and accessibility will substantialy increase the effective scope and utility of present files without a change in the size of the files or the attributes of the respondents they contain though additional resources might be required.

Second, although a significant part of the gap in information is a function of file incompatibility, there are also gaps which result from missing attributes. For example, some attributes of the population may appear only infrequently with the decennial census and be needed for intercensal periods for vital program planning and evaluation. Other attributes may not appear in any records under existing information programs. Some notable gaps appear in the fields of transportation and construction and in connection with some of the important welfare attributes of people.

The following observation is pertinent here: The system, as it currently operates, provides no authority or mechanism for the review of the statistical program and the allocation of its resources in the light of the most important changes in information requirements. The decisions about programs that determine the scope and quality and accessibility of the records are primarily made upon the basis of technical problems, cost considerations, respondent pressures, etc., that impact directly upon the production process of the individual program and agency. There is no systematic way for the requirements side of the problem to enter the decision process.

Potentialities for system development are being missed

The new technology is making feasible a number of possibilities for greatly improving the utility of existing records. 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, etc.

One aspect of the service potential inherent in the new technology deserves some elaboration. The association of records in analysis usually carries with it a computational burden. This may take the form, for example, of computing the ratio of two data sets or making seasonal adjustments or computing coefficients in the analysis of variance. All of these derived numbers form a latent set implicit in the original source data. The computational capacity of modern computers is such that computations of this kind can often be made as fast or faster than the tape can be passed through the machine. Once a system has been developed for providing such a service, the marginal cost of generating these numbers when the tapes are being passed for retrieval is close to zero. Adding this kind of system capability can have the effect of increasing the effective size of the files of the Federal statistical system tenfold with latent numbers involving some computation.

These potentialities require the expenditure of time and effort on system design and software accumulation that few agencies can justify in terms of their current program levels or even appreciate in terms of their existing individual program missions.

The problem of file fragmentation

This is not a problem that is conceptually distinct from the others. Instead, it intersects the entire problem set being discussed and forms a part of the explanation for some of these anomalies. Currently, files are being generated and managed by more than 20 different agencies. It is precisely this division of responsibility and framentation of resources that inhibits system development and generates many of the problems of file compatibility. But apart from the way this problem invades all of the others, it imposes additional constraints because of the procedural, bureaucratic, and sheer time and space restrictions upon file usage.

THE STAKES ARE HIGH

The stakes associated with even a partial resolution of these problems of file availability, accessibility, and compatibility are very high. This is apparent even if we restrict our view to the significant Federal programs mentioned above. There is manifest in these programs an impressive and unexploited joint demand for information. This jointness has two important aspects. Even where the attributes of the numerical files of importance to these agencies are disjoint, they require the same servicing capabilities in the statistical system in order to