COMMUNICATIONS

THE 1970 CENSUS START COMMUNITY*

BY JOHN BERESFORD

The 1970 census data base consists of two basic products: summary totals for areas and samples of individual records. Both of these are available on computer tape. Although there will be the traditional printed reports, it is the products on tape that are of interest to most researchers. These products will be the inputs to many social science projects in the decade of the seventies; problems in their use point to the value of cooperative efforts in which solutions are made available to users. There are certain technical goals shared by virtually all users of the 1970 Census data base. For example: the use should cost as little as possible, the use should minimize the time taken from the user's basic professional activities, the use should require as little calendar time as possible, the use should not duplicate something done by others, and the use should take advantage of the best procedures and techniques currently available. We believe that these goals relating the technical and mechanical aspects of the 1970 census use provide a basis for the cooperative development of systems for the census data base.

The community formed by National DUALabs is called the 1970 Census Summary Tape Assistance, Research, and Training Community (or START Community). The START community members are selected from about 50 universities, research organizations, governments, and commercial groups who share a need to understand and apply the census data base to their professional activities. They have convinced those who control budget decisions at their organizations that it is less expensive and more efficient to participate in a cooperative census data venture than to create a new and independent census data-use capability. The members have pooled their resources and commissioned DUALabs to develop techniques for summary tape and public use sample data acquisition, basic computer programs for summary tape and public use sample processing, training programs in data use, documentation, technical assistance, and a clearinghouse function which also distributes useful programs, data sets, or other items prepared by one member which he wishes to make available to all members.

To date, the joint-effort has resulted in a system of about 30 inter-related computer programs for summary tape processing which is installed and operating at about 40 sites on a variety of equipment. The START Community benefits from procedures for compression of the summary tape files as released by the Census Bureau. The First Count files, for example, were compressed from 163 to 15 reels, reducing the cost of acquisition from $9,780 to $675. The cost of acquiring the first five census counts can be reduced from over $100,000 to about $10,000 for the user who acquires the files in compressed form. The community shares in the exchange of information about the data base through regular bulletins and newsletters, and the community members have received considerable personal and

group training. The minimum reduction in net cost to the START Community member (software and data acquisition alone) is estimated at about $100,000 for those whose data needs are minimal up to $200,000 plus for the user of all census data. The time savings are difficult to estimate but the typical community member is producing reports and analyses of census data in less than a month from the day he decides to become a member (he can be in production in one day), whereas non-community members have reported lags of six months to two years from the day a decision is made to work with the data base. Community members have begun to make available to their fellows the results of their special programming efforts, and important data sets (such as 1960 summary tapes) are available for processing by the START software.

The installation of the START community programs is very simple. The programs are operating under a variety of conditions and user experience has been accumulated. This experience is being monitored to uncover problems and find out about the applications being made. These applications are being catalogued so that users can avoid doing jobs that have already been done. For example, creating a Fourth Count File of county records need be done only once.

The START community programs represent a common set of tools which can be shared by professors and students at many universities. Since the tools and techniques are available at many sites, a scholar can move from one research site to another and find that no time need be lost in communicating to a computer center staff the mechanics of census file handling. More important, by using these tools he finds that he can focus his efforts on research problems, with reasonable confidence that data base use problems have been solved. To date, the vast majority of "software problems" have been, in reality, problems in understanding the census data base or geographic files, and these problems are dealt with by individual training in census data use. Users with diverse backgrounds and objectives have applied the programs with considerable success to the analysis of small-area statistics across the nation.

The START Community cooperative effort has created a mechanism for cross-fertilization among various social science fields for technical and mechanical ideas. Ideas relating to measurements are beginning to emerge as the use of the 1970 $\frac{1}{100}$ public use samples increases and experience accumulates. The desire for cooperation is evidenced by the Microdata Workshop sponsored by NBER and by the formation of the *Annals of Economic and Social Measurement*. The basis for the cooperation is the shared acceptance of the common cost/benefit goals. Thus, the focus of START is on facilitating data base use : the complexity and universality of the census data base makes that resource a natural subject for DUALabs' experiment in cooperative social science research.[1]

National Data Use and Access Laboratories

---

[1] For further information, contact DUALabs, Suite 916 1411 Jefferson Davis Highway, Arlington, Virginia 22202; phone (703) 920–8300.