



- (5) Roll or scroll data is rolled up or down the screen, permits the user to scan a large volume of data
- (6) Paging - Data is stored on pages (a full screen) user is able to review any selected page.

Editing Functions

Editing features include:

- (1) Character deletion
- (2) Line insertion
- (3) Line deletion
- (4) Erase
- (5) Character repeat.

External I/O devices can add flexibility to the applications possibilities for display terminals. A cassette tape drive or diskette drive can be used to store display formats, data to be transmitted, or user programs. A printer can provide hard copy.

Selecting a Terminal

Some questions you should ask yourself when selecting a display terminal are:

- (1) What are the essential parameters for a display terminal that will satisfy your needs?
- (2) Who supplies the terminals with the features you desire?
- (3) Maintenance provisions?
- (4) Talk to users concerning problems encountered when installing it, failures that have occurred, and any incompatibilities.

DISCUSSION PAPER / alice robbin

The Issue of Confidential Data: The Need for Formulation of Policy by the Data Archive and Library

by

Alice Robbin
Data and Program Library Service
University of Wisconsin-Madison

The Issue of Confidential Data

I.

During the last decade there has been increased concern about the problems of confidentiality involved in the collection and dissemination of individual microdata. Concern has revolved around the government's perceived need to collect increasing amounts of information at a microdata level for social policy formulation and evaluation, types of information which potentially compromise

personal privacy; lack of good security measures for protecting access to personal information; availability of this information to individuals, private firms and public agencies unrelated in any way to the primary producers of the statistical data; growing restrictions placed on the access to statistical microdata by the statistical agencies, restrictions which appear to be in response to public pressure; and, the impact of restrictions on access to these data which are needed for research and public policy development by the scholarly community and others engaged in statistical analysis. This concern has been manifested in a very large literature, formal conferences and commissions, and legislation.¹ Recently, a meeting concerning the issues of privacy, confidentiality, and the use of government microdata for research and statistical purposes was held at the Rockefeller Foundation's Bellagio Study and Conference Center at Lake Como, Italy, August 16-20, 1977.^{2,3} It seems useful to summarize the draft Report on this conference because of the importance of this issue for the data archive.

II.

David Flaherty, author of the draft Report describes the origins of the Bellagio meeting. During the last three years he and Edward Hanis of the University of Western Ontario have been involved in the Privacy Project studying the problems of privacy and confidentiality involved in the collection and dissemination of individual microdata by central statistical agencies in Canada, the United States, the Federal Republic of Germany, Sweden, and the United Kingdom. The Privacy Project identified key personnel in each central statistical agency and found that they face common problems in the formulation of policy on confidentiality and data dissemination. The Privacy Project found that although these individuals face common problems they have had surprisingly little contact with one another on these issues. The Bellagio Conference brought together leading individuals from each of the five statistical agencies to discuss common problems and solutions. The Conference included sociologists, economists, and historians because it was felt necessary to demonstrate to the "custodians" of government data that a real need exists for access to individual microdata for research and statistical purposes. A series of general principles or propositions pertaining to research and statistical uses of individual information held by government agencies evolved during the meeting.

I have reproduced the 18 principles in toto, since I feel it useful to add to the public debate on the privacy/confidentiality issue. Many of the principles, while of potential interest to the data archive community do not directly affect data archivists at a local repository level; however, I have starred (*) those which I think may prove applicable to the archive's role.

1. *National statistical offices should provide researchers both inside and outside government with the broadest possible access to information within the bounds of accepted notions of privacy and legal requirements to preserve confidentiality.*
2. *Legal and social restraints on the dissemination of microdata are appropriate when they reflect the interests of the general public in an equitable manner. These constraints should be re-examined when they result in the protection of vested interests or the failure to disseminate anonymized information for statistical and research purposes without direct consequences for a specific individual.*



3. All copies of government statistical records should be rendered immune from compulsory legal process by statute.
4. In making data available to researchers national statistical offices should provide some means to ensure that decisions on selective access are subject to independent review and appeals.
- *5. The distinction between a research file in the sense of a statistical record (as defined in the 1977 report of the U.S. Privacy Protection Study Commission), and other micro files is fundamental in discussions of privacy and dissemination of microdata. All dissemination of government microdata discussed in connection with the Bellagio Principles is assumed to be a transfer of data to research files.
6. There are valid and socially-significant fields of research for which access to microdata is indispensable. One of the prime sources is government microdata, including statistical agencies.
- *7. Public use samples of anonymized individual data are one of the most useful ways of disseminating microdata for research and statistical purposes.
- *8. Techniques now exist that permit preparation of public use samples of value for research purposes within the constraints imposed by the need for confidentiality. Countries with strict statutes on confidentiality have prepared public use samples.
- *9. There are legitimate research purposes requiring the use of individual data for which public use samples are inadequate.
- *10. There are legitimate research uses which require the utilization of identifiable data within the framework of concern for confidentiality.
- *11. Other techniques of extending to approved research the same rights and obligations of access enjoyed by officers of the agency need to be considered in terms of better access.
- *12. There is considerable potential for development of more economical and responsive customized-user services, such as 1) record linkage under the protection of the statistical office, 2) special tabulations, 3) public use samples for special purposes. Such services must often involve some form of cost recovery.
- *13. Some research and statistical activities require the linking of individual data for research and statistical purposes.
- *14. Professional organizations or national organizations should have codes of ethics for their disciplines with respect to the utilization of individual data for research and statistical purposes. These ethical codes should furnish mutually agreeable standards of behavior governing relations between providers and users of governmental data.⁴
- *15. Users of microdata should be required to sign written undertakings for the protection of confidentiality.

- *16. *Considerable efforts should be made to explain to the general public the procedures in force for the protection of the confidentiality of microdata collected and disseminated for research and statistical purposes.*
17. *The right of privacy is evolving rather than static, and closely related to how statistics and research are perceived. Therefore, statisticians and researchers have a responsibility to contribute to policy and legal definitions of privacy.*
18. *Public concern with privacy in the collection and utilization of individual data can be addressed in part as follows:*
- 1) *voluntary data collection, whenever practicable*
 - 2) *advanced general notice to respondents and informed consent, whenever practicable*
 - 3) *provisions for public knowledge of data uses*
 - 4) *public education on the distinction between administrative and research uses of information.*

Flaherty summarizes the discussion on each of these principles in a concise form. Appendix I includes the conference participants. Appendix II summarizes the themes which emerged from the sessions on the United Kingdom (public relations, uses of information, and dissemination of data) and Canada (dissemination and uses of data). Appendix II describes the agenda for the General Sessions: goals, disseminating data, regulation of dissemination, forms of microdata dissemination, accountability of researchers, and privacy and data collection. Appendix IV lists the 18 principles agreed upon by the participants.

III.

Data archives are concerned with the preservation of information. Not only are there technical issues related to preservation and release of statistical information, but also ethical, moral and legal considerations about the release of confidential information on individuals. What should be the role of the data archive in accepting information of a confidential nature? Should the archive agree to preserve this information? What sorts of restrictions on access should archives adhere to? If the archive is to distribute confidential information, what sorts of protection should be applied to this information?

These and other questions concerning the maintenance and release of confidential information are questions which seem to have been raised by few data archives and organizations outside the federal statistical agencies such as the Bureau of the Census. Most of the information which the data archive has dealt with has been of an anonymous nature and with few exceptions little confidential data have been deposited with the data archive. [This discussion does not pertain to the National Archives, which has had a long history of concern with this issue and has developed various mechanisms for dealing with the problem.] Nevertheless, confidential information is being collected and archives as such do play a role as preservers of this information. In addition, there are a number of data organizations, integral parts of (survey) research organizations, which are responsible for maintaining the data files created by their researchers; there are data libraries, which although do not have a mission to preserve original data, sometimes (because of their experience with data generation and processing) become involved in projects where confidential data are collected.



At the same time public concern about the release of confidential information by statistical agencies has led to restrictions on access to information needed for research and public policy development by the data archive's clientele.

This suggests that the issue is of immediate and continuing attention by the data archiving community. What to do about confidential data, what is the role of the archive, and what are the archive's responsibilities in this area are difficult questions and have no easy solutions. The problem does suggest however that it would be wise for a data archive to formulate a coherent policy on the preservation and release of statistical information which potentially invades individual privacy.

While I have not suggested any guidelines that a data archive can follow for formulating a policy on the archiving and dissemination of confidential data, I hope that this discussion will stimulate further thoughts by the archive community and a continuing dialogue in this Newsletter.

Footnotes and References

¹Some introductory materials to the issue of privacy/confidentiality are:

- 1) The House and Senate hearings held during the middle 1960's, which dealt with the creation of a Federal (National) Data Center, contain a great deal of discussion on the confidentiality problems inherent in a government data center.
 - U.S. Congress. House. Committee on Government Operations. Special Subcommittee on Invasion of Privacy. The Computer and Invasion of Privacy. Hearings, 89th Cong., 2d sess. Washington: G.P.O., 1966. 318 pp. (Appendix 1: "The Ruggles Report," pp. 195-253;
 - U.S. Congress. Senate. Committee on the Judiciary. Subcommittee on Administrative Practice and Procedure. Computer Privacy. Hearings, 90th Cong., 1st sess., on S. Res. 25. Washington: G.P.O., 1967. 388 pp. ("The Kaysen Report," pp. 357-359.)
 - U.S. Congress. House. Committee on Government Operations. Privacy and the National Data Bank Concept. 90th Cong., 2d sess., H. Rept. 1842. Washington: G.P.O., 1968. 34 pp.
- 2) The Alan Westin book, Privacy and Freedom (New York: Atheneum, 1967), provides a basic introduction to the problem. See also his book Data Banks in a Free Society (New York: Quadrangle Books, 1972).
- 3) It is useful to consult the various pieces of legislation which have been passed in the United States, such as the Privacy Act of 1974, the Freedom of Information Act, Federal Reports Act, Fair Credit Reporting Act, and the Family Educational Rights and Privacy Act of 1974. Summaries of the contents of these Acts are contained in Kent Greenawalt's monograph, Legal Protections of Privacy, Final Report to the Office of Telecommunications Policy (Executive Office of the President). (Washington, D.C.: Superintendent of Documents, U.S. Government Printing Office).

- 4) Legislation has been considered and promulgated in other nations. Orjar Oyen, in his article, "Social Research and the Protection of Privacy: A Review of the Norwegian Development" (Acta Sociologica: 19 [1976], pp. 249-262), discusses what is occurring in Norway. The impact of the Swedish Data Act is discussed in the proceedings of a symposium on Personal Integrity and the Need for Data in the Social Sciences, held at Håsselby Slott, Stockholm, March 15-17, 1976, sponsored by the Swedish Council for Social Science Research. Other discussions have taken place and are reported in proceedings: Stig Stromholm, Right of Privacy and Rights of the Personality: A Comparative Survey (Stockholm: P.A. Norstedt and Söners Förlag, 1967), which is a working paper prepared for the Nordic Conference on Privacy organized by the International Commission of Jurists, held in Stockholm, May 1967. See also The First International Oslo Symposium on Data Banks and Society, by Universitetsforlaget in Oslo in 1972.
- 5) For discussions about statistical techniques to protect confidential information, see the Gwendolyn B. Moore et al. report, Accessing Individual Records from Personal Data Files Using Non-Unique Identifiers, prepared for the Institute for Computer Sciences and Technology of the National Bureau of Standards in Washington, D.C. (NBS Special Publication 500-2, U.S. Department of Commerce, National Bureau of Standards); Appendix A ("Confidentiality-Preserving Modes of Access to Files and to Interfile Exchange for Useful Statistical Analysis" by Donald T. Campbell et al.) in Protecting Individual Privacy in Evaluation Research, published by the National Academy of Sciences (Washington, D.C., 1975) (included in this monograph is an excellent bibliography); Proceedings of the Second Midwest Conference on Confidentiality of Health and Social Service Records: Where Law, Ethics, and Clinical Issues Meet (Chicago: University of Illinois, 1976). Robert Boruch, director of the Evaluation Research Program at Northwestern University, has been very kind to supply me with a number of articles and references to techniques for ensuring confidentiality. Some of the articles by him include "Relations Among Statistical Methods for Assuring Confidentiality of Social Research Data (Social Science Research I, 403-414 [1972]); "Strategies for Eliciting and Merging Confidential Social Research Data" (Nejelski, P. [Ed.], Research in Conflict with Law and Ethics (Cambridge: Ballinger, 1976); "Statische und methodische prozeduren zur sickerung der vertralulichkeit bei forschung" (Eser, A. and K.F. Schumann [Eds.], Forschung im Konflikt mit Recht un Ethik [Stuttgart: Ferdinand Enke Verlag, 1976]); "Educational Research and the Confidentiality of Data: A Case Study", Sociology of Education 44, 1971, 59-85; "Methodological Techniques for Assuring Personal Integrity in Social Research", September 1976 (prepared as background research for Evaluation Research Program's Project on Secondary Analysis).
- 6) Most of the articles cited above also discuss the ethical issues inherent in the confidentiality problem. In addition to the Boruch article published in the Sociology of Education, and the Nejelski (editor) book, also useful is the Boruch and Joseph S. Cecil article, "Is a Promise of Confidentiality Necessary? Sufficient?", which appears as chapter 3 of Methods for Assuring Confidentiality of Social Research Data, prepared for the American Psychological Association Task Force on Privacy and Confidentiality and National Academy of Sciences Committee on Federal Statistics, Panel on Privacy and Confidentiality as Factors in Survey Response (Background Research).



²David H. Flaherty. "Report of the Bellagio Conference on Privacy, Confidentiality, and the Use of Government Microdata for Research and Statistical Purposes, Lake Como, Italy, August 16-20, 1977." (Draft.) University of Western Ontario, September 17, 1977. (Mimeographed.)

³For a brief description of the Privacy Project, see Flaherty's paper, "Privacy and Access to Government Microdata," (revised, November 8, 1976) prepared for the Executive Seminar on "Expanding the Right to Privacy: Research and Legislative Initiatives for the Future," Washington, D.C., October 14-15, 1976, sponsored by the Washington Public Affairs Center, University of Southern California.

⁴The Inter-university Consortium for Political and Social Research (ICPSR) has developed a statement on the archiving, processing, and release of confidential data. I include it because it represents the fullest statement to date which describes what ICPSR processing and dissemination responsibilities are in this area.

PRESERVATION OF CONFIDENTIALITY

The issue of confidentiality arises primarily, but not exclusively, in the case of data collections that include information that is or is seen as potentially damaging or threatening to respondents, or when a promise of confidentiality has been given respondents in the process of data collection, and when information is included that allows or potentially allows identification of individual respondents. The confidentiality issue may appear most pressing in the case of data pertinent to elite populations. Such data are now increasingly available, and elite populations are more visible, are or are seen as at greater risk, and are more easily identifiable by means of a few key variables than are respondents in a mass survey. Also contributing to the gravity of the issue is the sensitivity of the public as a whole, and public figures more particularly, to the "potential" uses of data obtained through personal interviews. The increased research focus on elite populations and the resultant increased availability of data combined with the need to facilitate extended research use of data collections makes it necessary for the Consortium to develop policies and procedures for legitimate data dissemination without jeopardizing the rights of respondents to confidentiality.

It should be noted, of course, that the issue of confidentiality is not confined to information collected through personal interviews nor is it limited to elite populations or even individuals. Data collected through means other than personal interviews--through laboratory experiments, for example--also present the issue, and data collected from public record sources, such as court records, can be seen as unnecessarily and unjustifiably damaging to individuals unless anonymity is provided. Data on organizations or particular groups can also present confidentiality issues as can data relevant to deviant populations and the issue is also raised by mass surveys. While the following policies are intended to apply specifically to individual level data for elite populations, they are also seen as applying in more general terms to other categories of data such as those suggested above.

To protect the identity of respondents, the following guidelines will be applied:

1. The Archive will not accept any documentation, list or data files that explicitly identify respondents by name except in the case of data that are in publicly available sources and which do not present problems of confidentiality.
- 2) In cooperation and consultation with principal investigator(s) the Consortium staff will assist in the identification of "sensitive" variables but reserves the right to extend the list of variables to be deleted, aggregated or otherwise masked beyond those identified by the principal investigator(s)
- 3) Variables will be flagged when a) they allow identification of particular respondents, and b) when used in combination with other variables, they also identify particular respondents.

- 4) The list of sensitive variables will be presented to the principal investigator(s) with recommendations for which variables should be deleted, collapsed or grouped (such as age) or have the code category descriptions changed to generalize the category.
- 5) A version of the dataset which incorporates the above recommendations will be produced for general distribution.
- 6) The codebook for the version of the data to be distributed will include documentation for the masked variables, a description of the nature of the deletions and maskings and frequencies for the masked variables.
- 7) Requests for data reductions (for example, cross-tabulations) or aggregations involving the masked variables will be accepted, but the Consortium staff will judge whether the requested data reductions or aggregations will preclude identification of individuals and will supply only those results that do so. Decisions will be made in consultation with principal investigators in order to insure that protection of confidentiality does not unnecessarily limit researchers' access to analytic information.
- 8) The original version of the data will be maintained under security (see document entitled "Physical File Security").
- 9) Public record data will not be added or merged into a survey data file if that data allows identification of individuals. If public record data have been added to the survey data file and identification of individuals as a consequence is possible, these data will be removed from the survey data file and retained separately without a linking variable that would allow the two files to be merged again.

The proposed method for handling confidentiality attempts to protect the respondents without destroying the meaning of the data. If this process renders the data meaningless in any given case, the data could be maintained under security by the Consortium and not be generally distributed. The documentation could be distributed upon request and only data reduction requests would be accepted.

If none of the above is acceptable, the data cannot be archived.

The Data and Program Library Service has attempted to identify donor, archive and user responsibilities in a simpler form, since it involves itself less frequently in processing a data file. Appended is a first (working) version of the DPLS "authorization form" which is agreed upon by the donor when depositing data at DPLS, DPLS when accepting responsibility for a file and the user when receiving a copy of the data. Of particular relevance to the readers are lines #3-5, 10-18, 27-32. As the readers will note, DPLS considers it essential that restrictions may be "relaxed" upon review and approval of data access by the DPLS Faculty Advisory Committee.

DATA AND PROGRAM LIBRARY SERVICE
4451 Social Science Building
University of Wisconsin-Madison
Madison, Wisconsin 53706
608-262-7962

AUTHORIZATION FORM

The Data and Program Library Service serves as an archive for social science related quantitative machine-readable data for the University of Wisconsin-Madison. When authorized as a permanent repository for a data file, the archive assumes responsibility for the condition of the data, changes in the data, distribution, and protection against the release of information deemed confidential. Unless otherwise indicated by the donor of the data, distribution of the file shall be limited to individuals or institutions which shall utilize the data for statistical or research purposes.

The donor of a data file(s) is requested to supply DPLS with information on any restrictions placed upon dissemination of the file. DPLS adheres to these restrictions. DPLS is responsible for notifying the potential user of a file that the file cannot be reproduced without the written consent of the donor. DPLS prefers that a time limit of not more than three (3) years be applied to restrictions on the file, unless confidentiality constraints limit in any way distribution of the data. In that case, unless the archive is assured by the donor of the data and the archive's own evaluation of the file that confidentiality requirements can be met, public access to the file will remain restricted.



DPLS requires that the user of the data agree to the following conditions for receipt of the file:

- (a) Neither the transmitted file(s), listing(s), nor any copy thereof in whole or in part, shall be disseminated, or sold, to any further party;
- (b) The data provided shall be used solely for professional or other official purposes related to teaching, research, public service, public policy, and planning;
- (c) Publication or dissemination of data, or the results of analyses of the data provided, shall be based on a minimum of five individual subjects and/or three organizations per distributional or tabulated cell and neither any individual subject or organization shall be explicitly identified. Where exception to this requirement is requested, permission for access to the data shall be granted by the Data and Computation Center Faculty Policy Committee upon review of the research needs of a particular project;
- (d) An abstract or copy of any published document upon which the analysis is based to be provided to DPLS;
- (e) DPLS will not be responsible for losses or inconvenience resulting from delayed delivery or errors due to defects attributable to the requestor's own equipment. DPLS will bear the cost of replacing the data in the event that the defect is attributable to DPLS's work;
- (f) Responsibility for the accuracy of the data and documentation rests with the donor unless DPLS has itself been responsible for producing the file or documentation.

Please sign below if you agree to the above restrictions.

Date _____

Donor: Name _____
 Address _____
 Telephone # _____

User: Name _____
 Address _____
 Telephone # _____

DPLS _____

(1/2)

DISCUSSION PAPER/ donald f. harrison

On Work in Progress: Directory of Directories

Donald F. Harrison
 National Archives
 Washington, DC

The US Action Group on Process Produced Data Files, in its mandate to prepare a "Directory of Directories", (IASSIST Newsletter Vol. I, No. 3, May 1977), has prepared the following initial entries and desires feedback from the membership. With two exceptions, entries describe only social science data files, printed and available in the United States. Because of the nature of where files are created, plus the specialized knowledge of the list makers, this list is heavy on Federal directories.

The committee seeks information on: 1) additional directories not listed and available for description; and, 2) additional or different information that may be desirable in the entry format. Please send any suggestions to:

Donald F. Harrison
 Chairman
 Process Produced Data AG
 Machine-readable Archives Division (NNR)
 National Archives
 Washington, DC 20408