

INTERVIEW THE EXPERT

## An Interview with Kenneth Prewitt

Andy Peytchev

Keywords: survey practice

<https://doi.org/10.29115/SP-2008-0005>

---

### Survey Practice

Vol. 1, Issue 1, 2008

---

#### An Interview with Kenneth Prewitt

Kenneth Prewitt is a Professor of Public Affairs at Columbia University. He has formerly held positions as director of the U.S. Census Bureau, director of the National Opinion Research Center, president of the Social Science Research Council, and senior vice president of the Rockefeller Foundation. We asked him about his career and the major issues the survey field is facing. In particular, Dr. Prewitt has raised concerns about privacy and confidentiality, and declining cooperation in surveys, at a time with increasing amounts of information from multiple sources.

**SP :** What led you to a career in surveys? Did you have plans to do something different?

**Prewitt :** I was teaching political science and chairing the department at the University of Chicago. I was involved with NORC, though in a quite peripheral way. The Provost called and said that the University wanted to appoint me as Director of NORC. To which I replied, "I have no experience in running anything so big and complicated." In the best U. of Chicago tradition, the Provost replied: "Well that's the idea. This is a faculty-run university; we don't want professional managers making academic decisions." I became NORC director, though certainly not with the intention of initiating a "career in surveys," and in fact don't see my career that way. As it turned out, however, NORC was relevant to my appointment as Director of the Census Bureau. The Clinton administration was looking for someone who at least on paper had the right credential – and my earlier role at NORC provided that credential. It allowed the Democratic administration to present me to a Republican Congress (the position requires Senate confirmation) as a nonpartisan academic with relevant experience in scientific management. My career, however it might be characterized, is more an accident than a plan. I see myself as an academic who happened to do some other things in foundations, scientific organizations, and the government.

**SP :** What do you think are the most pressing problems facing surveys in the near future?

**Prewitt :** Public cooperation is a very serious problem and we'll talk in more detail about that. But there is a larger issue of which public cooperation is just one element. I believe that over the next quarter-century or so, the

government will increasingly merge administrative data and survey data. What we today understand as the “national statistical system” will more properly be thought of the “national information system.” Sample-based survey data will be part of that system, but less dominant than it has been in the previous half-century or more. For example, the new SIPP [Survey of Income and Program Participation] that is under consideration might be based on 50% administrative records and 50% survey data. If so, that is an indicator of where the whole system is going.

One reason for the turn to administrative data – and other data sources, such as commercially provided scanning – is the survey response rate problem: the unit cost of each respondent to a survey is high and getting higher. If, as some have suggested, we can control the response rate issue with incentive payments, there will be further cost increases – as well as data quality problems. Nonresponse, by the way, is not just how many respondents answer but also item nonresponse. There has been less attention to item nonresponse, but in the 2000 Census there was a sharp increase of item nonresponse, reaching into the 20% range on several questions.

**SP :** In your article in *Science* you point out that the public has serious privacy and confidentiality anxieties and that voluntary cooperation with surveys is declining. Why do you think that the two are linked, as opposed to other factors affecting responding to surveys?

**Prewitt :** There are two pieces of relevant empirical work. After the 1990 census and again the 2000 census, Eleanor [Singer] reported a correlation between levels of privacy concerns and responses to the census. Then, during the 2000 census, another study in which I was involved used Knowledge Networks (an Internet survey firm) in a design that took real-time surveys, six in all, as the census was taking place. When the privacy outcry erupted as the long-form reached households, we even added an experimental design on its impact on census cooperation. The results are presented in *The Hard Count*, Hillygus et al. (Russell Sage Foundation). Here we estimated that the privacy uproar over the long form depressed the mailback response rate by as much as 5 percent. This work, along with that of Eleanor Singer and her colleagues, offers strong evidence of the association between census cooperation and privacy/confidentiality issues.

You are correct to suggest that privacy concerns are not the only thing affecting fall-off in survey cooperation. Junk mail and push/pull marketing research turn the public off. Half the population, no doubt including you and me, have refused to cooperate with phone surveys. I’m sure you do what I do and try to find out how serious it is. If it’s serious, you cooperate, but most people aren’t going to find out if it is serious. So if half the population, and that’s the half that responded to the survey asking whether they had refused to cooperate, are saying they’ve already turned down surveys, it is highly likely that more is going on than privacy concerns. It may have to do

with disgust over the whole marketing agenda that disrupts our dinner hour. In terms of the larger picture there is a serious problem with response rate, some portion of which is attributable to privacy/confidentiality concerns, but how much of the variance we attribute to that factor is uncertain at present.

As I said, however, there is a larger, more complicated challenge to survey data. It will occupy a steadily decreasing role in the nation's information system. Already a number of European countries, especially the Nordic countries, will tell you that less than 25% of the information used by the government comes from surveys. The administrative data are already collected by the government for program management purposes; why not use it in lieu of survey data to understand the economy and society? Even if we were not facing a response rate problem, the sheer density of administrative and surveillance data presents a challenge to our traditional reliance on survey data as the platform for the national statistical system. By the way, by "surveillance," I do not have in mind the Patriot Act so much as the data we provide every time we use a credit card or book a flight. This is the digital footprint each of us leaves. The sheer amount of digitized data is enormous and we are at the early stages of its expansion and of the data mining methodologies used to extract information from it. We cannot be surprised if the government (now, for example, facing a full cycle 2010 decennial census, which includes the American Community Survey, that will exceed \$12b) asks "cannot it be much cheaper to see what we can learn from all of this administrative and digital data than to try to find people and convince them to answer what they see as our intrusive survey questions?"

As I wrote in the *Science* essay (or have written somewhere), it has taken nearly a century to get survey data to the level of quality we now expect – measuring sampling and non-sampling error, using cognitive science to improve question wording, etc. The amount of serious scholarship on the error structure of administrative data is miniscule in comparison, and even less on the error structure of scanned data or other surveillance sources.

Certainly one of the big challenges looming in front of us is the quality of administrative data. From the perspective of quality, administrative data have a troubling characteristic. Collected to administer a program, what matters is the accuracy of the variables that are germane to the particular program. The Social Security Administration really wants to get my age right, but does not need to be precise about my earnings above the threshold level that determines how much they collect and then will have to distribute. The IRS wants to get very exact data on my income, but can be more casual about my age. Neither of these programs has to be overly concerned about my current residential address. School records, in contrast, care less about my income and age, but if I am sending my children to a local school will want to know in which district I live. My address matters.

Administrative records are case-rich but variable-poor, that is, a large number of observations but only a small set of information about each observation. To be useful for population analysis, then, they have to be linked. This invites matching errors, and we know those to be serious. Nevertheless, a heavier use of administrative data is part of our future. All, or nearly all, of the questions on the American Community Survey could be addressed with federal and state administrative data or from private-sector data. There is an item on home mortgages – why not call banks for that information? Of course the blurring of the boundary between survey and administrative data, to say nothing of blurring the boundary between public and private-sector data, raises a host of issues other than data quality – issues of coverage and representativeness and, of course, of privacy and confidentiality. We are at the earliest stages of assembling scientific talent to take up these questions.

**SP :** You have been concerned with privacy and confidentiality as increasing threats to the US Census and the American Community Survey; why do you think these concerns are increasing?

**Prewitt :** For reasons unrelated to the Census and the ACS, there is a public reaction to the intrusiveness of the survey industry more generally. This intrusiveness irritates the public irrespective of data confidentiality concerns, but the Census can only rely on promising confidentiality. The person saying, “just leave me alone” is not going to be persuaded when the Census Bureau says, “your answers are confidential.” The irritation and the response don’t match.

This is not to ignore the problem of confidentiality, and here the Census is vulnerable to the more generalized anxiety over matters such as identity theft. Nearly half the public already discounts the pledges of confidentiality by the government. When asked, “Do you think your census data are being kept confidential?” about 40–45 percent of the population says no, and I fear that the percentage will increase because of news coverage of missing laptops from the VA, etc. The sheer volume of data collected in so many different places and via so many different methodologies guarantees an increase in incidents of leakages. There is a huge information market, largely driven by commercial interests. It is gathering up everything possible in order to sell products or, in this season, to win elections. There will be inadvertent as well as deliberate misuse of data. As the public experiences this, it will discount the privacy/confidentiality problem on grounds that there is not much that can be done anyway – short of throwing away credit cards, staying off the Internet, not visiting a doctor or catching a plane. But a public irritated by intrusiveness and knowing that there is risk that private information will not always be carefully handled can take it out on the census and other government surveys. It is easier to say “no” to a census-taker than to quit shopping on the Internet.

**SP :** What do you think we could be doing to protect the mission of a survey organization from the public withdrawal from cooperation?

**Prewitt :** Survey data, whether collected by the government or by reputable private organizations, are, we know, a public good. We can do a better job at packaging and presenting this public good data in a manner that is of value to the general public rather than just to government and commercial decision-making. For example, the new effort to create a national indicator system, under the leadership of a new non-profit, The State of the USA, is such an attempt. Those of us involved in that effort intend to design key national indicators that will be used by schools, churches, community organizations, local governments, and dozens of similar settings. The original data will largely come from the federal statistical information system, but will be returned to the public for its purposes and goals. This is providing a better tool than what is currently available, so that citizens can take advantage of the information that, after all, only exists because they answered government survey questions in the first place.

Put more bluntly, I don't think we can protect survey data by simply saying we're going to keep these data confidential. We've said that and said that, but even if census data are well-protected (and they are), it is not enough. We have a different educational project before us – to remind the public about the source of news coverage of the housing market or immigration issues or school reform. Hardly a day goes by in which the *New York Times*, for example, fails to have a story that does not cite the American Community Survey. We have to convince people that we can only tell you about your own community if you cooperate with surveys such as the American Community Survey.

I want to connect census cooperation less to a pledge of confidentiality and more to data accessibility and usability. This, I believe, is the basis on which to protect government surveys.

**SP :** Selzer and Anderson presented data recently that indicates the Census Bureau released identifying information during WWII. Do you think it's possible that any statistical agencies are currently releasing identifiable information as part of the war on terror?

**Prewitt :** I think not, though from the outside we cannot offer a definitive answer. If a statistical agency were releasing identifiable information to, for example, Homeland Security, the agency would have to deny it. But a statistical agency has much less useful data on the specifics of flight school enrollment or who is learning to drive large trucks or purchasing chemicals than many alternative sources. The Anderson-Selzer paper was a great piece of detective work, and an important corrective to the historical record. After 50 years of denying that there had been any release of micro-data in connection with the treatment of Japanese-Americans, we stand corrected. This makes the statistical agencies even more determined to prevent this misuse.

I worry less about statistical agencies and more about data mining of administrative data, an issue that gets more troubling as the boundary blurs between administrative data and survey data – which returns us to where we started. There is a new “information system” in the making. My plea is that we subject it to the same array of quality standards, principles of confidentiality, and accessibility practices that we have worked so hard to ensure for the survey-based statistical system.

## REFERENCES

- Hillygus, D.S., N.H. Nie, K. Prewitt, and H. Pals. 2006. *The Hard Count: The Political and Social Challenges of Census Mobilization*. New York: Russell-Sage Foundation.
- Prewitt, K. 2004. "What If We Give a Census and No One Comes?" *Science* 304 (5676): 1452–53.