

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