

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