

ISSUES IN ACCESSING AND SHARING CONFIDENTIAL SURVEY AND SOCIAL SCIENCE DATA

Virginia A. de Wolf, Ph.D.

U.S. Department of Transportation, Bureau of Transportation Statistics, Office of Statistical Quality, K-24, Room 3430, 400-7th St., SW, Washington, DC 20590, USA
Email: virginia.dewolf@bts.gov

ABSTRACT

Researchers collect data from both individuals and organizations under pledges of confidentiality. The U.S. Federal statistical system has established practices and procedures that enable others to access the confidential data it collects. The two main methods are to restrict the content of the data (termed "restricted data") prior to release to the general public and to restrict the conditions under which the data can be accessed, i.e., at what locations, for what purposes (termed "restricted access"). This paper reviews restricted data and restricted access practices in several U.S. statistical agencies. It concludes with suggestions for sharing confidential social science data.

Keywords: Confidentiality, Data access, Data sharing, Disclosure, Identifiability, Microdata, Statistical disclosure limitation, Statistical uses of data, Tabular data.

1 INTRODUCTION

Data sharing practices vary in the social sciences. For example, some social sciences encourage sharing, e.g., see Section 13.05 of the American Sociological Association's Code of Ethics (American Sociological Association, 2002) while others do not. In the U.S., both institutions that fund most of the social science research, the National Institutes of Health (2002b) and the National Science Foundation (2002a), have statements on data sharing.

And even among social scientists who propose to share data, many do not understand the potential risks of re-identifying research participants. That is, while they understand the need to remove names, addresses, etc., and other "obvious identifiers," they do not realize that the remaining variables could be linked with other data sources and possibly lead to the re-identification of those who participated in their research. Social science researchers need to be educated about the benefits of sharing and the protections that are needed. The purpose of this paper is to review practices and procedures employed by the U.S. Federal statistical agencies and suggest ways in which they can be used by the social science research community.

The U.S. Federal statistical system is decentralized and comprised of over 70 agencies (Office of Management and Budget [OMB], 2002). These agencies collect data from individuals and organizations (such as schools, hospitals, and businesses) to inform policy decisions and for research. They also have an obligation to disseminate these data in a responsible manner. Over the past 12 years, several groups in the U.S. have examined the role of the Federal government as a "data steward." The bulk of this paper is drawn from the reports of three such groups: the Panel on Confidentiality and Data Access (Duncan, Jabine, & de Wolf, 1993; Jabine, 1993a, 1993b); the Subcommittee on Disclosure Limitation Methodology (Federal Committee on Statistical Methodology, 1994); and the Federal Committee on Statistical Methodology's (FCSM) Confidentiality and Data Access Committee (CDAC).

The Panel on Confidentiality and Data Access was the first to provide generic labels for the two main alternatives that U.S. Federal statistical agencies use to protect the confidentiality of data that they collect. These are:

- "Restricted data" -- to restrict the content of the data prior to releasing it to the general public and
- "Restricted access" -- to restrict the conditions under which the data can be accessed, i.e., who can have access, at what locations, for what purposes.

Of course, sometimes a combination of these two approaches are used when sharing a data set that poses various risks of disclosure.

And, as noted in the 1994 report of the Subcommittee on Disclosure Limitation Methodology (called the "Subcommittee" below):

Regardless of the basis used to protect confidentiality, federal statistical agencies must balance two objectives: to provide useful statistical information to data users, and to assure that the responses of individuals are protected. (FCSM, 1994, p. 6)

At the outset I need to state that there is no such thing as a "zero risk" of disclosure (the only way to have no risk is to not collect data). However, most Federal agencies work hard to keep this risk as low as possible.

The next sections describe some of the restricted data approaches and restricted access procedures used by Federal statistical agencies. It concludes with some suggestions for the social science research community about sharing confidential data. All sections are quite brief -- see references for additional information.

2 RESTRICTED DATA

The type of data product to be released dictates the choice of statistical methods used to limit disclosure. Most data are released as either tables or microdata files. A microdata file is a computerized file that "...consists of individual records, each containing values of variables for a single person, business establishment or other unit" (FCSM, 1994, p. 3). Confidential data that are obtained from organizations are typically released in tabular format; rarely do the agencies release microdata from organizations -- the risk of re-disclosure is too great, given the amount of information that exists in the public domain. Data from individuals are released in either format. The Subcommittee's report contains a "primer" on these methods and is an excellent resource (see Chapter 2 in FCSM, 1994).

Once the tables and/or public use microdata are released to the public there are no restrictions on how the data may be used. For example, marketing firms often use decennial census data for nonstatistical purposes.

Prior to releasing a restricted data product, agencies assess the level of protection afforded the confidential information; this is done through a formally or informally designated unit called a Disclosure Review Board. For more information about these Boards, consult the papers presented at the Joint Statistical Meetings, August 17, 2000, Indianapolis, Indiana, in the "Panel on Disclosure Review Boards of Federal Agencies: Characteristics, Defining Qualities and Generalizability " (CDAC, 2000).

CDAC (1999) created a "Checklist on Disclosure Potential of Proposed Data Releases" based on the practices of several agencies. It contains three subsections: one for microdata files and two for tables (one for data collected from individuals, the other for data collected from organizations). The Checklist is one tool that can assist agencies or organizations in reviewing disclosure-limited data products. Completed Checklists should be submitted to the Disclosure Review Board for review prior to releasing the tables/microdata. CDAC encourages organizations to adapt and modify the Checklist as needed.

2.1 Tables

If information is collected on a census, one way of preserving confidentiality is to release only tables based on a sample. Regardless of whether the data are a census or sample, the cells in a table should not be "too" small (some agencies require a minimum of 3 entries per cell while others require 5). If a table that is to be released contains "small" cells, agencies often "suppress" these cells to protect confidentiality. Of course, when you suppress a value in a row, you must also suppress values in one or more other row(s) and column(s) so that the suppressed value can not be obtained by subtraction from the row/column totals. Appropriate statistical methods must be used (for more information see FCSM, 1994).

Sometimes the resulting "suppressed" table contains too many "blank" cells to be of value to data users. To provide more information, some agencies have developed alternative policies that would enable "small" cells to be published. For example, the National Agricultural Statistics Service (NASS) has a policy that allows its data providers to "waive" the confidentiality protection so that small cells can be published (NASS, 1996).

Agencies, such as NASS, produce "special tables" for data users. Once a special tabulation is approved and provided to the requester it is made available on NASS's web site (NASS, n.d.). This enables use of the table by a broader audience. It is not prudent to release special tabulations only to requesters for the following reason. That is, two individuals might collude and ask for special tables that are related in some fashion; once they obtain the tables then they could potentially subtract one from the other and "learn" about sensitive information that was suppressed.

2.2 Microdata Files

Creating a public use microdata file is as much an art as a science since the methods used to protect confidentiality are varied and often depend on the type of data that underlies the microdata files. The first step is to remove all personal identifiers and this seems obvious until you ask yourself the question: What is identifiable? CDAC (July, 2002) has provided recommendations for answering this question.

Even after personal identifiers have been removed the remaining information could be used to infer the identity of an individual. The next step is to alter variables so that the possibility of inferential disclosure is reduced. Examples of methods used by the agencies are:

- Releasing a random subsample;
- Limiting geographic detail;
- Reducing the number of "unusual cases" (examples of methods used include rounding, recoding categorical responses, using ranges for age rather than exact age or date of birth, and top- or bottom-coding variables); and
- Increasing the uncertainty associated with data (i.e., data swapping, adding random noise).

Computationally intensive statistical methods are also used, e.g., multiple imputation (Little and Rubin, 1987). For instance, the Federal Reserve Board's Survey of Consumer Finances uses multiple imputation as a disclosure-limiting technique (Federal Reserve Board, 2002; Rubin, 1993; Kennickell, 1998).

The proliferation of data available via the internet (Sweeney, 2001) and the low cost of high-powered computers make it imperative that re-identification assessments be conducted prior to the release of microdata. In fact, it is advisable to periodically examine earlier data releases to determine whether or not microdata files which were once deemed "protected" can inadvertently be re-identified (some agencies have contracted with "hackers" to do this). CDAC's (July, 2002) "Identifiability" paper, cited above, provides some useful starting points. For example, a recent article in the *New York Times* (Lee, 2002) described a case where the clerk of the court for one county in Ohio put county records on the internet -- including Social Security Number, financial data, and divorce proceedings.

2.3 Additional References

Domingo-Ferrer (2002) and Doyle et al. (2001) both contain an excellent series of papers on protecting tabular data as well as creating microdata. The Digital Government Project at the National for Statistical Sciences (2002), funded by several Federal statistical agencies, is examining a web-based disclosure system and its web site is a valuable resource. In addition, the European-based project, Computational Aspects of Statistical Confidentiality (2003), has developed software programs, *tau*-ARGUS and *mu*-ARGUS, to aid in the creation of disclosure-proofed tables and microdata files, respectively.

3 RESTRICTED ACCESS

Some of the confidential data collected by Federal statistical agencies cannot be released as either tables or microdata files in a format that would be useful for research. In such cases, agencies have developed a set of

administrative procedures to enable research use of these data. Agencies place restrictions on the use of the data (for statistical purposes but not for regulatory, judicial, or other administrative purposes); conditions of access (e.g., location); whether or not data can be linked (and if so, who does the linking); and so forth. Below are descriptions of three examples.

The Census Bureau (2003) pioneered Research Data Centers (RDCs) which were first used to enable researchers' access to its economic microdata. The National Science Foundation was involved in establishing this Census Bureau program. There are six RDCs at this time.

Other U.S. statistical agencies with RDCs include the National Center for Health Statistics (2002) and the Agency for Healthcare Research and Quality (2002). For information on these RDCs, as well as a comparison among the three agencies, see CDAC's (April, 2002) "Restricted Access Procedures" paper. An initiative of Statistics Canada (2002) created the Canadian Research Data Centres Program and nine RDCs are currently operating.

RDCs enable users to access the data at a site controlled by an agency and staffed by employees. Research projects must be approved by the agency; each researcher enters into a formal agreement with the agency and often covers the costs associated with the work (e.g., computer charges, rental of space). Other "typical" characteristics include:

- Use of "stand alone" workstations that do not have floppy disk drives or CD readers and are not connected to the internet or any agency network;
- Restrictions on linking data (in general if a linkage is approved it will be done by agency staff);
- Inspection of all materials removed from the Center;
- Limitations on the types of analyses; and
- Disclosure review of researchers' output.

Remote access to data is a second example. The National Center for Health Statistics' (NCHS) remote access system is handled by its RDC and has two components (NCHS, 2002). After a proposal is approved, RDC staff develops a "pseudo" data file which has the statistical properties of the actual data file. This fictitious file is then sent to the researcher who uses it to debug computer programs. Once debugged, the files are then sent to NCHS by email. NCHS requires users to submit programs in SAS and all programs are automatically scanned upon arrival for non-allowable commands (certain SAS procedures are disabled). The output is reviewed before it is emailed back to the researcher. Of course, remote access and on-site analyses at the RDC can be combined.

A final example is licensing or data use agreements that allow researchers to use non-public data at their home institution. Seastrom's paper (2001) is an excellent summary of the current status of the use of licenses in a number of U.S. agencies. Using the National Center for Education Statistics (NCES) as an example, she delineates the components of its licensing application that include:

- Formal letter of request (e.g., who will use the data, a description of the planned statistical use of the data, specification of the time period for the loan of the restricted use data file);
- License documentation (i.e., a legal agreement signed by the researcher, a senior official at the researcher's institution, and NCES's commissioner);
- Security plan at the home institution (NCES has specified a list of requirements); and
- Affidavits of nondisclosure that are signed by each data user.

Users must follow NCES publication requirements when publishing results from restricted use data, agree to unannounced and unscheduled on-site inspections by NCES's contractor, and return restricted use data files once the project is completed.

4 SUGGESTIONS FOR THE SOCIAL SCIENCES IN SHARING CONFIDENTIAL DATA

As was stated in the session overview of Track I-C-4 to CODATA 2002: "The behavioral sciences have not had a tradition of data sharing." The examples of data sharing in that session from several U.S. Federal agencies, such as

the National Institutes of Health (Kelty, 2002) and the National Science Foundation (Rubin, 2002), as well as the paper on the Murray Research Center (James, 2002), provide excellent illustrations of pioneering data sharing programs in disciplines where data sharing has been slow to take hold.

Below I offer suggestions for sharing confidential social science data to two broad audiences: professional associations and then educational institutions.

4.1 Suggestions for professional associations

Associations need to inform their members about the methods used to protect confidential data by providing resource materials (e.g., on the associations' web sites). Information about the restricted data and restricted access methods employed by the U.S. Federal statistical agencies are neither widely publicized nor well-known among behavior and social scientists. Information to be provided should include:

- Descriptions of restricted data methods used to protection confidentiality and links to Federal resources (ex., CDAC) as well as web sites from other countries (e.g., Canada, Eurostat, and Statistics Netherlands). This may require the modification of materials so that the examples are relevant to the discipline represented by the association;
- Explanations of restricted access approaches that are currently in use by Federal statistical agencies and links to relevant web sites (such as Census and NCHS); and
- Links to other pertinent social science resource material, e.g., National Science Foundation (2002b) has recently posted a document on regulatory and Institutional Review Board issues that contains a subsection on protecting confidentiality.

In addition, associations should sponsor short courses at their annual meeting that focus on "restricted data" and "restricted access" approaches. CDAC members have given short courses to various audiences. One suggestion would be to ask members of CDAC to give such a short course or to ask CDAC to work with members of the association to tailor CDAC's short course materials for the membership.

Another important educational role that associations could play is to keep abreast of Federal laws/regulations on confidentiality-protecting mechanisms as well as data sharing issues and provide information to their members. For example,

- Describe "Certificates of Confidentiality" which are used to prevent compelled disclosure in court of law and are available from the Department of Health and Human Services (National Institutes of Health, 2002a) irrespective of the source of funding for the project;
- Include links to the data sharing statements/policies by the two agencies that fund the bulk of social science research grants, National Institutes of Health (2002b) and National Science Foundation (2002a); and
- If many of the association's members get Federal grants to fund their research, include an explanation of the 1999 amendments to OMB Circular A-110, which governs grants to not-for-profit institutions (including institutions of higher education and hospitals). This change makes grant data vulnerable to a Freedom of Information request (OMB, 1999), i.e.,

Subpart C, 36(d)(1): "In addition, in response to a Freedom of Information Act (FOIA) request for research data relating to published research findings produced under an award that were used by the Federal Government in developing an agency action that has the force and effect of law, the Federal awarding agency shall request, and the recipient shall provide, within a reasonable time, the research data so that they can be made available to the public through the procedures established under the FOIA. If the Federal awarding agency obtains the research data solely in response to a FOIA request, the agency may charge the requester a reasonable fee equaling the full incremental cost of obtaining the research data. This fee should reflect costs incurred by the agency, the recipient, and applicable subrecipients...."

In addition, it might be useful to include examples of how such grant recipients are facilitating access. For instance, the University of Michigan's Health and Retirement Survey has restricted access agreements (University of Michigan, n.d.) and also supports a data enclave (Michigan Center on the Demography of Aging, n.d.).

Last, but not least, it would be important to describe "re-identification" assessments and to encourage members to conduct such assessments prior to releasing a new microdata file. The value of doing such assessment on microdata files that were released at an earlier point in time should also be explained.

4.2 Some suggestions for educational institutions

Many of the suggestions for professional associations would be germane to educational institutions. For instance, there is need to educate researchers on campus about restricted access and restricted data methods. These could be posted on a web site.

A related activity would be to create a cross-disciplinary Disclosure Review Board that could review tables and microdata created from confidential data collected from individuals and organizations. Such a Board could provide a valuable service to the academic community by assessing the level of protection and making recommendations to enhance the protection of confidential data. Expertise that is available in the various disciplines on campus should be pooled. The Board could also serve an educational function by making researchers aware of the various techniques that are available. The Board could adapt CDAC's (1999) Checklist or create its own tool(s).

Determine whether your university's Institutional Review Board (IRB) has formalized a process for review of output from data collected under a pledge of confidentiality. If not, then perhaps a cross-disciplinary Disclosure Review Board could serve as an ad hoc committee to make recommendations to the IRB about release.

Last, but not least, researchers who receive Federal grants (and, therefore, are governed by OMB Circular A-110, [OMB, 1999]) should learn what their universities' reaction/response to the 1999 changes to this Circular. What plan of action has the institution's legal office developed if faculties' data are subject to a Freedom of Information Act request based on use of grant data by the Federal government? How are the various disciplines on campus dealing with the changes to Circular A-110? Such information should be broadly disseminated across campus.

One final suggestion is to create a cross-disciplinary Research Data Center for on-site analysis of confidential data. Of course, one major question is funding. Perhaps, the institutions that fund most of the social science research (National Science Foundation and National Institutes of Health) could be persuaded to provide money to establish such Centers.

5 CONCLUSIONS

This paper reviewed several methods used by U.S. Federal statistical agencies to enable researcher access to confidential data. It provides illustrations of restricted data and restricted access procedures used by several agencies and offers suggestions for social science researchers to enhance the sharing of confidential data. As an overview, it does not provide detailed presentations of the procedures; such details can be obtained in the references.

6 NOTE

This paper reflects the views of the author and not necessarily the view of the U.S. Department of Transportation's Bureau of Transportation Statistics.

7 ACKNOWLEDGEMENTS

The author thanks Arthur Kennickell, Federal Reserve Board, Alvan Zarate, National Center for Health Statistics, and Laura Zayatz, U.S. Bureau of the Census, for their helpful comments on earlier drafts of this paper.

8 REFERENCES AND RESOURCES

Agency for Healthcare Research and Quality. (2002) CCFS Data Center. Retrieved February 10, 2003, from the AHRQ site: <http://www.ahrq.gov/about/ccfs/ccfsdata.htm/>

American Sociological Association (2002) American Sociological Association Code of Ethics. Retrieved February 10, 2003, from the ASA site: <http://www.asanet.org/members/ecostand2.html/>

Census Bureau (2003) Homepage of the Census Bureau's Research Data Center Program. Retrieved February 10, 2003, from the Census Bureau site: <http://www.ces.census.gov/ces.php/rdc/>

Computational Aspects of Statistical Confidentiality (2003) Homepage of Computational Aspects of Statistical Confidentiality. Retrieved February 10, 2003, from the CASC site: <http://neon.vb.cbs.nl/casc/default.html/>

Confidentiality and Data Access Committee (July, 1999) Checklist on Disclosure Potential of Proposed Data Releases. Retrieved February 10, 2003, from the CDAC site: http://www.fcsn.gov/committees/cdac/checklist_799.doc/

Confidentiality and Data Access Committee (2000) Panel on Disclosure Review Boards of Federal Agencies: Characteristics, Defining Qualities and Generalizability. *2000 Joint Statistical Meetings*, Indianapolis, Indiana. Retrieved February 10, 2003, from the CDAC site: <http://www.fcsn.gov/committees/cdac/DRB-Panel.doc/>

Confidentiality and Data Access Committee (July, 2002) Identifiability in Microdata Files. Retrieved February 10, 2003, from the CDAC site: <http://www.fcsn.gov/committees/cdac/IdentCDAC3.doc/>

Confidentiality and Data Access Committee (April, 2002) Restricted Access Procedures. Retrieved February 10, 2003, from the CDAC site: <http://www.fcsn.gov/committees/cdac/cdacra9.doc/>

Domingo-Ferrer, J. (Ed.) (2002) *Inference Control in Statistical Databases, From Theory to Practice. Lecture Notes in Computer Science, Volume 2316*. Heidelberg: Springer-Verlag.

Doyle, P., Lane, J.I., Theeuwes, J.J.M., & Zayatz, L.V. (Eds.) (2001) *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. Amsterdam, The Netherlands: Elsevier Science B.V.

Duncan, G.T. Jabine, T.B., & de Wolf, V.A. (Eds.) (1993) *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*. Washington, DC: National Academy Press.

Federal Committee on Statistical Methodology. (May, 1994) *Report on Statistical Disclosure Limitation Methodology. Statistical Policy Working Paper # 22*. Prepared by the FCSM's Subcommittee on Disclosure Limitation Methodology. Retrieved February 10, 2003, from the FCSM site: <http://www.fcsn.gov/working-papers/spwp22.html/>

Federal Reserve Board (2002) Homepage of the Survey of Consumer Finance. Retrieved February 10, 2003, from the FRB site: <http://www.federalreserve.gov/pubs/oss/oss2/scfindex.html/>

Jabine, T.B. (1993a) Procedures for Restricted Access, *Journal of Official Statistics*, 3(2), 537-589.

Jabine, T.B. (1993b) Statistical Disclosure Limitation Practices of United States Statistical Agencies, *Journal of Official Statistics*, 3(2), 427-454.

James, J.B. (2002) Acquisition Criteria at the Murray Research Center: A Center for the Study of Lives. *18th International Conference of CODATA*, Montreal, Canada.

Kennickell, A. (September, 1998) Multiple Imputation in the Survey of Consumer Finances. Retrieved February 10, 2003, from the Federal Reserve Board site: <http://www.federalreserve.gov/pubs/oss/oss2/papers/impute98.pdf/> and <http://www.federalreserve.gov/pubs/oss/oss2/papers/impute98.fig.pdf/>

Kelty, M.F. (2002) Data Sharing at NIH and NIA. *18th International Conference of CODATA*, Montreal, Canada.

Lee, J. (September 5, 2002) Dirty Laundry, Online for All to See, *New York Times*, p E1 & E7.

Little, R.J.A. & Rubin, D.B. (1987) *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Michigan Center on the Demography of Aging (n.d.) MiCDA Data Enclave. Retrieved February 10, 2003, from MiCDA site: <http://micda.psc.isr.umich.edu/enclave/>

National Agricultural Statistics Service, U.S. Department of Agriculture (October 1, 1996) *Standard for Suppressing Data Due to Confidentiality*. Policy and Standards Memorandum No. 12-96. Washington, DC: National Agricultural Statistics Service.

National Agricultural Statistics Service, U.S. Department of Agriculture (n.d.) USDA-NASS Data Lab. Retrieved February 10, 2003, from the NASS site: <http://www.usda.gov/nass/nassinfo/datalab/datalabs.htm/>

National Center for Health Statistics (2002). Homepage of the NCHS Research Data Center. Retrieved February 10, 2003, from the NCHS site: <http://www.cdc.gov/nchs/r&d/rdc.htm/>

National Institutes of Health (2002a) Certificates of Confidentiality Kiosk. Retrieved February 10, 2003, from the NIH's Office of Extramural Research site: <http://grants2.nih.gov/grants/policy/coc/index.htm/>

National Institutes of Health (2002b) NIH Draft Statement on Sharing Research Data. Retrieved February 10, 2003, from the NIH's Office of Extramural Research site: http://grants2.nih.gov/grants/policy/data_sharing/index.htm/

National Institute for Statistical Sciences (2002) Digital Government Project Web Site. Retrieved February 10, 2003, from the NISS site: <http://www.niss.org/dg/index.html/>

National Science Foundation (2002a) Data Archiving Policy of the Division of Social and Economic Sciences. Retrieved February 10, 2003, from the NSF site: <http://www.nsf.gov/sbe/ses/common/archive.htm/>

National Science Foundation (2002b) Frequently Asked Questions and Vignettes - Interpreting the Common Rule for the Protection of Human Subjects for Behavioral and Social Science Research. Retrieved February 10, 2003, from the NSF site: <http://www.nsf.gov/bfa/dga/policy/hsfaqs.htm/>

Office of Management and Budget (1999) Uniform Administrative Requirements for Grants and Agreements With Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations. OMB Circular A-110. Retrieved February 10, 2003, from the OMB site: <http://www.whitehouse.gov/omb/circulars/a110/a110.html/>

Office of Management and Budget (September, 2002) *Statistical Programs of the United States Government. Fiscal Year 2003*. Retrieved February 10, 2003, from the OMB site: <http://www.whitehouse.gov/omb/inforeg/03statprog.pdf/>

Rubin, D.L. (1993) Discussion: Statistical Disclosure Limitation, *Journal of Official Statistics*, 3(2), 461-468.

Rubin, P. (2002) Data Sharing at the Social and Behavioral Sciences at the National Science Foundation. *18th International Conference of CODATA*, Montreal, Canada.

Seastrom, M.M. (2001) Licensing. In Doyle, P., Lane, J.I, Theeuwes, J.J.M., & Zayatz, L.V. (Eds.), *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. (pp 279-296.) Amsterdam, The Netherlands: Elsevier Science B.V.

Statistics Canada (2002) The Research Data Centres Program. Retrieved February 10, 2003, from the Statistics Canada site: <http://www.statcan.ca/english/rdc/index.htm/>

Sweeney, L. (2001) Information Explosion. In Doyle, P., Lane, J.I, Theeuwes, J.J.M., & Zayatz, L.V. (Eds.), *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. (pp 43-74.) Amsterdam, The Netherlands: Elsevier Science B.V.

University of Michigan (n.d.) Obtaining Restricted Data from the Health and Retirement Study. Retrieved February 10, 2003, from the HRS site: <http://hrsonline.isr.umich.edu/rda/>