

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 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. 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.