Journal of Official Statistics
Vol. 9. No. 2, 1993, pp. 335–337

Discussion
Disclosure Limitation and Data Access

Thomas Plewes[1]

The Bureau of Labor Statistics has a derived, shared, and direct interest in the papers by Lambert and Reynolds. First, the Bureau has a typical user's interest in derived confidentiality because we are the major sponsor and user of the Current Population Survey. We also have a shared interest, in that we share data with state agencies. And, third, we have an original or direct interest in confidentiality because we collect data from establishments and publish data directly. These papers provide a point of reference for us in all of these areas, and, thus, make an important contribution.

I have two general comments on the Reynolds paper before I get to the subject on which he makes his most pronounced contribution.

1. I agree that the usual case in Reynolds's taxonomy of risk is the case of moderate risk and significant consequence. If you fill in the blanks on Reynolds's matrix, you could agree on consequence. Unfortunately, that does not lead us very far in solving the problem. You should realize that we are speaking of significant risk to bureaucrats – namely, the risk of lawsuits. What Reynolds's labels "significant" may, for many of us, be outrageous. In the agencies, we categorize these cases as "moderate risk, outrageous consequence."

2. The discussion of informed consent follows the usual course, though I found especially interesting his postulate that elected officials can provide informed consent on the part of their constituents. In that case, I would have preferred the term – uninformed consent – since no informing is done, and I would have cautioned that elected officials do not usually weigh the risks and consequences as he is so careful to do.

Reynolds's foremost contribution, where he strikes my greatest interest, starts about half-way into his piece. First of all, he differentiates between individual and collective rights to privacy. Organizations are, indeed, different from people, and for all of the reasons that Reynolds suggests. He advances the state of understanding of the issues when he puts forward the notion of individual rights to privacy and organizational rights to confidentiality. Thus, he asks us to think of these as different creatures, with different sets of social rules that apply, different problems, and different procedures that should be taken into consideration. If take his analysis a step further, perhaps these differences merit individual treatment in other papers, because they reflect very different and distinct issues.

This is an unfinished symphony because of the marked absence of a third dimension. We should be talking in terms of domains, organizations, and uses. Let me

[1] Bureau of Labor Statistics, 2 Massachusetts Avenue, N.E., Washington, D.C. 20212, U.S.A.

explain why uses should be added to make this a three dimensional matrix. Consider the following: Are businesses concerned about the confidentiality of the information about the products and sales they provide to the Bureau of the Census because (a) the data might be used by the Bureau of Labor Statistics to identify candidates for solicitation into another data collection, or (b) that the data might be used by their competition to learn something about the way that they are organized and operate, or (c) the data will be used by the Internal Revenue Service or a regulatory agency to identify and develop data that can be used in legal action against the business? Well, the answer certainly is "all of the above." But, even though all of these possibilities form a valid basis for concern, that concern is surely weighted by the perceived seriousness of the consequence of the use of the data. It goes without saying that the fear of being put on a mailing list is less of a concern than the fear of being liable for triple damages due to an anti-trust action. (Although the fear of being put on a mailing list is growing, with good cause.) If we do not consider the ultimate uses, we will establish confidentiality policies that equate the risk associated with each of the potential uses, and all of the domains. When we equate risks, we tend to adopt policies that preclude access to all users for all uses, except those users in our own agencies, for whom we waive the rules.

Reynolds also makes the point that you need to understand the role of small business, and suggests the need for microdata to aid in that understanding. Unfortunately, just any data will not suffice. We need a lot of intrusive information on the business to understand the business. We need successor/predecessor coding, auxiliary coding, and other data that are not necessarily available on tax and other administrative databases. The addition of these data types requires statistical supplementation to the basic administrative databases. In our experience, to get statistical supplementation from those industries requires a pledge of confidentiality.

Let me conclude my discussion of Reynolds's paper with an initial reaction to his implementation strategy. He proposes establishing an organization to control the process that, I believe, already exists – an organization called the Office of Management and Budget. This office already has a heavy workload borne by a limited number of staff. Furthermore, the addition of these duties would imbue the function with powers that are already around – swearings, certification, review by panels, and so forth. What Reynolds has failed to address is what we do to throw researchers in jail if the trust is violated.

Diane Lambert provides the next chapter in an experimental "how-to" manual for insuring access while avoiding disclosure. She contributed to the literature by modelling the perspective of the rational intruder who makes optimal decisions, coming to the important conclusion that masking data to protect against the rational intruder will usually mean stripping the data of analytical utility. Thus, she recommends placing some of the burden of protecting the data on the researcher. The premise is well-taken if we are only concerned about the rational, though unscrupulous intruder. Intruders are certainly a danger, but so are innocent researchers who transport their files from one place to another, hackers who are looking for interesting data, bumblers, and outright incompetents. In fact, I suspect that an incompetent bumbler is more of a danger to protect against than a rational intruder, against whom we can devise protections and penalties. We know of a number of cases in which originally-

protected data are provided by researchers to their colleagues and research associates. In this manner, copies of the originally-protected files proliferate. This is not a spy operation. Still, it is something that has happened, and must be protected against.

Reynolds's unfinished symphony and Lambert's measures of harm are complementary. They lead us first to consider domains, organizations, and uses. Then, they consider the important notation of the value of the consequence of re-identification. With both of these contributions, we can begin to figure the elusive concept that descision makers need to know – total harm. We may never devise a comprehensive mathematical depiction of risk and harm, but these papers take us a long way along that road.