

DISCUSSION

C. L. Kincannon, Bureau of the Census

All of these papers deal with the familiar dilemma of how to make more statistical use of a record without giving up vital protections. The recent legislative proposals for statistical data sharing and confidentiality--in shorthand, the enclave proposals--are discussed in the papers, not so much as the solution to the dilemma but as a vehicle for finding one. We should not lose sight of the fact that this vehicle has been under construction for a half century, and despite many interesting models off the assembly line for testing, there is still no gas in the tank.

We need to reexamine what we have been trying to do, whether there are workable alternatives, and even whether technological change and changes in public attitudes may moot the question. I will try to suggest an approach to this assignment by commenting briefly and selectively on the papers and introducing my own list of considerations that the papers seem to understate.

I do not at all underestimate the effort or the skill that has gone into the enclave proposal in order to find a legal framework for increased sharing of identifiable records while preserving all that is essential to public confidence in statistical and research endeavors. The proposal is, after all, by as well as for statisticians. It is not surprising that statistical agencies with Congressional mandates to produce data of good quality at modest cost should want greater access to large record systems, particularly if they contain good sampling frames and comprehensive general-purpose information about the universes of study. Nor is it surprising that the two agencies that have such holdings--the Census Bureau and the Internal Revenue Service--should resist encroachments upon their reservations that might impair their primary value through loss of public confidence, respondent cooperation, or taxpayer compliance.

In other words, within the Federal statistical environment, there are certain record seekers and record holders who are at an impasse on the merits of broad record-sharing proposals, though there is agreement in principle on two more limited proposals: first, for interagency use of the Standard Statistical Establishment List (SSEL); and second, on the importance of protecting the confidentiality of record sets that are not adequately protected now.

The Mugge paper, representing a view from the health statistics area, points up practical difficulties posed by the interaction of current laws, and suggests that much good work is getting done under present arrangements, but more could be done if the decennial census addresses were accessible. The Alexander paper describes the issues that must be addressed in any record-sharing scheme and focuses on the balancing of privacy interests and research needs. In particular, this paper gives a comprehensive account of the different levels of existing protection, and Alexander makes an excellent case for protecting statistical and research records from subpoenas or similar intrusions that may compromise the promise of confidentiality.

The Clark-Coffey paper, a characteristic OMB point of view, is understandably in favor of the enclave proposal. I was taken by the comment in the paper that protected statistical centers, in order to be credible, must be designated by law or by the President. This idea has its own dilemma. If the law designates the centers, the law must be changed to alter their numbers. Resorting to the legislative process is complex and runs the risk of unrelated issues determining the outcome. If, on the other hand, protected centers may be established by Presidential designation, an uncertain mixture of political, management, and statistical considerations may shape the outcome. In either case, the statisticians are not really in control of their enclave.

Recently I received the section of the paper entitled "History of Confidentiality Legislative Initiative." I'm not well equipped here to check historical facts, but I am concerned about how some of this history is set forth. For now, I have two comments:

1. 1954 recodification of census laws:
 - a. When Congress recodifies, it intends no substantive changes in laws--for example, penalty provisions apply only to Census employees.
 - b. All Census responsibilities are delegated in full to the Director of the Census from the Secretary of Commerce. This has been practiced faithfully by every Secretary.
2. As to the Paperwork Reduction Act:
 - a. This is not a legacy of the Bonnen project as stated in the paper.
 - b. It did not "give OMB extraordinary authority" to direct interagency disclosures; it merely retained the rarely used authority given to OMB in 1942.

The title of this paper is, "How Many People Can Keep a Secret?" A more relevant question, in my view, is how many people should decide who should keep a secret, and who are the people who decide? There is no serious question as to the integrity of statistical agencies in their adherence to privacy and confidentiality principles and statutes. The problem lies in the reality and the perception of the mechanisms by which the walls of secrecy can be penetrated, broken down, or moved around by persons who might be given delegated authority on record sharing and who have overriding interests that are not statistical. The importance of perception is noted in the Alexander paper, which reviews the present treatment of the use of IRS tax records.

The Wilson-Smith paper, representing a view of the statisticians within IRS, applauds the greater use of existing records for statistical and research purposes, and then hopes for some resolution that allows them to do their own work more effectively and perform services for others in a way that preserves both the credibility of tax return confidentiality and Treasury Department control over who gets what from its holdings.

One of the reasons given by Wilson and Smith for resolving the record-sharing dilemma is that "Good government calls for efficient management of available resources." It's a sound principle, but there is no doubt in my own mind that there are proper exceptions to the rule. The government is not merely a very large business, in which efficiency of operation is the sole or even principal objective. Some of the privacy accorded to individuals by government flows from the inefficiencies we tolerate in the interest of pluralistic values. Efficiency must give way, in this case, to protecting privacy and preserving public confidence that statisticians can keep secrets, even, sometimes, from each other.

One obvious reason the Federal statistical environment is not as efficient as it might be is that it is decentralized. One difficulty of the present enclave proposal is that it affirms the value of a decentralized system of data production and attempts to institute record sharing through a set of decentralized controls. I am not sure we can have it both ways. If the benefits of record sharing are as great as alleged, then, perhaps, it is time to get on with the consolidation of statistical agencies and centers.

I would suggest that public perception of confidentiality might be less of an issue if the individual agencies with statistical record holdings were brought closer together. In this way, strong, clear, easily understood control of record access would provide unambiguous protection of confidentiality. More importantly, the public would correctly perceive this to be so. In contrast, the enclave proposal, through the complex provisions of an omnibus statute, would move the individual records farther away from the agency-specific statutes that protect them. However carefully worked out, this sort of provision is difficult to understand. It does not convey clear and unambiguous protection without significant and protracted study. We cannot reasonably expect that of the general public.

There is probably no doubt as to the Census Bureau position on the enclave proposal, but I will state it just to be sure. The benefits of broader access to the SSEL are considerable and well established. Except for this limited sharing of the SSEL for statistical purposes, we oppose any scheme that carries a risk of loss of public confidence and respondent cooperation. We see this risk embedded in the current proposals. The animus that accompanies this position is frustrating to the empirically minded and research oriented, but I believe it is warranted. It is evident, in fact, in the IRS paper.

Simply put, Census and IRS have the most to lose if the wrong gamble is taken. In each case, an enormous information gathering enterprise serves a vital government function, has the sanction of law to enforce compliance, and depends, in fact, on credibility and voluntary cooperation to be successful at any reasonable cost.

The population and housing census is like the historic village common. It is community property and all may use what it produces, but if it is overgrazed, it disappears. We should take

heed of what has happened recently in two European countries. In both the Netherlands and West Germany, the overgrazing has cost them the cancellation or postponement of the census. It is easy to say the circumstances and cultures are different, and they are. But as you read the papers on the same topics we are discussing today that were prepared for the 1983 Conference of European Statisticians, the themes are almost universal: the rapid expansion of administrative and other record systems containing sensitive (personal and business) information, real or perceived increases in the sharing of information between agencies and levels of government, the credibility of government actions, and the forward march of computers into virtually every aspect of life.

As to our own American dilemmas, I do think there are alternatives to the enclave and that they are not being pursued with sufficient vigor. I want to mention several developments that are already changing the way we do business. First is that over the past decade, our capacity to produce microdata files and public-use samples has been getting stronger. Some of these files are now richer in content than when this development started, and the files are more accessible to private, as well as public, researchers.

Second, the development of improved techniques for designing area samples, telephone surveys through random-digit dialing, and more accurate administrative record systems from which samples can be drawn, all suggest to me the gradually diminishing importance of the decennial census as the paramount resource from which other agencies obtain samples and data for follow-on statistical and analytical work.

In fact, I see the substantial increase of data sharing in the years ahead, as distinct from record sharing, simply because of technological change. The pace may be different for business data, because of the greater complexity of disclosure avoidance techniques, than for individual and household data, but I think the trend is similar.

Third, we should be able to do more than we are doing now to enrich data files through masking techniques. I have saved my comment on the Spruill paper on data masking, to savor a bit its contrast with the others. This paper is barely on the margin of the policy debate that has been underway about record sharing, but it is right near the center of a search for creative alternatives. Spruill is addressing one of the key questions; namely, enriching the data file designed for multiple use through masking techniques and assessing the risk of disclosure in quantitative terms so the file may be constructed accordingly. Good progress in this type of analysis should begin to eliminate some of the subjective judgments and arbitrary rules we now apply to file construction through cell suppression, grouping, and so forth. The Census Bureau needs to contribute more to this line of research.

To sum up, I am willing to speculate that 5 years from now at a session of this kind there will be 5 papers of the Spruill kind and none of the other kind, and those who would have written the other kind will regard the broad-scale en-

clave concept as an idea whose time came and went. But I will keep the other papers on file, just in case.

EDITOR'S NOTE:

For space reasons some of the specific discussions cited above from the Alexander and

Wilson-Smith papers have been modified or eliminated by the authors from the papers prepared for the Proceedings. Also, despite the discussant's comments, access to the SSEL outside the Census Bureau is not a position with which the Internal Revenue Service agrees (as pointed out in the original draft of the Wilson-Smith paper).