

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Annals of Economic and Social Measurement, Volume 2, number 2

Volume Author/Editor: NBER

Volume Publisher:

Volume URL: <http://www.nber.org/books/aesm73-2>

Publication Date: April 1973

Chapter Title: Problems of Access: Some Comments

Chapter Author: Taeuber

Chapter URL: <http://www.nber.org/chapters/c9893>

Chapter pages in book: (p. 212 - 217)

PROBLEMS OF ACCESS: SOME COMMENTS

BY RICHARD C. TAEUBER*

This paper is limited to "some comments," since the subject is a broad one with too many ramifications to discuss fully in a brief paper. Nor do I feel that the general topic is one that lends itself to a definitive discussion in a single session. Thus, it is hoped that the dialogue begun today will be a continuing concern of the Conference on the Computer in Economic and Social Research, and any other organization which uses the Current Population Survey or other microdata sets.¹

INTRODUCTION

It is gratifying that the Bureau of the Census is continuing efforts to release microdata in machine-readable form, including the Current Population Survey (CPS). There is now the promise of a standard releasable form of the CPS data. At some point in time the community of users will have access to a standard data file which will permit research efforts to be compatible across studies as well as across time.

Nonetheless, the Census Bureau could have done more, and it is hoped that it will take major and continuing steps to improve data release and data use programs in the future.² In addition, the users of the microdata and machine-readable summary data should have more input into the future data programs of the Bureau, for it is only from the users of data that the real pressures for improvement in data series and data programs will come.

SOME COMMENTS ON PRIVACY AND CONFIDENTIALITY

Despite the legitimacy of popular concern over the loss of individual privacy, the risk of breaches of file confidentiality, and the threat of increased civilian alienation, there can be little question about the importance of most information activities conducted by the federal government. And certainly, without the CPS or

* Chief, Urban Research Section, Health Physics Division, Oak Ridge National Laboratory. The research done by the Section is sponsored by the Department of Housing and Urban Development and the U.S. Atomic Energy Commission under interagency agreement. The Laboratory is operated by Union Carbide Corporation for the U.S. Atomic Energy Commission.

¹ Additional disclaimers: First, the comments and views presented in this paper are strictly personal; they do not necessarily reflect the view of any organization with which I am or have been affiliated. Secondly, these views are from a sympathetic, friendly feeling for the Bureau of the Census, with a concern for future improvement rather than a criticism of the past - which we cannot change anyway. Third, the Census Bureau is not being singled out, for many of the comments herein cross a broader spectrum of microdata than just the CPS.

² The concept of a user and use support mission is something that has not been formally and officially acknowledged. But regardless of the Bureau's interests in such a formal mission, if it is to be, then it needs support and dollars at the Social and Economics Statistics Administration, Department of Commerce, and (especially) Office of Management and Budget levels. Such a formal mission acknowledgement would also permit intra-Bureau recognition and equitable priority access to available resources: personnel and computer time.

other Census information activities, many citizens simply would not be provided with many of the benefits of life in our organized society. Rather than contributing to the loss of individuality, the computer and numerous data-gathering activities associated with federal programs actually are preserving citizen autonomy by providing the necessary informational base to help assure each person his rights and privileges under governmental programs. In addition, without the type of microdata that is now generated by many federal activities, it would be impossible for the government to ascertain the needs of various segments of the population and allocate the nation's resources accordingly.

I. P. Fellegi, Director General, Methodology and Systems Branch, Statistics Canada, has a recent paper which discusses confidentiality problems with summary data, and the subsequent problems of defining cells so as not to permit or provide statistical, inadvertent, direct, or residual disclosure of data. Talking about samples of unit records, Fellegi states:

"Even though the release of Census data for a sample of individuals may, in a rigorous interpretation of the concept, be disclosure, it can be argued that the probable pay-off to anyone looking for information about a particular person is sufficiently small, while at the same time the benefit to users of such tapes (and, indirectly, to society) is sufficiently large that the cost-benefit ratio to society is highly favorable. Obviously, pragmatic considerations must be taken into account." (p. 7)

With continuing increases in hardware and software sophistication and capability, I would add to Fellegi's pragmatism that we cannot rest on yesterday's decisions without today's review of those decisions and consequent operational procedures. The definitions and interpretations of privacy, confidentiality, disclosure, and similar concepts are largely relative, and thus need continuing evaluation in the light of current and anticipated capabilities and needs.

One of "yesterday's" decisions regarding confidentiality which I would like to see be reviewed is the practice by the Bureau of the Census (under the existing or earlier versions of Title 13 of the United States Code) that no geographic area may be identified which has fewer than 250,000 people. Whichever version of the derivation of the 250,000 rule is true, and however eminently reasonable it was at the time it was chosen, I would like to suggest that the time may be appropriate to review the whole question of research needs for data from smaller areas within the absolute mandate for protecting confidentiality of individual responses. In addition to the 250,000 rule, it would seem that another area of investigation would be a possible tradeoff between geographic grossness and a recoding scheme which grouped or censored details of the distribution of the responses to individual questions.

METHODS OF ACCESS

The Federal Reports Act of 1942, Section 2, states, "It is hereby declared to be the policy of the Congress... that information collected and tabulated by any federal agency as far as is expedient be tabulated in a manner to maximize the usefulness of the information to other federal agencies and to the public." A variety of user-oriented issues are raised by this statement of general principles:

the subsequent development of the computer and the capabilities it provides the researcher for the manipulation and analysis of large amounts of data, as well as the Freedom of Information Act in 1967, compound our problems.

The quote from the Federal Reports Act mentions usefulness "to the public." In interpreting this phrase, must it be interpreted as the lay public in general or is one permitted to partition "the public" into various categories of users or various levels of sophistication of users of the information? Quite obviously, the public at large has very different requirements on information than do those researchers concerned with extracting information from federally published data or federally published microdata. And if we go into research usefulness then we, as a research community, are increasingly facing the problem of the affluence of the host institution or what might be called the problem of Large Science. Within the physical sciences the instrumentation required for research explorations often is so massive that only the very few affluent institutions can conduct meaningful research. The move towards computerization and the availability of large summary and unit record files is raising the specter of Large Science on the horizon of social science research. Care must be taken to insure that the less affluent users, e.g., staff members of smaller colleges and graduate students, are not discriminated against because of the lack of access to computing power or the lack of funds to access and process the data.

Should there be responsibility or obligation somewhere within the data acquisition establishments to further or broaden the uses of the data which are collected? A "yes" answer admittedly calls for a change in mission statement, for the acquisition of most data is justified for a specific purpose or a single program. Whatever the answer to the posed question, though, uses of most data series are broadening, as we have seen in the papers presented in this volume. And whatever the answer to the question of obligations or responsibility on the acquisition agency, it would seem desirable that an agency of the federal government be charged with direct responsibility for monitoring actual and potential research. If the social payoff proves to be substantial, the necessary support offices can be established and the necessary marginal budgetary amounts can be authorized. The CPS is a perfect example since, as described in other papers, its funding is for the narrowly defined purpose of measuring unemployment, but its uses impact on several other program or policy formulation areas. Certainly, it makes little sense to spend millions of dollars on a program of data acquisition and then hold back on a small incremental expenditure which would complete a totally useful data publication program, be it print or magnetic tape. Economists and other social scientists could press for such publication programs, but insuring such completeness should be a function of an office within the federal budgetary and program control processes.

Should research centers functioning with federal funds and currently manipulating files such as the CPS be put under some obligation to function in the nature of a service center? Such centers would be expected to serve a community of users affiliated with them—probably academic in the main, but not necessarily so limited. Such service might be expected to be on a marginal cost basis, in that the data acquisition and file manipulation and analysis program costs would have been funded by the initial projects, and the other, smaller research would not be

expected, necessarily, to bear a portion of those costs. Should the Government, or the Bureau of the Census in particular, foster the establishment of such centers which would permit an interplay of resident and visiting scholars and which would have the capability to service requests for processing runs for members of consortia?

Should federally funded research projects which develop computer programs for information retrieval, data extraction or basic analysis of the CPS mandate the expenditure of resources necessary to provide full documentation of the programs so that they will have at least major elements of exportability? Certainly such efforts cost, but the marginal costs of exportability are frequently much less than those of recreating the wheel.

If problems of utilizing the information contained in the output of the standard releasable version of the CPS exist, it would seem incumbent upon the federal government, in the form of the Bureau of the Census, to take the lead in coordinating the information exchange necessary to minimize these problems rather than leaving that solely to independent groups such as CCESR. Much has been done at the Bureau in the way of supporting data users, and the CPS operation has maintained a fairly close relationship with its sophisticated users. The latter is desirable, for it is through this interaction of the producing agency and its sophisticated users that optimal progression towards a fully debugged and completely documented research tool can proceed as well as produce results of optimal usefulness. The Bureau has also established a Data Users Service Organization (DUSO) primarily for the decennial Census of Population and Housing. However, it is giving signs of broadening into the total spectrum of census products. In moving beyond the 1970 decennial Census to the support of data operations such as the CPS questions can be raised. For example, will they include in the CPS service the resources necessary to serve general users of the file, in terms of providing very basic documentation and information support, even though the community of users will be much smaller than that of the decennial Census? Whatever level of user is supported, one can also ask if the Bureau will face the general question of how to further the dialogue among actual and potential users? To aid this, they could sponsor or co-sponsor an annual meeting of users as well as co-sponsor application area workshops or symposia, both in a manner similar to that of the Social Security Administration with regard to the actual and potential users of the Continuous Work History Samples.

As specific questions on access, one could ask:

- Should the Bureau pay for a flexible report generator for use with the CPS tapes, which would be made available at reproduction cost (fully documented)?
- What should be the responsibilities, or the obligations, of the Bureau as to machine-readable encoding of all acquired data elements, rather than the encoding of just those elements on the basic survey or census instrument which are needed for immediate purposes?³

³ This is not necessarily a criticism of the Bureau, but I do know of a case at another agency where a decision not to punch date of birth, which was acquired, prohibited a key portion of a major research study because age, although a mandatory informational input, was thus not available to the research analyst. Having been authorized to expend public resources for the acquisition of data elements, who should have the authority to "hide" any of these data from other or future users?

- What obligations will the Bureau assume to make the data tapes available in specifications which meet the processing requirements of individual users in terms of number of tracks, densities and coding schemes? Will ASCII coding be offered as an option since it is suggested federal standard?
- Does the Bureau have, or is it giving consideration to, tape maintenance and rewriting procedures to insure readability or guard against storage deterioration? It is not infrequent that a five-year-old tape is removed from a tape vault or library and is found to be unreadable.

And as a final issue on access to the CPS data, I urge the Bureau to make plans to release data from earlier years in the same standard format so as to enhance the usefulness of the CPS for trend or other studies over time. It has been mentioned that data prior to 1967 are now lost as far as such a publication program is concerned. However, I hope that all data subsequent to 1967 can be put into the standard releasable format and thus made available to researchers.

EPILOGUE

We should conclude on a congratulatory note, since the standard releasable version is a major, and much needed, move to further the dissemination of the informational content of the CPS, and to increase the value to society of the total operation. Nevertheless, the Department of Commerce and the Office of Management and Budget must still provide the marginal budgetary support necessary to implement this new aspect of the CPS program fully and effectively. It makes little sense to spend the substantial amounts of money necessary to acquire data, and then not spend the small additional amounts necessary for a full dissemination and publication program. An end objective of any statistical program should be getting the information content of the data into the hands of the research and program users; we should not have acquisition for the sake of acquisition, nor solely for a narrow program justification, when the data clearly have much broader potential in providing information for research and policy analysis.

Oak Ridge National Laboratory

BIBLIOGRAPHY

- Bureau of the Budget, Circular No. A-86, "Standardization of data elements and codes in data systems," September 30, 1967.
- Bureau of the Budget, "Principal laws relating to the Bureau of the Budget with amendments through December 30, 1963, the end of the First Session, 88th Congress."
- Bureau of the Census, "Concepts and methods used in manpower statistics from the Current Population Survey," BLS Report No. 313, Current Population Reports, Series P-23, No. 22, June 1967.
- Bureau of the Census, "Current Population Survey, annual March demographic supplement file, survey years, 1968-1970, preliminary data," internal preliminary draft, 1972.
- Bureau of the Census, "The Current Population Survey: a report on methodology, Technical Paper No. 7," U.S. Government Printing Office, Washington, D.C., 1963.
- Bureau of the Census, Title 13, United States Code - Census.
- COSATI Panel on the Legal Aspects of Information Systems, "The right of privacy," draft panel report, April 21, 1971.
- Fellegi, I. P., "On the question of statistical confidentiality," *Journal of the American Statistical Association*, March 1972, Vol. 67, No. 337.
- House of Representatives, 92nd Congress, HR 14153, "To Amend Title 13, United States Code."

- Martin, Margaret E., "The Current Population Survey as a statistical information system," draft for presentation to American Statistical Association Meetings, December 1970.
- Richardson, Elliot L., Secretary of Health, Education and Welfare, Statement for the Subcommittee on Constitutional Rights, of the Committee of the Judiciary, United States Senate, 92nd Congress, First Session, at hearings on Computers, Data Banks and the Bill of Rights, March 15, 1971.
- Schultze, Charles L., "Governmental and public data needs," unpublished paper, 1969.
- Tauber, Richard C., "Research usage of the 1970 Census information," Statistical Policy Division, OMB, internal memorandum dated September 8, 1970.
- "CPS data dissemination issues," Statistical Policy Division, OMB, internal memorandum dated November 10, 1970.