Security of information processing—Implications from social research*

by ROBERT F. BORUCH
Northwestern University
Evanston, Illinois

INTRODUCTION

Many social research programs are characterized by a stringent requirement that identifiable data collected on the subjects of research be kept confidential. This requirement, coupled with the increasing number of sensitive, sometimes controversial research efforts, has stimulated social scientists' interest in legal, administrative, and technical methods for assuring that confidentiality is maintained. We concern ourselves primarily with the technical methods in this paper, treating "security" as a partial operationalization of the notion of confidentiality.**

Specifically, we should like to sketch those problems met in social research which are relevant to security-oriented activities in information processing. In the following remarks, some of the distinctive features and needs of large-scale social research are outlined. Then, the research design, data collection, maintenance and dissemination stages of the research system are examined to discover how the interests of social research and those of security-oriented information processing might intersect vis-à-vis the problem of assuring confidentiality.

THE CHARACTER OF SOME SOCIAL RESEARCH PROGRAMS

Maintaining confidentiality and security of data are likely to be important objectives in a variety of social research efforts. In the section, examples of these are furnished and the factors which appear to be important in distinguishing research archives from other kinds of information systems are described briefly.

Focuses of the research

In order to establish a manageable topic area, suppose we restrict attention to large-scale social research which results in a computerized information system containing data on identified research subjects. Some form of identifiers (e.g., names and addresses) are essential when individual subjects must be tracked over time to investigate biological and social development, to appraise the cumulative impact of drugs or alcohol abuse, etc. These so-called "longitudinal studies" are frequently conducted, and although many are quite small, some involve repeated in-depth measures on over 100,000 individuals over a 10 or 20 year period.

The research topics which can be expected to generate some concern about privacy, confidentiality, and security cut across all the social sciences. In political science, for example, whether an individual voted or not is frequently a provocative topic for inquiry and a negative response usually constitutes "sensitive" information. Human factors psychologists, often involved in accident research, focus on seat-belt wearing behavior; in some highway surveys, spot checks are made of drivers' alcohol use. Each type of information may have a stigmatizing character. Epidemiologists, of course,

* Work on this topic has been supported by NSF Grants GS320-73X and GI29843. Naturally, the views expressed here do not necessarily reflect those of the sponsoring agency. Some of the observations made in this paper are an extension of earlier research reports, notably (1).

** Confidentiality here refers to the status of information, a condition under which access is formally restricted to certain agencies or individuals. Security refers to the administrative, technical, and legal devices used to assure that the formal restrictions are met; i.e., security is an operational definition of the concept of confidentiality.
frequently need to acquire longitudinal data on incidence and spread of venereal disease, on illegal abortion, and on other socio-biological deviations from the norm. Social psychology, traditionally concerned with relatively innocuous laboratory experiments, has become associated with research on white-collar crime, on mob violence, and on helping behavior in critical situations (e.g., bystander apathy to a streetcorner mugging). Large-scale research in economics and law has, in recent years, accumulated much longitudinal data on individual’s spending behavior, deviations between actual and reported taxable income and other sensitive topics. (For references to work in each area, see Reference 1.)

In the past, confidentiality has not been so crucial and generalized a concern because the size of the research efforts had been small and visibility of the studies low. Perhaps more importantly, the academic orientations seemed to have been associated with relatively innocuous data on anonymous individuals or subjects tracked over very short time intervals. During the past five years, the size and visibility of social research projects such as those described above has increased dramatically, particularly in the policy research and evaluation areas. As in commercial data collection activities, accidental disclosure and deliberate penetration of research files can have serious consequences: research subjects may be embarrassed or harassed and the research programs would undoubtedly suffer. Although the empirical risks here are sometimes no better documented or appraised than risks in commercial data collection enterprises seem to be, the issue is serious enough for both Federal and private grant agencies to develop guidelines on collection and maintenance of identifiable data on individuals (see references in Reference 2). The social researcher’s interest in establishing the security of information stems from increased visibility of research, from these formal legal requirements, and from the ethics and the realities of research. Our ability to collect data will suffer considerably unless we conscientiously and conscionably recognize the need for security.

Archival data: Functional distinctions relevant to security

How might we describe the functional character of social research data archives and those features which appear to be important for the sake of security? As a first approximation, we might consider a rough continuum of computerized data banks which contain personal records, defining the continuum such that one end represents an auditing function and the other represents a research function. Personnel records and intelligence systems typify the first extreme, where each identifiable record serves as a basis for making evaluative judgments about the individual on whom the record is kept, and for taking direct and personal action which directly affects the individual.

The research-oriented systems generally serve not as a vehicle for decision and action about an individual, but for appraising the group’s condition with respect to some social theory or with respect to the effectiveness of a program with which the group is involved. The American Council on Education’s Higher Education Data Bank, and Project Talent exemplify this activity. Each collects identifiable data on thousands of students annually. Most of the data are innocuous by any standards, but some pertain to campus protest activity, alcohol use or other sensitive behavior. Identifiers serve as an accounting device, and the data are not meant for use as a basis for evaluative decisions about individuals.

The functional distinction—audit versus research—has some rather important implications for minimizing the likelihood of disclosure or the utility of data should the data be deliberately appropriated for nonresearch purposes. Identifiers, even if collected, do not need to be as accessible as statistical data for research purposes. Special strategies for separate handling of identifiers and statistical records can be developed and have been used to minimize risk of disclosure (see Intrasytem Linkages, below). Statistical records in audit systems usually must be quite accurate, but in the social research systems, imperfections generated by the method of data collection are recognized and estimated, not for the individual, but for the group as a whole. In fact, to undermine the utility of individual records, without jeopardizing the integrity of the total data seriously, random error whose parameters are known can be inoculated into the data. This strategy, evidently inappropriate for commercial record systems, seems to hold some promise in research concerning topics such as use of contraceptives and illegal abortion.

* Some research designs can be set up such that each respondent injects his response with random error in a manner prescribed by the researcher. For example, in a question requiring a yes-no response, the researcher might instruct the subject to roll a die and to lie if a “1” shows and to tell the truth if 2, 3, or 6 shows. The known likelihood of false positive and negative responses in the paradigm can be used to obtain unbiased estimates of parameters in data analysis. The presence of randomized error in the record system would presumably reduce embarrassment, and threats of unauthorized or legal disclosure, since individual records cannot be used for unambiguous judgments about individuals on whom records are kept.
Linkage problems also differ a bit depending on function. In the research systems, one often wishes to merge identifiable data collected by different agencies. Unlike merges in many audit systems, the separate agencies each may have their own rules and practices regarding disclosure of individual records but may be willing to share data if rules about confidentiality are not compromised. The researcher must then devise special strategies to link data without breaching these rules and without compromising the promises of confidentiality made to individuals on whom records are kept. Specialized methods have been developed (see remarks below on Intersystem Linkage) but more work needs to be done.

The legal status of information in social research also differs from data in the audit system. In some states, socio-medical research records, some educational and psychological records are protected from even legal interrogation by a testimonial privilege. More often, however, they are not so protected and some mechanisms have been devised to undermine the data's legal utility or to minimize its legal accessibility. The inoculation of random errors probably meets the first objective; specialized froms of data linkage and maintenance (Intersystem Linkages, below) help to meet the second. These legal differences are related to security needs in general, and since processing is typically conducted with computing machinery some particular features of information processing technology may also be relevant here.

Each of these differences imply some of the specialized needs of the research data archive in contrast to the audit information system. In the next sections of this paper, the collection, processing, and maintenance stages of the research system are described in a bit more detail and linked to methods for assuring security of data.

DATA COLLECTION

In the simplest case, data are elicited by the researcher and an individual’s response transmitted back to the researcher through various intermediary groups. The intermediaries often include local administrators, staff members of scanning/mark sense processing units, and key punch operators as well as the researcher’s personal representatives. For the sake of security, many social researchers are attempting to reduce the possibility of disclosure to intermediaries, particularly by reducing the number of intermediate stages between eliciting information and the provision of response.

Questionnaire surveys

In order to eliminate the possibility of disclosure during survey administration, some plans require the respondents to put the completed questionnaires into locked and addressed boxes which would be sent directly to the data processor. In some cases, representatives of various interested and disinterested groups can and do monitor the collecting, packaging, and mailing of completed questionnaires.

Even more simply, questionnaires or interview documents have been designed so that one section, containing identification and code number, can be detached from the other, containing responses and an identical code number. Either the respondent at the site of the survey or the researcher at later stages of the survey process can actually separate the two components of information. The identifying information can then be held by the respondent or by a monitoring agency (e.g., group of respondents or representatives of the host agency) and submitted to the researcher after the statistical information is compiled. The code numbers permit later linkage of statistical data with information collected later in the research process.

Rather than require individuals to respond directly on a questionnaire, some researchers are making more use of perforated, but otherwise standard EAM cards as a vehicle for recording data. In requiring that the respondent merely punch his responses out on the card and return it by mail, any intermediate handling of identified records is reduced. And, we can couple this strategy with the use of nominal or numeric aliases to further enhance security. The principal problems with this approach seem to be subjects’ reaction to the cards and limitations of the card format on permissible response options. Human engineering studies would probably help to ameliorate some of these problems.

Remote terminals

One idea which seems to have some merit involves the use of remote terminals as a kind of voting booth for repeated surveys of certain groups of individuals. That is, rather than have respondents furnish data via questionnaire or telephone, we might require that they do so through “social reporting units” in which opinions and self-descriptions can be input directly to storage by an individual. Remote input devices might be particularly useful in organizational settings where continuous monitoring of individual’s attitudes, activities, expectations, and status are essential for research on the effects of policy changes or of organizational innovations.
The voting booth or other remote input methods might, for example, be applied usefully to public housing appraisal where good data on resident's status is essential to economic studies. Other applications may include welfare recipient's reporting, transportation depot surveys, or surveys of any well-defined group (e.g., hospital, military, prison or student groups), whose members can provide useful input data to the social research reservoir. In many such reporting systems, a guarantee of anonymity is necessary for honest and continued reporting; however, tracking the development of individuals is also a frequent requirement. These two needs suggest creation of systems in which the technically unsophisticated respondent can make inputs easily and without being jeopardized by the opinion or factual information he offers. The numeric alias or password systems already developed appear to be relevant here. Some are persuasively secure, e.g., permitting the respondent to form his own transform of a random number of identifiers supplied by the computer. The human factors problems in getting people to use and to adhere to their personal, private transforms will probably outweigh the technical problems in implementing such a system, but these do not seem to be intractable.

DOCUMENT PROCESSING

Anonymous reporting, responding under alias identifiers, and using specially constructed questionnaires (or having respondents incolette their response with random error), usually minimize if not eliminate the likelihood of unauthorized disclosure at later stages in the research system, including document processing. But these strategies may be inappropriate or too expensive for particular kinds of research. Very large and very expensive field experiments, for example, are an important means of evaluating economic and other governmental programs; intensive and long term longitudinal surveys of small samples contribute to our knowledge of human development. Both kinds of studies typically require exhaustive cross-checking capabilities, very complex merge operations, and other activities which appear to justify the joint processing of statistical and identifying information. The use of aliases in these cases may be completely inappropriate and the use of specially constructed documents may make cross-checking the validity or completeness of response very expensive.

In these circumstances, the social researcher usually meets several problems. For one thing, document processing agencies often have neither written policies nor formal administrative regulations regarding the treatment of sensitive data. Similarly, the paucity of information on the establishment of and adherence to codes of ethics in the document processing industry is serious concern to many researchers; since the document processing is frequently (perhaps necessarily) tied to computer operations, the concerns apply to this area as well.

When no administrative or ethical codes are espoused by the service agency that the researcher must employ, it may become necessary for the researcher and the data processor to reach some formal contractual agreement on the treatment of data. At a minimum, such agreements should require that identifying data and response data be separated at an early stage, that the documents be destroyed soon after processing, and that the responsibilities and consequences of negligence on the part of the service agency be carefully defined. At present, insuring that such a prescription is adequate can be difficult because legal precedents and specification of negligence and liability in a document or data-processing environment have not been fully established. The current explorations of these legal problems may clarify the situation (see references in References 2 and 11).

MAINTENANCE AND DATA LINKAGE

When identifying information must be collected with data, the device most frequently used by social researchers for minimizing accidental disclosure or deliberate interrogation of identifiable records is physical separation of identifiers and statistical data. Each separated file usually contains code numbers which permit later merging operations and the identifier file is often kept in vault storage. A few social research agencies have applied some of the Department of Defense administrative and mechanical requirements for security, and the agencies often require computer service groups with which they deal to use the same regulations where feasible.

More elaborate schemes for minimizing the likelihood of disclosure have been developed and are being used. Many of these strategies can be divided into three groups depending on the purpose of maintaining identifiers: schemes for intrasystem linkages, for intersystem linkages, and for combined audit-research systems.

Intrasystem linkages

IntraSystem linkages refer to a single agency's collecting and merging data on the same sample of individuals over an extended time period. In longitudinal
studies of students' political activism, for example, data are frequently collected in identified form. It is reasonable to expect that nonresearchers may be interested in examining identifiable data. The researcher with no legal testimonial privilege (i.e., without the ability to resist subpoena), would normally like to minimize or eliminate the possibility of disclosing sensitive data to even legal authorities when he has promised confidentiality to his respondents.

An interesting operational resolution of this problem is the American Council on Education's LINK FILE SYSTEM.4,11 The strategy was developed to assure the confidentiality of longitudinal data on college students, data which includes limited but identifiable information on disruptive campus protest activities. It works in the following way.

After identified questionnaires are returned by students, the researchers split the information into two segments. The first contains statistical data with one of arbitrary numerical codes attached to each record; the second contains students' names and addresses linked to a second set of code numbers. A third file matches the first and second set of numerical identifiers (aliases). This code linkage is kept in a foreign country with an agency contracted to maintain the linkage for later data merges; the agency is also required by the contract not to return the linkage to the researchers under any circumstances. In followup studies, the researcher's name and address file is used to distribute questionnaires. The associated numerical aliases are substituted for names during document processing and this file is then shipped to the contract agency. The agency replaces the numerical identifiers in this file with the first set of identifiers, using its code dictionary. Then, this follow-up file is returned to the research agency for merging follow-up data with the original data, using the numerical identifiers common to both files (i.e., the first set of arbitrary numerical identifiers).

The system is certainly flawed in that it can be undermined in some cases by the research staff, by the agency holding the code linkage file, and by legal agencies with international ties (see the Hoffman and Turn critiques described in Reference 11). But it is a useful prototype which may help us learn a bit more about how to design and implement a system which will assist in protecting social research data. It does provide a concrete target for the check list strategy given by Peterson and Turn,13 to determine susceptibility of the data to legal interrogation and corruptibility of the system by its creators as well as by outside agencies. The difficulty of using encoded identifiers (arbitrary identification numbers) in physical protection for files and of protecting against indirect disclosure overlap considerably with problems in intersystem linkages which we consider next.*

**Intersystem linkages**

Intersystem linkage refers to the researcher's merging his own identifiable research records with records maintained under other auspices. As an example, consider the (true) example of an economist who obtains data on spending behavior and wishes to correlate these with items from income tax returns. The linkage of both sets of records raises difficulties of two kinds. On one hand, the researcher's provision of identifiable information to the IRS for merge purposes may violate his promise to his respondents assuring the confidentiality of the data. On the other hand, the researcher cannot obtain identifiable data from the IRS for merging because IRS regulations generally prohibit such disclosure.

The so-called insulated linkage process for merging data is an illustrative resolution of these two problems. To link the files, the researcher first cryptographically encodes all statistical data in his own records. He then supplies the joint records (encoded statistical data coupled with identifiers), to the other archival agency. The latter then merges its own files with the researcher's file, basing the merge on the identifiers appearing in both files. When the merge is complete, identifiers are deleted and the resultant file, consisting of unidentified statistical records from both agencies, is then returned to the researcher. This system has been used in actual merge operations with some success and is one of a general class of strategies for linking data under security restrictions.11

Again, a linkage strategy of this sort can sometimes be rather vulnerable, and additional mechanisms must be invented to minimize deliberate efforts to interrogate identifiable data in either file. To corrupt the system, the researcher could, for example, encode a duplicate set of identifiers in his file, allowing identifiers to masquerade as statistical data. Presumably this strategy can be rendered useless by having the archival agency not only merge the data but also summarize it. The provision of summary data then may permit only indirect disclosure efforts by the researcher. But if the researcher uses a very simple encryption scheme, such as systematically substituting one character for another in the records, the archival agency may be able to penetrate

---

* Indirect disclosure involves using a twenty questions strategy to deduce new information about an identified individual when the interrogator has statistical as well as identifiable records on the individual.
the substitution scheme and in fact examine the researcher's identifiable records.

In both the intrasystem and intersystem linkages, social researchers need more guidance on appraising the vulnerability of the strategies. Aside from making more thorough appraisals as outlined, for example by Peterson and Turn, we should obtain better insights into more systematic ways of detecting and inhibiting the likelihood of indirect disclosure, and the utility of cryptographic encoding in these applications.

**Combined audit-research systems**

Some organizations, governmental ones especially, have both audit and research missions and the information maintained in their computers reflects this dual objective. Trust may be a reasonable basis for assuring that researchers will not improperly explore identifiable administrative records or that nonresearchers will not interrogate identifiable research records. More formal restrictions on access and disclosure may be warranted, however, particularly where the data vary considerably in sensitivity, and administrative or personnel monitoring procedures are difficult to implement.

Some of the researcher's needs here can be characterized as having two dimensions. On one hand, he has some need for a flexible, hierarchical system of protection for his own data which can be tailored to the pyramidal nature of its sensitivity. Innocuous and public data might then be kept secure with the cheapest form of protection possible, e.g. existing administrative checks on personnel and the physical plant. More sensitive information such as sources of income, psychiatric and hospitalization records, personal habits and beliefs would justify more secure (and presumably more expensive) mechanisms including those reviewed by Hoffman, say, in his state-of-the-art survey. Some flexibility is essential if the researcher is to keep pace with both changing public opinions regarding the sensitivity of stored information and the changing substance of research. These requirements may be met with the development of hardware modules or micro-coded instructional sets which the researcher himself can use as building blocks for made-to-order protection of data with different levels of sensitivity.

On the other hand, the audit portion of a combined audit-research system may warrant authority hierarchies for access to data which are geared to administrative and researchers' needs. Normally the social researcher wishes to meet his research objectives without incurring the responsibility or liabilities associated with access to joint information and without forcing a compromise of the original conditions (e.g., a promise of confidentiality) under which information was originally supplied to an audit agency. The Shared File System (APL) developed by David Booth appears to have some relevance to this problem; it involves the use of access authorization codes associated with particular primitive (and unmodifiable) commands and particular roles. Presumably, research needs can be accommodated well by tailoring the system so that the researcher can operate with restricted functions in restricted work spaces and arrays, while locked out of his administrative or research colleagues' work spaces, and unable to examine or modify other functions and files stored in the same equipment.

**DISCUSSION: POTENTIAL USE OF A DATA BANK REGISTRY AND DEVELOPMENT AGENCY**

A paper as brief as this one must be cannot hope to give a detailed appraisal of the social scientists' needs in their efforts to maintain the confidentiality and security of the data they maintain. As a framework for summarizing those needs, suppose we consider the current proposals for a national registry of computerized data banks. The proposals are in the interest of developing mechanisms for solving problems in the security area and they may be helpful at the design as well as implementation stages of social research.

It has been suggested that such a registry, coupled with a development agency, be created for the purpose of documenting the nature of computerized information systems, the kinds of personal data maintained in such systems, and the rules and practices which pertain to storage of data. Alan Westin's proposed "data bank on data banks," John Kemeny's plan for a National Computer Development Agency, and other suggestions for monitoring large-scale data collection seem to imply documentation functions of this sort. We can anticipate that such plans, if implemented, will be of considerable interest and use to social researchers, especially if they include the kinds of information listed below.

**POLICY AND PRACTICES IN DATA COLLECTION**

Given the diversity of social research programs, no single policy or managerial practice is likely to satisfy all public and private requirements for assuring confidentiality of data. Statistical methods for minimizing likelihood of identification, legal constraints against
access as disclosure and administrative methods for assuring confidentiality have been developed, but they have been organized and appraised in only a few instances.²,⁴,²⁰ Regrettably, these strategies have not been tied well to more computer-bound technical devices such as those described by Hoffman,¹⁴ Peterson and Turn,¹² and Goodfellow.²¹ An agency with an information clearinghouse function, coupled with a development mission, would be quite helpful in documenting, consolidating, and organizing information in the following categories.

Legal solutions

Local, state, and Federal statutes relevant to privacy and confidentiality of data; court precedents, administrative regulatory powers; empirical data on problems in enforcement of codes, and adherence to guidelines furnished by government agencies to social researchers regarding rights of privacy and conditions of disclosure.

Administrative approaches

Link file systems,⁴ insulated data banks,¹¹ and other similar strategies for eliciting and merging sensitive data; vulnerability, utility, and frequency of the strategy’s use; cost data.

Statistical/Mathematical Solutions

Documentation on applications of error inoculation⁵ and other approaches to depreciating probability of indirect disclosure;¹³,²⁰ costs and benefits of applications.

Technical mechanisms

Types of cryptographic encoding appropriate for computer applications; their cost and vulnerability; catalogs or listings of hardware and software security devices; possible relevance of new devices to specialized research needs (such as remote terminal application mentioned earlier; see also Reference 22).

Empirical studies

There is some real value in consolidation of data on people’s resistance to data collection and to social research. Complaints about the collection of information and against organizational disclosure practices, concerns about the magnitude of data maintained, etc., need to be well-documented. Although some empirical data exist, there is currently no single source on which the researcher may draw to establish the likelihood of privacy problems in the conduct of his research and to anticipate the costs of resolving them.

In many cases, questions can be phrased to minimize embarrassment and/or threats of sociolegal action against a respondent. Some of the relevant strategies—elimination or generalization of the inquiry, approximations to direct questions—are fairly well documented. Small “item pools” or computerized retrieval systems containing questions which pertain to the same general behavior, but with varying levels of sensitivity and intrusiveness do exist. But data on both strategies and item pools are widely dispersed. There is still a great need for large, accessible item pools which have been tested for objectionability, intrusiveness, and susceptibility to error.

Validity appraisals and secondary analysis

Frequently, social researchers elicit anonymous information from previously identified samples or require research subjects to use an alphanumeric alias (in short term longitudinal studies), so as to minimize if not eliminate any risks that data will be used for nonresearch purposes. An information registry would be of considerable help in appraising validity of sampling and credibility of reporting in such efforts. Suppose, for example, a medical sociologist, who usually has no testimonial privilege for the data he collects, relies on mailed or telephoned responses to his questionnaire on illegal methadone use. He might encourage the use of aliases to assure that his data are not appropriated (legally or otherwise) for harassment of his subjects, but he still needs to anticipate the redundancy of his data, and to appraise its validity since he does depend on voluntary responses. The researcher could do so if a data bank register furnished information about the existence of medical records, census data, police intelligence systems, etc., which contained relevant statistical data on the population from which subjects were sampled. And if identifiers were actually obtained he could merge his own data with existing files without violating access restrictions using some special administrative strategies which might also be documented in the same registry.

SUMMARY

The objective of maintaining security of social research data is an operationalization of the concept of “confidentiality” in social research. The problems in meeting
the objective depend on where the research falls on a hypothetical audit-research continuum for the data, on the kinds of process being used to elicit the data, and on the level of identifiability of records necessary in the research. Major differences between audit and social research approaches to security problems stem from the social researcher’s infrequent need to maintain joint identifying and statistical records, and the opportunity to use modified (alias) identifiers and modified response data (i.e., inoculated with random error in a controlled process).

Aside from benefiting from systematic appraisal methods such as those described by Peterson and Turn, social researchers might do well to capitalize on other research efforts connected with security in information processing. Linkage systems and similar devices mentioned earlier depend very much on encryption schemes for assuring integrity of the system. The encryption transforms used in the examples cited have been limited to simple substitution of one character for another or simple linear transforms of original numerical characters. Perhaps certain kinds of transposition or additive transforms, as yet unfamiliar to the social scientist, can be adapted to this kind of problem to assure greater security. Certainly, the development of algorithms which help in checking whether indirect disclosure is possible or likely would be well received by managers of the research data banks. Translating the structure of data sets into simple algebraic equations is a skill which is usually beyond the social scientists’ own expertise. Judging from Fellegi’s work and current activities by Turn, such algorithms are likely to require a great deal of technical attention to efficiency, to heuristic alternatives to searching large sets of equations (data sets), to determining the likelihood of indirect disclosure, tasks in which the social must be educated by the computer technologist.

Certainly, if proposals for national data registries and development centers are implemented, social scientists will have the opportunity to reduce redundancy in collection and maintenance of identifiable data. A centralized information source may help to stimulate more interest and expertise in technical solutions to problems in this area. Since most social research involves data which are heterogeneous with respect to their sensitivity and publicity, the researcher will benefit most from technological developments which associate more protection with increasing levels of sensitivity, and authority access designs which recognize these levels.

REFERENCES

1 R F BORUCH
An annotated bibliography of randomised field experiments

2 R F BORUCH
Maintaining confidentiality in educational research: A systemic analysis
American Psychologist 1971 26 pp 413-430

3 T K GLENNAN
Using experiments for social research and planning
Monthly Labor Review February 1972

4 A W ASTIN R F BORUCH
A "link" file system for assuring confidentiality of research data in longitudinal studies

5 R F BORUCH
Administrative, statistical, and legal solutions to the problem of assuring confidentiality in social research
Paper presented at Statistics Department Colloquium University of Chicago 1972

6 J ROTHENBERG
Urban economics

7 H BLACK E SHAW
Detroit’s social data bank
In A F Westin Information technology in a democracy Cambridge Harvard University Press 1971

8 E B SHELDON
Social reporting for the 1970’s

9 W D WALL H L WILLIAMS
Longitudinal studies in the social sciences
London Heinemann 1970

10 D T CAMPBELL
Administrative experimentation, institutional records, and nonreactive measures
In W M Evan (Ed) Organizational experiments Laboratory and field research New York Harper and Row, 1971

11 R F BORUCH
Strategies for eliciting and merging confidential social research data
Policy Sciences September 1972 (in press)

12 H E PETERSEN R TURN
System implications of information privacy

13 I P FELLEGI
Question of statistical confidentiality
Journal of the American Statistical Association 1972 67 pp 7-18

14 L J HOFFMAN
Computers and privacy: A survey
Computing Surveys 1969 1 pp 84-103

15 D F BOOTH
File security for a shared file, remote terminal system
Paper presented at the Conference on Computers, Privacy, and Freedom of Information (Mimeo) Queen’s University 1970

16 A F WESTIN
Civil liberties and computerized data systems

18 President's Commission on Federal Statistics
Report of the President's Commission

19 G B F Niblett
Digital information and the privacy problem
Paris Organization for Economic Cooperation and Development 1971

20 M H Hansen
Insuring confidentiality of individual records in data storage and retrieval for statistical purposes
Proceedings of the 1971 Fall Joint Computer Conference

21 B B Goodfellow
Projections of the impact of technology on the development of large data base information systems

22 N M Bradburn
Survey research in public opinion polling with the information utility—promises and problems
In H Seckman and N Nie The information utility and social choice Montvale (New Jersey) AFIPS Press 1970

23 R Turn N Z Shapiro
Privacy and security in data banks: Measures of effectiveness, costs, and protector-intruder interaction
Proceedings of the 1972 Fall Joint Computer Conference American Federation of Information Processing Societies 1972