14 April 1972, Volume 176, Number 4031

SCIENCE

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

EDITORIAL CORRESPONDENCE: 1515 Massachusetts Ave., NW, Washington, D.C. 20005. Phones: (Area code 202) Central office: 467-4350; Book Reviews: 467-4367; Business Office: 467-4411; Circulation: 467-4417; Guide to Scientific Instruments: 467-4480; News and Comment: 467-4430; Reprints and Permissions: 467-4483; Research News: 467-4321, Reviewing: 467-4440. Cable: Advancesci, Washington. Copies of "Instructions for Contributors" can be obtained from the editorial office. See also page xv, *Science*, 24 December 1971. ADVERTISING CORRESPONDENCE: Room 1740, 11 W. 42 St., New York, N.Y. 10036. Phone: 212-PE-6-1858.

Who Should Ultimately Own the Data?

I have just been informed that, in the next meeting of a National Institute of Mental Health review committee on which I serve, 31 research proposals will be assessed. If the committee has a good weekend, we shall allot $1 million or more of public funds. Scores of other such committees, within federal, state, and city agencies, will award much larger amounts. A good part of the tax monies involved will end up paying for the collection or production of data. The question is what happens to this data, and what ought to happen to it, once the original researchers lose interest.

Often the data rots, disintegrates, or is otherwise lost or inaccessible. Decks, tapes, and records are stored in attics or basements at home or on the campus. In some cases, the bewildered widow of the deceased researcher asks his colleagues to "clear away the junk"; in others, it is thrown away with yellowing blue books, early drafts of manuscripts, and books that came apart when the professors moved, as they frequently do.

Data stored on the campus often fares little better. Data decks warp and get moldy; tapes may erase. Most important, code books are not kept with the material or are not updated as the decks and tapes are changed, making the data unusable. Records, unprotected, turn into dust.

More than the original investment is lost. Often the data cannot be duplicated (for example, public opinion polls of views people held in the 1940's). Continuity is undermined, since one researcher cannot build on the efforts of earlier researchers. Also, as the published portions of the studies often do not contain sufficient data to allow for replication, the mutual check by colleagues on the validity and quality of the research findings suffers.

I therefore suggest that government agencies and foundations which finance mass collection or preparation of quantitative data should require that the data generated will eventually be made available to other researchers, by depositing a copy of the data (tape, decks, and so on) in a data bank or library. It would be up to the recipients of the funds to decide when they no longer need to keep the data exclusively for their use (that is, when they have published all they hope to get out of it or their interest has moved elsewhere) and which data bank or library will benefit from their copies. It will also be up to each researcher to decide if his data is sensitive (for example, highly personal or political), in which case some omissions may be called for (for example, no personal identification of the subjects should be provided in medical data about venereal disease).

The data should be deposited in a way that renders it usable by others—by including code books, for example. Appropriate and specific budgetary provisions for preparing such copies (which are not costly) should be made on all grants and contracts.

If such measures are taken, perhaps initially only for those studies in which the investment in the data is $100,000 or higher, it must be expected that the public's investment in research will yield fruit that can be shared more widely, enjoyed for longer periods, and, not least, be of better quality.—Amitai Etzioni, *professor of sociology, Columbia University, Director, Center for Policy Research, 475 Riverside Drive, New York 10027*