

Session 9

Providing Adequate Technical Support

Confessions Of A Survey Guy
Stephen Dienstfrey
Schulman, Ronca, & Bucuvalas, Inc.

I'm a survey guy. I learned how to do surveys during my graduate studies in Sociology. It seemed I had a knack for statistics and methodology. A professor asked me to help him with a study he was doing of residents of public housing. This led to a job with a political polling firm, to a short stint at the Bureau of the Census working on a survey of Inmates of Local Jails, to AMTRAK doing consumer research, to the Energy Information Administration doing a survey of Industrial Energy Consumption, to the Department of Veterans Affairs where I was responsible for surveys of the veteran populations, and for the last 5 ½ years I have been back in the private sector.

I just covered 35 years of my professional life – and until I worked for VA documentation was never a problem: either no one ever asked me a question about a data set I put together, or I wasn't there long enough for people to ask me a question.

Let me step back a minute. I know meta data is more than documentation about data files – it's archival material, it's bibliographic material, it is a whole host of other types of data about data in a format that can be searched using common terms and I know that there are a lot of people that are working on developing standards to standardize meta data formats. But I'm a survey guy, and to me it is documentation.

(And when did 'documentation' become 'meta data'? I can't decide if this is like doctors calling 'anti-histamines' 'histamine antagonists' or religious leaders calling 'sin' 'transgression'.)

So while I know we are talking about meta data, my experience is with documentation and I am going to confine my remarks to that specific area.

My first contact with the vagaries of documentation came while I was at VA doing some work with the 1980 Census Public Use Microdata files. VA had ordered some special tabulations of veterans from the Bureau. These tabulations were used as the basis for the re-calibration of the Veterans Population Model. The results of this recalibration had been distributed throughout VA and to states. However, the counts we ran from the PUMS tapes did not match the counts from the special tabs. After much trial and error we realized that the definition of "veteran" in the PUMS included individuals who were still on active duty. These individuals could easily be removed by using the current employment questions where those on active-duty were identified. This information may have been available in the PUMS documentation, but it seemed a little intimidating and no one bothered to read it.

One would have thought that this experience would have instructed me on the need for adequate documentation. But, alas, this was not the case.

The 1980 Census was the first time that women were asked if they had served in the military. This important group of veterans was further studied in the 1987 Survey of Female Veterans.

Prior to releasing copies of the data we prepared SAS and SPSS data files where all the variables and values were labeled. So much for documentation.

As the file was used outside VA people would call with questions. While everything was relatively fresh in my mind, it was not a problem. But as time went on and the calls became less frequent, it became more and more time consuming to run down the answers. At no time during this experience did I bother to keep notes about what was asked and what the answers were.

One would have thought that I learned from this experience, but a similar pattern took place with other VA studies.

More recently I have been more of a user of public data sets than a provider. We have downloaded files from various health related surveys as well as from the Bureau of the Census. Generally, the documentation is both voluminous and written for those who are initiated in the theology of the survey not those of us who are novices. Specifically, the variables have names that sometimes defy understanding, and computed variables are identified but the computations were not there. One data set had 6 different weights with the instructions to use the one that was most appropriate without giving the user who may not understand the subtle difference between “household” and “individual” weights.

So if we can agree that there is a need for complete and understandable documentation, how do we go about making sure that it happens? We could give the task to OMB, and while that might work for data produced by federal agencies and the key word here is “might”, I think it would be unwise to create yet another onerous bureaucratic requirement. Further, this would do nothing for data that was created outside of the federal umbrella.

Another possibility is work on people’s guilt. But this hasn’t worked in the past and I can see of no reason for it to work in the future.

It seems to me that if there is to be adequate documentation it needs to be a transparent by-product of the entire survey process from start to finish.

Using my current job as a model, let’s walk through the steps that take place when we conduct a study. From my experience in a variety of jobs, I think these are fairly standard.

- Develop the questionnaire.
- Program the questionnaires for Computer Assisted Telephone Interviewing (CATI) with range checks, automatic skips, internal codes, etc.
- Edit the data – code open ends, collapse data, etc. – to final data set.
- Create analytical variables by combining and/or recoding data from the previous steps during the reporting phase.

What I suggest is a software suite that takes the final questionnaire (along with any notes about specific questions or series of questions that can be carried along with the document) and feed that into a CATI program that is very close to final. As the CATI program is modified, notes are automatically kept on what was changed. As the data are edited, more notes are kept. The codes

used for open-ends are also attached. All of this is carried along to the data set similar to the information SPSS displays with the variable tab. The syntax used to create an analytical variable is also attached to the data set.

The result may not be the complete documentation that archivists might like, but it is a very good first step. More importantly, it might help the casual data user, such as myself, to get a good idea of what is going on. And when I have to call and ask for further help I sound less like a complete idiot.

But what do I know; I'm a survey guy.

Training and Staff at the U.S. Census Bureau to Create Metadata and to Provide Consultation with Users

Pat Doyle
U.S. Census Bureau¹

There is widespread corporate and project level support for documentation and user assistance across the Economic, Decennial, and Demographic Directorates of the Census Bureau. This paper focuses on the programs in the Demographic Directorate. It covers the formal and informal procedures for training staff to provide documentation and technical assistance, guidelines for preparing documentation for demographic programs, corporate level support for documentation preparation and tools to support documentation efforts, organizational support for staff to provide technical assistance, and methods of obtaining technical assistance. The paper concludes with a discussion of issues related to survey documentation and technical assistance and some thoughts for the future.

Staff Training

The Census Bureau strongly supports staff development and encourages staff to provide technical assistance. The formal training program consists of a 55-hour classroom training program spread over six weeks, which is targeted to new staff in the Demographic Surveys Division but available to all interested staff. This program covers all aspects of the design, development, administration, processing, and documentation of demographic surveys. Its intent is to give survey staff specializing in one area of the survey process an overview of the other aspects of the survey process and to introduce them to the various Census Bureau organizational units that play a role in the survey. This gives survey staff a broad perspective on the Census Bureau and on the range of expertise employed in conducting surveys. This will enable staff to effectively route specific questions from users and customers that require expertise in a particular part of the survey or its design and execution.

As part of this demographic survey training program, a three-hour session is devoted to documentation and to the importance of documentation to the wise use of Census Bureau data products and to the continued support for Census Bureau surveys from other agencies. This training program offers examples of how secondary data users and funding agencies rely on data documentation to carry out their work and illustrates both good and bad examples of documentation. It is an interactive, hands-on class intended to motivate the Census Bureau staff not only to develop good documentation but to provide users with methods of access to that information that is helpful and efficient.

The demographic directorate actively participates in the Joint Program in Survey Methodology (JPSM program) offered by the Universities of Maryland and Michigan and Westat. The Census Bureau cosponsor this program along with other federal agencies, and the demographic program encourages its staff to take advantage of the various seminars and classes offered through that

¹This paper reports the results of research undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope to that given to official Census Bureau publications, and is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

program. A number of staff have received masters degrees in survey methodology through that program with sponsorship from the Demographic Directorate, and many more have taken (and continue to take) courses and attend seminars.

Guidelines

The demographic directorate developed guidelines for documenting all its surveys, to govern the type and extent of documentation needed for the data products. These guidelines encourage the staff to prepare documentation at three-levels: summary, user, and technical. The summary level material is intended to provide a quick and easily accessible overview of the survey and its data products. The user-level information is a comprehensive set of documentation for individuals using the data products, written in a style that is informative but not overly technical and thus does not require the user to be an expert statistician or methodologist to comprehend its content. The technical information is the most comprehensive description of the survey and its processes but can be fairly dense and difficult for a novice to follow. It is devoted to individuals with a need to know on a particular point and to experts in various components of the survey process.

The general guidelines cover all aspects of the survey design, development, and execution of the survey. They focus on producing documentation for all users, both internal and external to the Census Bureau. They cover content of the data collection instruments as well as data products, procedures for data collection and data processing, and survey and data quality.

One specific product included in the overall documentation guidelines is the quality profile. The demographic area has separate guidelines devoted explicitly to the organization and routine production of quality profiles for demographic surveys. The emphasis in these guidelines is to focus the reader's attention on the relative quality of Census Bureau surveys within the survey industry, emphasizing the high response rates and other high quality features. The guidelines also stretch the scope of quality profiles to touch on areas not routinely included, such as the outcome of instrument pretesting and the field procedures used to minimize nonresponse. These guidelines (along with similar ones developed for the Economic Directorate) are in the pipeline to become formal standards.

Corporate Support and Tools

Traditionally, survey documentation has been prepared by technical staff who are not always skilled in the fine art of writing for a broad audience (like users of public use files). With the new documentation guidelines, the Demographic Surveys Division provides a technical writer to assist the projects in refining the presentation of technical materials, so that they meet the needs of secondary data users. Thus, we encourage staff to use their technical staff to write the technical-level documentation and to use the technical writer to prepare the summary-level material.

The Census Bureau recognizes that production of good documentation can be handled most efficiently if there are good support tools. In fact the agency has sponsored and continues to sponsor a number of documentation tools.

One of those is the instrument documentation system (IDOC), the first attempt to address the serious gap in the automation of complex survey questionnaires, which was (and is) the lack of interpretable instrument documentation. IDOC grew out of an effort by the Association of Public Data Users to focus attention on the impact this lack of documentation would have on intelligent use of the data. Its concept was broad, in that it used the notion of generating documentation directly from the instrument: storing it in a neutral format that could then be converted to a format like HTML, so that all users with a web browser could access it and use all the features of a web browser to search and excerpt from the documentation. The system includes hypertext links to facilitate following all the possible paths an instrument might take.

Another tool is the survey development and documentation system (SuDDS), for the development and documentation of instruments. This system relies on a Microsoft Access database system to house question- or item-level documentation and specification; and an interface to facilitate development of question text, fills, flows, universes and response categories. This will enable developers to more easily specify and develop automated instruments, track their evolution over time through the development, testing, and refining process, and generate instrument documentation in the form of items booklets. It will also enable the Census Bureau to integrate results from the survey (such as the frequency of responses to a particular item by type of response) with the documentation. Finally, its long-term goal is to support the specification of the post-collection processing system within the same system that documents the instrument.

Another tool supported by the Demographic Directorate is Data FERRETT, a metadata driven data access system and a component of the American Fact Finder (AFF). It is a data access tool that encompasses and relies on item-level documentation to make it easier for analysts to gain access to survey microdata either through quick small tables or big extracts. Data FERRETT is distinguished from the decennial census component of AFF by its emphasis on survey rather than Census data and by its facility to support preparation of extracts with accompanying data file documentation.

Finally, the Demographic Directorate actively supports the Data Documentation Initiative (DDI), an international effort to establish standards for formatting data documentation to enable data archives and the like to establish generic documentation access systems applicable to all of their holdings. The DDI standard exists for both microdata files and for tabular data stored in machine-readable form. As such, it promises to significantly reduce the cost and resources required to share increasing complex information derived from surveys.

Technical Assistance

The Census Bureau encourages staff to give user questions high priority. Contact information for appropriate staff is widely-available. The staff are organized in different ways, depending on what phase of the survey process or demographic program they support: survey-specific experts, subject area experts, and methodologists. Each group has a different set of skills and knowledge, a different organizational structure, and slightly different approach to training. However, all groups are required to provide assistance to those who need it.

The survey-specific experts make up the operational staff that manages surveys, supports their administration, and guides the production of files and documentation. This group includes survey statisticians and programmers who are organized by survey and, within survey, by function (instrument development and administration versus data processing). Technical questions on file development, access, format and documentation are often fielded by staff from this group.

Training of this staff is a mixture of classroom training (the 55-hour training program developed and administered by the operational division for demographic surveys), courses like those at JPSM, and on-the-job training. They take phone calls, emails, and other inquiries from users; answer questions they can; and refer other questions to those who can provide answers.

For Census-sponsored surveys (and some reimbursable surveys) and for other demographic programs, the Census Bureau has subject area experts who develop the survey or program content and scope, recommend questions to be administered, review the data, develop the approach to post collection processing, and produce Census Bureau reports and technical papers in their area of expertise. Individuals in these groups are content specialists, typically with degrees in one of the social sciences, and their responsibilities span multiple surveys. Questions about specific variables and how to analyze specific concepts are referred to this group. They are organized first by subject area into two divisions (one focused on housing and household economic information and the other on demographic information) and by content (rather than by survey) within those divisions. Their work extends beyond demographic surveys to include the decennial census and its related studies. As with the operational staff, the subject area staff take phone calls, emails, and other inquiries from users; answer questions they can; and refer other questions to those who can provide answers.

Training opportunities for this staff include the 55-hour training program developed for the operational staff, as well as on-the-job training, Census Bureau seminars (by Census Bureau and non-Census Bureau researchers), and participation in external conferences and seminars—both as presenters and attendees.

To support state-of-the-art survey and sample design and analytic techniques, the Census Bureau has a number of methodologists specializing in a wide array of technical topics—including sample design, questionnaire design, and research methods. This group tends to be organized by the type of methodological work and by survey, but this is not exclusively the case.

Staff in this group provide assistance within the Census Bureau on the proper interpretation of statistical information in analysis and assist users in understanding the technical design features of the sample. As with the operational staff, they take phone calls, emails, and other inquiries from users; answer questions they can; and refer other questions to those who can provide answers. Some (particularly the sample design group) are actively involved in the preparation of documentation on surveys, routinely preparing the source and accuracy statements and periodically preparing quality profiles.

This staff participates in the 55-hour on the job training program, actively participates in the JPSM program, attends seminars and conferences, and presents technical papers on the procedures used by the Census Bureau.

Access to Technical Assistance

Users can get their questions answered by phone, through list serves, directly by E-mail or regular mail, through the internet, or through bulletin boards. There are also a variety of user groups, working groups, and advisory committees through which information can be obtained. Examples of contact information follow:

By Email:

Pop@census.gov (for domestic demographic questions).
Hhes-info@census.gov (for household and household economic questions).
Ipc@census.gov (for international demographic questions).
Cpshelp@info.census.gov (for questions on the Current Population Surveys).

Web Sites:

www.census.gov (general Census Bureau home page).
www.sipp.census.gov/sipp (information on the Survey of Income and Program Participation (SIPP)).
www.sipp.census.gov/spd/ (information on the Survey of Program Dynamics (SPD)).
www.census.gov/hhes/www/ahs.html (information on the American Housing Surveys).

List serves:

ACS Alert (for information on the American Community Survey (ACS)).

There are information groups within the subject area divisions whose primary job is to field questions on the demographic surveys (and other programs). They answer routine questions directly when they can and refer more technical questions to the appropriate expert. There are also some survey-specific groups within the Census Bureau to provide direct support to users of the survey. For example, the outreach and analysis branch within the Housing and Household Economic Statistics Division provides support for SIPP and SPD, and an area within the Demographic Surveys Division that advocates for the ACS, reaching out to users to determine their needs

The Demographic Directorate promotes and participates in the Census Bureau's user conferences which are heavily focused on the Decennial Census and related programs that produce a large amount of tabular information. General information about demographic surveys is included in these conferences and provided by the staff from the demographic directorate.

The directorate is also exploring the possibility of establishing user conferences for users of microdata files, which would be heavily focused on demographic surveys. The project is in the concept stage at the moment, and approval and implementation will depend on costs, resource requirements, and expected levels of participation.

Issues

Federal agencies in general are facing a number of difficult issues with documentation of demographic surveys, and the Census Bureau plans to continue its lead role in attacking these. The first issue, alluded to earlier, is the loss of good instrument documentation with automated surveys. Generally, a small or linear character-based automated instrument can be documented manually by compiling the image of the screens and describing the navigation possibilities. However, this is too resource-intensive for complex instruments and nearly impossible for instruments designed to take full advantage of graphical user interfaces (such as the Windows operating system). Two attempts have been made so far to address this problem (one of which is the IDOC system noted above), but neither of these has reached a level of maturity that yields an adequate substitute for the paper instrument that existed for surveys which were not automated.

A related issue is that the industry is not actively tackling the problem of the loss of good instrument documentation, leaving each survey project to fend for itself to compensate. This yields varying quality instrument documentation that is often incomplete in its explanation of all the possible wording variations and sequences of questions. The National Academy of Sciences has recently called attention to the problem (through a workshop sponsored by the Demographic Directorate and subsequent seminars) and encourages computer specialists to join the survey industry in tackling it but, as yet, no substantial source of funding has arisen to support these efforts.

The ever present issue is that there is never enough documentation for everyone, and there are never enough resources to do good comprehensive documentation—particularly for large complex surveys. Any pressure to save resources will affect documentation first, because the theory is that you can't document what you don't collect and we never have enough data to answer all of society's important questions. This is somewhat inconsistent with our goals, however, in that it is difficult to insist on quality uses of quality data if the documentation is inadequate. But that is a weak argument when there isn't enough money to collect all the data needed in the first place. It's also misguided in the sense that it is really an inefficient shift in the costs of using the data that are collected. If the data collector does not adequately document the data, the data user will have to do it him or herself. That is an inefficient use of resources, unless there is only one user and, hence, only one occurrence of the documentation effort.

To proceed to produce good documentation on a routine basis at minimal cost within the current budget environment requires a lifestyle change, accompanied by strong management to enforce that change. The lifestyle change is to view documentation as an integral part of the survey, that should be generated alongside the other survey processes rather than saved until the end. It is also to recognize that preparation of the documentation is most efficiently accomplished by a mix of technical staff and technical writers, rather than exclusively by one or the other. Finally, it is a recognition that developing and conforming to standards for documentation can support a more efficient and cost-effective documentation production system.

Another issue which may be unique to the Census Bureau is that we tend to have a different set of data accessible to staff than to external users. This is largely due to the disclosure avoidance measures applied to data files before they can be released. However, in some instances it is due

to a preference for using a more efficient file format for the data stored internally than for data distributed to a broad set of users. The different format and content make it difficult for the staff to answer technical questions that the users generate directly from observed anomalies in the data. The Demographic Directorate moved away from the use of the differently-formatted files, but it is unlikely to restrict its use to the disclosure-proved data because of the analytic restrictions imposed by the limited geography that is required for the disclosure avoidance.

One issue implicit in this paper is that the training provided to our staff does not include on-the-job training in the type of people skills that would make the technical assistance process go more smoothly. We hire good people and expect they can convey answers to technical questions in a manner that users can appreciate and understand. For the most part, we are lucky and this works.

The final issue to note is that expertise tends to reside with individuals rather than groups, so that the absence of one individual can create a void in the information available for users. More cross-training of topics and cross-fertilization of staff would be very helpful. The Census Bureau has begun to recognize the importance of this cross-fertilization in terms of staff development and succession planning, and any efforts that come to pass in that vein should benefit efforts to provide technical assistance.

The Future

The demand for documentation and technical assistance will continue to grow with the increasing complexity and sophistication of surveys and data collection programs. As a result, the federal statistical system (in general) and the Census Bureau (in particular) need to address the issues just noted. For its part, the Demographic Directorate intends to continue to support the development of documentation tools, continue to participate in the development of documentation standards, and hopefully will be able to implement a series of user conferences to assist users in the wise use of our products and to share information between us and the user community on uses of the data.

