

Discussion

*Cynthia Z.F. Clark*¹

Dr. Habermann has provided a careful analysis of issues relating to ethics and confidentiality for a National Statistics Institute (NSI) as it seeks to fulfill its dual role of providing statistics on society while protecting the confidentiality of individuals or institutions providing data. He has noted the inherent tension of the dual role in that the more information that is provided, the more likely is the risk of an inadvertent disclosure or release of individual information. NSIs develop statistical disclosure control or access procedures to minimize this risk. This risk has been addressed systematically through refinement and automation of these procedures at the U.S. Census Bureau for over thirty years and subsequently at other NSIs.

The issue of group identification through release of statistical information, however, has not had as much attention. This issue was raised in both the Japanese relocation (1942) and the Arab-American (2002/2003) cases cited by Habermann. The potential for group identification is a by-product of two factors – the data variable requested and the geographic level or domain anticipated for tabulation.

When a new or changed data variable is proposed for a census, the U.S. Census Bureau and other NSIs engage in discussion with the data user community that includes advocates for many groups – ethnicity, privacy, civil rights, government, etc. – to assess the need for that “particular good” and the issues inherent in the release of statistics relevant to that data item. The expectation is that relevant issues are raised in the debate. Such was the case with the decision to include the category “Arab-American” on the 2000 U.S. Census long form. At the time there was no reason to anticipate September 11th and its repercussions on the Arab-American community. Data dissemination plans through the American Fact Finder were also well defined prior to September 11th. As noted in the Habermann article, information requested by the Department of Homeland Security could have been extracted by their own staff without making a request to the U.S. Census Bureau. In fact, the information was publicly available as part of the overall agreed upon dissemination plans for census information. The issue that this now raises is to what extent an NSI should assist any organization whose mission is not both in the general and particular public interest.

As an example of an issue of potential group harm that was not resolved through public dialogue, I note the decision to collect data on multiple races in connection with the five races – white, African American or black, American Indian or Alaska Native, Asian,

¹Office for National Statistics, London, UK. Email: cynthia.clark@ons.gov.uk

Acknowledgments: The views expressed are those of the author and do not necessarily reflect the views of the U.K. Office for National Statistics.

Native Hawaiian or other Pacific Islander – on Census 2000. The Office of Management and Budget's Statistical Policy Office held extensive public discussions on the issue of race data collection in the U.S., concluding that race data should be collected by allowing the respondent to check as many categories (of the five) as were applicable. No decision was made on presentation of the multi-race information. The U.S. Census Bureau faced the issue of how this information was to be provided for block level data on race required by Public Law 94–171. Public Law 94–171 addresses issues relevant to the decennial creation of congressional districts. The ultimate decision was made to provide count data for the population older than 18 years using 63 categories tabulated by Hispanic/non-Hispanic. Data swapping procedures were applied to the data set prior to release to reduce the likelihood of disclosing identifiable information.

These two cases provide information on the difficult challenge confronting an NSI meeting the public need for particular goods and yet protecting the information that is released to meet statistical purposes from identification of an individual or that individual's attributes. The Office for National Statistics (ONS) in the UK recently found itself in a difficult situation for protecting the confidentiality of abortion statistics. Although the ONS does not produce these statistics, the UK National Statistician has oversight of the compliance with the professional Code of Practice and ONS produces the complementary vital statistics. Abortion statistics released by the UK Department of Health were highlighted in the media in 2004 following identification of a particular incident with the provider of the service. As some information is in the public domain it was possible to identify the medical practitioner performing a particular abortion. This incident drew attention to a group that had the potential for doing harm. Recent guidelines for statistical disclosure control procedures were issued that provide a higher level of protection to groups affected by the provision of this "particular good." It did not, however, eliminate the production of these statistics nor did the protection imposed have a detrimental effect on the data now available.

In all these examples a statistical agency has the ethical responsibility for assessing the "need" for a particular information good relative to known issues in the production of such a good to ensure quality of the resulting product. In the case of the race data and the abortion data, the relevant issues were not adequately addressed prior to data collection decisions – the first situation did not result in high-quality information; the second caused known harm that might have been anticipated to a group. If any ethnicity data is to be collected, there is potential harm to individual groups, particularly if they are small as was the case with the Japanese and Arab-American data. The same rationale holds for the collection of religious affiliation information, but there is broad recognition of the sensitivity of this information so the public need for the information is tempered by the perceived future potential of harm. What a statistical agency may not be able to do is to foresee potential harm to groups due to future societal events.

Thus, a statistical agency places high priority on engaging in public debate regarding its data collections. However, as Morris Hansen so eloquently stated: "A primary function of the census is to provide statistics for small groups to be applied in whatever ways users desire to use them for or against various programs." If statistical information is to be a

public good – a common good of a democracy – the NSI will never be able to control how the information that it releases is used. What it can do is to be transparent about its methods and procedures, what it collects, and how the resulting information will be released (i.e., as tabulations, microdata, etc.). Credible NSIs hold transparency as a “high” ethical value to ensure credibility in the public arena. They also engage in public dialogue to gain support for their procedures and publication plan.

Received January 2006