

## Discussion

*Dennis Trewin*<sup>1</sup>

### 1. Introduction

First I should congratulate Hermann Habermann on a very thought provoking article. He raises a very important issue particularly for heads of the National Statistical Offices. We are all likely to meet similar problems if we have not already done so. As Habermann says at the end of his article,

“Those who direct statistical agencies must consider and prepare for how they will react to different situations, including ethical challenges, before they happen.”

His article certainly helps that.

As an aside, I should say I am pleased that the U.S. Census Bureau did not release data about individual Japanese during the Second World War, and Habermann’s article dispels that myth. Folklore among privacy advocates in Australia was that individual data was released. But the incident still raises group privacy issues that they will no doubt latch on to if given the opportunity.

What Habermann describes in his article, particularly the release of data about Arab Americans, is what is known as a wicked problem. But it cannot be ignored. Many of us will have to face similar problems. It is often referred to as group privacy and I believe it is an issue that statistical offices will have to increasingly deal with.

How do we handle these group privacy issues? For sensitive cases, it is not sufficient to hide behind our legislation, i.e., to argue that we have not broken legal confidentiality provisions. The reaction of the media and others to the Arab American case is confirmation of that. They want ethical issues to come into consideration when we decide what to release. And we do have some discretion over what to release. In the next section, I will explore some options. I will be commenting from the point of view of a head of a National Statistical Office.

### 2. Group Privacy

Habermann’s analysis looks at the Arab American case from three perspectives:

- The law that protects confidentiality,
- Statistics and harm, and
- Sensitive information and public trust.

I will do the same.

<sup>1</sup>Former Australian Statistics, 38 Mirning Cres Aranda ACT 2614, Australia. Email: dennistrewin@grapevine.net.au

With respect to (i), we should never breach the confidentiality provisions of our statistics legislation even if we believe the law is unduly restrictive. If this is the case, our task is to persuade the legislators to change the law. If we want to maintain public confidence, there should be some public debate as part of the process.

So not breaking the law is a necessary but not sufficient condition for assuring public confidence in our release practices.

With respect to (ii), I agree with Habermann that it would be inappropriate if we decided not to release statistics just because some individuals or groups of individuals might be harmed by the release. If statistics are to be policy relevant, they are likely to have some impact on policy or its implementation. If that is the case, there are likely to be businesses or persons who benefit from the policy and others who do not benefit. But I think “statistics and harm” is still a relevant criterion for using our discretion in deciding what to release. We use quality and other criteria for deciding what to release (even though the data may not breach confidentiality provisions) and I would suggest “harm” another consideration. But it should be regarded as a “net” concept. I am not suggesting we should not release statistics just because they might harm some individual or group. Rather, if the net effect of the release of the statistics is that there might be more costs (including to the reputation of the national statistical office) than benefits, their nonrelease should be considered. I will discuss how below.

With respect to (iii), I think the arguments are similar to those in the previous paragraph. Here the loss is the reduction in the public trust of the national statistical office. This can have a severe impact on the work of the Office. It can be difficult to recover from. Therefore, it has to be a consideration in deciding what to release.

What are the options when it comes to deciding what to release? I will discuss three broad options. I do not discuss selective release of statistics to particular users as an option. Although we might have this option for certain types of statistics (e.g., microdata in the case of Australia), it is not the way to solve the type of problem posed by Habermann. It would only foster suspicion if special tabulations were available for some users and not other users.

The three broad options I would like to discuss are:

- Not collecting the data at all;
- Only agreeing (preferably in advance) to publish at a specified level;
- Publish at the detailed level (i.e., what the U.S. Census Bureau did with the 2000 Census data set for Arab Americans).

Before discussing these options, it is worth emphasising that privacy is not an absolute concept. Most statistical work involves some invasion of privacy (e.g., data collection). There are trade-offs. The key criterion is whether the benefits of the statistical activities outweigh any privacy costs. We implicitly make this decision every time we decide to proceed with a statistical collection, i.e., the benefits outweigh any costs including the privacy costs.

Returning to Option 1, if the trade-off mentioned in the previous paragraph suggests the privacy issues are paramount; we should not collect the data. In the U.S. Census, this decision has been made on a religion question. In Australia, a slightly different tack is

taken. Religion is included in the Census but in an overtly voluntary way. For us, that has meant that the religion question has become noncontroversial.

Decisions on the content of the Census, particularly if they involve ethical issues, are decisions that statistical agencies should not make in isolation. They need to consider the views of key stakeholders, both the users of the information and the groups whose privacy might be most affected. Sometimes there will be considerable overlap between these two groups. For example, in the case of ethnic communities, they will be key users of the information as well as those whose (group) privacy is most at risk from the release of detailed data. Discussions with these groups could be very revealing. If privacy issues are a source of such angst that this would argue against release at a small area level (not necessarily the most detailed level), it suggests that the data should not be collected at all in the Census. After all, one of the main purposes of the Census is to produce detailed data. Of course, there are other users who might want to justify the collection of the data, even though there were significant privacy risks. But these uses would have to be substantial to obtain public support. Without that support, the success of the Census will be at risk if we know that there will be groups who will mount strong contrary arguments on privacy grounds.

In Australia, in the lead-up to the Population Census, we had earnest talks with the various ethnic councils to make sure we had a common understanding, particularly in messages given to the media. They are strong supporters of the Census and the need for detailed data. They need it for their own purposes. I understand that in the UK approaches to ethnic councils were successful in obtaining a mutual understanding of the level of detail to which Census data should be released.

On Option 2, a statistical agency does have discretion over the level at which it releases data. It does not need to release all data at the most detailed level. In fact, from Australian Censuses, we only release a limited amount of data at the most detailed geographic level, i.e., what we refer to as the mesh block. We will not release religion or ethnicity data at this level. Confidentiality is one consideration but implicitly perceived harm to population groups may be another consideration. But the ABS will use these variables at the mesh block level to generate more aggregated data at user specified levels. Option 2 is one of the ways we will use to address the problem raised by Habermann.

Option 3 should not be used universally. It may be the appropriate option for core demographic variables, for example, but not I believe for other, more sensitive variables.

The trade-offs involved in deciding how to implement these options are not straightforward. Statistical agencies may not have the knowledge or perspectives to make the appropriate judgements in isolation. For Option 2, it will be the statistical agencies that are making decisions on tabulations. For Option 1, in some countries it is governments not statistical agencies that make the final decision on census content. But I would argue that they would benefit from expert advice on privacy and sensitivity issues to help them with their decisions. So a key element with regard to the implementation of these options is to set up Advisory Committees to help statistical agencies or possibly the government in these important deliberations.

It is not just Advisory Committees that should be involved. The community should be given an opportunity to comment. As Habermann says, "openness and transparency in the

proposed content of surveys and the decennial census allows affected segments of society to participate in the process and make their views known.”

### **3. The Common Good, the Particular Good, and Individual Ethics**

This section of the article is particularly important for heads of national statistical agencies who are confronted with issues where resignation may be an option. I will quote parts of this section which I found particularly informative and which summarise the logic that should be used in making such decisions.

“Statistical agencies derive their statistical authority from being a member of the government. Statistical agencies are part of the government in general and as such are subject to all the laws and bureaucratic procedures that bind government agencies.”

“The director of a statistical agency is at variance with the common good as supported by a superior or the legislature. If the laws are broken or there is an attempt to break the laws, it is an easy case. Sometimes, however, it is not an easy case. It is often not a question of legality – and, therefore, the law does not provide sufficient guidance. If it is a question of conscience, the decision may not be so clear.”

“These decisions cannot be made in a fit of pique, nor should they be made to garner public attention – although public attention can be an effective tool if the conflict rises to a matter of conscience.”

“Adherence to bureaucratic procedures may, in fact, be a denial of the basic responsibility of those who work in statistical agencies. The common good at any particular time may negate the particular good of the statistical agency and/or particular groups.”

There are three basic options available to the heads of national statistical agencies. These are personal judgements although discussions with others will often be useful in helping them to decide what to do.

- Accept that the common good outweighs the particular good (e.g., public trust in the official statistical system). Both short-term and long-term views of the particular good should be taken.
- Argue with the government that the particular good is so important that it outweighs the common good in these circumstances. The possibility of resignation could be suggested as part of these arguments but only if you are prepared to follow through if your arguments are unsuccessful.
- Resignation.

None of these options should be taken lightly or, as suggested by Habermann, decided in a fit of pique.

### **4. Conclusion**

Hermann Habermann is to be congratulated on his article. It raises some very important issues. It deserves thorough attention and debate as I believe the public is expecting more of statistical agencies than just protecting confidentiality. There is a risk that the public will not cooperate with the Census unless we address these issues of public concern in an appropriate way. We need to look at these issues from the point of view of the public, and

those like the media who can influence public opinion, not just the paradigm of official statisticians.

I believe there are more options available to statistical agencies than suggested in Habermann's article. In particular, we have the discretion to decide the detail with which to publish particular variables. This is a decision that should be made after seeking the advice of others, particularly representatives of the groups whose (group) privacy might be affected.

Received January 2006