

back from the initial phases of such an operation cannot be used to alter and improve later phases. The final tables are essentially cross-tabulations at a highly aggregated level. If the statistical work is repeated periodically, a time series results.

Attempts were seldom made to match a reporting unit in one period to the same reporting unit in later periods. Comparability over time was seriously impaired, since the tabulations reflected differences in classification procedures, definitions, and coverage.

Matching of data from one set of information to another was really not feasible. The Census Bureau points out that even with the population census arranged in the best possible form, the cost of matching information for a specific individual, to be used as proof of citizenship, was \$4 to \$5 per match, and that the match was unsuccessful on average 15 percent of the time. Matching of different sets of data either over time or between Government agencies was seldom even attempted. When a Government agency wished to use information which was contained in existing sets of data processed for another purpose, it was generally cheaper to completely redo the work rather than try to use the existing data.

From the point of view of the social scientist, the increasing flood of information has been both welcome and disheartening. Data is as important to the social scientist as laboratories are to the scientist or libraries to the humanist. The social scientist has become increasingly aware that progress in his discipline is closely tied to his ability to analyze, explain, and understand the empirical information on the behavior of the economic and social system. But the information available has tended to swamp empirical research workers. Until recently the cost of data handling and processing on any significant scale has been prohibitive.

The lack of consistency between sets of data and the lack of comparability of classification systems further complicated the work. Even where suitably disaggregated data existed, the individual scholar rarely could gain access to it. As a result, the economist has generally taken refuge in macroeconomic data such as the national accounts which are manageable and are presumably comparable and consistent. Unfortunately, the use of macroeconomic models has methodological disadvantages and limitations. But until very recently the social scientist wishing to engage in empirical research had little alternative.

In view of these considerations, it does not seem relevant to criticize the adequacy of the data base. The basic limitation in the past has been the cost of processing and handling of data, which restricted users to a partial and fragmentary basis. Even the imperfect and incomplete statistical system yielded greater amounts of data than could be efficiently used. The problem was not a lack of data, but rather the inability to use efficiently all the pieces which did exist.

THE IMPACT OF THE COMPUTER

With the introduction of the computer, a new set of forces was set in motion. Starting in the 1950's, the Bureau of the Census pioneered in the use of electronic equipment for data processing. UNIVAC I, now in the Smithsonian Institution, was a monumental step forward, although it was only a modest beginning of what turned out to be a completely new technology.