

There are a number of characteristics of existing programs and procedures that stand in the way of an effective association of numerical records for purpose of analysis.

(1) Important historical records may be lost because of the absence of consistent policy and procedure for establishing and maintaining archives.

(2) The absence of appropriate standards and procedures for file maintenance and documentation lead to low-quality files that contain many technical limitations to effective association of records.

(3) Many of the most useful records are produced as a byproduct of administrative or regulatory procedures by agencies that do not recognize a general purpose statistical service function as an important part of their mission.

(4) Record association requires a good deal of intelligence about the compatibility of records in several dimensions and the circumstances that condition their availability. There exists no organized reference capability for performing the kind of reference service essential for the Federal system or even within individual agencies.

(5) Production procedures for collecting, coding and tabulating data that were appropriate when developed now lead to several types of record incompatibility that block the kinds of record association in usage that is required by current policy problems and made possible by computer technique.

(6) There are identifiable gaps in existing data records that stand in the way of bringing together records of greatest relevance for today's problems. Some of these gaps are more apparent than real and reflect the effect of the other obstacles to effective record association.

(7) The structural problems of concern to today's policymakers and the effort to bypass problems of record incompatibility force the utilization of data at levels of disaggregation that place severe strains upon regulations restricting the disclosure of information about individual respondents. Technical possibilities for using the computer to bypass these disclosure constraints have not been generally developed and made available by the agencies.

(8) There are new possibilities for more efficient management of large-scale numerical files in terms of storage and retrieval; new possibilities for rearranging files in more useful form; new possibilities for retrieving in the form of maps, graphs, charts and other media in addition to the traditional tabular forms; new possibilities for building in disclosure controls and disclosure bypasses; new possibilities for matching records to assure compatibility. These potentialities require the expenditure of time and effort on system design and software development that few agencies can justify.

THE STAKES ARE HIGH

The stakes associated with even a partial resolution of these problems of file availability and compatibility are very high. This rests in part upon an unexploited joint demand for information and an information service capability. In the Federal domain alone large amounts of money are being ineffectively spent in an effort to deal with these problems. The amount of overlapping and resource waste is substantial. The stakes are also high because the improvement in the utility of the information base could have an unmeasurable but substantial effect upon the quality of public administration.

THE SOLUTIONS ARE NOT SIMPLE

The solution to these problems do not rest, as some think, in bringing a large number of tapes into a common repository. Nor does it rest upon the fact that many different uses impact upon the same data sets making them "general purpose" records susceptible to central management. General-purpose files are always put to special-purpose uses. What makes a record a general-purpose record is for it to be 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 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 solution to the problems will require program modification on a broad front involving all of the agencies as well as an emergent data service center.