

**STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE**

**COMMISSION OF THE  
EUROPEAN COMMUNITIES**

**CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROSTAT**

**Joint ECE/Eurostat Work Session on  
Statistical Data Confidentiality**  
(Skopje, The former Yugoslav Republic of Macedonia,  
14-16 March 2001)

Room Paper No. 3  
English only

Topic III: Attitudes of respondents towards statistical confidentiality

## **ATTITUDES OF RESPONDENTS TOWARDS STATISTICAL CONFIDENTIALITY**

### **List of key issues for discussion**

Julia Lane

Professor of Economics, American University  
Principal Research Associate, The Urban Institute

#### **I. BACKGROUND**

1. There is a fundamental tension at the heart of every statistical agency's mission. Each is charged with collecting high quality data to inform national policy and enable statistical research. This necessitates dissemination of both summary and micro data. Each is also charged with protecting the confidentiality of survey respondents. This often necessitates blurring the data to reduce the probability of the re-identification of individuals. The tradeoff dilemma, which could well be stated as protecting confidentiality (avoiding disclosure) but optimizing access, has become more complex as both technological advances and public perceptions have altered in an information age.

2. Topic III addresses a core issue faced in this statistical mission. In particular, statistical agencies enter into an implicit pact with their respondents to protect their data – both because they are mandated to do this by law and because it is necessary to maintain the trust necessary for high quality responses (Sundgren and Greenia/Jensen). While most of the technical discussion at this workshop is targeted at addressing the technical issues associated with data protection, the papers in this topic shed new light on the ability of statistical agencies to reassure the public that statistical agencies have, in fact, delivered on their confidentiality promises.

3. Statistical agencies have historically relied on two approaches to do this: legal/ criminal penalties for the unauthorized data access and/or disclosure and technical protections for released data. However, as Gates points out, these approaches will only enable statistical agencies to maintain high quality responses if the public both knows about and believes in their protection methods. Historical practices may no longer suffice in the wake of new events such as technological change, especially computational progress, and changes in public attitudes to the privacy and confidentiality of their records.

#### **II. NEW EVENTS**

4. What are the other new events that are likely to have changed the landscape? A number of changes are cited by the authors: the increasing importance of administrative data (Greenia/Jensen), the propensity of statistical agencies to disseminate information over the internet (Gates), an increasing demand for micro-data by academic researchers (Kamen), an increasing demand for timely, relevant information by policy makers (Kamen and Sundgren), and possibly most important, the increasing data collection by the private sector (Greenia/Jensen and Sweeney).

5. It is worth expanding on this last because of the complex ways in which it may have changed the landscape of public perception. On the one hand, the vast expansion of private data collection creates a much more sophisticated master file for potential intruders. Sweeney (2001) gives a sense of just how vast private collections are in her example of Catalina Marketing. This company began in March 1996 to store the shopping patterns of 143 million shoppers each week from 11,000 supermarkets nationwide. By July 1998, they had amassed a 2-terabyte database with 18 billion rows of data by July 1998. The public may thus be more aware of and more concerned by the increased potential for re-identification. On the other hand, the widespread public awareness of private sector profiling may act to actually reduce privacy and confidentiality concerns among the public, if they believe that all information about them is already known. Without further research, the net effect of this change is not known.

6. The papers document a number of different current approaches that statistical agencies have taken to address these challenges.

### **III. CURRENT APPROACHES**

7. Two of the papers summarize current research on public perception (Gates and Greenia/Jensen). The U.S. Census Bureau has initiated research on household perceptions of the confidentiality protections in both the 1990 and 2000 Decennial Census, continuing a research agenda long associated with Singer. However, almost nothing is known about the perceptions of businesses in the treatment of their data – this gap is being filled by the new survey and resulting research by Greenia/Jensen.

8. Two other papers describe probably the most important new access modality that has been developed in the past decade – that of restricted access sites (Thorogood and Kamen). These permit statistical agencies to respond to the microdata needs of researchers, avoid the linkage problems posed by the internet, and address potential perception problems that might be associated with other access modalities. Indeed, the success of the Israeli site, and the Eurostat proposal to institute secure sites may prove to be pathbreaking endeavours to demonstrate to other statistical agencies how such sites can be operationalized.

9. A third approach is to examine the protection safeguards that are in place. These can be categorized in three ways: technical/legal, organizational and methodological (Sundgren). Sundgren describes some very innovative technical advances that are underway at Statistics Sweden to automate some of the decision making processes associated with data release. Kamen describes some interesting organizational approaches – particularly the institution of a confidentiality committee in Israel (which appears to be similar to the CNIL in France and the American Disclosure Review Boards).

### **IV. WHERE DO WE GO FROM HERE?**

10. Statistical agencies should routinely take stock of public perceptions of the way in which they protect different types of data – household, business and administrative data. The usefulness of this information is potentially great. If some data are less sensitive than others – either inherently, or as time goes by - then the public might be willing for those data to be less protected, and valuable resources can be used to focus on those data elements that are most sensitive. If the public has become more sensitive to data privacy concerns in response to private sector actions, then statistical agencies need to respond accordingly.

11. The assessment of public perceptions can be an important tool in guiding data dissemination decisions. For example, new access modalities such as restricted access sites, are potentially an important new dissemination mechanism. However, their establishment should be evaluated, not only for the technical protection and data quality issues raised as a result, but also in the context of public perception of these approaches. It may well be that the public perceives these as being much safer than the release of public use files, particularly given the wealth of private data available, and statistical institutes can tailor their response appropriately.

12. The new organizational methods can also be used not only to protect the data but to assist in public perceptions of the safety mechanisms used by statistical agencies. The use of either automated processing techniques (as is being proposed in Sweden) or relative small, but well structured, confidentiality committees (as has been implemented in Israel), could certainly be used to further reassure the public that their data are fully protected by their national statistical agencies.

## References

Gates, G “A Holistic Approach to Confidentiality Assurance in Statistical Data”, *2<sup>nd</sup> Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality*, Paper no. X 2001.

Greenia, N and B. Jensen “Business’ Perceptions of Confidentiality” *2<sup>nd</sup> Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality*, Paper no. , 2001.

Kamen, C. “Control of Statistical Disclosure versus the Needs of Data Users in Israel”, *2<sup>nd</sup> Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality*, Paper no. X, 2001.

Singer E., J. VanHoewyk and S. Presser “Public Attitudes Toward Data Sharing by Federal Agencies”, Chapter 8 in “ “, National Academy Press, 1996

Sundgren, Bo “ Statistical Microdata – Confidentiality Protection vs. Freedom of Information” *2<sup>nd</sup> Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality*, Paper no. , 2001.

Sweeney, Latanya “Information Explosion” mimeo, Carnegie-Mellon University, 2001 (b)

Sweeney, Latanya “Towards all the Data on All the People”, mimeo, Carnegie-Mellon University, 2001(a).