

wants to identify a variety of characteristics of the firms and households that form his principal markets often faces the same problem. The State Governor who wants to evaluate elements of his own program or the problems of his State by comparing them with the problems and programs of other States faces the same problem.

While the general problem is one of associating numerical records in use, it is understood better if we can visualize it in terms of its problem subsets. There are a number of distinct obstacles to file association that need to be identified before we can talk about solutions and program options. Consider the following:

The archival problem

This is the problem that initially interested the Ruggles Committee and the door through which they entered to their concern with some the broader issues of file management. The problem arises from the fact that the statistical agencies are oriented primarily to producing data publications and often leave their records improperly documented for further proceeding and analysis. Worse still, useful records are sometimes destroyed. These things occur because the existing system has no standards for identifying the files significant for preservation or for essential levels of file documentation. It provides no financial or organizational mechanisms for their maintenance. The decisions about the significance of archives is left to functionaries with little knowledge of their value in use and who must allocate funds for their documentation and preservation in competition with agency missions defined by previous policy in more restrictive terms and considered primary by agency personnel.

This, obviously, constitutes a major obstacle to the association for records for anyone who needs to work with data with any significant historical dimension.

Problems of file maintenance

Closely associated with the archival problem are some of the more fundamental problems of file maintenance. The utility of a file and its capacity for association with other records rests on more than the existence of a tape and a document that identifies its content sufficiently well for the data to be retrieved. Many additional problems stem from the low quality of file maintenance.

A couple of the more gross and obvious defects are associated with the fact that there are still important records that do not exist in machine readable form. Amongst the files that do exist on tape, some are in a mixed binary mode and some in a decimal mode making data association impossible without expensive and time-consuming mode translation. This is often true even between records of the same agency.

More subtle defects in file maintenance are uncovered when the need for data association requires bringing the detailed data in to accord with summary or published data. Often in the rush to meet production goals, agencies have pushed work through the processing stages of screening, reconciliation, estimation and summary in great haste without correcting prior files whenever errors or discrepancies are found. For example, corrections made "at the summary level are not carried back into the micro-detail. Indeed in some instances corrections have been made only in the published data, leaving both the summary and the detailed machine records uncorrected. Occasionally, summary data may no longer be in machine records and must be recreated by reprocessing the detail files or by keypunching and processing the new records. In some surveys, standards for screening data for creditability may have been coarser than appropriate for other uses.

The urgency to release results may also leave a disarray from the viewpoint of good file record and format management required for the files to be reused. There may be no uniform position in the records for like data and duplication of the same data can occur. For archival purposes a uniform record for the same data is essential and elimination of duplication economical. Also, the tape records may contain excess information. Certain codes and indicators used in the initial processing have no meaning in the archival context and the files must be purged of the excess information. Files may have interspersed alphabetic information useful only in the narrow survey context and which add complexity to programming efforts when used in other contexts. The absence of clear identifiers as part of the tape must be corrected to facilitate use of such tapes.

The requirements for simple access as well as the association of records are often stymied by the limitations of standards and procedures for routine file maintenance. Mr. Mendelssohn, who was loaned to OSS by BLS to conduct a