

The selection should recognize the importance of data acquired in the administration of regulatory programs and welfare or benefit programs. A primary purpose of the Federal Statistical Data Center is the organization of information in such way as to permit the use of data from various sources in the same analysis. The byproduct information from many agencies must now be reconsidered to determine how these data can best be combined with those of other reporting systems to contribute to socioeconomic analysis. Both program data (amount of grant, number of grantees, geographic location of program elements, etc.) and information about applicants can enrich the existing store of socioeconomic statistics from major statistical agencies.

The above paragraph is addressed to only one of the general principles discussed in an earlier section. Issues of inclusion or exclusion of various classes of data must be reviewed with all four general principles in mind and also with appreciation of the remaining items on this list itself.

2. Methodology and principles must be developed for the conversion of present files and production data to suitable archive form and quality, and their maintenance in proper form and quality, supported by codebooks, manuals, etc. It is not to be presumed that complete and consistent records will arise routinely from the collection, adjustment, analysis, and publication of data. Resources must be made available and priorities assigned. Above all, standards of form and quality must be prescribed and checked in some regular manner. Procedures must also be prescribed for work with the archive collection to prevent loss or contamination of the master records by tape erasure, statistical adjustment, aggregation or reclassification.

(a) The principles established for archives must presume that data will be transferred from the collecting agency to the Center in full available detail. The decisions about the lowest level of detail—other than the separate record for each respondent or analytical unit—will often be arbitrary. They will reflect notions of the finest detail that analytical purposes are likely to demand. In principle, there is no such ultimate disaggregation for many reporting units. For example, the use of the establishment as the reporting unit in many standard statistical systems is frequently dictated by the inability to define or obtain information for subestablishment entities; the choice is not based upon satisfaction with the level of detail obtained. The term "full available detail" must be read with a rule of reason.

(b) The principles established for archives must provide for the systematic completion of all corrections and adjustments to all data affected and all levels of detail, resulting in a fully reconciled and consistent body of data.

(c) The principles established for archives must provide for data to be transferred to the Center without screening for confidentiality; all confidentiality audits and checks would be applied to the formats and information content of the output of the Center.

(d) The principles established for archives must provide for the preservation of the identity of the reporting unit and the association of all information about the unit without regard to the agency or manner of the information; the rules and the economics of matching of existing records will be very different from those intended for future production of data.

3. Standards must be developed for definition, coding, classification, and aggregation with the intent of maximizing the ability to use different kinds of data in the same analysis, and of minimizing the loss of information. The reasons for this requirement are set forth in an earlier section of this report. A large number of standards would have to be developed, beginning with such seemingly elementary concepts as a household, a structure (a building), a business organization, an establishment (industrial), a populated place, a county; and proceeding to a school pupil, a hospital day, a recipient of (some particular) welfare service, etc. Some such standards now exist, although they are neither wholly satisfactory nor uniformly observed. These existing standards should be reexamined and many new standards developed. All standards need to be more rigorously defined and more effectively enforced. Ideally, a close matching in many dimensions of classifications, hierarchical aggregation, timing, and spatial extent should apply to all information in the basic record; but a system of practical compromises would unquestionably have to be accepted.

4. Standards must be developed for formats in order to facilitate the management, housekeeping, and retrieval of records and to avoid the loss of information. This item presumes that the characteristics of the information have, in principle, been defined. In practice, formats greatly influence the effectiveness, economy, and error rates of the whole operation.