

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-