

*General principles for the organization of data for a Federal Statistical Data Center*

The consequent reconsideration of the organization of socio-economic data in the Federal Government is based upon two general guides. The first is to review the implicit informational requirements of the whole range of analyses and formal models proposed by social scientists, rather than to restrict the statement of information requirements to those needs which have been given principal attention in the past. The second is to consider the whole range of relevant tasks that the computer can assist, even if the manner of proceeding is radically different from current practice.

From this reconsideration four general principles emerge for constructing specifications for a Federal Statistical Data Center:

1. Maximum ability to exhibit the interrelations among various kinds of data;
2. The unification of all information about the individual reporting unit or analytical unit;
3. The preservation of detail in the basic records and the avoidance of loss of information in the storage, manipulation, and retrieval of information; and
4. The ability to produce the full measure of inherent information which is computable from the basic records.

These four principles will now be developed as groundwork for specifications for the information organization and the services of a Federal Statistical Data Center.

One of the greatest deficiencies of the existing Federal statistical system is its failure to provide access to data in a way which permits identification and measurement of functional interrelationships among interdependent activities. Identification and measurement of such interrelationships are essential to a wide range of economic and social analyses. It is also the chief problem in the design of mathematical models of economic and social processes suitable for appraising the impact of alternative policies and programs as well as possible changes in environmental factors.

Such appraisal is, in turn, prerequisite to effective benefit-cost analysis of proposed and ongoing programs. The essence of rational benefit-cost analysis is the tracing of indirect as well as direct effects of programs and the evaluation and summing of these effects. Typically, the methodology for tracing all but the most obvious linkages is entirely lacking or fails to use the relevant information.

Until recently, economic model builders have been restricted to relatively aggregative economic and resource flow models, and to inferring interrelationships among very few aggregative variables. Such relationships often have considerable predictive value where other conditions remain relatively stable or continue to change at a constant rate. But the essential relationships are correlative or associative rather than structural. Hence, they generally fail to give acceptable prediction when other conditions change markedly, as a result of changes in major program, policy, or environmental factors.

Acceptable prediction under changing circumstances requires analytical models which give much more detailed and explicit recognition to interrelationships among the criteria and variables which will be affected by the changed conditions. Such analytical models generally describe the mechanisms in greater detail than the associative models; they use more information, and they often rely less heavily on trends or the postulation of only slow changes among the variables in the model. The present and prospective accelerated pace of technological and statistical change now requires the development and use of more detailed and complex models than can be created or supported by the present Federal statistical system.

The rapidly developing tools of automatic data processing and systems analysis now make possible—and necessary—both the development of more advanced models and the elaboration of the Federal statistical system which is needed to support them.

Many of the data needed for establishing causal interrelationships among related economic variables are contained in the existing Federal statistical system. But present collection methods, tabulation procedures, and disclosure rules combine to make it difficult and often impossible to extract such data. Where samples are largely enough, it may be possible to cross-tabulate in a way which permits determining the interrelationships between two variables or, rarely,