

Discussion

*Tim Holt*¹

Habermann provides a thoughtful, and thought-provoking, account of the controversy surrounding the use of detailed small-area Census data by the U.S. Department of Homeland Security. He draws a distinction between the issue of data confidentiality and the use to which the data are put, correctly reminding us that virtually no use of statistical information in support of public policy is free from detrimental effects for some citizens. Even the most socially responsible policy aimed at supporting the economically or socially disadvantaged will carry with it a (fiscal) disbenefit for the main body of tax-paying citizens. One could take this argument much further by arguing that such social policies have long-term benefits (such as social cohesiveness) for all citizens if the result is to improve economic and social advantage for those who need it and help them to play a wider role in the nation's economy.

Nonetheless I agree with Habermann's conclusion that it is not the role of a statistical agency to assess the social benefits and disbenefits of any use and to regulate access on this basis. The role of the statistical agency is to provide high-quality statistics, objectively and scientifically produced, and to support as wide a range of users' needs as possible.

The data accessed by the Department of Homeland Security were generally accessible via the publicly available website, American Factfinder, maintained by the U.S. Census Bureau and as such had met the Bureau's internal criteria for general release while maintaining confidentiality. The remainder of my comments explores what this means in practice.

I focus on a breach of confidentiality, by which I mean that one or more Census respondents are specifically identified (and as a consequence may be subject to administrative actions, although this is not essential for the breach of confidence to have occurred).

A direct breach of confidence would occur if the U.S. Census Bureau were to release data with the identifiers attached. This is unthinkable and in practice Statistical Organisations are more concerned with indirect breaches of confidence whereby sufficiently detailed data are released so that a third party can marry this with other information or can take a feasible action to identify to whom the data belong. For example a family of certain ages and races living in a small neighbourhood with known occupations for the adults in specific sectors of the economy may be more than enough to be linkable to a unique family and a breach will have occurred. Thus the Statistical Organisation has to

¹University of Southampton, Department of Social Statistics, Highfield, Southampton, S017 1BJ, UK. Email: tholt@socsci.soton.ac.uk

consider the potential for third parties to have the capability to breach the confidentiality of the respondents.

It may be helpful to think about three situations:

- A large organisation (commercial or government) which has large data bases that might be used to match onto released data and hence breach confidentiality. If the released data contain 'population uniques' (people or families that are unique within the known population – e.g., neighbourhood) then the potential for a breach exists. The problem here for the Statistical Organisation is that the size and potential of data bases maintained by third parties is unknown. However, these data bases have been transformed by the advance of electronic technologies and data gathering, and what is certain is that data that it may have been safe to release in 1905 or 1955 may be at greater risk of breach in 2005. This is perverse since the potential for a breach is increasing as the demand for more and more detailed statistics for small neighbourhoods is also increasing.
- The second situation is an extension of the first when data is released with additional area level administrative measures appended. For example each individual record may have appended to it a neighbourhood level measure of the average household income (from IRS) or the proportion of people receiving one or more specific social payments. These additional measures are not disclosive for general users since the additional measures cannot be attributed to any detailed location, but this does not apply to the government agency providing the administrative measure. For this specific organisation the additional measures could be used to help identify named respondents.
- The third situation is rather different from the previous two and applies more readily to tabulations as in the case in question. For a government agency to try to identify people of a particular race or background throughout the country is a formidable task. But if the statistical tabulations identify a relatively small number of localities in which a significant proportion of the target group live then the use of alternative methods to identify the individuals may be much more feasible. In this case the Statistical Organisation has unwittingly assisted actions that may lead to a breach of confidentiality. I have no knowledge about whether such actions were taken to breach confidentiality in the case in question.

The question is whether any Statistical Organisation could take legitimate steps to protect confidentiality further in this situation. An obvious step would be to transfer some of the responsibility from the provider to the user by extending existing practice in the case of individual records to include all statistical tabulations and other outputs. In effect anyone accessing any of the Statistical Organisation's outputs would be required to enter into a legal obligation not to use the accessed data in any way that could lead to a breach of confidentiality. This step would place all users under a legal obligation and failure to comply would carry legal sanctions. Entering into such contracts when using many commercial websites is commonplace and would not impede access to the statistics. From the Statistical Organisation's perspective this

could be cited publicly as an extra protection for the confidentiality of the data and may help to strengthen public confidence in the confidentiality assurance and the integrity of the Statistical Organisation in honouring this. It would have the additional benefit of focussing concerns onto the user organisation if a breach of the contractual obligation was suspected.

Received December 2005