

Why It Matters to Distinguish Between Privacy & Confidentiality

A comment on Gerald W. Gates's *How Uncertainty about Privacy and Confidentiality Is Hampering Efforts to More Effectively Use Administrative Records in Producing U.S. National Statistics*

Kenneth Prewitt*

1 Introduction

The article by Gerald W. Gates is a deeply informed and near-exhaustive treatment of how privacy and confidentiality practices and concerns need to travel from their primary home in the domain of survey data to a domain that, inevitably, will blend survey and administrative data far more extensively than is presently the case in government statistics.

If the treatment is informed and exhaustive, is there anything left to say? Yes, there is a critical conceptual distinction that Gates largely overlooks. Privacy and confidentiality are not terms to be used interchangeably, as if they describe the same problem facing population data. To use the terms interchangeably, as generally characterizes the research literature and Gates's use of it, seriously hampers a research-based understanding that can chart a course forward. My plea, then, is that the new research that is recommended start from a fresh conceptual foundation—one that considers rather than blurs the distinction between privacy and confidentiality.

To offer one example, Gates correctly notes that:

Attitudes of program participants and survey participants are always in the back of agency decision makers' minds when deciding to share information. Negative public reaction (frequently related to privacy and confidentiality) can have dramatic impacts on the agency's ability to function by reducing participation, increasing program complexity, and fostering greater oversight.

Insofar as these “negative public reactions” differ depending on whether it is privacy *but not* confidentiality that is at stake, much is lost by overlooking the distinction, and the proposed research agenda (Section 7.2) and agency actions (Section 9) will not take us as far as will be needed. For example, disclosure and informed consent policies are primarily about confidentiality, not privacy.

*Former Director of the U.S. Census Bureau (1998–2001); currently Carnegie Professor of Public Affairs, School of International and Public Affairs, Political Science, Columbia University, New York, NY, <mailto:kp2058@columbia.edu>

The Distinction

At the most simple and common-sense level, the distinction is between “*don’t ask*” and “*don’t tell*.” The citizen saying “I won’t answer that question because it is none of your business” is sending a different message from the citizen who is saying “I won’t answer that question because you may share it without my knowledge or permission,” and neither response is the same as “I won’t answer because I don’t trust the government not to use my answers against me.”

Privacy, confidentiality, trust, and fear of misuse mingle in complicated ways, but this does not relieve us of the responsibility to sort them out as best we can. To treat these public concerns in a largely undifferentiated fashion, as much of the research literature has, makes for a poor guide to nuanced policy and practice responses by statistical agencies.

I develop my argument in two steps: first, from the perspective of a participant observer; second, citing research findings not considered in Gates’s otherwise comprehensive review.

A Participant Observer’s Perspective¹

I was Director of the Census Bureau when the 2000 decennial census was fielded. In mid-March census forms were mailed to approximately 120 million American households, sixteen percent getting the census long-form. Unexpectedly, immediately, and vigorously, the census long-form was broadly attacked as a government “invasion of privacy.”

Initial complaints were voiced by conservative talk show hosts and editorial writers. Late night comics were quick to chime in. Soon political leaders, from small town mayors to the congressional leadership, added their voice. George W. Bush, an announced candidate for the presidency, told the press that he understood “why people don’t want to give over that information to the government. If I had the long-form, I’m not so sure I would do it either.”

With a presidential election approaching, the census became a poster child for an “invasive federal government.” Colorado Congressman Robert W. Schaffer, whose household received the long-form, announced that some questions were “too nosy,” adding: “I’m happy to voluntarily cooperate with the government in areas where I decide it makes sense. Beyond that, it starts to meet the definition of intrusive.” The Republican Senate majority leader, Trent Lott, urged voters to skip any questions they felt violated their privacy. His Senate colleague, Chuck Hagel similarly advised the public to, “just fill out what you need to fill out, and [not] anything you don’t feel comfortable with.” Republican Tom Coburn of Oklahoma said he was “appalled and outraged by the

¹This section borrows heavily from Prewitt, K. (2003/2005). *Politics and Science in Census Taking*. New York: The Russell Sage Foundation and Washington D.C.: The Population Reference Bureau. Reprinted in R. Farley and J. Haaga (eds.). *The American People: Census 2000*. New York: The Russell Sage Foundation.

intrusiveness” of the long-form questions. He argued that “It is ridiculous for the Census Bureau to ask personal questions that have nothing to do with their constitutional mandate to count the citizens of the United States” (Cohn, 2000).

Six congressional bills were introduced that would have seriously affected the ability of the Census Bureau to collect long-form data, an indication of how much attention was drawn to the controversy. The Senate did pass a nonbinding resolution urging that “no American be prosecuted, fined, or in any way harassed by the federal government” for not answering certain questions on the census long-form. The Senate was in effect telling the public it was acceptable to ignore what was a legal obligation—to complete the census form, short or long.

The controversy erupted at a point of maximum vulnerability for the census. The Census Bureau went into scramble mode fearful that the strong criticism of the census, extensively covered by the press, would depress mail-back cooperation (which it did, as documented below). A cleverly constructed media event was immediately put in place, involving emergency response teams with a firehouse as a backdrop. The Fire Chief eloquently said, “we need census data to do our job.” My talking points were prepared, essentially saying, “not too worry; the Census Bureau really does protect the confidentiality of your answers.”

But confidentiality was not at issue; privacy was. The public outcry was not about disclosure, it was about intrusiveness. There is no other way to interpret the quotes above. George W. Bush, Trent Lott, and Tom Coburn were not worried that the Census Bureau would disclose the size of their household or whether they rented or owned. They were not worried that their answers would somehow be used against them. They were voicing a widespread view that some things are simply “none of the government’s business.” Thus to promise confidentiality was beside the point. I knew this, but having no better answer, gamely repeated in press interviews, speeches, and congressional testimony “not to worry, the Census Bureau won’t tell anyone what you tell us.”

Political debate blurs the distinction between privacy and confidentiality, trust and fear, as generally does the survey data cited by Gates. But from my participant observer perspective in 2000, I have no doubt that it was privacy, not confidentiality, concerns at the core of the unprecedented 2000 census. The Senate Resolution made no reference to disclosure issues; it, instead, stressed “objectionable questions.” A very widely-cited statement, issued by the Libertarian Party, was unambiguous: “Real Americans don’t answer nosy Census questions. You can strike a blow for privacy, equality, and liberty by refusing to answer every question on the Census form except the one required by the Constitution: How many people live in your home?”

The Libertarian Party is not without sound reasons for citing the Constitution. A constitutional right to privacy was famously framed by the Senate as it debated the confirmation of Robert Bork to the Supreme Court in 1987. Then the Senator from New York, Patrick Moynihan justified his vote against Bork’s appointment with the following reasoning: “it is his restricted vision of privacy which troubles me most. I cannot vote for a jurist who simply cannot find in the Constitution a general right to privacy...” When

Bork was rejected by the Senate, Anthony Kennedy was nominated. In his testimony before the Senate Judiciary Committee, which was to confirm him unanimously, he insisted “that the concept of liberty in the due process clause is quite expansive, quite sufficient, to protect the values of privacy that Americans legitimately think are part of their constitutional heritage.” Whether Senator Moynihan and Justice Kennedy got the constitution right is not our issue; certainly they were correct, as Kennedy put it, that Americans consider privacy to be part of their constitutional heritage.

It was this heritage, claimed the critics, that was at stake in Census 2000. If information is central to a functioning democracy and economy, so is privacy, and we expect government to protect it and certainly not to itself infringe on it. Across American history there has been an unspoken but working consensus that it is possible to have both population-based information and privacy protection. Essentially the government has avoided the inherent contradiction by saying to the public, “yes, you give up a bit of privacy in answering government questions, but this pledge we make to you. If you cooperate, we will never share your answers. They are confidential.”

As Gates correctly emphasizes, the pledge of confidentiality is one that the Census Bureau is careful to honor and on which it can point to a remarkable track record. But it does not help to treat confidentiality protection and respecting privacy as though they reference the same thing. The government, in fact, recognizes the distinction. In 1971, the President’s Commission on Federal Statistics defined privacy as “the individual’s right to decide whether and to what extent he will divulge to the government his thoughts, opinions, feelings, and the facts of his personal life,” and defined confidentiality as prohibiting disclosure “of data in a manner that would allow public identification of the respondent or would in any way be harmful to him” (The President’s Commission on Federal Statistics, 1971, pages 197 and 222).

An Attempt to Untangle Matters

In 2000, two large-scale survey projects investigated public responses to the 2000 census experience. The Census Bureau funded an NORC conducted study, which is cited in the Gates review especially noting the careful and informative analysis of Eleanor Singer. The other project, not cited by Gates, was privately funded and fielded by Knowledge Networks. In all important respects, analysis from these separate projects came to similar conclusions. But a design feature of the privately funded survey allowed its principle investigators to respond to the unexpected privacy controversy. In this brief note, I summarize only a few findings from this study that directly bear on the distinction between privacy and confidentiality. For a more detailed discussion, see *The Hard Count*.²

The Hard Count is based on seven monitoring surveys conducted from late February into early June 2000. The surveys addressed issues relating to the census mobilization campaign and included items on privacy and confidentiality. The data structure com-

²Hillygus, D. S., Nie, N., Prewitt, K., and Pals, H. (2006). *The Hard Count: The Political and Social Challenges of Census Mobilization*. New York, The Russell Sage Foundation.

bined longitudinal and panel data, which allowed attention to opinion change during the course of the census experience. When the privacy debate erupted in mid to late March, two additional and separate surveys were taken. The first was designed to gauge exposure to and reaction to the privacy debate. The second was an experiment to evaluate the impact of elite and media rhetoric on census cooperation, with a specific focus on privacy issues.

Three agree/disagree questions were used to capture different dimensions of privacy and confidentiality:

“The Census Bureau promise of confidentiality cannot be trusted.”

“My answers to the census could be used against me.”

“The census is an invasion of privacy.”

Though all three concerns increased over the course of the census data collection period, there was a sharp difference in the amount of change between the first two questions and the third (privacy) question—a 3.6% increase on whether the promise of confidentiality was trustworthy and a 3.9% increase on whether answers could be used against me, indicating that the public had largely made up its mind before the census got underway. Agreeing that the “census is an invasion of privacy,” doubled, from less than 10% to 20%, indicating two things. First, the controversy made a difference; second, the public made a distinction between the three dimensions measured.

The panel data also showed that privacy concerns heightened over the course of the census collection period had a larger negative impact on census mail back cooperation than preexisting concerns. The privacy controversy sparked by talk shows, editorials, political speeches, and the like, depressed census cooperation rates by reinforcing, heightening, and even creating privacy concerns. In *The Hard Count*, we estimate that respondents who entered the census period unconcerned about privacy, but who were persuaded differently during the controversy, were 23% less likely to cooperate than those whose views were unchanged. This is a non-trivial finding if our concern is cooperation with government data collection efforts, whether based on survey or administrative data.

Other analysis from *The Hard Count* confirms this basic finding. In the experimental data, respondents were asked to complete a version of the long-form. Item non-response was consistently higher in the treatment group that had been negatively primed with respect to privacy issues than in the control group, not primed at all, and in a group that was positively primed. And on a general attitude question of whether refusal to cooperate with questions that “violated privacy” was warranted, the treatment group more readily agreed than the control group.

Item non-response and response rates in general are, of course, a growing concern to our statistical agencies. Understanding item non-response is likely to offer insight into public attitudes toward the use of administrative data. The public, of course, cannot refuse to supply administrative data—there is no general response rate issue—but it can be selective in how it is willing to have those data used for statistical purposes (“item

non-response”).

If the public’s reasoning is based on privacy-reasoning as much or more than confidentiality-reasoning, “the government doesn’t share data” will not be a persuasive reply.

Analysis in *The Hard Count* is far from definitive on the issue addressed in this note. It is, however, indicative in important respects about the merit of more carefully distinguishing among the dimensions of privacy and confidentiality.

The Census Bureau has long operated with an exceptionally strong law guaranteeing that answers to its surveys are to be used for statistical purposes only. This guarantee now extends to all government statistics. The Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) prohibits the release of any information collected by a federal agency for a statistical research purpose from being disclosed in an identifiable form for non-research purposes.

This government-wide policy is critically important in an era when many Americans believe that answers they give to the government or to their hospital, bank, or local supermarket are likely to be disclosed without their consent or even knowledge. Government surveys have to contend with serious collateral damage when, for instance, the Social Security Administration reveals an ad hoc policy of cooperation with the Department of Homeland Security, or when a huge store of digitized credit card records go missing from a commercial site. Already, as Gates notes, a significant proportion—half in some surveys—discounts government pledges of confidentiality. Strong laws, policies, and practices are the best answer to these worrisome levels of public disbelief about inappropriate disclosure by the government.

It is my claim that even if the entire public were confident in the security of their answers in government data efforts, a number—and probably a growing number—will refuse to answer questions they view as intrusive. It is this, of course, which motivates members of Congress to demand that the American Community Survey be voluntary. A voluntary ACS is not proposed as an answer to confidentiality concerns; it is an answer to privacy concerns.

A distinction between confidentiality and privacy can guide research in ways that can help design appropriate policy and practice responses, which is the purpose of this note.

Conclusions

I believe the nation’s statistical system is on the verge of major change. One dimension of the change is reflected in the greater incorporation of administrative data. A second dimension will be the increased use of digital data.

Hovering over these changes are three major challenges, one being the politics and practicalities of public views toward privacy and confidentiality. A second is data quality; here, I believe, we can count on government statisticians to successfully address quality issues in public sector administrative data. The quality challenge with digital

data will be much more difficult, not least because it will come primarily from proprietary sources in the private sector.

The third challenge is data representativeness. Sampling theory has been a powerful protector of the basic value that the government's information system has to be fully inclusive, and sample weights based on the census are of critical importance to realizing that goal. Administrative data are, by definition, a "census" of program participants, not of the full population. Digital data are a "census" of those who leave a digital footprint, again not of the full population. Blending survey, administrative, and digital data in a new government information system will require attention to representatives, less those outside the reach of administrative record-keeping and "off the digital grid" be marginalized.

If, as I believe likely, the challenges of data quality and representativeness can be met, we will be left with our third challenge, privacy and confidentiality. Here the complication is that public attitudes will differ across the three data domains. It is comparatively easy for the public to say "no" to government surveys; not easy to opt out of administrative records but easy to refuse data linkage proposals, and nearly impossible to forego the conveniences of a digital economy and communication system. Moreover, privacy violations in one arena will spill-over into anxieties in another. In particular, when commercial firms make light of privacy protections, the government may bear the brunt of public anger.

I take away from this the sobering thought that scientifically understanding privacy or confidentiality concerns as they interact with government surveys is insufficiently robust to meet the coming challenges. I applaud the important step forward represented in Gerald Gates's article. I urge that the next research phase exercise more care in distinguishing between privacy and confidentiality. I anticipate many more phases to come. There can be no population statistics in a democracy without the consent and cooperation of the public. And the prospects for that consent and cooperation are poorly understood, by scholars and, no doubt, by the public itself.

