

THE SOLUTIONS ARE COMPLEX

It seems clear from the foregoing problem characterization that the solutions to this set of problems will have to be multidimensional.

When it made its proposal for a national data center as a solution and developed its justification, the Ruggles committee revealed an understanding of many of these dimensions. Its proposal was a constructive one and intended to be interpreted with some flexibility. However, the representation of the solution in this way has had some unfortunate consequences not anticipated by the committee.

In designating the center as a national data center and placing considerable emphasis upon the collection of tapes—growing out of its concern with the archival problem—the proposal became quickly translated in the minds of many as another data bank proposal.

The data bank idea is enjoying a considerable fad at the present. Many people have grasped this as the solution to their information problems. They have been encouraged by the substantial success that some fairly restricted and specialized information systems have had. A number of businesses, for example, have enjoyed some success in pulling their management records into a compatible and useful information system. The impression is widespread that bringing machine records together into some kind of central file will be instrumental in resolving the data problems of the broad class of users who attempt to use the files of the Federal Government.

This notion is supported by a general misunderstanding about the character of the files and their use. Those supporting this view are impressed with the fact that many different users have intersecting requirements for the same sets of data produced by the Federal statistical system. These records are, accordingly, viewed as general-purpose files. The convenience and economies of scale of bringing these records together into a common repository seem obvious. The obstacles to effective use under the present system are interpreted as technical and bureaucratic limitation amenable to this kind of technical solution.

What is not often adequately appreciated is the fact that general-purpose data are always used to fill special-purpose needs. This means that, while there are many intersecting interests in the same files, the impact on the file of each use may be quite different in terms of the organization, the levels of disaggregation required, and, most importantly, in the way the file needs to be associated with other records. It is this need for record association that is paramount and the source of most of the difficulty as was represented above. What makes a record a general-purpose record is not the fact that many users have an interest in its dimensions. It rests upon the file being constructed on the basis of standards, maintained in effective condition, and serviced by institutional arrangements and a technical system capability that will allow it to be reprocessed and combined successfully with other records in a wide variety of ways that will meet the special requirements of a wide range of users.

Thus, the key to solving these problems does not reside in the assembly of the records in a center but in the capacity to provide certain forms of file management and utilization services to the user. The effective provision of these services may require the assembly of some of these records into an integrated file, but this is defined by technical system requirements and is not the central issue it is made to be by many representations. It is important to characterize such a program as a data service center. The proposal is too important and fundamental to be burdened with its association with the naive data bank concept.

The Ruggles committee explicitly formulated at least a part of this rationale in their report and, hence, were putting forward a constructive proposal worth serious consideration. However, the committee never made explicit the way in which the problems of file compatibility rest upon the collecting and tabulating procedures of the agencies. It needs to be emphasized that these are important dimensions of the problem. Extending the mission of the Federal statistical system to provide user servicing capabilities based upon the new technology can do a great deal to extend the utility of existing records. However, the logic of a flexible service capability rests upon the ability to manipulate statistical building blocks. The development of these building blocks is a production task not contemplated in the suggestions for a data center. Some modification of current production practice will be essential for success. Indeed, if this problem is not tackled on a broad front, the generation of the servicing capability will fail to provide the kind of service intended and aggravate the sources of friction and dissatisfaction vis-a-vis the producing agencies.