Data Center. The original intent to explore the potentialities of modern large-scale computers is constantly in evidence in the present formulation.

In preparing the report, the problem of describing the customer population was considered first. Something is known of the kinds of specialists who use data originating in the Federal Government to solve problems in economic policy, public administration, business economics, business administration, and a great range of social science subjects. But it is also apparent that this present user population reflects the capabilities and logistics of present ways of organizing and purveying data. At least equal importance attaches to those needs which are not met by present practices. How can these unmet needs be characterized?

An adequately specified information system would have to be based upon a broad review of the types of analyses that a wide spectrum of social scientists propose and upon the quantitative models that they build. For the present purposes and the limited scale of effort, it was decided to restrict the review to several classes of economic models directed at problems of national economic policy. Even this limited review revealed a variety of possible requirements for socioeconomic information which are not now being met, although many of the basic data are collected and compiled in some form by some Federal agency.

The review of economic models and their needs for statistical information was conducted at a 4-day conference at Fort Ritchie, Md. on August 26–29, 1965. Participants were M. K. Wood, D. Rosenblatt, and E. Glaser of the National Bureau of Standards and E. S. Dunn and P. F. Krueger of the Bureau of the Budget.

Subsequent conferences and staff work built upon the Fort Ritchie conference by developing (a) an enumeration of the services to be rendered, and (b) a description of the Federal Statistical Data Center in terms of its functions and principal characteristics. A summary of these is given below.

A. Nature of the services to be rendered

An integrated Federal Statistical Data Center appears necessary to perform the following functions:

1. To provide data in cases where the primary agency in possession of the data is not capable of making it available in the required format, detail, flexibility, or quality.

2. To provide data where the information originates in two or more reporting systems or agencies, in order to make available information about interrelationships in maximum feasible detail, without restrictions resulting from screening for improper disclosures at the time of transfer into the Center and through association of information from multiple sources relating to the same individual reporting unit or analytical unit.

3. To maintain an archive of statistical data, complete in the sense described in 2 above, with all corrections and adjustments carried through in a consistent manner, and with a collection of the accompanying codebooks and manuals.

4. To provide information outputs (responses to queries) in a variety of forms at the customer's option: printed tabulations, machine readable tapes, graphs, diagrams, etc., either locally or through telecommunications.

5. To establish, maintain currently, and operate a reference and referral service for the Federal statistical system.

The creation of such a Federal Statistical Data Center also should provide the following additional services an corollary benefits at minimum cost:

6. ADP equipment would be available for computation and data reduction in response to queries of customers: cross tabulations, averages, distribution statistics, smoothed curves, trend fittings, seasonal adjustments, periodic analyses, correlations, regressions, and more advanced analyses in order to give access to the full range of information computable from the collection.

7. Confidentiality audits would be performed by machine upon the information intended for release to customers.

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 from 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.