

Of the \$3 to \$3.5 million needed to supply the data center over the 5-year period, about \$700,000 would be needed to bring data not now machine accessible into usable form. About \$500,000 would be needed to transfer punched card data to magnetic tape. And about \$1.5 million would be needed to reprocess data now on magnetic tape. Looked at another way, about 100 man-years of professional aid will be required to review and correct the records, develop computer systems and programs to process the data, and provide reference documents to show file contents to the serious scholar from both the substantive and machine processing points of view. Also, about 100 man-years of keypunching and a very rough estimate of 10,000 computer hours are indicated (table 2). In contrast, about 13 man-years of professional work and 1,200 hours of machine time would stock the center with half the total in a year.

The major resource requirement of the 5-year effort is for the Census Bureau were \$1¼ million is requested. Over half the amount reported for that Bureau, about \$700,000, is needed to bring the 25- and 5-percent population samples for 1960 to acceptable levels. On the other hand, the Internal Revenue Service and the Bureau of Old Age and Survivors Insurance, both among the giants of data processing, have requested relatively limited amounts \$17,400 and \$14,300, respectively.

In considering the course of stocking the center several key factors should be kept in mind. First, the cost of additional historical reels after the initial storage of 9,000 is quite high—about \$300 per reel, compared with about \$27. The cost per reel is considerably higher for some of the files. For example, the SEC Quarterly Financial data and the FRB Report of Condition of Insured Banks would cost \$20,000 and \$8,000 per reel, respectively. Obviously, a careful review and justification for high-cost files is needed before their improvement can be supported. Second, the costs indicated in this report refer only to those needed to make data accessible within the responsible agency. I am assuming that the proposed data center would defray the costs of tape copying and would supply its own blank reels. Such costs are not inconsiderable. The 9,000 reels which could be made ready in about a year would cost the center over a half million dollars for blank reels and for copying.

GENERAL COMMENTS

I have the general impression that the larger the volume of data and the higher the frequency of processing the greater is the tendency for the files to be in acceptable order. That is to say, large files like those of the IRS and BOASI and the high-frequency operations in the BLS manpower field and in the FRB are in good shape while some decennial and annual operations at the Census Bureau and the relatively small files at SEC, FTC, OBE, and the Office of Education are either not well maintained with the computer or are not well mechanized at all. In other words, the degree of accommodation to the computer seems to be a function of the work pressures to use it.

The vast majority of available information is already in machine form. The small amounts of significant data not machinable are found in OBE, the Office of Education, and in the Department of Agriculture. The files in OBE and OE are not large, would total about 200 reels, in my estimation, and are not in machine form in appreciable amounts. As you know, efforts are underway to correct this in both agencies. About 25 percent of the Agriculture data are now machinable, according to Department representatives. It was asserted that the remaining 75 percent are significant and useful data and should be available to a data bank for research in agricultural economics.

AGENCY COMMENTS

Three general comments tended to be made by agency representatives. First, and least frequent, assertions that the Budget Bureau survey and the Ruggles Committee report has brought an increased awareness of the need for more effective file maintenance and that efforts to achieve this end would be incorporated in ongoing operations where feasible. These good intentions should be supported with funds where appropriate and the dilemma in which the Census Bureau finds itself with respect to the 1960 population samples should not be allowed to be repeated. I have no doubt that the \$700,000 now required would have been far less if the job had been done as a part of the 1960 census work.

Once the records have been brought to acceptable levels through new financial support, I doubt that programs operated at high frequencies will require more