This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

Contextualizing Social-Science Data

Tom W. Smith
NORC, University of Chicago

Human societies are complex, multi-faceted entities and social-science research designs need to measure that complexity. A key requirement is contextualization. Most people live in households which are nested in neighborhoods, communities, and countries. They are not isolated individuals, but interact with and are notably influenced by their families, neighborhoods, communities, nations, etc. Surveys need to collect information on each of these levels from the individual to the nation so the contexts in which people live are understood and that through multi-level analysis the impacts of these different levels can be examined. Such a geographically-contextualized dataset can be compiled by combining together information from surveys with geographically-linked data from other sources. Surveys should naturally collect information on respondents and their households and should augment the survey data with both observational and process paradata that are generated as part of the activity of conducting the survey. Then the survey-based data should be further enriched with data from other sources (e.g. US Census, economic databases, voting records, environmental readings, etc.) which are linked to the survey data via GIS and other geographically-coded data sources (Smith and Kim, forthcoming). Such aggregate-level data can notably enhance our understanding of individual-level attitudes and behaviors.

The geographically-contextualized data should also be augmented with other sociologically-contextualized data. The idea of sociological context is used broadly to cover social, political, economic, and related aspects of human society. A particularly powerful way of doing this is through hypernetwork sampling. A national sample of adults can be used to generate a sample of any activity or entity that is connected to such individuals. For example, the General Social Survey (GSS) has used hypernetwork sampling on three occasions to generate samples of and then to study employers, on two occasions to research congregations, and on one occasion to examine voluntary associations. Once a sample is generated, a follow-up survey is conducted of the units identified from the initial survey. Hypernetwork sampling both generates representative samples of the entity under investigation (e.g. employers, congregations, voluntary associations, discussion networks, social movements) and collects information about these entities that can be linked back to the respondents in the survey that originally created the sample. Thus, hypernetwork sampling sociologically contextualizes in two ways. First, it produces studies of important social units beyond individuals and households. Second, it links specific people to the other groups and institutions that they are members of and/or interact with. Their attitudes and behaviors can then be related to attributes of the groups they belong to or associate with. For example, one can study the role of employers from the unlinked hypernetwork sample and one can link people in the adult sample to their specific employer and examine how attributes of employers influence individuals.

Another valuable way to sociologically-contextualize data is through network analysis. Egocentric, social-network analysis has been employed on several GSSs (McPherson, Smith-Lovin, and Brasear, 2006). A criterion for defining a network tie (e.g. discussing important matters, friendships, socializing together) is selected and people are asked to name individuals who qualify. Then in turn one collects information about the characteristics of these alters, the nature of the interaction with and connection to them (and amongst them), and/or other relevant information. Social-network data can provide much information on how people live their lives including whether they are socially isolated or

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

well-connected, whether their networks are homogenous or diverse, how they utilize their networks (e.g. for social support), etc.

Combining together detailed survey-based information on individuals and their households with geographically- and sociologically-based data contextualizes people's lives and gives social scientists a much deeper understanding of the complex ways in which individuals and societies function.

Besides being contexualized, social- science data need to be reliable and valid. As Linnaeus noted, "method is the very soul of science." The total-survey-error perspective should be applied in the design, execution, and assessment of surveys (Groves, 2005). Surveys are easy to do poorly and hard to do well. All of the components – study design, sampling, content, wording and order of questions, scale construction, training of interviewers, data collection, data entry, data cleaning, verification, documentation, analysis, etc. need to be both carefully planned and competently conducted. This is especially challenging since there are many components, error interacts across components, and high-quality work is both technically-demanding and expensive. But survey research is the classic example of the adage "if it is worth doing, it is worth doing well." Good methods lead to good data and strong science. Poor methods lead to the opposite.

References

Groves, Robert M. (2005). *Total Survey Error: Past, Present, and Future.* Paper presented to the International Total Survey Error Workshop. March. Washington, DC.

Miller, McPherson; Smith-Lovin, Lynn; and Brashears, Matthew E. (2006). "Social Isolation in America: Changes in Core Discussion Networks over two Decades," American Sociological Review, 71:353-375.

Smith, Tom W. and Kim, Jibum, (forthcoming). "An Assessment of the Multi-level Integrated Database Approach" in The Future of Social Science Surveys: Opportunities and Challenges, edited by Douglas S. Massey and Roger Tourangeau. Washington, DC: National Academy of Sciences.

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.