

substance of these recommendations, and in some respects went beyond them. Dr. Dunn was assisted in this report by a group of experienced professionals drawn from various parts of the Federal Statistical System, as well as by experts in automatic data processing of the National Bureau of Standards.

As it is presently operated, the statistical system is both inadequate—in the sense of failing to do things that should and could be done, and inefficient—in the sense of not doing what it does at minimum cost, or getting less for what it spends than might be possible.

The inadequacy of the present statistical system has three major aspects. The first is the lag between the receipt of information and its availability in usable form. This is most striking in the case of the Statistics of Income for Corporation Income Tax Returns. There is a one-and-a-half year lag between filing of returns and preliminary summary publication, and a two-and-a-half year lag before final detailed publication. A large part of the problem arises from the variation in filing dates of corporations filing on a fiscal year basis: some may file as much as 10 months after the end of the calendar year under which their returns are compiled. But part of the problem does reflect questions of priority and availability of facilities, and though these reports provide a basic source of economic data of great importance, their reporting function cannot be given first place in the administration of the Internal Revenue Service.

A second and deeper source of inadequacy in the present system is its widespread suppression of micro-information, and its orientation toward publication of necessarily aggregated and tabulated information as its major goal. These are of course intimately related: restrictions on disclosure to the general public or unauthorized persons within the government of information on individual reporting units is a necessary and desirable legal constraint on any official agency collecting information under the sanction of law. So long as publication is thought of as the basic process that makes information available for use, aggregation and the suppression and ultimate permanent loss of micro-information cannot be avoided. The consequence, however, is the necessity of substituting worse for better information, and cruder for more refined analyses, by those who use the data for research and policy purposes. In particular, much ingenuity and effort is spent in the construction of rough estimates of magnitudes and relations that could be measured with much greater accuracy, if the micro-information that present statistical records originally contained was preserved in usable and accessible form. Present technology makes it possible to do this economically and consistently with desirable limits on disclosure.

The growing decentralization of statistical programs has led to another major inadequacy. At the present time different agencies view the problem of the right to privacy very differently. In some agencies the policy of protecting the privacy of the information reported by individuals and businesses is formally stated and protected by law; in such instances the enforcement of such policies has also been found to be very good. In other instances, formal policies regarding disclosure have not been set up, and in many of these cases the protection depends on the judgment of those who are in charge of the different programs involved. Understandably, the growing decentralization of statistical programs has thus led to considerable unevenness in the nature and enforcement of disclosure rules. It is quite possible that without some overall policy which can be responsibly supervised major violations of individual privacy may take place. It should be the function of some group within the Federal Statistical System to ensure that data gathered for statistical purposes or obtained as a by-product of the administrative process is not to be used against an individual or enterprise. Thus at the present time information about individual persons or businesses collected by the Census Bureau cannot be used by the Internal Revenue Service or the Department of Justice against individuals or enterprises in the investigation or prosecution of such things as tax evasion or antitrust violations. This type of protection must be preserved in order both to protect the rights of individuals involved and to avoid falsification of information which might develop if individuals were not given assurance against disclosure.

The major elements of inefficiency to which decentralization has led are of three kinds. The first is duplication in the collection of information. Although the Office of Statistical Standards controls duplication, it is not always successful in eliminating it entirely. Avoiding duplication is especially important in that it needlessly spends not only money but the even scarcer resource of cooperation by the public; households, business firms, and other respondents, in answering enquiries. While duplication within single agencies is not serious, the great degree of decentralization leads to overlaps between programs of different agencies. The problem is less the collection of exactly the same information by two agen-

cies, and more the collection in two surveys or reports of data that could be collected in one. Failure to make the maximum use of each occasion for collecting information may well lead to a burden on respondents which becomes intolerable with growing needs for data. An example of the problem is provided by current practice in connection with sample data on retailing. The Bureau of the Census collects data on retail sales from one sample of retail stores and the Bureau of Labor Statistics collects data on employment, wages, and hours from another. As a result, there are doubts about the comparability of these input and output data at various levels of publication detail. These doubts arise not so much from the differences in the two samples as from differences in the two Bureaus' methods of assigning industry codes and definitions of reporting units. If both input and output data were collected on the same report form and processed by the same agency, these differences in comparability would be eliminated. This situation applies not only to retail sales but also to manufacturing data, where the Bureau of the Census collects monthly figures on sales, orders, and inventories, while the Bureau of Labor Statistics surveys manufacturing employment, man-hours, and wages each month. There is little doubt that a single consolidated reporting system, using one sample, would be both less burdensome, and less costly, and yield better information.

The second source of inefficiency is failure to use as a statistical resource all the information potentially available in the data collected. This, in turn, has a number of sources. (1) Collection of the data on the same reporting units by different collecting agencies operating with different classification systems, unit definitions, and the like, results in inability to match all the relevant available information on a responding unit for analytical purposes. Information on groups of respondents of different, and to some extent imperfectly known, composition cannot properly be compared and correlated. Census, IRS, SEC, and FTC data on business enterprises exemplify this problem. These incompatibilities in definition often reflect the different purposes of the several agencies that collect the data; yet effort directed to resolving these problems can be fruitful and is worthwhile. (2) After separate collecting and processing, agencies assemble data in summary form; the original individual reports are all but unavailable for further use, or available only at prohibitive costs. This effectively prevents different summaries and analyses of the data for other purposes by the same agency or by different agencies. In particular, the efficient use of data for intertemporal comparisons over any but a short time period becomes difficult, as the classifications change over time, and thus much information is irretrievably lost. (3) Confidentiality restrictions as interpreted by different agencies often act as a barrier to the full use of data for statistical purposes inside the government and within the legal boundaries of use.

The third source of inefficiency is that many of the smaller agencies operate on too small a scale to make fully efficient use of modern techniques, professional specialists, and economical large-scale machines. Only further centralization, rather than better coordination, can cure this situation.

The degree of decentralization in the system, and its predominant orientation toward publication as a means of making information available, correspond to a now-obsolete technology of handling and storing information, as well as to a much lower level of demand for detailed quantitative demographic, economic, and social information by policy-making agencies of all levels of government. Our present organization and mode of operation does not take advantage of modern information processing technology, and is not capable of meeting the variety and scale of present day information needs. The deficiencies of the system, and the gap between what it can provide and what would be technically possible under appropriate organizational arrangements will grow rapidly in the near future. As we have already pointed out, the demand for detailed quantitative information will continue to increase at a high rate. Further, the nature of the demand is changing in qualitative terms in ways that are only just becoming clear. The degree of disaggregation now demanded in the data relevant to economic policy has changed greatly in the last decade, even though the policy continues to focus on objectives stated in terms of such aggregate magnitudes as employment, unemployment, output, and the general wholesale and consumer price indices. The demand for comprehensive micro-data will grow explosively as policy becomes increasingly concerned with the micro-effects of the economic system, in terms of particular localities, income, and occupational, age and ethnic groups; as policy instruments become increasingly capable of sensitive and selective application to particular needs, and include a broader range of government actions in such areas as education, research, health, housing, transportation, and resource development. Further, the need for coordination of data collected

at state and local levels, in answering questions concerning specific small geographic regions, is also growing rapidly.

On the side of information processing technology, the last decade has seen great developments in machine processing, storage and transmission of information in machine-readable form. This progress is continuing both in the computing equipment itself (hardware) and in the programmed instructions for directing the machines (software). These developments place the problems of large-scale storage, integrated data files, rapid access, and confidentiality in an entirely new light. In particular, it is now possible, with sufficient effort, to create the capability for combining centralized processing and storage of large bodies of data with decentralized analytical use, subject to the restraints of a uniform system of limits on the disclosure of data on individual reporting units.

In pointing to the shortcomings of the Federal Statistical System as presently organized, the Committee does not wish to suggest either that those charged with its operation are unaware of these problems, or that they are making no attempt to find remedies for them. Quite the contrary. The Committee has the highest regard for the professional competence and dedication of the senior personnel of the major statistical agencies. We can say the same of the Office of Statistical Standards in the Bureau of the Budget, which is now charged with the coordinating responsibilities for the Federal Statistical System, as adviser to the Director of the Budget. These two groups are now making serious efforts to deal with the kinds of problems we described. Recent increases in the level of Census work performed for other agencies on a reimbursable basis is an example of one method of meeting these problems. The cooperative efforts of the IRS and the Census Bureau to use data from the income tax returns for the Economic Censuses is another. Further, the heads of the statistical agencies and the officials of the OSS have been of the greatest assistance to the Committee in making this report. However, the Committee believes both that insufficient resources are being devoted to dealing with the problem, and that the present organizational framework cannot generate improvements in the existing situation fast enough to cope with the growth of the problem. We conclude, therefore, that significant organizational change as well as increased effort are necessary conditions for a successful attack on it.

The building of an integrated body of data combining presently available sources of the appropriate kinds of data, which preserves in usable form the maximum detail of information, stored on tape or other machine-readable form, coded, organized, and indexed so as to be readily accessible, is the minimum step which must be taken to cope with problems sketched above. The existing agencies are now approaching this task slowly, with a scale of effort too small to ensure that it can be completed, and under a variety of inhibitions and constraints. The Census has taken a commendable lead, and already has done a number of useful tasks. However, this job is viewed both by Census and the other agencies as a second-priority activity, which cannot compete for personnel, machine time, or funds with ongoing current programs. This is natural, and indeed inevitable. Simple inter-agency jealousies and rivalries have also created inhibitions on prompt and full cooperation.

The Office of Statistical Standards is too remote from operating responsibility to move forward at the pace at which such integrating activities need to be carried on. Its negative powers are, at least theoretically, great, but its ability to promote new programs is limited to what it can accomplish by persuasion in the face of institutional pressures which go in the opposite direction.

Finally, none of the constituent agencies of the System has given the problems described above or the opportunities to deal with them provided by advancing technology, the importance that this committee—following in the footsteps of several predecessors—assigns them.

3. WHAT IS TO BE DONE?

Were the Committee to be designing a Federal Statistical System *de novo*, it would clearly recommend the creation of a single Central Statistical Agency with the following responsibilities:

- a. Collecting all systematic, general-purpose, large-scale quantitative information of a demographic, economic, or social nature, insofar as it is not produced as a by-product of the administrative operations of the Government.

The qualifying adjectives are meant to preserve the freedom of operating and policy agencies to make, when necessary, occasional surveys or special-purpose studies for their own purposes when the Central Agency was unable to provide the requisite information, although the Central Agency would have the facilities

and capacity to carry these out as well, on request. Further, the limitation of the Central Agency to dealing with general purpose information is made explicit in order to recognize the variety of operating needs for gathering and processing information of the several branches of Government, which they must be in a position to meet directly in order to carry out these tasks.

b. Receiving and integrating into its general information stock, data which are the by-product of administrative operations of the other Federal Agencies. In this connection, the Central Agency would, through liaison with the other agencies, help to design tax, regulatory, social insurance, and other report forms in such a way as to produce the maximum information consistent with reasonable burdens on respondents.

c. Developing and maintaining appropriate standards of confidentiality in the release of any data, as determined by law, using the basic data within the bounds of these standards so as to minimize the loss of analytically useful information, while at the same time ensuring protection to the privacy and identity of individual reporters.

d. Coordinating its activities to the greatest possible extent with those of the information collecting of states, cities, and other governmental units so as to arrange as far as possible a rational division of labor, a maximum integration of information, and a free flow of useful information in both directions.

e. Organizing and storing information in such a way as to provide maximum legitimate accessibility, for both governmental agencies and other users.

f. Providing computing, tabulating, and analytical capacity for all government users. Insofar as it proves economical, these facilities could be drawn on by other agencies for any special computations, analytical studies, etc. In other words, the Agency should maintain the central general-purpose large-scale data-processing center for general Government needs. Non-confidential data from the Center in standard documentary or machine-readable form should also be available, on a suitable compensatory basis, for the research uses of academic and other private groups, when such use serves a public purpose.

Neither this function, nor that of *a.* above is intended to preclude other agencies from maintaining independent computing facilities and independent analytical capabilities of an appropriate sort. In particular, it is obvious that every policy agency, as well as many administrative units, will require analytical capabilities for program planning and program evaluation. But current technologies, including distant consoles connected with a central computer facility by telephone links, make compatible decentralized use of data for analytical purposes, with highly centralized data storage and processing.

g. Studying methods of improving the protection of individual privacy and the confidentiality of data while at the same time providing use of it for legitimate analyses. Both the screening of analyses before release, and the camouflage of the basic data itself offer promising paths for exploration.

h. Improving methods of data collection, techniques of sampling, and opportunities for maximum use of by-product information, both for economy's sake, and to minimize the burdens on respondents of increasing demands for information.

i. In cooperation with the analysts who use the data, both within and without the government, defining and refining the standards and bases on which information is collected, and determining the probable development of information needs.

j. Improving the techniques of data handling, storage, and computation, in cooperation with appropriate technologically competent public and private agencies.

k. Securing the research and development contributions of university, business and other groups to the effectiveness of all these functions. This can be done both by making grants and contracts and by providing facilities and capacity which such research personnel could use on a variety of financial bases. Outside research and development assistance might usefully cover the whole range from long-term basic research to assistance in the solution of an immediate problem.

The Committee is not starting with a clean slate. Realistically, the question before us is how to proceed from the present situation of too much decentralization and insufficient coordination. We have not attempted to judge either the wisdom or the feasibility of attempting to create, at one blow, the kind of Central Statistical Agency we have described above. Rather we have sought to make a step, of sufficient magnitude to inject a genuinely needed new element into the system to help it adapt more rapidly to the growing problems it faces. Following this step we envisage further adaptive evolution in the direction of a stronger and more centralized system by an experimental process.

What first step is sufficiently large so as to promise a good prospect of further development? The Federal Statistical System has three basic functions; namely, (1) collection, (2) integration and storage in accessible form, and (3) analysis, tabulation, and publication. It is reorganizing the second that offers the most promise. This function is now the least well-performed of the three, and it is the one which is most easily separated out from the present organizational structure. However, it must be done on a substantial scale, and in such a way as to recognize the interaction of this function with the other two. Further, the new organization must not be confined to a merely archival function. If it is defined along the lines suggested below, it offers the best promise, in the judgment of the Committee, for starting the development of the Federal Statistical System toward a more integrated and efficient form.

Accordingly, the Committee proposes the creation of a National Data Center. This Center would be given the responsibility for: (1) assembling in a single facility all large-scale systematic bodies of demographic, economic, and social data generated by the present data-collection or administrative processes of the Federal Government, (2) integrating the data to the maximum feasible extent, and in such a way as to preserve as much as possible of the original information content of the whole body of records, and (3) providing ready access to the information, within the laws governing disclosure, to all users in the Government and, where appropriate, to qualified users outside the Government on suitably compensatory terms. The Center would be further charged with cooperation with state and local government agencies to assist in providing uniformity in their data bases, and to receive from them, integrate into the federally generated data stock, store, and make accessible, the further information these agencies generate. The funding and staffing of the Center should recognize both these functions.

In more detail, the functions of the Center would be:

(1) To establish and maintain an inventory of all available data in the relevant categories in the Federal System.

(2) To set and enforce uniform disclosure standards so that the legal requirement of confidentiality can be met with no unnecessary sacrifice of analytically useful information.

(3) Similarly, in cooperation with the state and local government units, to perform similar tasks for information generated at those levels of government.

(4) To assemble centrally the data from all these sources, integrate it to the maximum feasible extent, and preserve it in usable and accessible form. This will involve:

The maximum ability to exhibit the interrelations of various kinds of data.

The preservation of detail in basic records, and the avoidance of the loss of information in the storage, manipulation, and retrieval of information.

The ability to produce the full measure of inherent information which is computable from the basic records.

(5) In cooperation with users in and out of government and collection agencies, to set the standards for further collection efforts, so as to make maximum use of administrative information and provide maximum cross-linking of different bodies of data.

(6) To provide facilities—from working space to linked input-output consoles—for major users within government to facilitate their access to the data and improve their ability to work with it.

(7) To develop software and hardware, especially input and output devices.

(8) To define the regulations and compensation arrangements under which non-government users would have access to data in the Center. In general, subject to disclosure restrictions, standard tabulations and tapes could be made available at cost to private users for research and analytical purposes. However, the Center should not become a service bureau or data-processing agency selling special order analyses to private users in competition with firms and individuals in the information processing industry.

In full operation, the National Data Center would provide the following benefits:

(1) Reduce the collection effort and particularly the burden on respondents required to secure a given amount of information.

(2) Improve the protection of individual privacy by developing standards of disclosure, techniques of preserving confidentiality and supervision of enforcement of disclosure rules.

(3) Preserve for continued use all or nearly all the relevant detailed information contained in the original data, as compared with the present situation in which much of the detailed information is irretrievably lost, or become retrievable only at prohibitive cost.

(4) Reduce the processing costs associated with the use of a given amount of information.

(5) Store information in more accessible forms at lower per unit costs and with a comprehensive index or bibliography.

(6) Make much information accessible to non-Government users which now is too expensive or too cumbersome for them to use, even though it is legally available and its use would benefit the general public. This is especially relevant to users in state and local governments, academic and other non-profit research users, and business users. In this connection, the Center should develop extensive working relations with academic users, of the sort which the Census has done to a much greater extent than other agencies. Even these are limited and currently are handicapped by lack of physical facilities, programming capacity, and organizational capacity for dealing with them. As the working relations of the physical science establishments of the Federal Government with the academic and industrial scientific communities demonstrate, such cooperation is of great benefit to the Government in performing its tasks effectively.

(7) Provide improved analyses of existing data for all users. The facilities for cooperative efforts are highly relevant to this point as well.

(8) Facilitate greatly improved coordination of statistical data between the Federal Government and the states and localities, and internationally as well.

(9) Create a repository of technical competence in statistical services, and computer software and hardware, that would be available to the whole Government establishment.

4. PROPOSED ORGANIZATION

In order for the National Data Center to function properly, it must be given a proper position in the Federal Statistical System, and sufficient authority, leadership, trained personnel, and funds to perform its mission. The Committee has given special attention to the problem of finding the organizational arrangement most conducive to the successful functioning of the Center, and attaches great weight to its organizational recommendations. We recommend the creation of a new position, Director of the Federal Statistical System, in the Executive Office, and the placement of the Census Bureau and the National Data Center as coordinate units under his direction.

The Bureau of the Census is the largest, most widely experienced, most professionally competent, and broadest in scope of all the present statistical agencies. To the extent that any agency in the System attempts to perform the functions described above, it is the Census. The data Center will require close cooperation and support from the Census in order to function effectively. For all these reasons, it appears desirable to put the Center in close organizational and physical proximity to the Census. On the other hand, the Center's tasks are not the present tasks of the Census; the Center will be a new organization with the difficult problem of establishing itself as a going concern and making its way in the complex of agencies producing and using large bodies of quantitative information—its suppliers and customers, so to speak. Thus it does not appear appropriate to subordinate the new agency to the existing Census organization. Further, the establishment of smooth working relations between the Center and the other elements of the Federal Statistical System might well be easier if the Center is a new, coordinate agency rather than a part of the Census Bureau.

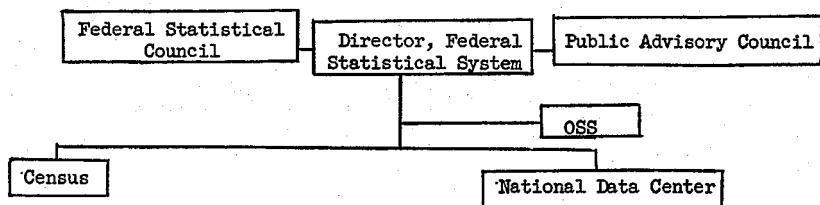
If the new data Center is to have specially close but coordinate relations with the Census and similar, if organizationally less intimate, relations with other data collecting and using agencies, some method must exist to regulate and oversee these relations. This coordination function is now assigned to the Office of Statistical Standards, but in the Committee's judgment, that Office is not placed so as to be able to carry it out effectively. We propose that a new position be created with this function, entitled Director of the Federal Statistical System, to be filled by presidential appointment. The new Director would exercise, by delegation or new legislation, as seemed appropriate, the coordinating powers over Federal statistical programs provided for in Sec. 103 of the Budget and Accounting Procedures Act of 1950 (P.L. 784, 81st Congress). The Office of Statistical Standards would accordingly be transferred from the Bureau of the Budget to become a staff office of the new Director, to assist him in carrying

out these responsibilities. The Census Bureau and the National Data Center, each under its own director, would report to the Director of the Federal Statistical System.

In addition, we propose two councils advisory to the director. The first is a Federal Statistical Council, representing the major data producing and using agencies in the System and reporting directly to the Director. Thus the Commerce Department might be represented by the Assistant Secretary for Economic Affairs, the Council of Economic Advisers by one of its members, etc. The second would be a public advisory council, with members from outside the Federal government representing both the public in general and particular users of information such as business, labor, state and local governments, and the academic community. This council would advise the Director particularly on such matters as the burden on respondents, the protection of confidentiality, and the satisfaction of user needs. The uses and possible abuses of information collected by the government are so important in our society that continuous public scrutiny of these problems at a high level in the Federal System is desirable.

The proposed new office, with its two operating elements and its government-wide coordinating functions could not readily fit into any of the existing Cabinet departments. Rather, its natural home would be the Executive Office of the President, and the Committee recommends that it be placed there.

The organizational relations of these elements are shown in the chart below.



The National Data Center itself might be organized in two main branches: an operations division, and a research division which would consider sampling methods, analytical and computing techniques, methods of protecting confidentiality, and liaison with extra-governmental research enterprises and with users. The research division could also cooperate with the relocated Office of Statistical Standards for data collection and for integration of existing bodies of data. The operations division might be so organized as to provide for the inclusion of sections on assignment from both the major data producing agencies, such as the Census itself, the Bureau of Labor Statistics, the Internal Revenue Service, etc., as well as from the major users within the government, such as the Office of Business Economics, the Council of Economic Advisers, the Tax Research Division of the Treasury, etc. These transplanted sections would greatly facilitate the free flow of information to and from the Center, on which its entire function would depend. They could also take responsibility, as a transition measure, for seeing that the confidentiality requirements of their own agencies were properly applied to the data collected by them.

The provision of space, funds, technical personnel, and machine capacity in the Center must from the first be such as to allow a good deal of flexibility in meeting demands on it by both governmental and non-governmental users. Too close a calculation of capacity, especially in machine time and programmers could prove fatal to the Center's ability to establish itself. So could too early an expectation of visible results.

It would be the major responsibility of the Director of the Federal Statistical System to see that the proper division of labor, coordination of information, and utilization of the Data Center were made by the constituent agencies in the System. In the furtherance of this responsibility, he might be asked by the Budget Director to review for him the statistical budget of each agency, in much the way that the Director of the Office of Science and Technology assists the Budget Director in reviewing the science budgets of each agency. Yet the basic functions in the System would be determined more by what might be called "market forces" than by the fiat of the Director. To the extent that the statistical agencies under his direction provided quicker, cheaper, and better sources, to meet rapidly expanding demands, the rest of the government would increasingly rely on them, rather than on the expansion of their own in-house

capabilities. The Committee is strongly convinced that the proposed new organization would encourage such a path of development. We do not believe that a mandatory reshuffling of the existing tasks of the other statistical agencies is either necessary or desirable, in the face of what we are confident will be great increases in demands on all statistical agencies, provided only that the new organization is started on a sufficiently firm basis to permit it to survive the inevitable birth trauma.

5. INITIAL STEPS

A prerequisite to the creation of the National Data Center is a re-examination and consolidation of the laws and rules governing intragovernmental disclosure of information on individual respondents, and their recodification in terms which will make the operation of the Center feasible. The Committee is neither sufficiently competent nor informed to judge whether this will require new legislation, or whether it can be done by executive authority within the framework of existing legislation.

As soon as the Center is created, it should be put to work with existing personnel, and machine capacity borrowed chiefly from Census, to deal with those tasks of inventorying data and creating integrated files which are already most advanced. A sufficient beginning has been made—as detailed in the Dunn report and the appendices thereto—to permit the Center a running start.

Initial funding for the Center must be large enough to ensure its viability, and to attract to it good people from within and without the government. As a crude guess, the Committee suggests an initial budget of the order of at least \$2 million per year, with the prospect of rising to \$5 million within three years, exclusive of buildings and computing equipment. A generous allotment of supergrades is as important to the new agency as money, since a large proportion of its tasks, especially in the initial stages, will require high technical and professional skills.

ANNEX: THE RIGHT TO PRIVACY, CONFIDENTIALITY AND THE NATIONAL DATA CENTER

After this Committee was convened and well into its work, Congressman Cornelius E. Gallagher, as Chairman of a Special Subcommittee on Invasion of Privacy of the Committee on Government Operations of the House, has raised questions about the possible threats to privacy and freedom that a National Data Center might present. These are serious questions, that deserve to be met squarely.

In general, our Committee believes that the problem of the threat to privacy can be met best by Congressional action, which defines a general statutory standard governing the disclosure of information that is collected on individuals either as a by-product of administrative, regulatory and taxing processes, or through Census or sampling procedures. The Director of the Federal Statistical System should then be given the responsibility for monitoring compliance with this standard, not only by the Data Center, but by all the parts of the System.

The problem of disclosure of confidential information about individuals and businesses is not new. It has long been recognized that the information which individuals and businesses provide under law to the Bureau of the Census, for example, is confidential. This means that no other Federal agency is permitted to see or use the individual records, and even Congress itself cannot obtain census information on any individual or company. In fact, this confidentiality has been guarded so zealously that Congress and the other agencies of the Federal government have been enjoined from obtaining from companies duplicate copies of those records which were submitted to the Census Bureau. The disclosure rules are meant to safeguard individuals so that they can feel sure that information which they give to the Census Bureau will never be used against them for such purposes as tax enforcement, antitrust, or Congressional investigations. The disclosure rule has not been interpreted, of course, as preventing the use of Census information for analyzing policy or providing information about specific groups, regions of the country, performance of industries, etc. In making tabulations of data, however, the Census Bureau carefully omits those classifications which might enable anybody to figure out information about individual firms or persons.

There are, of course, other Federal agencies which must by their very nature use information about individuals and firms for their operations. Thus for example the Internal Revenue Service not only must collect information about people's income and the taxes they pay, but this information can and should be used to

prosecute tax fraud or tax evasion. Similarly, the Social Security Administration must process information about each individual over a period of years, recording his job status, family status, etc. This information is necessary for the determination of social security payments. Such use of individual information is of course justified, necessary, and legal. On the other hand, it is a real question whether tax returns or social security records should be turned over to other groups who may wish to use them for other purposes if the persons or firms to whom the records refer may individually be affected thereby. The question of the proper or improper use of information by different agencies is indeed a ticklish one, and procedures should be developed by both the executive branch and the legislative branch which will protect confidentiality and insure the privacy of the individual. In a great many instances, agencies may wish to obtain information not for operating purposes, but in order to make policy decisions and to guide future operations. Thus the Office of Education has a real interest in knowing how college enrollments may be expected to develop in the future. Those concerned with questions of poverty wish to know the dimensions and structure of this problem. In a great many of these instances, the agencies in question have contracted with the Census Bureau to provide them with such general information based upon sample surveys. In these instances, a disclosure and confidentiality rule must be developed which will protect the individual and yet yield the general information which is required.

The enforcement of a statutory obligation as the primary method of dealing with the problems of safeguarding privacy can work excellently, as the experience of the Census Bureau shows. Indeed, the present situation, in which there exist a variety of different disclosure standards, some statutory and some executive, is much less conducive to protection of individuals' privacy than would be a situation in which, as our report suggests, the Director and the Data Center would have the obligation of enforcing a uniform standard over the whole system.

The Subcommittee has also raised the question of the creation of a vast file of individual "dossiers" incorporating police and FBI information, Armed Service and government personnel records, and the like. This is not the purpose of the proposed Center at all, and it is clearly within the power of Congress to distinguish between the collection and organization of general economic, social, and demographic information of the sort that Federal statistical agencies have traditionally collected—much of it on a sample basis—to which our proposed National Data Center is directed, and assembly of the sort of personal history information on named individuals that is contained in a personnel file or police file.

Finally, the Subcommittee has raised certain questions as to the technical security of data stored in machine readable form, and accessible through machine operations. Here again, this is not a new problem, and both organizational and technical means are available to control and limit the risks. Though bank robbers have not been totally eliminated, we have not on that account abandoned banks and banking, and the analogy seems to be perfectly appropriate. We think that the maintenance of privacy against both unwitting and illegal disclosure of information made available to the Government are real problems, to which our proposed new Center must direct attention and effort. However they are neither insoluble problems, nor ones of such magnitude as to make the organization and effective functioning of a National Data Center possible only at the expense of significant inroads on liberty and privacy.

APPENDIX III

COLGATE-PALMOLIVE COMPANY,
New York, N.Y., May 23, 1967.

HON. HERMAN E. TALMADGE,
Subcommittee of Economic Statistics,
Joint Economic Committee,
Congress of the United States,
Washington, D.C.

DEAR SENATOR TALMADGE: The testimony of Mr. John H. Aiken of the Federal Statistics Users Conference and Professor Frederick F. Stephen of Princeton University on May 18, 1967 before your Subcommittee made it apparent that a clear understanding and definition of available economical statistical resources, their uses, economic statistical needs and potential uses are needed.

Accordingly, I suggest that your Subcommittee consider proposing to Congress a comprehensive survey of:

1. Economic Statistical Resources and Services.
2. Description of actual uses and actual application
3. Description of users
4. Statistical needs
5. Description of potential users and potential applications among non-users

A survey of this kind would in part follow through the requests received by various agencies to determine the use to which the statistics were applied and possibly their contribution to a specific action. In order to sample properly the various services, institutions, population groups and resources I recommend that Census Bureau statisticians be consulted in view of their proficiency in sampling.

National Data Center

The need for a central file for statistical data is obvious to me for it would make available in relatively convenient form the statistical resources of the nation and minimize, if not eliminate, any duplication that may exist. A National Data Center would perhaps be best as a coordinating body rather than a controlling one. The individual agencies would continue to collect, analyze and control their data while filing with the National Data Center.

The Library of Congress would seem the most appropriate governmental body for maintaining the Central Computer file for data.

Individual Confidence

Much of the data collected for government and other institutions are dependent upon the voluntary cooperation of individuals and the organizations they represent. Legal safeguards should be provided to insure cooperation for surveys to be used for statistical purposes. The Bureau of Deceptive Practices of The Federal Trade Commission can provide you with further background information on the potential dangers when these confidences are violated.

With the increased electronic data processing and growing sophistication in statistical techniques applied to government, military, business, education and welfare and agriculture our needs are becoming such that the costs of conventional collection and analysis of data may far exceed our ability to meet them. For this reason and the pressing problems in every sector of the nation, dependent upon impartial data, in need of solution I cannot stress enough the importance of your committee and how appreciative many of us are of your efforts.

Very truly yours,

ARTHUR KOPONEN.

