

the author's report to the Bureau of the Budget and then move to a more general perspective of the issue. Because of the overriding importance of realistic time dimensions in the evaluation of this problem we also need to make a distinction between the short-run and the long-run and we shall progress, in our treatment, along this time path. (It will help to bear in mind that the author's concept of the short-run implies something like ten to fifteen years.) The treatment of this issue in the press and in public hearings has confused the particular and the general and the short-run and the long-run. We will begin by reviewing briefly what the "Dunn report" does and does not say. First, contrary to reports, it does not constitute a formal plan in any sense. It is primarily an informal review of certain problems and prospects associated with the management and organization of statistical information generated by public, general-purpose statistical programs within the federal government. It was a preliminary review carried out with limited time and resources. The conclusions of this review were circulated within various administrative agencies as a basis for discussion and taken for evaluation by the administration.

The report did state that certain obstacles to effective use of federal statistical files might require some centralization of function. It did not at any point recommend what changes should be made or what agencies or files should be involved. It contented itself with a generalized treatment of the problem and some indication of the general direction in which the solutions might lie. It was not presented as a final program design.

Turn now from what the report did not say that people think it said to what it did say that few people have noticed. First, it pointed out that the new technology has made possible a new kind of capability in servicing the requirements of public policy and public management for statistical information, and that this capability is an order-of-magnitude different from any capability we have had in the past. It identifies a major "pay-off" in improving the public and private decision process, and in supporting research aimed at improving our understanding of the social process. This emerging potential of statistical programs received short shrift in congressional hearings and was totally ignored by the press.

Consider the benefit side of the equation for the moment. This benefit is not to be identified in terms of the direct savings and costs for some given number of traditional requests or inquiries made upon statistical files. It derives from the fact that there is available a new dimension of information service for the decision process. The source of this benefit stems from the fact that the central problem of using statistical records for purposes of policy, management, or research has always been one of associating statistical records. No number conveys any information by itself. It acquires meaning and provides useful insight only when compared with other numbers. The policy maker wants to know the answer to questions such as this: "What proportion of the residents of Appalachia have incomes under \$3000 a year and how do their age, race, sex and educational characteristics differ from those with higher incomes both within and outside Appalachia." In this one typical question we can identify dozens of separate statistical attributes of interest. Categories of income, age, race, sex, education, geographical designation, etc. all have to be specified and *related in some way that is relevant for the problem at hand*. This often requires that attributes of different and separate statistical records be brought together to reveal the particular collection of characteristics of use in analysis. In the not too distant past this kind of record association could be done only if the attributes of interest existed in traditional publications and conventional tabulations in a form from which a statistical clerk could, without exorbitant expense, extract a new tabulation that met the program requirements. In the more recent past there has been a growth in the possibility of doing special tabulations from machine records where the attributes of interest happen to be found in compatible records—usually within the same agency or program.

Commonly, however except for a limited portion of the statistical uses of interest to policy, many policy decisions have to be made on the basis of very sketchy information because there has never existed the kind of capability that could fulfil these needs.

Until recently the principal reason for this has been the lack of technical capacity to manipulate large quantities of numbers fast enough and with sufficient economy to make it feasible to meet the information requirements of policy in a flexible way. That technical capacity now exists and is growing. This has tempted many people to the conclusion that we should computerize all the data we have in the back room so that we can match any number instantaneously with