—Continued

File name and data type	Current number of tape reels (1)	(2) Improvement through added resources				Cost of documentation (3)	Total costs (4)
		Substantive personnel (a)	Systems analysis and programming (b)	Machine processing (c)	Number of tape reels after additional resources (d)		
Total, all agencies—Continued							
Federal Reserve Board	74				50	\$150	\$150
1963 Survey of Financial Characteristics	25				1		
Reports of condition of all insured commercial banks	35				35		
Reports of income and dividends, all insured banks	11				11		
Daily money supply	1				1	150	150
Industrial production index	2				2		
			\$4,200	\$10,125	215	5,050	19,375
National Center for Health Statistics	220						
Annual questionnaire	35		400	1,000	35	1,800	3,200
National Natality Sample Survey, 1963	2		400	25	1	200	625
National Mortality Sample Survey, 1961-65	4		500	75	1	350	925
National Mortality Sample Survey, 1960	2		400	25	1	200	625
Annual birth and fetal death statistics	87		1,000	4,000	87	1,000	6,000
Annual death statistics—detached data	90		1,500	5,000	90	1,500	8,000
			2,740	8,000	1,915	3,360	14,100
Bureau of Old Age and Survivors Insurance	2,379						
1 percent continuous work history sample, 1937 to date	196		870	4,000	120	1,080	5,950
0.1 percent continuous work history sample, 1937 to date	33		870	1,000	31	1,080	2,950
Name and address file tape	34						
Summary earnings tape	1,600				1,600	200	200
Regular transcript master benefit tape	450		1,000	3,000	70	1,000	5,000

TABLE 2.—Partial data file inventory—Totals by agency and type of data

Agency	Current number of tape reels	Professional (substantive, system, and documentation) costs	Transcription costs	Machine processing costs	Total costs	Number of tape reels after additional resources
	(1)	(2)	(3)	(4)	(5)	(6)
Total, all agencies.....	18,831	\$1,239,610	\$544,300	\$1,017,825	\$2,788,235	16,269
Source.....	18,338	999,410	503,000	939,440	2,441,850	15,988
Summary.....	493	240,200	41,300	78,385	346,385	282
Bureau of Labor Statistics:						
Total.....	385	218,100	32,000	81,400	318,000	124
Source.....	148	68,900		44,100	113,000	102
Summary.....	237	149,200	32,000	37,300	205,000	22
Bureau of Employment Security:						
Total.....	32	6,900	1,000	22,650	30,550	74
Source.....		400	1,000	850	2,250	1
Summary.....	32	6,500		21,800	28,300	73
Bureau of the Census:						
Total.....	10,047	566,550		715,050	1,281,600	8,106
Source.....	9,838	545,800		705,050	1,250,850	7,994
Summary.....	209	20,750		10,000	30,750	112
Office of Business Economics:						
Total.....	1	10,900	300	2,250	13,450	5
Source.....		6,000		1,500	7,500	2
Summary.....	1	4,900	300	750	5,950	3
Department of Agriculture:						
Total.....	155	119,500		37,290	156,790	151
Source.....	155	91,500		35,780	127,280	141
Summary.....	0	28,000		1,510	29,510	10
Internal Revenue Service:						
Total.....	5,303	17,400			17,400	5,303
Source.....	5,303	17,400			17,400	5,303
Summary.....						
Federal Trade Commission:						
Total.....	0	8,000	2,000	3,000	13,000	7
Source.....	0	7,000		2,000	9,000	6
Summary.....	0	1,000	2,000	1,000	4,000	1
Office of Education:						
Total.....	10	2,315	0	2,570	4,885	21
Source.....	0	2,115	0	2,370	4,485	10
Summary.....	10	200	N.A.	200	400	11
Bureau of Old Age and Survivors Insurance:						
Total.....	2,538	32,650		110,500	143,150	2,074
Source.....	2,538	32,650		110,500	143,150	2,074
Summary.....						
Securities and Exchange Commission:						
Total.....	0	213,500	56,500	9,400	279,400	12
Source.....	0	193,500	52,000	9,250	254,750	11
Summary.....	0	20,000	4,500	150	24,650	1
Federal Reserve Board:						
Total.....	75	30,050	452,500	15,640	498,190	125
Source.....	71	20,400	450,000	15,090	485,490	109
Summary.....	4	9,650	2,500	550	12,700	16
National Center for Health Statistics:						
Total.....	285	13,745	0	18,075	31,820	268
Source.....	285	13,745	0	12,950	26,695	235
Summary.....	0	0	0	5,125	5,125	33

TABLE 2, PART I.—Partial data file inventory—Data on magnetic tape, by agency and type of data

Agency	Current number of tape reels (1)	Improvement through added resources (2)				Cost of documentation (3)	Total costs (4)
		Substantive personnel costs (a)	Systems analysis and programming costs (b)	Machine processing costs (c)	Number of tape reels after additional resources (d)		
Total, all agencies.....	18,831	\$211,200	\$304,915	\$840,880	15,717	\$126,680	\$1,484,675
Source.....	18,338	172,300	267,615	814,830	15,545	110,380	1,365,125
Summary.....	493	38,900	37,300	27,050	172	16,300	119,550
Bureau of Labor Statistics:							
Total.....	385	29,400	59,700	46,100	101	20,000	155,200
Source.....	148	2,900	26,100	31,600	86	10,400	71,000
Summary.....	237	26,500	33,600	14,500	15	9,600	84,200
Bureau of Employment Security:							
Total.....	32	1,000	1,000	3,000	32	-----	5,000
Summary.....	32	1,000	1,000	3,000	32	-----	5,000
Bureau of the Census:							
Total.....	10,047	176,100	201,600	658,500	7,810	66,850	1,103,050
Source.....	9,838	165,000	199,000	649,000	7,700	65,800	1,078,800
Summary.....	209	11,100	2,600	9,500	110	1,050	24,250
Office of Business Economics:							
Total.....	1	300	100	50	1	500	950
Summary.....	1	300	100	50	1	500	950
Department of Agriculture:							
Total.....	155	4,400	7,500	10,780	100	5,400	28,080
Source.....	155	4,400	7,500	10,780	100	5,400	28,080
Internal Revenue Service:							
Total.....	5,303	-----	-----	-----	5,303	17,400	17,400
Source.....	5,303	-----	-----	-----	5,303	17,400	17,400
Office of Education:							
Total.....	10	(¹)	(¹)	(¹)	10	(¹)	(¹)
Summary.....	10	(¹)	(¹)	(¹)	10	(¹)	(¹)
Bureau of Old Age and Survivors Insurance:							
Total.....	2,538	-----	29,290	110,500	2,074	3,360	143,150
Source.....	2,538	-----	29,290	110,500	2,074	3,360	143,150
Federal Reserve Board:							
Total.....	75	-----	-----	-----	51	5,150	5,150
Source.....	71	-----	-----	-----	47	-----	-----
Summary.....	4	-----	-----	-----	4	5,150	5,150
National Center for Health Statistics:							
Total.....	285	-----	5,725	12,950	235	8,020	26,695
Source.....	285	-----	5,725	12,950	235	8,020	26,695

¹ Not available.

TABLE 2, PART II.—*Partial data file inventory—Data on punched cards, by agency and type of data.*

Agency	Cost of professional services (1)	Machine processing costs (2)	Tape reel equivalents (3)	Total costs (4)
Total, all agencies.....	\$367, 415	\$139, 120	478	\$506, 535
Source.....	292, 715	106, 285	376	399, 000
Summary.....	74, 700	32, 835	102	107, 535
Bureau of Labor Statistics, total.....	64, 000	18, 800	21	82, 800
Source.....	29, 500	12, 500	16	42, 000
Summary.....	34, 500	6, 300	5	40, 800
Bureau of Employment Security, total.....	4, 500	18, 800	41	23, 300
Summary.....	4, 500	18, 800	41	23, 300
Bureau of the Census, total.....	122, 000	56, 550	296	178, 550
Source.....	116, 000	56, 050	294	172, 050
Summary.....	6, 000	500	2	6, 500
Office of Business Economics, total.....	6, 000	1, 500	2	7, 500
Source.....	6, 000	1, 500	2	7, 500
Department of Agriculture, total.....	102, 200	26, 510	51	128, 710
Source.....	74, 200	25, 000	41	99, 200
Summary.....	28, 000	1, 510	10	29, 510
Federal Trade Commission, total.....	7, 000	2, 000	6	9, 000
Source.....	7, 000	2, 000	6	9, 000
Office of Education, total.....	2, 315	2, 570	11	4, 885
Source.....	2, 115	2, 370	10	4, 485
Summary.....	200	200	1	400
Securities and Exchange Commission, total.....	57, 500	6, 775	5	64, 275
Source.....	57, 500	6, 775	5	64, 275
Federal Reserve Board, total.....	1, 900	490	12	2, 390
Source.....	400	90	2	490
Summary.....	1, 500	400	10	1, 900
National Center for Health Statistics, total.....	0	5, 125	33	5, 125
Summary.....	0	5, 125	33	5, 125

TABLE 2, PART III.—*Partial data file inventory—Data not on machineable forms, by agency and type of data*

Agency	Professional services cost	Transportation costs	Machine processing costs	Tape reel equivalents	Total costs
	(1)	(2)	(3)	(4)	(5)
Total, all agencies	\$229, 400	\$544, 300	\$23, 325	75	\$797, 025
Source	156, 400	503, 000	18, 325	67	677, 725
Summary	73, 000	41, 300	5, 000	8	119, 300
Bureau of Labor Statistics, total.....	45, 000	32, 000	3, 000	2	80, 000
Summary	45, 000	32, 000	3, 000	2	80, 000
Bureau of Employment Security, total.....	400	1, 000	850	1	2, 250
Source	400	1, 000	850	1	2, 250
Office of Business Economics, total.....	4, 000	300	700	2	5, 000
Summary	4, 000	300	700	2	5, 000
Federal Trade Commission, total.....	1, 000	2, 000	1, 000	1	4, 000
Summary	1, 000	2, 000	1, 000	1	4, 000
Securities and Exchange Commission, total.....	156, 000	56, 500	2, 625	7	215, 125
Source	136, 000	52, 000	2, 475	6	190, 475
Summary	20, 000	4, 500	150	1	24, 650
Federal Reserve Board, total.....	23, 000	452, 500	15, 150	62	490, 650
Source	20, 000	450, 000	15, 000	60	485, 000
Summary	3, 000	2, 500	150	2	5, 650

APPENDIX C

THE DESIGN OF A FEDERAL STATISTICAL DATA CENTER

(A report to the Bureau of the Budget, prepared by E. Glaser, D. Rosenblatt, M. K. Wood, National Bureau of Standards)

SUMMARY AND CONCLUSIONS

This report was prepared in response to a letter from R. T. Bowman, Assistant Director for Statistical Standards, Bureau of the Budget, to A. V. Astin, Director of the National Bureau of Standards. The original request was focused principally upon "the possibilities for developing new capabilities in computer hardware systems that will improve their flexibility and economy in specialized file storage, management and retrieval functions" in connection with a national data service center. It directly became evident that any such study would be substantially conditioned by the characteristics of the information system to be mechanized: the scope and content of the economic and demographic data to be included; the degree of detail for each kind of information; the ability to use (a) data of more than one kind in a single analysis or mathematical model, (b) data derived from different reporting systems, (c) data collected by different agencies, (d) data referring to different time periods; the kinds of access to the files and the formats of acceptable queries; the nature and extent of computational and manipulative services to be provided; and other system specifications. Since there was no definite specification with regard to these characteristics, it was decided that a preliminary description of such a system was prerequisite to the requested analysis of hardware characteristics.

Mr. Edgar S. Dunn, consultant to the Office of Statistical Standards, Bureau of the Budget, worked with the staff of the National Bureau of Standards in reinterpreting the original request. As a result, the group's efforts were directed toward the issues that would govern the design of an effective Federal Statistical

Data Center. The original intent to explore the potentialities of modern large-scale computers is constantly in evidence in the present formulation.

In preparing the report, the problem of describing the customer population was considered first. Something is known of the kinds of specialists who use data originating in the Federal Government to solve problems in economic policy, public administration, business economics, business administration, and a great range of social science subjects. But it is also apparent that this present user population reflects the capabilities and logistics of present ways of organizing and purveying data. At least equal importance attaches to those needs which are not met by present practices. How can these unmet needs be characterized?

An adequately specified information system would have to be based upon a broad review of the types of analyses that a wide spectrum of social scientists propose and upon the quantitative models that they build. For the present purposes and the limited scale of effort, it was decided to restrict the review to several classes of economic models directed at problems of national economic policy. Even this limited review revealed a variety of possible requirements for socioeconomic information which are not now being met, although many of the basic data are collected and compiled in some form by some Federal agency.

The review of economic models and their needs for statistical information was conducted at a 4-day conference at Fort Ritchie, Md. on August 26-29, 1965. Participants were M. K. Wood, D. Rosenblatt, and E. Glaser of the National Bureau of Standards and E. S. Dunn and P. F. Krueger of the Bureau of the Budget.

Subsequent conferences and staff work built upon the Fort Ritchie conference by developing (a) an enumeration of the services to be rendered, and (b) a description of the Federal Statistical Data Center in terms of its functions and principal characteristics. A summary of these is given below.

4. Nature of the services to be rendered

An integrated Federal Statistical Data Center appears necessary to perform the following functions:

1. To provide data in cases where the primary agency in possession of the data is not capable of making it available in the required format, detail, flexibility, or quality.

2. To provide data where the information originates in two or more reporting systems or agencies, in order to make available information about interrelationships in maximum feasible detail, without restrictions resulting from screening for improper disclosures at the time of transfer into the Center and through association of information from multiple sources relating to the same individual reporting unit or analytical unit.

3. To maintain an archive of statistical data, complete in the sense described in 2 above, with all corrections and adjustments carried through in a consistent manner, and with a collection of the accompanying codebooks and manuals.

4. To provide information outputs (responses to queries) in a variety of forms at the customer's option: printed tabulations, machine readable tapes, graphs, diagrams, etc., either locally or through telecommunications.

5. To establish, maintain currently, and operate a reference and referral service for the Federal statistical system.

The creation of such a Federal Statistical Data Center also should provide the following additional services an corollary benefits at minimum cost:

6. ADP equipment would be available for computation and data reduction in response to queries of customers: cross tabulations, averages, distribution statistics, smoothed curves, trend fittings, seasonal adjustments, periodic analyses, correlations, regressions, and more advanced analyses in order to give access to the full range of information computable from the collection.

7. Confidentiality audits would be performed by machine upon the information intended for release to customers.

8. ADP equipment would also support a battery of services to the statistical system of the Federal Government: computations essential to the conduct of test adjustments on statistical series and collections, computations for test reconciliations of data from two or more sources or for two or more time periods, detection of errors in primary collections or derived statistics through consistency tests and anomaly detection routines, computations necessary for the study of error propagation through the Federal statistical system, combinations of the above computations in support of validation studies for Federal statistics and in support of procedures for certification of the accuracy and consistency of Federal statistics.

9. ADP equipment would also be used to provide service agencies with large-scale adjustment and reconciliation tasks (which is already being done by some agencies) in the production of standard series and to assist in the creation of new series through the reconciliation, adjustment and transformation of standard series.

B. Character and organization of the data in a Federal Statistical Data Center

1. Principles must be developed for the initial selection and future accession of data to be included in the Center's collection; they should reflect a broad range of uses and full utilization of basic information rather than a codification of present uses, present practices, and present compromises.

2. Methodology and principles must be developed for the conversion of present files and production data to suitable archive form and quality, and their maintenance in proper form and quality, supported by codebooks, manuals, etc.

(a) The principles established for archives must presume that data will be transferred from the collecting agency to the Center in full available detail.

(b) The principles established for archives must provide for the systematic completion of all corrections and adjustments to all data affected and all levels of detail, resulting in a fully reconciled and consistent body of data.

(c) The principles established for archives must provide for data to be transferred to the Center without screening for confidentiality; all confidentiality audits and checks would be applied to the formats and information content of the output of the Center.

(d) The principles established for archives must provide for the preservation of the identity of the reporting unit and the association of all information about the unit without regard to the agency or manner of the collection of the information; the rules and the economics of matching of existing records will be very different from those intended for future production of data.

3. Standards must be developed for definition, coding, classification an aggregation with the intent of maximizing the ability to use different kinds of data in the same analysis and of minimizing the loss of information.

4. Standards must be developed for formats in order to facilitate the management, housekeeping and retrieval of records and to avoid the loss of information.

5. Standards must be developed for quality of data (consistency of definition, error rates, etc.) and for means of assuring maintenance of quality.

6. Automatic data processing (ADP) equipment and systems must be available to the Center to provide economy, timeliness, and flexibility of access to the information in the records.

(a) ADP must be available

(1) to compute statistics that are inherently computable from the records,

(2) To provide answers in the required form, and

(3) To avoid unnecessary withholding of information as a result of using inefficient and redundant disclosure criteria.

(b) ADP must be available for the conduct of confidentiality and other disclosure audits, such rules and procedures to be applied to the data in the form and content intended for release from the Center.

7. Criteria must be developed for assuring that the Center is established, and is maintained, in a manner that is responsive to a broad base of potential users, rather than in a manner which seems to suit the present habitual users at any time.

8. Criteria must be devised for periodic review of the value of the data contained in the archive followed by a selective purging of the data whose retention is no longer justified.

DESCRIPTION OF A FEDERAL STATISTICAL DATA CENTER

Introduction—Federal statistics and the computer

The statistical services of the Federal Government were initially created in response to a variety of unrelated needs. The census of population, in its early simple form, arose in response to a specific constitutional provision. The first census of manufacturers was a purposeful study of the existing status and likely potential economic development of the young Nation. Other collections of data were directed at a continuing study of the operation of the economy: prices,

employment, sales and inventories, production of specific minerals, etc. A large number of statistical collections arose in connection with specific Federal programs, the statistics themselves being largely byproducts: activities and finances of regulated industries, internal revenue statistics, health and educational defects among selective service registrants, grants to scientific researchers, veterans' benefits, etc.

Two major influences have been at work to give improved, quality and cohesiveness to this initially piecemeal collection of information about all aspects of the Nation and its people. One was primarily organizational and the other technical.

The Federal Reports Act of 1942, building upon such earlier programs as that of the Central Statistical Board of the National Recovery Administration, created a coordinating mechanism for the improvement and rationalization of the Federal statistical system. The Director of the Bureau of the Budget was given staff and responsibility for the development and introduction of standards for the collection, processing, and dissemination of data through much of the Government's activities. The standard industrial classification, the standard metropolitan statistical areas, the standard sample week for monthly surveys, and the standard base periods for economic time series are examples of this standards-setting function. The financial reporting program and the current population and labor force program are examples of the coordination of the work of several agencies to produce data useful for a variety of purposes. The Bureau of the Budget had become an instrument for promoting systematic cooperative efforts among the many Federal, State, and private collectors and processors of information.

The technical base for improved quality of information also has roots in the past. The development of improved techniques for acquiring information, analyzing it, preparing it for publications, and using it for economic, social, and political studies has been active for over a century. With the growth of applied social sciences and the progressive elaboration of the Government's statistical activities, the pace of research in technical methods was greatly accelerated. Trained statisticians brought improved techniques to many aspects of their work: sampling, the design of experiments, seasonal adjustment of time series, the construction of national and regional accounting models, the study of non-sampling errors in surveys and censuses, the development of quality control and other sequential methods, and the interpretation of data in complex situations. The consequences of these technical improvements have been far reaching.

During and directly after World War II, the design and construction of the first electronic computers foretold a potential for vast improvements in many aspects of statistical technology: the recording and editing of field survey data; the compiling, tabulating, and publication of data; the analysis of data and their use in problem solving. The Federal Government pioneered in exploiting these new capabilities. However, much more can be done, particularly in the design of better ways of organizing economic and social data, more thorough integration of information from the many separate statistical programs, and the reduced loss of information in utilizing data for analytical purposes and surveying it to various classes of customers.

The very general logical powers, the great storage capacity, the high speed of manipulation, and the low unit cost of modern ADP systems combine to promise great potential improvement in information resources and problem-solving capabilities. A number of Federal agencies have learned the advantages of mechanizing their routines. Indeed, the Bureau of the Census contracted for the development of the UNIVAC, the first commercial internally programmed computer, and it acquired the first and fourth units produced. Many of the frequently cited Federal statistics are more promptly and more satisfactorily produced than would be possible without computers, whether the statistics arise from a primary function of the agency or as a byproduct.

Yet, the improvements were typically made within the context of a single agency—usually a single reporting system—and without the possibility of raising broad questions about the fundamental organization of the Federal statistical system as a potentially unified and cohesive collection of intelligence. Nor was there any practical way of applying the explosively growing power of computers to general questions of preventing loss of information once it had been brought into the system by one or another Federal agency. Finally, there has been no serious attempt to assess the consequences of the computer for improved access to Federal data or for meeting the need for providing information in the form, degree of summarization, format, and physical output desired by various classes of customers.

The present report is addressed to these questions.

General principles for the organization of data for a Federal Statistical Data Center

The consequent reconsideration of the organization of socio-economic data in the Federal Government is based upon two general guides. The first is to review the implicit informational requirements of the whole range of analyses and formal models proposed by social scientists, rather than to restrict the statement of information requirements to those needs which have been given principal attention in the past. The second is to consider the whole range of relevant tasks that the computer can assist, even if the manner of proceeding is radically different from current practice.

From this reconsideration four general principles emerge for constructing specifications for a Federal Statistical Data Center:

1. Maximum ability to exhibit the interrelations among various kinds of data;
2. The unification of all information about the individual reporting unit or analytical unit;
3. The preservation of detail in the basic records and the avoidance of loss of information in the storage, manipulation, and retrieval of information; and
4. The ability to produce the full measure of inherent information which is computable from the basic records.

These four principles will now be developed as groundwork for specifications for the information organization and the services of a Federal Statistical Data Center.

One of the greatest deficiencies of the existing Federal statistical system is its failure to provide access to data in a way which permits identification and measurement of functional interrelationships among interdependent activities. Identification and measurement of such interrelationships are essential to a wide range of economic and social analyses. It is also the chief problem in the design of mathematical models of economic and social processes suitable for appraising the impact of alternative policies and programs as well as possible changes in environmental factors.

Such appraisal is, in turn, prerequisite to effective benefit-cost analysis of proposed and ongoing programs. The essence of rational benefit-cost analysis is the tracing of indirect as well as direct effects of programs and the evaluation and summing of these effects. Typically, the methodology for tracing all but the most obvious linkages is entirely lacking or fails to use the relevant information.

Until recently, economic model builders have been restricted to relatively aggregative economic and resource flow models, and to inferring interrelationships among very few aggregative variables. Such relationships often have considerable predictive value where other conditions remain relatively stable or continue to change at a constant rate. But the essential relationships are correlative or associative rather than structural. Hence, they generally fail to give acceptable prediction when other conditions change markedly, as a result of changes in major program, policy, or environmental factors.

Acceptable prediction under changing circumstances requires analytical models which give much more detailed and explicit recognition to interrelationships among the criteria and variables which will be affected by the changed conditions. Such analytical models generally describe the mechanisms in greater detail than the associative models; they use more information, and they often rely less heavily on trends or the postulation of only slow changes among the variables in the model. The present and prospective accelerated pace of technological and statistical change now requires the development and use of more detailed and complex models than can be created or supported by the present Federal statistical system.

The rapidly developing tools of automatic data processing and systems analysis now make possible—and necessary—both the development of more advanced models and the elaboration of the Federal statistical system which is needed to support them.

Many of the data needed for establishing causal interrelationships among related economic variables are contained in the existing Federal statistical system. But present collection methods, tabulation procedures, and disclosure rules combine to make it difficult and often impossible to extract such data. Where samples are largely enough, it may be possible to cross-tabulate in a way which permits determining the interrelationships between two variables or, rarely,

among three. But generally when more than a two-way relationship is involved, it is impossible to tabulate necessary totals in a way which will define the desired relationship without disclosure of proprietary data. Such relationships among many variables can be extended, however, at a much greater level of detail, if it is possible to apply standard statistical analysis techniques to the observations for individual respondent units over the whole range of the relevant variables. It is possible in this way to extract much more useful structural information and still insure that no disclosure of individual respondent data is contained in the results of such analyses.

In many cases, the data necessary to such an analysis require the matching of items from two or more statistical sources. One important class of analyses involves the matching or reports by the same respondent for different time periods. Some data files are so organized as to make this possible, but many are not. An even more complex problem arises when it is necessary to match data from the same respondent collected as parts of different statistical programs, by different agencies. This can be extremely difficult or impossible, though substantial progress has been made in some areas, as, for example among Census, Bureau of Old Age and Survivors Insurance (BOAST), and the Internal Revenue Service. Problems of disclosure are most difficult in this context.

There are several fundamental problems dealing with the coding and classification of original source data. Most serious is the need for uniform identification, definition, and coding of the respondent unit. Unless this is done, matching of data from diverse sources is generally impracticable if not impossible. A uniform system of classification and coding for geographic area is another major deficiency.

In general, the classification and grouping of data are dictated by the problem environment, the basic logic of the analytical model, and the kind and degree of detail in which the results must be expressed and interpreted. In practice, there is frequently need to compromise the ideal classifications and aggregations of data for several reasons; the basis and criteria of classification in the collection agency being inconsistent with the ideal requirements of the model; the lack of sufficient detail (industry, process, product, geographic location, etc.); the withholding of detail under proprietary confidentiality or security restrictions; the noncompatibility of the definitions of the respondent units in the several collection systems which could otherwise provide the information specified by the model, which can be reconciled only by coarse aggregation but with accompanying loss of information and structural detail; the noncompatibility of classification of the data by several collection agencies and information systems also capable of specious resolution by aggregation; the difficulty and cost of identifying and matching the reporting units from two or more reporting systems, so that the information about the reporting unit can be pooled; the absence of technique, staff, funds, and machine time to use large-scale data processing equipment to recode, recompile, reconcile, reclassify, and aggregate data and to perform all manner of statistical procedures upon the data.

Since there are very large numbers of ways in which most economic variables might reasonably be classified and aggregated, it is not practical to prepare the data in all of these formats in anticipation of possible requests. Nor is this necessary. The same results can be achieved with favorable logistics and great flexibility by providing for the basic records to be maintained in machine-receivable form and in as fine detail of classification as is practical. The low unit cost and high speed of modern computers can then be exploited to meet requests for data with little loss of the available information inherent in the combined resources of the participating agencies.

The availability of modern computers can meet two important requirements in this context. The first is discussed above: the conversion from finely disaggregated classes to all manner of special purpose classifications and aggregations (and, indeed, conversion to publishable forms). The second requirement is to avoid unnecessary loss of information because of proprietary and confidentiality restrictions. The fundamental rule in this case is to perform all edits and checks relating to unwanted disclosure upon the fully processed data (aggregations, summaries, averages, correlation coefficients, regressions, fitted curves, etc.) rather than upon the detailed raw data. This will assure full use of information consistent with disclosure rules. The logical capability of the computer also provides the key for the necessarily elaborate systems of rules essential to the prescribed protection.

Another major class of problems arises from the fact that errors and inconsistencies in the data as reported, transcribed, and coded are always discovered

in the process of editing tabulations for publication. These errors are generally corrected at the levels of aggregation at which data are published, but often are not carried back to the basic records for the individual respondent unit, in machine sensible form. Carrying back such corrections to the basic files is prerequisite to the kind of analysis of interrelationships which is here proposed. Some method of insuring that this is done, and that the basic records meet appropriate standards as archives, is essential. Such tasks generally receive a low priority in the statistical agency whose primary task is production and publication, rather than analysis, of data.

As a technical device, the use of master samples can achieve a high degree of unification of information about the individual reporting unit. Moreover, proper experimental designs provide for the straightforward estimation of sampling variances; differences among subpopulations can be measured with specified precision if this requirement is stated in advance; variances due to differences among samples can often be eliminated; costly matching of units at the later stages can be avoided. Inconsistencies arising from many kinds of differences between surveys can also be avoided.

The use of the current population survey for special questions (veteran status, duration of unemployment, preferred number of hours, work, etc.) permits a number of useful comparisons with standard information about labor force status. Similarly, it may often be practical to use master samples to obtain information about subjects vital to some of the newer Federal welfare programs. Hence, an integrated system of master samples of households could be used to collect information about income, education, health, crime, employment, social services, housing, demography, voting registration, and the effects of opening or closing industrial plants. Not only could information be compiled about each of these subjects, but analyses could be performed which inter-related several of the subjects: education-income-crime rates, health-housing-education, etc., without loss of information or the introduction of uncertainty arising from variances between samples or from different survey practices. Indeed, the judicious use of master samples can lead economically to conformity with the general principles stated above.

The fourth principle is ability to produce any information computable from the basic records. The principle acquires new power when combined with the other three principles, because a great deal more becomes computable. The concept of "inherently computable" is taken literally, and includes kinds of statistical operations not now widely used. The paucity of current use derives from unfavorable economics, unsuitable organization of data, insufficient available detail, failure to use known techniques, obstacles growing out of confidentiality restrictions on data intended for input to the analysis, and current habits and practices deriving from all of the other obstacles. In short, the current ways of doing business fall far short of the potentiality of advanced statistical techniques applied to a well-organized body of Federal data. The present report suggests the means for mitigation or elimination of the shortcomings of the statistical system built before modern computers became available.

In these terms, the notion of "inherently computable" takes on new meaning. Obviously included are the routine computation of averages, cross-tabulations, correlations, curve fittings, time series analysis, seasonal adjustments, distribution statistics, and the application of other techniques of mathematical statistics. But it would also now be possible to test the reconciliation of one series of data against others. Test adjustments of all sorts, even very detailed and burdensome adjustments involving manipulation of very large matrices, could be countenanced. Errors could be studied, including those for whose estimation there is little theoretical foundation—the myriad kinds of inconsistencies of definition, practice, error rate, personnel, etc.—when data from two or more sources are used. Propagation of errors through the system, especially in the major synthetic series (national income and product accounts, Federal Reserve Board production indexes, price indexes) could also be studied and estimated. One set of objectives would be error detection and measurement in the primary collections. Another would be consistency testing and anomaly detection in two or more collections from different agencies, geographic regions, time periods, etc. Computations of this sort could also be used to assist in the setting of quality standards for Federal data and for validation or certification of particular bodies of data.

There is already a praiseworthy trend toward the use of computers in the production of standard series of data by several agencies. The notion of "inherently computable" includes the generation of new series for special purposes

through the adjustment of standard series, limited only by the techniques and imagination of social scientists.

Nature of the services to be rendered

This section discusses the services which a Federal Statistical Data Center could render. The characteristics are enumerated as they are in the summary of this report. An information system capable of providing these services is described in the next section.

The services which are proposed for this suggested system are discussed below:

1. The Federal Statistical Data Center would provide data in cases where the primary agency in possession of the data is not capable of making it available in the required format, detail, flexibility, or quality. Primary agencies would continue to provide data which they can furnish in the needed form, even though they had previously delivered the relevant basic data to the Center. For example, an agency might produce statistics as a byproduct of its principal mission, having no resources to organize the information for flexible or rapid access. Or the data might require adjustment or reconciliation which the collecting agency cannot perform as well as the Center.
2. The Center would provide data where the information originates in two or more reporting systems or agencies, in order to make available information about interrelationships in maximum feasible detail, without restrictions resulting from screening for improper disclosures at the time of transfer into the Center and through association of information from multiple sources relating to the same individual reporting unit or analytical unit. The intent of this specification and its improvement over present characteristics of the Federal statistical systems are discussed in an earlier section.
3. The Center would maintain an archive of statistical data, complete in the sense described in (2) above, with all corrections and adjustments carried through in a consistent manner, and with a collection of the accompanying codebooks and manuals. The intent of this item is discussed in an earlier section.
4. Outputs (responses to queries) would be provided in a variety of forms at the customer's option: printed tabulations, machine readable tapes, graphs, diagrams, etc., either locally or through telecommunications.
5. The Federal Statistical Data Center would establish, maintain currently, and operate a reference and referral service for the Federal statistical system. This service is not concerned with the actual provision of data. It deals more with those matters that a user might need before he can formulate a proper query. The reference and referral center would give information about various concepts that lie behind the statistics: general imports in contrast with imports for consumption; total employment and employees in establishment; value of product and value added; industry and product statistics, etc. Questions that could not be answered at the Center would be referred to specialists in the various agencies; the Center would identify and locate the specialists. It would also protect the experts from inquiries that could satisfactorily be managed at the Center. Personnel at the Center would be equipped with reference documents to show dates for which each kind of data is available, changes in coverage, changes in definition, changes in quality, schedules for availability of future statistics; materials available in book or report form both for data and information about their definition, method of collection, adjustment, etc. Reference services would also be provided for information not in the Federal collection: statistics from trade associations, industrial institutes, State and local governments, international organizations and foreign governments. Statistical data that can be obtained directly from the primary collection agency would be known to the Center, which would act as a referral agent for the agency.
6. ADP equipment would be available for computation and data reduction in response to queries of customers: cross tabulations, averages, distribution statistics, smoothed curves, trend fittings, seasonal adjustments, periodic analyses, correlations, regressions, and more advanced analyses in order to give access to the full range of information computable from the collection.
7. Confidential audits would be performed by machine upon the information intended for release to customers. It is recognized that this raises complex and difficult issues which require intensive study. However, there are strong reasons to believe that these issues can be resolved with the aid of modern tools of the mathematical and computer sciences.

8. ADP equipment would also support a battery of services to the statistical system of the Federal Government: computations essential to the conduct of test adjustments on statistical series and collections, computations for test reconciliations of data for two or more sources or for two or more time periods, detection of errors in primary collections or derived statistics through consistency tests and anomaly detections routines, computations necessary for the study of error propagation through the Federal statistical system, combinations of the above computations in support of validation studies for Federal statistics and in support of procedures for certification of the accuracy and consistency of Federal statistics. Much of the work referred to here is not done at present. Ordinarily, the larger synthetic statistical series are prepared by gathering data from many sources and adjusting them in various ways including their reconciliation to benchmarks of higher quality. In many cases, the source series themselves are compounded from smaller elements, sometimes in several stages before reaching down to the point of primary collection from the respondents. Computers are used for convenience and economy to speed up the processing in most of the more elaborate systems. However, in this statistical production network, there is practically no feedback of information from this process to the primary collection agencies. The adjustments required to maintain the larger synthetic series are sufficiently burdensome and closely scheduled that there is neither time nor staff for research on adjustments or the conduct of test adjustments no matter how desirable this might be in the view of the interested agencies. The combination of the comprehensive unified data system and adequate ADP equipment would create a favorable climate for this work. In addition, all manner of test comparisons across different statistical series, and many kinds of consistency tests, could readily be performed. With much of the synthesis of major statistical series on compatible computers, the effects of errors in all stages of collection, estimation, and adjustment could be studied. Hence, studies of the quality of Federal Statistics could add such techniques to existing appraisals which are based on information about the collection (completeness, sampling variance, quality checks), size of adjustments to benchmarks, and a very limited kind and number of consistency checks.

9. ADP equipment would also be used to service agencies with large-scale adjustment and reconciliation burdens (which is already being done by some agencies) in the production of standard series, and to service the creation of new series through the reconciliation and adjustment of standard series. Specialized users could define new synthetic series based upon adjustment of the standard series. However, at present, such adjustments could be applied only to highly aggregated forms of the statistics because of the cost and cumbersome nature of the process. What is contemplated here is a much more complete reprocessing designed to retain a large measure of the detail available for the standard series. For example, the input-output transactions matrices (which are now embedded in the national income and product accounts) could be transformed from the present industry-based sectoral definitions to an activity basis (in which there are no secondary products).

Character and organization of the data in a Federal Data Center

This section presents and discusses principles governing a well-integrated body of statistics arising from the work of the Federal agencies. The items discussed below are numbered as they are in the summary of this paper. While there are intimations of services that the Center might perform, there is no attempt to describe the services as such in this section; the preceding section is devoted entirely to that end. This section relates to the internal structure and operation of the Center—in matters of information—and the preceding section views the same Center from the outside, as a series of capabilities to assist the customer to obtain data.

The principles are discussed in numerical order below. While it is convenient to set forth the seven separate items for exposition and reference, the entire characterization is conceived as a single entity: no item is to be read out of its context with the other items. It is the interaction of the points taken two, three, or more at a time that characterizes this report, in contrast with possible studies of the distinct issues one at a time.

1. Principles must be developed for the initial selection and future accession of data to be included in the Center's collection; they should reflect a broad range of uses and full utilization of basic information rather than a codification of present uses, present practices, and present compromises.

The selection should recognize the importance of data acquired in the administration of regulatory programs and welfare or benefit programs. A primary purpose of the Federal Statistical Data Center is the organization of information in such way as to permit the use of data from various sources in the same analysis. The byproduct information from many agencies must now be reconsidered to determine how these data can best be combined with those of other reporting systems to contribute to socioeconomic analysis. Both program data (amount of grant, number of grantees, geographic location of program elements, etc.) and information about applicants can enrich the existing store of socioeconomic statistics from major statistical agencies.

The above paragraph is addressed to only one of the general principles discussed in an earlier section. Issues of inclusion or exclusion of various classes of data must be reviewed with all four general principles in mind and also with appreciation of the remaining items on this list itself.

2. Methodology and principles must be developed for the conversion of present files and production data to suitable archive form and quality, and their maintenance in proper form and quality, supported by codebooks, manuals, etc. It is not to be presumed that complete and consistent records will arise routinely from the collection, adjustment, analysis, and publication of data. Resources must be made available and priorities assigned. Above all, standards of form and quality must be prescribed and checked in some regular manner. Procedures must also be prescribed for work with the archive collection to prevent loss or contamination of the master records by tape erasure, statistical adjustment, aggregation or reclassification.

(a) The principles established for archives must presume that data will be transferred from the collecting agency to the Center in full available detail. The decisions about the lowest level of detail—other than the separate record for each respondent or analytical unit—will often be arbitrary. They will reflect notions of the finest detail that analytical purposes are likely to demand. In principle, there is no such ultimate disaggregation for many reporting units. For example, the use of the establishment as the reporting unit in many standard statistical systems is frequently dictated by the inability to define or obtain information for subestablishment entities; the choice is not based upon satisfaction with the level of detail obtained. The term "full available detail" must be read with a rule of reason.

(b) The principles established for archives must provide for the systematic completion of all corrections and adjustments to all data affected and all levels of detail, resulting in a fully reconciled and consistent body of data.

(c) The principles established for archives must provide for data to be transferred to the Center without screening for confidentiality; all confidentiality audits and checks would be applied to the formats and information content of the output of the Center.

(d) The principles established for archives must provide for the preservation of the identity of the reporting unit and the association of all information about the unit without regard to the agency or manner of the information; the rules and the economics of matching of existing records will be very different from those intended for future production of data.

3. Standards must be developed for definition, coding, classification, and aggregation with the intent of maximizing the ability to use different kinds of data in the same analysis, and of minimizing the loss of information. The reasons for this requirement are set forth in an earlier section of this report. A large number of standards would have to be developed, beginning with such seemingly elementary concepts as a household, a structure (a building), a business organization, an establishment (industrial), a populated place, a county; and proceeding to a school pupil, a hospital day, a recipient of (some particular) welfare service, etc. Some such standards now exist, although they are neither wholly satisfactory nor uniformly observed. These existing standards should be reexamined and many new standards developed. All standards need to be more rigorously defined and more effectively enforced. Ideally, a close matching in many dimensions of classifications, hierarchical aggregation, timing, and spatial extent should apply to all information in the basic record; but a system of practical compromises would unquestionably have to be accepted.

4. Standards must be developed for formats in order to facilitate the management, housekeeping, and retrieval of records and to avoid the loss of information. This item presumes that the characteristics of the information have, in principle, been defined. In practice, formats greatly influence the effectiveness, economy, and error rates of the whole operation.

5. Standards must be developed for quality of data (consistency of definition, error rates, etc.) and for means of assuring maintenance of quality. It is essential to know—and to issue with the statistical data—information on the quality of the data. After standards have been agreed upon, appropriate quality-control procedures would have to be instituted.

6. Automatic data processing (ADP) equipment and systems must be available to the Center to provide economy, timeliness, and flexibility of access to the information in the records. This report does not deal with configuration of computing equipment. It may be noted in passing, however, that various units might be geographically scattered. This would allow inquiries to be made from points distant from the basic record stores and the replies or outputs to be received in these same remote locations. Probably more importantly, computer laboratories in universities, research institutes, business organizations and governmental agencies could be used to transmit requests for information over long-distance lines and to receive and store information. This would provide a convenient location for trial manipulations by those making the inquiries without disturbing the rest of the communication network. Such an arrangement would give increased service and analytical power to the participating analyst.

(a) ADP must be available—

(1) to compute statistics that are inherently computable from the records,

(2) to provide answers in the required form, and

(3) to avoid unnecessary withholding of information as a result of using inefficient and redundant disclosure criteria.

Note that, for the item immediately above, the confidentiality audit would have to apply before the information was transferred to an off-line computer under the control of the user. This discussion implies that disclosure rules would retain their essentially logical character without taking into account the possibility of introducing elements of probabilistic inference in determining whether or not an undesirable disclosure might be made.

(b) ADP must be available for the conduct of confidentiality and other disclosure audits; such rules and procedures to be applied to the data in the form and content intended for release from the Center. The comment on item 6 (a) applies here as well.

7. Criteria must be developed for assuring that the Center is established—and is maintained—in a manner that is responsive to a broad base of potential users, rather than in a manner which seems to suit the present habitual users at any time.

8. Criteria must be devised for periodic review of the value of the data contained in the archive followed by a selective purging of the data whose retention is no longer justified.

In sum, the main purpose of a Federal Statistical Data Center is to create a better integrated information network, for use by Government, industry, and the research community, which will provide better understanding of interdependencies within our pluralistic society, leading to better informed choices among alternative policies and programs, and more effective program implementation.

APPENDIX 3.—THE NEW COMPUTERIZED AGE

[From Saturday Review, July 23, 1966]

Few technological developments are formidable enough to mark turning points in human history. Two such phenomena have occurred in our time: the atomic bomb and the computer.

The implications of the bomb are beginning to be understood—its capacity for instant and total destruction has been demonstrated. The implications of the computer as yet are only faintly comprehended. That they will be awesome is already apparent. Indeed, as Dr. Jerome B. Wiesner, Dean of Science at the Massachusetts Institute of Technology and former science adviser to President Kennedy, wrote recently in *The New York Times*:

"The computer, with its promise of a millionfold increase in man's capacity to handle information, will undoubtedly have the most far-reaching social consequences of any contemporary technical development. The potential for good in the computer, and the danger inherent in its misuse, exceed our ability to imagine. * * * We have actually entered a new era of evolutionary history, one in which rapid change is a dominant consequence. Our only hope is to understand the forces at work and to take advantage of the knowledge we find to guide the evolutionary process."

The following special section is an attempt to identify some of these forces and to consider their implications. Nine authorities of diverse backgrounds discuss the possibilities and dangers of a computerized age. As their reports make clear, ultimately no area of human life will remain untouched by it. In the words of Automation Consultant John Diebold, whose article, "The New World Coming," introduces the section, "A complete new environment will exist."

The changes in business, government, science, education, and communications are occurring at a time when our technological capacity already has outstripped our understanding of many of its ramifications; when, as Marshall McLuhan, University of Toronto professor who often is quoted on the influence of electronic media (see Erik Barnouw's article), has said, the tumultuous pace of change already has resulted in an "information overload."

Decades ago, W. B. Yeats wrote, "The visible world is no longer a reality, and the unseen world is no longer a dream." More and more this will be true in the computerized age.

In addition to Mr. Diebold and Professor Barnouw, contributors to the section are: Gen. David Sarnoff, chairman of the board of the Radio Corp. of America; John W. Macy, Jr., Chairman of the U.S. Civil Service Commission; Patrick Suppes, director of the Institute for Mathematical Studies in the Social Sciences, Stanford University; Don D. Bushnell, associate director of the Brooks Foundation, and past president of the Association for Educational Data Systems; the Reverend Vernon F. Miller, pastor of the Goshen City, Ind., Church of the Brethren; John Tebbel, New York University journalism professor and author; and John Lear, Saturday Review's science editor.

The editors wish to acknowledge the valuable contribution of William L. Schubert, of the McCall Corp., in the basic planning and preparation of this issue.

—The Editors.

THE NEW COMPUTERIZED AGE—1: THE NEW WORLD COMING

Tomorrow's computers will revolutionize business, education, communications, science—in ways only dimly foreseen

(By John Diebold¹)

It is an extraordinary era in which we live. It is altogether new. The world has seen nothing like it before. I will not pretend, no one can pretend, to discern the end; but everybody knows that the age is remarkable for scientific research into the heavens, the earth, what is beneath the earth; and perhaps more remarkable still is the application of this scientific research to the pursuit of life. The ancients saw nothing like it. The moderns have seen nothing like it, until the present generation * * * The progress of the age has almost outstripped human belief.

Those words were not spoken today—though I choose them to set today in perspective—but were used in 1847 by Daniel Webster when he opened a new stretch of railroad track in New Hampshire. A greater parallel exists between that era and our own than we normally realize. In that earlier era, science first began to be applied on a wide scale and out of that process came an entirely new society—an industrial society. Out of it, too, came problems, many of which still plague us. When we look back at that great technological upheaval, the real significance of those then-wondrous machines is the human and social change that accompanied their industrial use.

Just as yesterday's innovations proved to be moments in history—way stations leading to newer technology—so today the conception of the computer which we have learned to accept is becoming a thing of the past. Up-to-date systems are no longer glassed-in, carefully isolated accounting machines. Instead they perform an almost limitless variety of functions, and vary with individual requirements.

For example, the newest computer systems may appear as input-output units in individual desks; small televisionlike screens with keyboards and copying devices. When you ask a question you see the answer almost simultaneously on the screen. If you want a copy of the answer, you can make it immediately. The heart of the system is a switching center rather like the telephone system. Computers, storage elements of many varieties, and many other devices used as part of the system are accessible as you need them, connected through the switching center to the terminal unit at your fingertips. Thousands of people may use such systems at the same time, and each need know no more about the operation of the system than the average person knows about the telephone. In the next decade the typical computer system is going to be of this kind.

Another radical change stemming from these new computer systems involves the relationship between man and machine. One no longer need carry data down to a computer center, or go through a laborious process of getting it into the machine and then waiting for results. Each technological development is moving us toward an easier, more productive relationship between man and machine. Already, for example, a computer can transpose a rough design into exact specifications. If an engineer makes a free-hand drawing of a bridge on such a system's television-like screen, the computer will convert the drawing into exact engineering specifications, will calculate and display materials and stress, and show the design in whole, in part, or in any perspective, in immediate response to the engineer's requirements.

Looking ahead, we see important changes in technology such as chemical memories; fluid and pneumatic systems that have instantaneous response; ability to store images, graphs, drawings, and photographs, and to transmit them around the world. All these will be important elements of future computer systems. Graphic elements and the ability to communicate with TV screens are already becoming influential in progress being made in computer design. Yesterday these elements were undreamed of.

Work is being done on language translation by machine. Some document-translation is already on a regular production basis—in fact, people are now attempting to digest articles by machine. This work is still in its beginning stages and there are many problems to be overcome. But the history of this

¹ The author, who generally is credited with coining the term "automation," is head of the Diebold Group, Inc., management consultants.

technology is that what seems impossible today becomes an accepted part of our lives tomorrow.

Development of voice recognition by computer, while rife with problems, also is yielding results. Despite all the difficulties, voice-recognition equipment can be purchased today. No serious forecast about computer systems in the 1970's can omit voice recognition systems with several-thousand-word vocabularies. If this sounds unpromising, remember that only a few years ago people used to have 2- and 3-day meetings to discuss the problem of keeping records on magnetic tape. How naive that seems to us now. Today, we already have machines that learn (they are called heuristic machines), that devise their own route to a goal or solution; machines that recognize patterns; and machines that can devise their own strategies—for example, winning at games with the men who design them.

Adding tremendous impetus to the technological explosion is the fact that, as computer capabilities are increasing, costs are decreasing. Between 1963 and 1972—a single decade—there will be a decrease of 85 percent in the cost of completing a typical data-processing job. During this period, the cost of storage by magnetic tape will go down by 97 percent; the cost of image storage by 96 percent; and communication line costs, because of increased speeds of transmission, will decrease by 50 percent. These changes in economics will mean that we will be able to do more with information technology than we now can even imagine.

Let me turn now to the problems of putting these machines to work.

Nowhere is the turn toward technology more obvious than in the way we manage. When we first started to apply computers to business operations in 1954, we went through a very difficult experimentation period and were faced with the most puzzling kinds of problems. We have largely emerged from that period, however, and today we are using computers in business for almost everything conceivable—and much that was not just a few years ago. Senior management has begun to realize that the application of this technology is too important to leave to technicians, and that dramatic things can be accomplished if people who know the objectives of a business will take the responsibility of putting these new capabilities to work. When this happens, you find remarkable achievements.

But along with this progress have come new questions and problems. There are, for instance, union negotiation questions. Throughout the country, a number of owners of newspapers have been willing to stake the very existence of their enterprises on the right to install a computer to prepare punch tape to drive linecasting machines. Just over the horizon, it is clear that this entire process will be bypassed. Is it worth risking an enterprise on a process that is disappearing?

There are many similar questions. What kind of men, for example, should be trained as managers in the new technological environment? How do we create an atmosphere that is conducive to creative people?—for more and more of our businesses must be staffed by highly educated and creative personnel. These are only a few of the problems we face.

Most important are the human aspects. They are related to every problem we have in this field: questions of fear and uneasiness when faced with technological changes; questions of education; questions of identification with an enterprise, with a profession.

But along with the question of how we manage are questions concerning what we manage—of new areas of business opportunity. Here, I will speak of four main new entrepreneurial opportunities. The first is the obvious one that has already taken form—the industry that supplies the systems and the equipment. It is already a multibillion-dollar industry, and this is only the beginning.

The second example, as yet nonexistent but about to bloom as an important basic industry, is the data utility field. This is analogous in some ways to the electrical utility industry: It is cheaper for many people to use a central utility than for each individual to have his own generator. The same economic reasoning applies to the data utility industry, where many people can use a machine simultaneously. The technology of real-time processing, time-sharing, and communication will allow this to happen. Small- and medium-sized businesses—and for some purposes large businesses—will just plug in for data processing as we now do for electricity.

The third example is the one now being called the inquiry industry—in some ways, the publishing field of the future. This will allow the sale of proprietary data over a communications system in answer to a query placed by the customer. The possibilities are unlimited; practically any information can be provided.

We have already started to see the purchase of publishing firms by electronic companies, and this is just the beginning. There will be major changes in ownership in this area in the near future as businesses begin to position themselves to offer such services.

The fourth example is an industry of computer-based educational systems. As technology allows a dynamic or "alive" relationship between a student and a machine system that answers questions as they are posed and discerns gaps in a student's basic grasp of a subject, the much-heralded but until now disappointing teaching machines (better, I think, called learning machines) will begin to mean something. Such systems are already at work in some industrial situations—IBM's maintenance training being a good example. Other precursors can be seen in mentally handicapped children's use of computer-driven typewriters to help them overcome some of their handicaps.

If there is one salient fact about information technology, it is that it is going to produce enormous social change. As the quality of life is changed, as the rate of learning, information, travel, and communications all change, we will see a major change in living patterns, in hopes and desires. In short, a complete new environment will exist.

THE NEW COMPUTERIZED AGE—3: NO LIFE UNTOUCHED

By the end of the century computers will affect every field in innumerable ways; some specific predictions

(By David Sarnoff¹)

In our increasingly complex world, information is becoming the basic building block of society. However, at a time when the acquisition of new scientific information alone is approaching a rate of 250 million pages annually, the tide of knowledge is overwhelming the human capability for dealing with it. So man must turn to a machine if he hopes to contain the tide and channel it to beneficial ends.

The electronic computer, handling millions of facts with the swiftness of light, has given contemporary meaning to Aristotle's vision of the liberating possibilities of machines: "When looms weave by themselves, man's slavery will end." By transforming the way in which he gathers, stores, retrieves, and uses information, this versatile instrument is helping man to overcome his mental and physical limitations. It is vastly widening his intellectual horizon, enabling him better to comprehend his universe, and providing the means to master that portion of it lying within his reach.

Although we are barely in the second decade of electronic data processing, the outlines of its influence on our culture are beginning to emerge. Far from de-personalizing the individual and dehumanizing his society, the computer promises a degree of personalized service never before available to mankind.

By the end of the century, for the equivalent of a few dollars a month, the individual will have a vast complex of computer services at his command. Information utilities will make computing power available, like electricity, to thousands of users simultaneously. The computer in the home will be joined to a national and global computer system that provides services ranging from banking and travel facilities to library research and medical care. High-speed communications devices, linked to satellites in space, will transmit data to and from virtually any point on earth with the ease of a dial system. Students, businessmen, scientists, government officials, and housewives will converse with computers as readily as they now talk by telephone.

In the health field, computers will be employed to maintain a complete medical profile on every person in the country from the hour of birth. The record will be constantly updated by a regional computer for immediate access by doctors or hospital personnel. The computer also will maintain files on every known ailment, its symptoms, diagnosis, and treatment. A doctor will communicate a patient's symptoms to the computer center and within seconds receive suggestions for treatment based both on the symptoms and the patient's history.

Computers will handle the Nation's fiscal transactions from a central credit information exchange, to which all banks, business enterprises, and individuals will be connected. Purchases will be made, funds invested, and loans issued by transfers of credit within the computer without a dollar or penny physically

¹ Gen. David Sarnoff, who this year is celebrating his 75th birthday anniversary, is chairman of the board of the Radio Corp. of America.

exchanging hands. Even the soil will be computerized. The long-range outlook for agriculture includes new sensing devices that will be placed on larger farms, feeding information to the computer on soil moisture, temperature, weather outlook, and other details. The computer will calculate the best crops to plant, the best seeding times, the amount of fertilizer, and even the correct harvesting time for maximum yield.

Some of the most profound changes wrought by the computer will be in education. Here, the machine will do more than assist students to solve problems and to locate up-to-date information: It will fundamentally improve and enrich the entire learning process. The student's educational experience will be analyzed by the computer from the primary grades through university. Computer-based teaching machines, programmed and operated by teachers thoroughly trained in electronic data processing techniques, will instruct students at the rate best suited to each individual. The concept of mass education will give way to the concept of personal tutoring, with the teacher and the computer working as a team. Computers will bring many new learning dimensions to the classroom. For example, they will simulate nuclear reactors and other complex, dangerous, or remote systems, enabling students to learn through a form of experience what could formerly be taught only in theory.

The computer's participation in the field of learning will continue long after the end of formal education. The government estimates that 50 percent of the jobs to be held 10 years from now do not even exist today. With this tremendous rate of occupational obsolescence, future generations of Americans will aid in developing career mobility by providing continuing self-instruction.

Just as it is recasting the educational process, the computer is also fundamentally changing the production and distribution of the printed word. Five centuries ago, Gutenberg broke words into individual letters. Electronic composition now breaks the letters into tiny patterns of dots that are stored in the computer's memory. Any character can be called up by the computer, written on the face of a cathode ray tube, and reproduced on film or paper in thousandths of a second. Nothing moves except the electrons.

When the electronic computer first appeared in composition rooms and printing shops several years ago, its job was to hyphenate words and justify text. But the computer, working at speeds of thousands of words a minute, was driving mechanical typesetting devices capable of setting only a few words per minute. Now, the development of computerized composition makes it possible to set text at hundreds of lines per minute. Photographs and drawings will be set the same way. Since the printed picture is itself a dot structure, the computer can electronically scan any photograph or drawing, reduce it to dots and store it, then retrieve it and beam it on a cathode ray tube for immediate reproduction.

In the future, electronics will develop processes that will make it possible to go from final copy and illustrations to printing in one integrated electronic process. One result will be that newspapers, in the foreseeable future, will no longer be printed in a single location. Instead, they will be transmitted through computers in complete page form to regional electronic printing centers that will turn out special editions for the areas they govern. Local news and advertising will be inserted on the spot. Eventually, the newspaper can be reproduced in the home through a small copying device functioning as part of a home communications center.

Basic changes also will come to other areas of the printed word. For example, of the more than one billion books published every year, almost half are textbooks. The growth of knowledge and the factor of obsolescence mean that these texts must be supplemented by a professor's mimeographed notes. Today, these notes have a small distribution of only a few hundred copies. Computers will make it possible to catalog this information and thus broaden its availability.

At the turn of the century, most large universities will not only have electronic composition systems that allow them to reprint original research, theses, or course notes upon demand; they will also have a computerized information retrieval library. This process of information retrieval can be duplicated in almost any other field. The scientist will have the latest technical papers culled by the computer and reproduced in the laboratory or home. The computer will bring to the attorney all the pertinent laws, decisions, and precedents on any case that concerns him. The business executive need not rush to the office every morning; most of the information he will need to conduct his business will be run off for him at home, and he will have a two-way national and global closed-circuit television, via satellites, for meetings and conferences.

Some of these developments are probabilities, some of them are certainties, and all of them are or soon will be within the capabilities of the computer art. But one fact is absolute: the incredible growth of the computer in numbers, power, and availability.

In just 10 years, the typical electronic data processor has become 10 times smaller, 100 times faster, and 1,000 times less expensive to operate. These trends will continue, and our national computing power, which is doubling every year, will soon be sufficient to make the computer a genuinely universal tool.

In 1956, there were fewer than 1,000 computers in the United States. Today, there are 30,000, or more than \$11 billion worth; and by 1976 the machine population may reach 100,000. And these figures will, of course, be greatly increased through the growth of data processing in other nations.

A decade ago, our machines were capable of 12 billion computations per hour; today, they can do more than 20 trillion, and by 1976—a decade from now—they will attain 400 trillion—or about 2 billion computations per hour for every man, woman, and child. Quite evidently, the threshold of the computer age has barely been crossed.

Nevertheless, for all its potential to stretch the mind a thousandfold, it is perhaps necessary to point out that the computer is still a thing—that it cannot see, feel, or act unless first acted upon. Its value depends upon man's ability to use it with purpose and intelligence. If his postulates are wrong, the computerized future can only be a massive enlargement of human error.

Ramsay MacDonald once warned against "an attempt to clothe unreality in the garb of mathematical reality." Computers echo this warning. For they cannot usurp man's unique ability to blend intuition with fact, to feel as well as to think. In the end, this remains the basis of human progress.

The task ahead will be to assign to the machine those things which it can best do, and reserve for man those things which he must provide and control. It is my conviction that society will adjust itself to the computer and work in harmony with it for the genuine betterment of life.

THE NEW COMPUTIZED AGE—9: WHITHER PERSONAL PRIVACY?

Computer Technology May Enlarge Man's Liberty or Inhibit It; New Rules Must Be Made; New Questions Answered

(By John Lear)

Between my resignation as an editor of the slowly dying Collier's and the inauguration of SR's Science and Humanity Supplement, I was for a short while a minor adviser to Thomas J. Watson, Jr., son of the founder of the International Business Machines Corp. My experience there contradicted two popular beliefs about IBM.

The first belief was that all IBM employees were required, as a condition of employment, to wear white shirts on the job, stay sober at home, and maintain upright on their desktops identical copies of a small sign bearing the personal command of Thomas J. Watson, Sr.: "Think." I wore light blue shirts, drank cocktails at lunch, and put the "Think" sign on the windowsill of my office whenever my secretary put it back on my desk; yet I had a standing invitation to young Tom's sanctum.

The second of my working conditions that ran against supposed IBM tradition had to do with that then-new phenomenon, the so-called "giant brain," or high-speed electronic computer. IBM has sold an enormous number of electronic computers. According to legend, a loyal IBM salesman would leave his wife if necessary to clinch another sale. The truth about IBM computer sales, as I experienced it, was that an immense share of IBM's sales investment went into persuading eager customers to delay the purchase of computers.

Delaying was sound business practice because a computer can do only what it is told to do; it must follow instructions literally; until the instructor himself is sure where literal pursuit of a long series of tiny steps will lead, turning the task over to a computer can be dangerous. The machine may complete its assignment before its owners realize that the outcome isn't really the one they seek.

The period of my IBM experience dates back roughly a dozen years. At that time, computers could do only one thing at a time, in sequence. Computer programmers—the people who break everyday English into binary arithmetic messages (consisting entirely of numerical zeros and ones) comprehensible to the ma-

chines—were such valuable individuals that IBM allowed them to sit at the computer control consoles as long as necessary to figure out flaws in the translation process. Since then, the speed of the machines has risen, the size of components has shrunk, and computer sophistication has grown to such a state that a machine can do a number of different things at a time, not necessarily in sequence.

As a result, the computer's time is worth upward of \$300 an hour—too valuable to be wasted by a programmer's headscratching; and each programmer now must figure out unexpected problems in a separate place while the computer goes on with other people's problems. In other words, it is now possible for wrongly instructed computers to make more disastrous and far-reaching mistakes in a shorter time than ever.

Paul Baran, of the Rand Corp. in Santa Monica, has studied this dilemma more searchingly than most observers of the phenomenon.

"As we pass through life," he reminds us, "we leave a trail of records, widely dispersed and generally inaccessible—except with a great deal of effort and diligence. Beginning with a birth certificate, we accumulate hospital and medical records. We become deductions on our parents' income tax. In school, we generate records of our grades, attendance, IQ tests, personality profiles, etc. (Automated teaching will add to this recordkeeping. The volume of data recorded per child may be expected to increase even more markedly.) After school we start accumulating employment, social security, and selective service records. We may get a driver's license. Most of us will apply for marriage licenses, and some of us will collect divorce decrees which will end in voluminous court records. If we are lucky, we will be able to avoid having arrest and jail records.

We move from job to job in a mobile economy creating moving-company inventory records of our goods. Even as we move from place to place we leave behind short records of our airplane reservations and, for some reason, every hotel makes a ritual of acquiring and preserving the alleged names and addresses of its guests for posterity. This is only a partial list. Think of all the records you leave as you go through life.

"Behind all this creating of records is the implicit assumption that they will some day be of use. In order to be of use, there must be some means of interrogating the files to resurrect the information sought.

"An Internal Revenue Department investigator might wish to have immediate access to the tax returns of each of the associates of a man who is being audited, in order to check on consistency of financial relationships.

"A company may wish to have rapid access to its personnel files to know whether to give a good reference to a former employee.

"A doctor may wish to trace the entire medical history of a patient to provide better input into a diagnostic computer.

"The Veterans' Administration may wish to examine a man's complete military record and possible other previous medical records to see whether the ailment claimed as being service connected really is service connected.

"A lawyer for the defense of a man will wish to search for jail and arrest records, and possibly credit records of all witnesses for the plaintiff.

"Professional licensing boards may want to delve into any records to determine if an applicant has an unblemished character.

"The military in filling extremely sensitive positions may even wish a record of all books borrowed by a prospective applicant to insure that his interests are wholesome and he possesses the proper political bias desired.

"Today it is difficult to gather such information about a prospective examinee. If one went through direct channels and asked most sources for their records about a person, he would most likely be rejected, if for no other reason than that the information is not available—cheaply. Even if the records were publicly available, the investigator would have to spend a great deal of time and effort delving through to discover pertinent data. Today, as a practical matter, if one wishes to obtain certain information about a person, he hires a private detective who charges a great deal of money and expends a great amount of time obtaining a little information available from a portion of these potential records. The price for a fishing expedition for information is high and most of the fish are inaccessible."

Having thus summed up the "the pleasant past," Rand Analyst Baran looks into the future through a three-step review of established processes of computer storage of information. Step 1: Manual records are kept by human clerks. Step 2: Some of the clerks are eliminated by putting all the records into a central

computer file with readout of the records controlled from a single point. Step 3: Information is read into and out of the file from a large number of different points.

Baran envisages connection of one remote-access computer with other similar computers, and through this, "danger of loss of the individual's right to privacy as we know privacy today." "The composite information data base may be so large and so easily accessible that it would permit unscrupulous individuals to use this information for unlawful means," he warns. "Modern organized crime should be expected to have the financial resources and skills necessary to acquire and misuse the information."

He expresses concern not only over the possible creation of "automated black-mail machines" but over the potential addition of "inferential relational retrieval techniques" now being developed which, "when fully refined, could determine relationships of any person, organization, event, etc., to any other person, organization, or event." Noting that "humans, by their day-to-day necessity of making decisions on totally inadequate evidence, are innately prone to jump to conclusions when presented with very thin chains of inferred relationships," he predicts an increase in the already growing practice of unearthing defaming information about candidates for political office.

The Baran forecast of computer hazards is fortified by the studies of another Rand researcher, M. R. Maron.

"Consider," suggests Maron, "what could happen as machines are used to make decisions about people. For example, consider a situation where a computer is programed to decide who should get a security clearance from the Government, or who should get an education loan, or whether someone's driver's license should be suspended, or who should get a passport, or who should be accepted for the Peace Corps or the Job Corps, etc.

"As larger files (of machine-language data, stored in computer memories, linked cross-country by telephone) become accessible there will be a natural tendency to use machines for the automatic selection (or rejection) of people according to some preprogramed set of criteria. Supposedly these criteria will have been carefully thought out before programing the machine. Even so, the implications are dangerous.

"In such a mechanized situation, how does an individual get an opportunity to 'tell the system' that its selective criteria don't apply to his own special case? Each individual is different, each has certain extenuating circumstances, each has information which he believes to be relevant to the selection decision and which the system does not consider relevant. And so on. If an individual does not have the opportunity to be judged on the circumstances of his own special (individual) situation, then he is being treated as a machine.

"Will there be a tendency in the future to create an environment where we treat each other as machines; i.e., where there is no opportunity to 'change the system's mind'? How can we create a society where we treat our citizens as people and not as machines? How can we create a society where each individual has the opportunity to explore and unfold his own special potentials—to realize what he is?

"These questions lead to further questions—to questions about who we are and what it means to be a person. And this brings us to the problem of values. What kind of a life do we want? What kind would we value—ought we to have? How can we create a society that fosters those actions and goals that we value? How define and explicate values? How measure and compare and rate values? How select among competing values? How can we estimate the impact of computers on our values?

"And if our projections into the future suggest that we are heading toward a future society which is not conducive to a 'good' life, what can be done to isolate the trouble spots and to influence those changes that will prevent the possible 'evils'? Such analysis of future prospects implies prediction, evaluation, and then some attempt at control. Can the process of control be made democratic so that a small professional elite does not dominate in influencing the shape of the future?

"Finally, there is the problem of time—the time that it takes to initiate and complete corrective action. Given an analysis of the impact of computers on society and given some corrective action that must be taken in order to avoid some future situation, how long a timelag will occur between corrective action and modification of the situation?"

The positive cultural potential of computers was emphasized last January in a report to President Lyndon B. Johnson by the National Commission on Tech-

nology, Automation, and Economic Progress. Although this report dissented from the "almost" * * * commonplace (opinion) that the world is experiencing a scientific and technological revolution" of sufficient power "to make our economic institutions and the notion of gainful employment obsolete," it proposed serious consideration of development of a computerized "system of social accounts" capable of analyzing accurately in advance the benefits and costs of any sociopolitical experiment. Such a system theoretically could grapple competently with complex problems such as water and air pollution, urban blight, the transportation tangle, integration of the Negro into American society, and the continuing spread of crime.

The Presidential Commission report defined the phrase, "system of social accounts," to include mixtures of systems analysis, simulation, and operations research in proportions required for particular cases. Systems analysis and operations research are now in wide employment in military planning and extraterrestrial space exploration. Simulation techniques are part of current plans for global weather observation and forecasting.

How close are we to a workable "system of social accounts"?

One of the best informed men on earth on the subject of computer development is Dr. Cuthbert Hurd, chairman of the board of Computer Usage Co., Inc. In addressing the National Automation Conference of the American Bankers Association in Chicago last month, Dr. Hurd observed that no computed manufacturer today markets an "operating system" flexible enough to apply all the diverse talents of computing machines to any complex problem.

"I suppose," Dr. Hurd told the bankers, "that as much as 200 man-years of effort might be required to produce a modern operating system, costing say \$5 million."

If such a system were to be perfected, Dr. Hurd said, "it is still unclear whether proprietorship (of the system) could be maintained under the existing patent or copyright laws."

There are two ways, then, to state the challenge of computerized society. One was succinctly put in a recent issue of the American Scholar by Lynn White, Jr., professor of history at the University of California in Los Angeles: "Must the miracle of the person succumb to the order of the computer?" The other statement comes from Paul Baran: "What a wonderful opportunity awaits us to become involved in such problems as to exercise a new social responsibility."

APPENDIX 4.—SPEECH BY VICE ADM. H. G. RICKOVER, U.S. NAVY,
ENTITLED "LIBERTY, SCIENCE AND LAW"

[Copyright 1966, H. G. Rickover]

LIBERTY, SCIENCE, AND LAW, BY VICE ADM. H. G. RICKOVER, U.S. NAVY, AT THE
ATHENS MEETING OF THE ROYAL NATIONAL FOUNDATION ATHENS, GREECE, JUNE
2, 1966

This speech reflects the views of the author and does not necessarily reflect the views of the Secretary of the Navy or the Department of the Navy.

I deeply appreciate your invitation to address this meeting. It is an honor and a moving experience—especially for an American—to speak here where the ancient *Ecclesia* had its seat, where men first practiced the difficult art of self-government, succeeding brilliantly for a time but failing in the end. My country, as you know, picked up the torch of liberty they had lighted and established the first *representative* democracy in modern times, even as Athens had established the first *direct* democracy in all history.

Twenty-four centuries separate these two great innovative acts in time, over 5,000 miles in space. One took place in a small city-state possessing few material resources, the other in a huge country of great natural wealth. Yet there is a close inner link between them. They had the same objective. The principles they adopted to achieve their purpose were similar. Both sought to create—and did create—the political framework for a society of free men.

Even as Solon, Cleisthenes and Pericles before them, the framers of the American Constitution of 1789 were political thinkers, as well as experienced practical politicians. They drew upon Greek political theory and practice with which they were thoroughly familiar, adopting what had proved successful, ingeniously improving where the earlier structure had shown weakness. They were men of the enlightenment, when classical rationalism sparked a new Age of Reason throughout the Western World; when philosophers were inspired to mount an attack on every custom and institution that shackles the mind of man and arbitrarily restrains his actions—from superstition to class privilege, from tyranny by an established church to tyranny by a secular autocrat. The political institutions of all the nations of the free world today—beginning with my own—had their inception in the turmoil of that last phase of the Renaissance.

Western civilization is set apart from civilizations elsewhere, both past and present, by its dynamism, its extraordinary creativity, its intense preoccupation with things of the mind. All this started with the Renaissance. Not until modern Western man rediscovered and retrieved his classical heritage did he begin to outstrip the rest of the world.

To borrow a Churchillian phrase, it can be said of Athens, of Greece in general, that never before or since did so few human beings leave so deep and lasting an imprint on so many others, differing in race and faith, distant in time and space from this cradle of Western civilization. Their mark is on all our science, our art, architecture, literature, theater, and on our political thinking and practice as well. Here in this city, on this hill where I am privileged to stand, the Athenians proved that free men could govern themselves; that it was possible to live in a civilized society without having to relinquish personal freedom.

This was an epochal achievement. In all his long life on earth, man has had but brief moments of freedom. His own nature is the cause of the paradoxical situation that *civilization and liberty are interdependent, yet at the same time antithetical*. One cannot be had without the other, yet reconciling them remains to this day what it has always been—the most difficult political, social and economic problem.

Civilization and liberty are *interdependent* because basic to freedom is exercise of mind and spirit, of the faculties that set us apart from other living things and make us fully human. For this there must be a modicum of leisure which

comes only with civilization, when men no longer need devote all their time and energy to appeasement of hunger and protection against the elements—as must the animals.

But release from endless toil for mere survival does not automatically set men free. Indeed, the very opportunity to cultivate mind that spirit which civilization opens to man lies at the root of the *antithesis* between civilization and liberty, for this opportunity is not seized to the same degree by everyone.

Always and everywhere, civilization results in much greater enlargement of the scope of human thought and action among the minority possessing high intelligence than among the majority of average people. Nature endows men with unequal capacities for acquiring knowledge and competence. More so in the realm of the intellect, which is all-important in civilized life, than in the realm of physical strength and courage, which counts most in primitive society. Men become, as it were, more unequal as civilization advances.

When life is simple, it can be understood by nearly everyone, and the competencies needed to function effectively are within the grasp of all. This makes for the rough equality of status that is so favorable to mutual respect of one another's personal liberties. There can be no freedom unless it is mutually conceded.

With civilization, life grows complex, harder to understand for ordinary people, demanding competencies many are unable to acquire. In understanding and competence, the gifted swiftly forge ahead. What they achieve is beyond the capacity of the average. The result is that men grow apart, their interests diverge. Society tends to divide into segments according to superiority of competence or superiority of numbers. The temptation is great for each segment to use the power its particular superiority confers to bend the whole of society to its will, thus putting an end to freedom.

The Athenians were first to devise a political system that preserved the citizen's liberty by counteracting the natural human inequalities which are the root cause of segmented power centers. So precise and clear was their thinking, that the basic principles of their system remain to this day the best protection of individual freedom. Government of the people, by the people and for the people was their great invention; political equality their crowning achievement. They inaugurated the reign of mind over force by providing for resolution of differences in point of view and interest through public dialog leading to consensus, instead of by the exercise of power. Perhaps the most remarkable feature of their policy was that it engaged the continuous participation in public business of a large part of the citizenry—somewhere between one-fifth and one-fourth at any given time. It was obvious to them that only when the people are personally involved with their government will public officials be responsive to the popular will. Citizens who shunned public service were called *idiotes* and considered useless; in some cases, failures to do one's public duty resulted in loss of civil rights.

To quote Edith Hamilton, the American classicist who was made an honorary citizen of Athens, "the idea of the Athenian state was a union of individuals free to develop their own powers and live their own way, obedient only to the laws they passed themselves and could criticize and change at will." This is the political ideal that to western man spells liberty and that is rejected in toto by all autocracies, modern as well as ancient.

The Greeks, I think, understood better than most of us what it means to be free. In his play, "The Persians," Aeschylus who fought at Marathon puts his finger unerringly on what distinguished free Greece from unfree Persia. He has the Queen of Persia ask about the Athenians: "Who is their master?" To which she received the answer, "they are not subject to any man"; they obey only the law. When she is told of her son's defeat, the Queen remarks: "Even if he fail, there is no law can call *him* to account." How better could one express the contrast between the protagonists in today's cold war?

Action wrote that "power corrupts and absolute power corrupts absolutely." The Greeks penetrated more deeply and saw that power erodes man's reason. One sees cool contempt in Herodotus' report of the wrath of Xerxes when the bridge he ordered built across the Hellespont was torn apart in a storm. Straightway he "gave orders that the Hellespont should receive 300 lashes, and that a pair of fetters should be cast into it," and he "commanded that the overseers of the work should lose their heads." Here stands revealed the totalitarian mind—the same today as in the past.

Liberty, never gained without enormous effort and sacrifice, is all too easily lost. Those who enslave their own people seem irresistibly driven to extinguish

freedom everywhere. When we understand them, we are better prepared to ward off their aggression. More important still is awareness of the forces *within* free societies that endanger liberty. In both respects, there is still much we can learn from the Athenians.

It seemed to me, therefore, that the setting here would be eminently suited to a discussion of certain developments in modern democracies that have an adverse effect upon the liberties of the individual and the social and moral values cherished by free men. The causative factor of this new threat to liberty is science and science-based technology.

This new science-technological threat is but the latest version of the age-old conflict between civilization and liberty—a conflict that has no permanent solution but reappears perennially in new form.

Liberty is never gained for once and for all. Each generation must win it anew. Each must defend it against new perils. These perils arise because men, being endowed with free will, continually alter the conditions of life. Countless decisions made in pursuit of private objectives may so transform society that institutional safeguards once adequately protecting human liberty become ineffective. It is then necessary to return to first principles and to adapt them to altered circumstances.

The title of my speech "Liberty, Science and Law" expresses my conviction that unless certain practices in the technological exploitation of scientific knowledge are restrained by law, they will cost us our liberties.

Science and technology are, of course, of immense benefit to man. They are so highly regarded that no one would, or for that matter could, prevent their spreading to areas that at present are retarded in this respect. But they may bring about changes in our physical environment of greatest potential danger. Certain technologies admittedly injure man, society, and nature. Yet, even in countries where the people are sovereign and where they recognize the danger, efforts to bring these technologies under social control have had little success. Those who have the use of technology are powerful enough to prevent legal restraint, the main prop of their power being the esoteric character of modern science.

Much of it is incomprehensible even to intelligent and educated laymen. When scientific-technological considerations enter into public issues—as is often the case today—the issues cannot be understood by the electorate, frequently not even by the public officials who are directly concerned. There is then no recourse but to call on scientists for expert advice. In effect, the issue will be decided by them, yet they have not been elected, nor are they accountable to the people. What is left of self-government when public policy no longer reflects public consensus? And, when the public finds that it cannot judge and evaluate issues involving science, will it not become apathetic toward *all* public issues? Does this not spell the doom of self-government, hence of freedom for modern man? Though all the institutions established to safeguard his liberties may remain intact, the substance of freedom will have been lost.

By one of those ironies of fate beloved of Greek dramatists, this new threat to liberty has its source in the noblest Greek achievement, the freeing of the human mind to roam at will in pursuit of truth and knowledge. All things are to be examined and called into question, said the Greeks. Unless men understood the world in which they lived, and because of this understanding felt at home in it and could be useful citizens, they were not truly free. Never before or since was intellectual freedom valued so greatly. "All things were in chaos when mind arose and made order," said Anaxagoras, the mathematician and astronomer.

Everywhere else, the domain of the intellect was the special preserve of powerful priesthoods who jealously guarded their monopoly of knowledge. "To teach the people so that they would begin to think for themselves would destroy the surest prop of their power," wrote Edith Hamilton. "Ignorance was the foundation upon which the priest power rested." The legends of most people are replete with stories of divine punishment for trying to know more than was deemed proper—clear evidence of the determination of this priestly elite to discourage ordinary people from seeking knowledge. Not so in Greece. There curiosity and search for knowledge were held to please the gods, for through these the marvels of the gods were revealed to man. Wisdom and intelligence had their own protective deity—Athena.

When Renaissance man recovered his classical heritage, the most precious treasure he found was freedom of the mind. With his mental powers set free, it took him but three and a half centuries to build on foundations laid in classical

foreseen that in its ultimate consequences the scientific revolution might displace the whole magnificent edifice of modern science. No one could have diminished human liberty.

But it has brought us back full cycle. Science—the vital area of knowledge today—is for most of us virtually a closed book; again it has become the monopoly of a small elite. This is not the fault of the scientists. Unlike ancient priesthods, they have no wish to bar others from knowledge or to use it to enslave the ignorant. Many scientists make strenuous efforts to explain science to the lay public. Nevertheless, we find ourselves in much the same position as the ancient Egyptians whose very lives depended on knowing when the waters of the Nile would rise and fall—knowledge possessed by their priesthood alone.

As in the past, it is not the knowledge gap per se that is most detrimental to freedom, not the fact that the majority cannot follow scholars into the realm of higher mathematics and science; rather it is the effect ignorance of science has on public attitudes toward science and science-based technology. The impact of technology, in particular, on the individual and on society at large is profoundly affected by prevailing concepts of what technology is and what purpose it should serve.

If people understood that technology is the creation of man, therefore subject to human control, they would demand that it be used to produce maximum benefit and do minimum harm to individuals and to the values that make for civilized living. Unfortunately, there is a tendency in contemporary thinking to ascribe to technology a momentum of its own, placing it beyond human direction or restraint—a tendency more pronounced in some countries but observable wherever there is rapid technological progress.

It manifests itself in such absurd statements as that technology *demand*s some action the speaker favors, or that "*you can't stop progress.*" Personalizing abstractions is a favorite means of semantic misdirection; it gives an air of authority to dubious statements. Most people are easily pressured by purveyors of technology into permitting so-called progress to alter their lives, without attempting to control it—as if they had to submit meekly to whatever is technically feasible. If they reflected, they would discover that not everything hailed as progress contributes to happiness; that the new is not always better, nor the old always outdated.

The notion is also widespread—doubtless fostered by users of technology—that, having wrought vast changes in the material conditions of life, technology performs renders obsolete traditional concepts of ethics and morals, as well as accustomed ways of arranging political and social relationships. Earnest debates are currently taking place whether it is possible to act morally in the new technological society, and proposals have been made—quite seriously—that science must now replace traditional ethics! We have here a confusion that must be cleared up.

Through technology we are relieved of much brutal, exhausting, physical labor as well as boring routine work; we are provided with numerous mechanical servants who do certain kinds of work faster, cheaper, and more efficiently than people. Why should the ease and affluence technology makes possible affect moral precepts that have guided Western man for ages? This may brand me as old fashioned but I have not yet found occasion to discard a single principle that was accepted in the America of my youth.

Technology is tools, techniques, procedures, things; the artifacts fashioned by modern industrial man to increase his powers of mind and body. Marvelous as they are, let us not be overawed by these artifacts. Certainly they do not dictate how we should use them nor, by their mere existence, do they authorize actions that were not anteriorly lawful. We alone bear responsibility for our technology. In this, as in all our actions, we are bound by the principles governing human behavior in our society.

Does it make sense to abandon principles one has lived by because he has acquired better tools? Tools are for utilizing the *external* resources at our disposal; principles are for marshaling our *inner*, our human resources. Tools enable us to alter our physical environment; principles serve to order our personal life and our relations with others. The two have nothing to do with each other.

This should be obvious, but erroneous concepts of science and technology abound because people tend to confuse the two. Not only in popular thinking but even among the well-informed, science and technology are not always clearly distinguished. Characteristics pertaining to science are frequently attributed to technology, even as science itself is confounded with ethics.

Science has to do with discovering the true facts and relationships of observable phenomena in nature, and with establishing theories that serve to organize masses of verified data concerning these facts and relationships. By boring into the secrets of nature, scientists discover keys that unlock powerful forces which can be made to serve man. It is through *technology* that these forces are then put to human use.

Science is a body of systematized knowledge; technology is the apparatus through which knowledge is put to practical use. The difference is important.

Because of the care scientists take to verify the facts supporting their theories, and their readiness to alter theories when new facts prove them imperfect, science has acquired great authority. What the scientific community accepts as proven is not questioned by the public. No one disputes that the earth circles the sun, or that atomic fission produces energy.

Technology cannot claim the authority of science and is therefore properly a subject of debate, not alone by experts but by the public as well. Little thought is customarily given to the possibility of harmful aftereffects by those responsible for technological exploitation of scientific knowledge. In consequence, technology has proved anything but infallibly beneficial. Indeed, much damage has been done because *no* thought was given to the interaction of technology with nature. More of this presently.

A certain ruthlessness is encouraged, in the mistaken belief that to disregard human considerations is as *necessary* in technology as it is in science. The analogy is false.

Rigorous exclusion of the human factor is *required* by the methods of science. These were developed to serve the needs of scientists, whose sole interest is to comprehend the universe; to know the truth; to know it accurately and with certainty. The searcher for truth cannot pay attention to his own or other people's likes and dislikes, or to popular ideas of the fitness of things. What he discovers may shock or anger people—as did Darwin's theory of evolution. But even an unpleasant truth is worth having; besides one can choose not to believe it. *Science, being pure thought, harms no one.*

Technology, on the other hand, is *action*, often potentially dangerous action. Never has man possessed such enormous power to injure his fellow humans and his society as has been put into his hands by modern technology. This is why technology can have no *legitimate* purpose but to serve man—man in general, not merely some men; future generations, not merely those who currently wish to gain advantage for themselves; man in the totality of his humanity, encompassing *all* his manifold interests and needs, not merely some one particular concern. Technology is not an end in itself; it is a means to an end, the end being determined by man himself in *accordance with the laws prevailing in his society.*

A word may be in order concerning the disparate meaning of the term *law*, depending on whether it is used in the ordinary sense—which is also the original sense of the word—or by scientists.

Law, as commonly understood, refers to those rules of human conduct prescribed and enforced by society. Its purpose is to resolve human conflicts by the application of definitive rules. These rules are always debatable and can be changed when there is demand for a change.

The scientists have appropriated the term law to describe regularities exhibited by physical phenomena—the rules by which the universe governs itself. In the transition, the word has taken on a new meaning.

From the layman's point of view, what the scientist calls law is fact, rather than law—immutable fact. Or, if you prefer, it is law operating in a sphere where human beings can exercise no influence. We cannot alter the laws of the cosmos; we can only discover them. A law of science expresses mechanical regularity where no choice of action, no free will comes into play; it deals with constancy of behavior in nature. It has relevance for us because it makes the universe comprehensible and so enables us to utilize the forces of nature for human purposes.

We are bound by the laws that science has disclosed when we exploit these forces by means of technology. Likewise we are bound by the manmade laws of our society, for our actions affect fellow human beings. Technology straddles, as it were, the law of the universe and the law of man; it is subject to both.

Much confusion in popular thinking arises from this fact. The two laws are confounded. Or, to put it differently, *they are thought to be part of a single system of law so that one or the other must perforce take precedence.*

Ever since scientists discovered that the earth is not the center of the universe, as had been maintained by the highest human authorities, we have been learn-

ing painfully that the laws of nature cannot be overturned by human fiat. It has taken a long time to attain this rational attitude; we are now conscious of the consequences of intolerance in the past. Perhaps this is why we are so tolerant toward those who claim the right to use technology as they see fit, and who treat every attempt by society to regulate such use in the public interest as if it were a modern repetition of the persecution of Galileo.

The right to be protected by law against injurious action by others is basic to civilized society. Yet, opponents of legislation intended to restrain use of potentially dangerous technologies are often able to prevent or delay enactment of such laws by playing upon the layman's respect for science. It is their common practice to argue as if at issue were a law of science when, in fact, what is being considered is not science but the advisability or legality of the technological exploitation of a scientific discovery. The public would not be deceived by such arguments if it clearly understood the fundamental difference between science, which is *knowledge*, and technology, which is *action* based on knowledge.

To guard against being misled, one should cultivate an attitude of skepticism whenever the word "*science*" is used. Is it science that is being discussed or is it technology? If technology, the question at once arises whether the proposed action is legally permissible and socially desirable. These are matters that lie outside the domain of science. Just as the law of the cosmos cannot be overturned by human fiat, so is human law supreme within its own proper sphere of operation. Technology must therefore conform to that most basic of all human laws, the maxim of the "mutuality of liberty," the principle that one man's liberty of action ends where it would injure another. Without this maxim, freedom would be a barren privilege.

Whether or not a particular technology has harmful potentialities should not be decided unilaterally by those who use it. For the user, destructive technologies are often highly profitable. He is, therefore, an interested party to the conflict between private and public interest that every potentially harmful technology poses. Nearly always he is also a practical man.

I think one can fairly say that the *practical* man's approach to a new scientific discovery and its technological exploitation is *short-range* and *private*, concerned with ways to put scientific discoveries to use in the most economic and efficient manner. Rarely will he give thought to the *long-range* and *public* consequences of his actions, that is, to the effects that a new technology may have on people, on the nation, on the world; on present and future generations.

To illustrate the disastrous consequences of a narrow practical approach, let me give some examples of technological damage to our national environment.

Carelessly emitted, the waste products of new technologies create a massive problem of soil, water, and air pollution. We may be permanently damaging the atmosphere by changing its chemical composition. New products, profitable to manufacturers and useful to consumers, are often themselves intractable pollutants. For instance, detergents which unlike soap do not dissolve in water, or pesticides and weed killers which, carelessly applied, will poison soil, crops, birds, animals, fish, and eventually man.

Other technologies enable man to alter the very contours of the land—as with new strip mining machinery. Because it cuts the cost of extraction, such machinery, is used in some places. Huge chunks of earth and rock with their topsoil and vegetation are gouged out, changing fertile country into a desolate lunarscape—a land robbed not only of its irreplaceable mineral wealth but of its fertility as well.

Man now has the means to slaughter all the wild animals on earth and he is well on his way of doing so. Consider what has been done to the vast riches of the seas.

With modern techniques, deep-sea fishing is so efficient that a few enterprises could rapidly sweep the oceans free of commercial fish. And this is what fishermen of all nationalities wish to do. As practical men they have no other interest than to use the latest technology that will increase their catch, preserve it and get it to market speedily as possible.

We witness at the moment the end of one of the saddest cases of misuse of technology by greedy fishing interests. Unless these interests are curbed by truly effective international action, the great whales—the blue, the finback, the sperm—will soon disappear, victims of man's "practical" folly.

These and other whales once populated the high seas in immense numbers. For hundreds of years whaling remained a reasonably fair contest between man and the intelligent, swift-moving mammals he hunted. Modern technology has turned it into brutal genocide. Blindly pursuing what they doubtless consider

an eminently *practical* objective—maximum profit *today*—the whalers are wiping out the very resources that could insure them a profit *tomorrow*.

In April of this year Japanese ships had to return home after only 3 of the normal 5 months at sea because they could find no whales.

Practical considerations aside, is anyone justified in using technology to exterminate a species that has existed on this earth for eons—the largest animal the world has ever seen? Are we certain our descendents may not at some future time have need of these mammals?

How we use technology profoundly affects the shape of our society. In the brief span of time—a century or so—that we have had a science-based technology, what use have we made of it? We have multiplied inordinately, wasted irreplaceable fuels and minerals and perpetrated incalculable and irreversible ecological damage. On the strength of our knowledge of nature, we have set ourselves above nature. We presume to change the natural environment for *all* the living creatures on this earth. Do we, who are transients on this earth and not overly wise, really believe we have the right to upset the order of nature, an order established by a power higher than man?

These are complicated matters for ordinary citizens to evaluate and decide. How in future to make wiser use of technology is perhaps the paramount public issue facing the electorates of industrial countries. It will tax their mental resources and challenge their political acumen. Certain measures suggest themselves:

Experience shows that by itself, the legal maxim of “the mutuality of liberty” will not prevent commitment to technologies that may later prove harmful. The maxim must be implemented by preventive public action—action of the kind that has long been operative in the field of public health. There is need for laws requiring that *before* a particular technology may be used, reliable tests must have been made to prove it will be useful *and* safe. A few such laws have been enacted; more are needed.

I suggest that, as a special public service, lawyers take on the task of working for better protection against technological injury. This is a new and fruitful area in which they could make important contributions to human welfare—an area which requires no revolutionary change in the political or economic structure of society, merely greater precision and fuller implementation of the traditional principle that injuring the health or causing the death of human beings is unlawful. The term “health” should not be limited to physical health but should include psychic health and protection of the human personality as well. New technologies based on the uncertain “science” of the social sciences involve snooping into the inner recesses of the human mind, personality testing and pseudo-scientific manipulation of human beings. When they are imposed as conditions of employment or otherwise partake of an element of compulsion, these technologies should be regulated or outlawed entirely.

Much more thought should be given to technological interference with the balance of nature and its consequences for man, present and future. There is need of wider recognition that government has as much a duty to protect the land, the air, the water, the natural environment against technological damage, as it has to protect the country against foreign enemies and the individual against criminals. Conversely, that every citizen is duty bound to make an effort to understand how technology operates, what are its possibilities, its limitations, its potential dangers. The leisure modern technology makes available to ever larger numbers of citizens could not be better spent than in a determined effort to narrow the knowledge gap between those who understand science and technology and those who do not.

Since law and public opinion always lag behind the swift development of new technologies, there is need for more informed and responsible thinking among those who control technologies. This might be achieved by professionalizing the decisionmaking process in technology. Experience has shown that in the hands of professional persons technology is managed with greater concern for human welfare than when it is controlled, as at present, by nonprofessionals. The classic example is medicine.

Of all technologies, that of the physician has benefited human beings most and harmed them least. The stringent standards set by the profession and by society for the education and professional conduct of physicians accounts for this happy circumstance. Not only is no one permitted to practice who has not given proof of his competence, but physicians must also be broadly, liberally, humanistically educated men and women. This gives them perspective in evalu-

ating their professional actions, an ability to see these actions against a humanistic background. Moreover, they operate under a code of ethics which requires them to place the needs of patients above all other considerations—a code incorporated 25 centuries ago in the Oath of Hippocrates, an oath still taken by young men and women embarking on a medical career.

To Greece we owe the noble idea that special knowledge and skill ought to be used to benefit man, rather than for personal aggrandizement or power, or as a means of extracting maximum gain from those in need of the services of men possessing special expertise. This concept of a *trusteeship of knowledge* could well be applied to *all* whose knowledge of science and technology surpasses that of the lay public, as it now is to physicians and surgeons. I have long advocated that engineering pattern itself after medicine and law, thus becoming a truly "learned" profession. It has, I believe, attained that status in some countries, though not in mine.

These are my suggestions; others may have better ones to offer. What seems to me of utmost importance is that we never for a moment forget that a *free society centers on man*. It gives paramount consideration to human rights, interests and needs. Society ceases to be free if a pattern of life develops where technology, not man, becomes central to its purpose. We must not permit this to happen lest the human liberties for which mankind has fought, at so great a cost of effort and sacrifice, will be extinguished.

QUESTIONS OF INVASION OF PRIVACY RELATING TO THE ESTABLISHMENT OF A NATIONAL DATA CENTER

[Reprinted from the Congressional Record, Aug. 18, 1966]

Mr. GALLAGHER. Mr. Speaker, under the direction of the Honorable William L. Dawson, chairman of the House Government Operations Committee, the Special Subcommittee on Invasion of Privacy, made up of my distinguished colleagues Congressmen Benjamin Rosenthal, of New York, and Frank Horton, of New York, has just conducted a series of hearings on the proposed establishment of a National Data Center by the Federal Government.

As chairman of this special subcommittee, it is my desire to present to the membership today some of my thoughts and findings upon the conclusion of those hearings.

Testimony before the subcommittee has illustrated first of all the great technological progress which has been made in the field of computer science and data processing. The potential of this technology and its value to our modern society are certainly impressive.

But the hearings have indicated as well an urgent need for a corollary study to determine the direction which our Nation will permit technology to take and the great responsibility we face to protect the public interest and rights of the individual.

Modern scientific achievement goes far beyond the full comprehension and knowledge of most of us. Yet its influence upon the life of each citizen is ever-increasing. As we realize our own inadequacy to evaluate an issue involving scientific technology, we seem to move toward an ever-increasing submission to the domination of those who are expert in the scientific disciplines.

Society borders on forgetting that technology is its own creation, to be guided and directed along the course which will provide its members most with the full benefits of scientific knowledge. The people seem dangerously prepared to surrender their age-old respect for the vast capabilities of the human mind and personality to the impressive and sometimes overwhelming knowledge which the scientific elite alone have mastered. Somewhat intimidated by the mystery of science, the average citizen in our Nation often seems reconciled to the sacrifice of individual liberties in the awesome name of "progress."

Moreover, this malaise threatens to distort our traditional concept of law and its meaning to society. The forefathers of this Nation defined a legal code to protect the rights of Americans against government encroachment. The ultimate value of this code, which we know as our Constitution, centers in its flexibility, its adaptability to the needs of each new generation.

In science, however, the term "law" takes on new meaning. It defines the regularity of physical phenomena and its definitions seem synonymous with cold

fact. Scientific law is not concerned with the multiple aspects of free will and individual personalities; it deals solely with nature's constancy, or if you will—mechanical regularity.

In today's world, however, these two totally distinct definitions of one term seem to have fused and been confused. And in this commingling of definitions, it is indeed unfortunate that the scientific appears to have taken precedence.

This precedence is well illustrated in the case before us at present. The individual's right to privacy has always been recognized in the full course of our Nation's history. The protection of the individual by law against infringements attempted upon this right by others is incorporated in our Nation's Constitution. And yet in the name of scientific advancement, this right is now potentially threatened.

Although the proposal to establish a national statistical data center, if adopted, promises greater efficiency in many Government operations, the possibility that such a center might become a depository for extensive personal information on every citizen raises questions fraught with serious implications.

Should the Government establish a centralized statistical data center with its interchangeable counterpart a personal dossier bank, there would be a tremendous store of data already available to feed it. That list includes tax returns, census responses, social security data, military records, security files, fingerprints, FHA and VA mortgage guarantees, credit records, health data, and research involving individuals. If State and local governments were tied into the proposal, such data as school records, police files, driving violations, and property holdings would also be on file.

The Bureau of the Budget contends that no one has proposed such a personal dossier bank. That is quite true. But it is also a fact that detailed information on millions of individuals and corporations would be poured into the national data center. Group data is made up of individual data. Testimony by computer experts before our subcommittee shows clearly that a data center could easily become a dossier bank. Simply stated, our concern is what an innocent statistical center could turn into as the years roll by and pressure mounts to program into the computers more and more information on individuals. Computer experts stressed that the same technology that put the information in for statistical data could be retrieved instantaneously on any individual.

At present, the confidentiality of some of this information is protected by the law. Centralization, however, would create the need for a new set of safeguards to protect the privacy of the material on file. It would appear obvious that the Federal official who has the authority to press the button to produce a dossier on any individual in the United States would possess a power greater than any ever before known in America.

We must remember that our citizens give the Government personal information on a confidential basis and for a specific purpose. Americans deserve the assurance that this information will not be used for any other purpose in the future. Our Government must decide now before we embark on this new and dangerous course whether we can properly protect the civil rights and civil liberties of each citizen.

Without carefully established safeguards, there exists a very real threat of great injustice. Safeguards, which incidentally, do not now exist in a technological sense. It is certainly conceivable that a potential Big Brother—in the frightening Orwellian tradition—might make excellent use of a big button on a dossier bank for his own purposes and for the sake of increasing his own power.

Writing on this subject, the Wall Street Journal, August 5, 1966, stated:

"We do not suggest that many officials would attempt to abuse the power. Yet the fact is that even as it is, Federal agencies have been known to harass individuals or businesses, just as some of them have not been above electronic prying and other violations of privacy.

"It is a cardinal requirement of a free society that the people do not entrust their liberties to the whims of men in power but rely rather on wise laws to protect them from oppression."

It seems evident that if the proposal to create a national data bank is adopted, we will have to rely only on the hope that benevolent people with benevolent purposes will operate the system. History, however, has already taught a terrible lesson illustrating exactly what can happen when large stores of information become available to nonbenevolent powerseekers.

The detailed European census, long in effect even before the advent of the Nazi Party, provided a most convenient and efficient tool for Hitler's use when he led

the party to control Germany. The census information provided a central data system from which the dictator could draw detailed information on any German citizen, thereby facilitating the power surge of his totalitarian regime.

Prof. Charles A. Reich, Yale University expert on constitutional law, stated in his testimony before the Subcommittee on the Invasion of Privacy:

"Real protection in this world comes not from people's good intentions, but from the law."

The risk involved now in entrusting the liberties of the American people to the men of power in the future, the names of whom we do not even know and whose benevolence we cannot presume to guarantee, is too great for us to take.

We must consider now whether our laws are keeping abreast of our rapidly expanding technology and whether with the advance of that technology, our present safeguards of the constitutional rights of our citizens are still adequate.

In his address before the Royal National Foundation at Athens, June 1966, Vice Adm. H. G. Rickover, stated that in areas of technology, man must consider first whether any proposed action can be allowed legally and secondly, whether the action will be beneficial to his society. Admiral Rickover continued:

"These are matters that lie outside the domain of science. Just as the law of the cosmos cannot be overturned by human fiat, so is human law supreme within its own proper sphere of operation. Technology must therefore conform to that most basic of all human laws, the maximum of the 'mutuality of liberty,' the principle that one man's liberty of action ends where it would injure another."

In our modern society, therefore, we must meet the challenge of providing legal safeguards today against the possible violation by technology tomorrow of the maxim of the "mutuality of liberty." The supremacy of human law has been seriously questioned; it must be reasserted now if our society is to remain free.

The technology of computerization has raised new horizons of progress, but it also brings with it grave dangers. The greatest of these is that we may allow ourselves to drift into a course of action that will ultimately substitute a computer for man's free will and his human consciousness of what is ethical and what is not.

The 19th century novelist, Samuel Butler recognized the fallacy of such a course when he wrote in his novel, "Erewhon":

"I cannot think it will ever be safe to repose such trust in the moral sense of any machine."

Americans of the 20th century cannot think it any safer. As Professor Neisser has written in a study entitled "The Imitation of Man by Machine":

"If machines really thought as men do, there would be no more reason to fear them than to fear men. But computer intelligence is indeed 'inhuman': it does not grow, has no emotional basis, and is shallowly motivated. These defects do not matter in technical applications, where the criteria of successful problem-solving are relatively simple. They become extremely important if the computer is used to make social decisions, for there our criteria of adequacy are as subtle and as multiply as human thinking itself."

The temptation to utilize computer technology in more and more areas of public concern is great, of course, because it provides for a greater efficiency and accuracy in statistical studies and for a sense of scientific authority in solving difficult problems involving complex information.

However, in their article in a 1962 issue of Science magazine, Professors Johnson and Kohler warned against turning over to computers questions with which man is better able to cope:

"It (computer technology) is being called on to act for man in areas where man cannot define his own ability to perform and where he feels uneasy about his own performance—where he would like a neat, well-structured solution and feels that in adopting the machine's partial solution he is closer to the 'right' than he is in using his own."

There is certainly no doubt that computerized centralization of information gathered from the files of Federal agencies would facilitate many government operations. In the words of Johnson and Kohler, it would provide a "neat, well-structured solution" to the present unordered and amorphous task of maintaining complete and current files on information gathered from widely scattered sources for widely scattered purposes.

At the same time, however, in their article on "Privacy and Behavioral Research" that appeared last fall in Columbia Law Review, Oscar M. Ruebhausen and Orville G. Brim, Jr., noted:

"Computerized central storage of information would remove what surely has been one of the strongest allies of the claim to privacy—the inefficiency of man and the infallibility of his memory."

Thus, greater efficiency in Government operations would be paid for at the far greater expense of weakening the right to privacy of all American citizens. Surely this is too exorbitant a price to pay for an economized filing system.

In his essay entitled "Machinemade Justice: Some Implications," Prof. Joseph J. Spengler discussed certain implications of the use of computers in the definition and the administration of justice. His essay illustrated how and why certain biases may be introduced along with the use of computers in the area of law.

His final warning seems to the point:

"Great care must be exercised to avert the biases, the distortions, and the probable miscarriages of justice . . . If computers and other instruments are to be used, every precaution must be taken lest the mechanical servant become master, and a tyrannical one at that."

We are particularly concerned with the threat which a possible future Federal dossier bank represents to our Nation's basic Judeo-Christian doctrine which provides man with the unlimited right to make amends for his mistakes and to begin again. Man can forgive and forget the trespasses of his fellow man, but computers do not forget and they are incapable of forgiving. Rash actions one may have taken in one's youth and which one regrets in his maturity would be recorded for posterity on the computer's tape. Mistakes in judgment which are later regretted and repaired in an individual's personal and business life would remain imprinted indelibly on his computer data card. Thus, computerized files could become a bank of static, petrified and tyrannical information which can be used again and again to the disadvantage and harm of the American citizen.

We are now on the brink of making a fundamental change in our society which will destroy the basic philosophy of letting a man start anew, his record unblemished by past mistakes for which he has paid his just debt to society. We appear to be moving slowly but steadily toward a doctrine of complete scientific objectivity which will categorize and catalog each aspect of individuality, leaving as an end result a stack of computer program cards where once were human beings.

During the last decade, tremendous advancements have been made in the fields of psychology, sociology, political science and economics. Emerging under the new title of the "behavioral sciences," these disciplines have delved deeply into the complex problem of how man affects his society and how social institutions affect man. It is predicted that within 30 years, the behavioral scientists will be able to produce the achievers in our society at will. The August 15 issue of Newsweek reported that current research indicates that—

"The achieving child will be the product of order, home and school environment, and other factors under the control of parents—or the state."

Newsweek also predicted that within this time period, sociologists will have developed the "complete picture of manpower flow in our Nation." Orville Brim, president of the Russell Sage Foundation—a private institution for research—was quoted as saying:

"Incentives like money, and educational opportunity could be controlled so that people are properly distributed."

The prospect of such social manipulation is a fascinating one, but it also raises serious questions, the answers to which are at present unknown. Who will define a "proper distribution" of financial and educational opportunity and who can guarantee that society will use such manipulative techniques constructively?

Experts in the field of computer science have joined with behavioral scientists, combining the knowledge of both disciplines in conducting extensive research into the question of man's relationship to the society in which he lives. The widespread use of computers has greatly facilitated the accumulation of data and the transformation of varied and unrelated information into clear and meaningful statements on man's behavior and the pattern of his actions in the past.

But it is essential to remember that the computer cannot predict infallibly what a man will do in the future; it cannot set down axiomatic laws to govern a man's action at any given moment in the next hour, the next day or the next year. A computer cannot measure courage, loyalty or love.

More than 2,000 years ago Plato wrote:

"The differences of men and actions, and the endless irregular movements of human beings, do not admit of any universal and simple rules. And no art whatsoever can lay down a rule which lasts for all time."

It is our greatest fear that modern computer technology will attempt to do just that—to establish on the basis of compiled data on man's past actions axiomatic principles for predicting what he will do in the future—and that these principles will become accepted by society as nearly infallible. The final result would be the restriction of a man's future based upon the statistical pattern of his actions in his youth.

I believe that our Government and our legal structure must act now to retain their flexibility, if they are to remain useful and beneficial to society. Both must continue to recognize and respect the infinity of choices which a single man may make during the course of a single life-span. This is why we voice so strongly our opposition to the present proposal to establish a national data bank which would store for all time private and confidential information on all citizens which might later be used to restrict their free choice of action in the future.

Therefore, it is essential that we begin to determine now the potential of today's technology and how our traditional liberties and beliefs can be protected from a technological onslaught in the future. Most specifically, we must chart today the course we want that will allow computer science to follow in contemporary society, and in our society of tomorrow.

We must call upon the scientific community, which is responsible for the development of this technology, to bear an equal responsibility for its control, in order to guarantee adequate protection of the freedoms we now enjoy. For this reason, it is of vital importance that representatives from all of the disciplines involved in the development and implementation of the computer join in creating a symposium independent of any Federal agency to examine the potential of computer science and its effects on the rights of the individual.

The need for such a symposium was well indicated during the course of our hearings last month. Spokesmen for the Bureau of the Budget came to the subcommittee to discuss the establishment of a Statistical Data Center. Yet, under extensive interrogation, the witnesses proved to be at a distressing loss of words when pressed for a detailed explanation of the system and for specific safeguards that could be built into the center to provide for the protection of the individual's right to privacy.

They seemed unable to comprehend the ease with which a statistical data bank could be converted into a personal dossier center, and they failed to realize the potential power of such a center. This is, we believe, the crux of the problem with which we are faced. There appears to be a basic lack of communication among the computer scientists, the behavioral scientists and experts in constitutional law and civil liberties.

It is unfortunate that while the American Orthopsychiatric Association was discussing data banks and invasion of privacy at its annual convention in San Francisco in April, the American Statistical Association was planning its convention for later this month in Los Angeles with a panel discussion on the "design and use of statistical data banks." The New York City Bar Association's Committee on Science and Law was studying the "impact of science and technology on privacy," and the American Bar Association was planning a special section of the association to deal with legal problems concerning individual rights and particularly the relationship of these rights to modern science. Yet, apparently no effort has been made on the part of these individual associations to combine their interests in a joint symposium.

We can no longer afford isolated contemplation in this area. Certainly social scientists, computer technologists and experts in constitutional law are all equally concerned with the problem of achieving a balance between advancing technology and the preservation of individual liberties. The computer has made all of us partners in the development of modern technology and we must continually educate each other, if we are to achieve this crucial balance. A symposium conducted by experts from each of these fields could serve as an arena for discussion of all major aspects of the problem.

It is of utmost importance that the symposium consider, independently of the Federal Government, the technical, legal, and sociological aspects of the proposal which the Government seems intent upon rationalizing.

The symposium might be conducted through a series of seminars organized by professional associations such as the American Economic Association, the

Association for Computing Machinery or the American Bar Association. Sponsorship of such a symposium might also be offered by a consortium of American universities, for certainly the role of the scholar as independent social commentator has long been traditional in our universities and colleges. Those private foundations and institutions which have financed research in these areas during the past decade might be called upon to fund the symposium.

In seminar fashion, the symposium would consist of a series of papers submitted and discussed by panels comprised of experts from each of the disciplines involved. It is hoped that the proceedings of the seminars could be published and widely circulated in order to provide for even more extensive examination by other members of the computer industry, the academic world, and the public in general.

The findings of such a symposium would then represent the recommendations formulated by the most qualified experts in our Nation. Only with the benefit of their accumulated knowledge could Congress properly evaluate such proposals as the Bureau of the Budget's, calling for the creation of a National Data Center. Only when all of the implications raised by such a proposal have been examined objectively can Congress act effectively to insure its legality.

In this era which has been so greatly affected by the machine, we must call upon our greatest natural resources—that is, the wisdom and knowledge of America's scientists, scholars and legal experts—to aid us more than ever before in achieving a balance between the advancement of scientific technology and the preservation of constitutional liberties.

The New York Times of August 9 in commenting on our hearings and expressing their concern over the National Data Center summed it all up in two lines:

"Perhaps in the long run the fight to preserve privacy is a vain one. But, like the struggle to preserve life it must be continued while any shred of privacy remains."

In the future no responsibility of our Government will be greater than the preservation of privacy and the protection of our fundamental human values. Like the problem of nuclear warfare, it is a time to reflect on how far we have come before we drift into a course that is beyond our capacity to navigate.

Mr. HORTON. It is a privilege for me to join the gentleman from New Jersey [Mr. GALLAGHER], chairman of our Special Subcommittee on the Invasion of Privacy, in pointing out some of the findings of our recent hearings on the proposed National Data Center. Chairman GALLAGHER has competently reviewed the fears and feelings these hearings have fostered: Scientific technology is the creation of man, and he, as creator, must also be controller. Before we allow ourselves to be mastered by machines, we must consider our rights guaranteed by the Constitution and take the necessary action to protect them.

Though a dossier bank is not proposed, testimony in our hearings substantiates the supposition that a statistical center, which is proposed, does have the potential to hold privacy-invading information on the citizens of this country. The spectrum of information already contained in computers within the separate agencies, if brought together, could reveal with the push of one button every record made on an individual from his birth certificate to the present. In fact, there is no reason to doubt that such a system could trace back and even bring together information on an individual's parents, grandparents, aunts, uncles, cousins, friends, and associates.

As I pointed out in my opening statement at our hearings, a central data service bank would require:

First. That confidential information now in Government files would be forwarded to a new group and used for other purposes than it was originally given; and

Second. That a new group would have the code and would have access to the names, addresses, and background of the people to whom this confidential information relates.

Tying the two together would be an easy matter—and it would be an outright denial of our right to privacy. As Vance Packard brought to our attention during the hearings, when the social security number was originated, it was a confidential reference. Now it is requested and given on practically every form an individual completes in his lifetime. It seems to grow easier to give out information, whether or not it was once confidential, than it is to protect confidentiality. One of the greatest safeguards now protecting information possessed by various agencies is its fragmented nature. Retrieval is impractical and often impossible. A central data bank removes completely this

safeguard and threatens to make available entire dossiers, rather than simply a social security number.

An injustice which deepens the privacy invasion constituted by such an information center is the individual's complete ignorance of what is contained in his dossier—whether it is scattered throughout various agencies as it is now, or available in one bank. Certainly it is not unreasonable that individuals be given the opportunity to see their files and to rebut information they feel is fallacious or, as it may often be, the result of prejudiced reports.

On July 28 I issued a statement supporting the idea for a symposium of educators, social scientists, lawyers and computer technologists to study methods to prevent large computerized data banks from invading the privacy of the American citizen. I would like to share with my colleagues the text of my remarks regarding this Government-sponsored symposium:

"Computer technology is advancing so quickly that we are in danger of allowing machines to bare the most private facts in our lives.

"Our hearings have pointed out a tremendous need for knowledge in this area of computers and privacy. This symposium would fill that gap and allow us to work out adequate safeguards against unauthorized release of private information to people with no right to have it.

"Vance Packard, a witness at our hearings, warned that 'Big Brother, if he ever comes to the United States, may turn out to be not a greedy power-seeker but rather a relentless bureaucrat obsessed with efficiency. And he, more than the simple power seeker, could lead us to that ultimate of horrors, a humanity in chains of plastic tape.'

"Packard, whose best-selling books have documented some of the country's social ills, said, 'The filekeepers of Washington have derogatory information of one sort or another on literally millions of people. The more such files are fed into the central files, the greater the hazard the information will become enormously tempting to use as a form of control.'

"Computer technology is so complex that we can't pass adequate laws immediately to block release of private information. This symposium must study the question with the aim of defining the problems involved.

"This committee should be commended, since it is one of the first groups to recognize the dangers inherent in the use of the giant, information-storage devices. For, as another witness at our hearings, Paul Baran of the Rand Corporation, predicted, computers of the foreseeable future could be 10,000 times the size of those available today.

"As I said during the hearings, there is a danger that computers, because they are machines, will treat us as machines. They can supply the facts and, in effect, direct us from birth to death. They can 'pigeonhole' us as their tapes decree, selecting within a narrow range the schooling we get, the jobs we hold, the money we earn and even the person we marry.

"It is not enough to say 'it can't happen here'; our grandfathers said that about television."

It is clear that a computer bank offers great potential for advancements in efficiency; yet, it also offers great potential for invasion of privacy. I do not propose to limit the progress of science or knowledge, but to control the use of knowledge scientific advancement makes available so readily. The concept of such control is not new; man's greatest helpers can also be his greatest destroyers if not controlled—we need only consider basic elements as water and fire to confirm this.

Mr. Speaker, I want to conclude by inserting an editorial entitled "To Preserve Privacy," which appeared in the August 9 edition of the New York Times. Following the editorial is the text of a letter I wrote to the editor commending his editorial and reviewing the necessity to protect the freedom of individuals with private lives:

"TO PRESERVE PRIVACY

"Can personal privacy survive the ceaseless advances of the technological juggernaut? Many in public and private life now fear to use telephones for conversations they would keep confidential, while the variety of electronic 'bugs' available to eavesdrop on even whispered communications staggers the imagination. And young lovers would be well-advised to remember that the skies are increasingly full of sputniks equipped with cameras capable of taking extraordinarily detailed pictures of what transpires under the moon as well as on it. George Orwell foresaw the logical end of this trend in a device that would enable 'Big Brother' to keep an eye on everyone anywhere.

"The Orwellian nightmare would be brought very close indeed if Congress permits the proposed computer National Data Center to come into being. We already live with the fact that from birth to grave Federal agencies keep tabs on each of us, recording our individual puny existence, monitoring our incomes and claimed deductions, noting when we are employed or jobless, and—through the F.B.I. and similar agencies—keeping all too close watch on what we think or say, what we read and what organizations we belong to.

"If this situation is still somewhat tolerable, it is because each agency keeps separate files and it takes some considerable effort to find and bring together all that is known about a particular individual. What is now proposed is the amalgamation of these files, and the creation of a situation in which the push of a button would promptly dredge up all that is known about anyone.

"Understandably, this idea has brought vigorous protest, in which we join. Aside from the opportunities for blackmail and from the likelihood that the record of any single past transgression might damage one for life, this proposed device would approach the effective end of privacy. Those Government officials who insist that the all-knowing computer could be provided with safeguards against unauthorized access are no doubt of the same breed as their brethren who 'guaranteed' that last November's Northeast electric blackout could never occur. Even the Swiss banks have learned to their own and their clients' sorrow that the device of numbered accounts is inadequate to frustrate determined would-be blackmailers.

"Perhaps in the long run the fight to preserve privacy is a vain one. But, like the struggle to preserve life, it must be continued while any shred of privacy remains."

AUGUST 10, 1966.

LETTERS TO THE EDITOR,
The New York Times,
New York, N.Y.

DEAR SIR: As a citizen and Congressman, and particularly because of my service as the ranking and only minority member of the Special Subcommittee on Invasion of Privacy of the House Government Operations Committee, I commend your August 9 editorial, "To Preserve Privacy."

Our Subcommittee, chaired by Congressman Cornelius Gallagher of New Jersey, just completed hearings on the Federal Government proposal to establish a National Data Center. The testimony removed any doubt as to the foundation of your "1984" fear.

However, I am not willing to resign myself to the fateful suggestion that "Perhaps in the long run the fight to preserve privacy is a vain one." Rather, I intend to continue, as your editorial also challenges Congress, with my struggle for its protection.

The problems posed by the National Data Center proposal are in carload lots. Like nuclear energy, there is nip and tuck competition to keep computers working for us, not against us.

There is undeniable value from the standpoints of economy and efficiency in allowing agencies to pool data. But, if it is to be done, must there not be ground rules and clear-cut standards?

Thus, I recommend the calling of a symposium of computer men, sociologists, educators, lawyers, and others to identify a correct course. These men and women should suggest safeguards: coding, surprise audits, inter-computer interrogation limits, interrogator identity, abnormal interrogation detection, illegal disclosure penalties, and many more. We also need to consider the individual's right to know the contents of any government dossier on him.

Computers can give us longer life, teach our students, design better transportation, provide our statesmen with facts and figures for sound decisions, capture criminals, diagnose disease, and add new dimensions to every element of society. But, making it possible means maintaining man as the master of the machine.

Advanced technology must not be paid from the accounts of individuality. For as Justice Brandeis said, we all are entitled "to be let alone."

Sincerely,

FRANK HORTON,
U.S. Congressman, 36th District of New York.



