David, M.H., & Robbin, A. (1981). The great rift: Gaps between administrative records and knowledge created through secondary analysis. *Review of Public Data Use, 9,* 153-166. (with Martin David)

The Great Rift: Gaps between Administrative Records and Knowledge Created through Secondary Analysis

Martin David and Alice Robbin
University of Wisconsin—Madison

Abstract: *Law, mission, and information management practices inhibit access to computerized administrative records produced by state government. Research use or secondary analysis is not on the agenda of the agency administrator. Computerized records are not routinely maintained or preserved. Records managers and archivists for public records do not participate in decisions about retaining or destroying computerized records. These findings emerged from a recently completed cooperative study conducted by the University of Wisconsin—Madison and the State Historical Society of Wisconsin on the impact of automation on state agency records keeping practices. In addition, changes in rules for access, computer-based technologies, pressures to maintain routine administration in the face of high turnover in data processing staffs, reduced budgets, and legislation to reduce paperwork pose a threat to the retention of administrative records. This article discusses the implications of the findings and trends, provides examples of data delivery failures, and recommends changes in law and administrative behavior. The authors conclude that the social scientist has a role to play in assisting government agencies in improving access to computerized administrative records.*

Address correspondence to Professor Martin David, Department of Economics, Social Science Building, University of Wisconsin—Madison, Madison, Wisconsin 53706.

Biographies

Martin David, Professor of Economics at the University of Wisconsin—Madison, has long been interested in research using data collected from surveys and administrative records. Early research at the Survey Research Center of the University of Michigan dealt with problems of the low-income population and distributional effects of government. Later studies on income tax data revealed deficiencies in the tax-averaging provisions and provided measures of the instability of income. More recent survey work has included studies of attitudes toward tax reform in Wisconsin and the provision of credit to Kenyan farmers and businessmen. Dr. David has been director of the Data and Computation Center of the University of Wisconsin, which includes the Data and Program Library Service, a pioneer in the development of data as a library service to Wisconsin faculty and students. He served as coprincipal investigator of the project on which this article is based.

Alice Robbin, Head of the Data and Program Library Service at the University of Wisconsin—Madison, has written extensively on the need to develop an information infrastructure and strategies for improving access to statistical data for social research. She serves as President of IASSIST, the International Association for Social Science Information Service and Technology, is a lecturer in the Department of Economics and School of Library Science, and is currently investigating the impact of privacy, confidentiality, and open-records laws on access to confidential records produced by state government. She served as

Review of Public Data Use 9:153–166 (1981)

0092-2846/81/0315314$02.75

codirector for the project on which this article is based.

Another title for this article might be: Is there life after death? What happens when administrative records die? Results of a 15-month cooperative research effort by the Data and Program Library Service (DPLS) of the University of Wisconsin and the Archives Division of the State Historical Society of Wisconsin (SHSW) indicate that the terrain through which bona fide researchers must pass on their way to the exploitation of state-government produced data for the production of beneficial knowledge is indeed a great rift marked by uncharted procedure, bureaucratic barriers, sinkholes in which data disappear, and cliffs of irrelevancies that only the most tenacious researchers can hope to scale.[1] This article will focus attention of the research community on three major points concerning exploitation of administrative data for basic social science research: (1) In many cases research uses of data are not on the conceptual agenda mandated by law and agency mission. (2) Even when research use and access to identifiable information is not in question, attention to maintenance and preservation of machine readable records (MRR) has not been given. Much information has been lost as a consequence of bad housekeeping and neglect. (3) The delivery of data to researchers requires inputs from the academic community at many stages other than the ultimate torture of data that is generally contemplated.

This article begins with a review of recent trends that affect secondary analysts as potential clientele of a delivery system for research data bases related to administrative records. Examples of aborted data base delivery follow. Analysis of the reasons for failure then gives insight into the recommendations that flow from the DPLS-SHSW undertaking. The argument should convince the research community that ancillary assistance needs to be given to administrative data gatherers to assure research access. It should convince data gatherers that developing adequate control over MRR and co-opting the research community into structuring archival data structures will be in the long-run interests of good government and public service (see also, U.S. Department of Commerce; The President's Reorganization Project for the Federal Statistical System 1981).

TRENDS

The operation of government is being automated. Data are being captured in machine readable records (MRR). Forms are being encoded on terminals; forms are being eliminated by the direct entry of applications on CRT terminals; and interviews ara being automated by the use of computer-assisted prompting and on-line data entry. In various degrees, paper is being eliminated. Administrative records are shifting from documents to computer records. (The survey of Wisconsin agencies revealed substantial duplication of paper documents by MRR. All indications are, however, that this duplication is temporary and will give way to a situation in which only MRR record available data.)

The capture of information as MRR makes it possible to improve the efficiency of ongoing operations in the agency, at least in principle. MRR can be electronically screened to select cases for tax audit or generate information on entitlements to economic assistance. Search of large record systems makes it possible to retrieve data using powerful algorithms devised by computer scientists. Registration of motor vehicles, holdings of libraries, and vacancies listed with the Job Service can all be searched by generic key word instructions. A file of MRR can also be statistically interrogated to produce reports on the welfare case load, number of delinquent taxpayers, handicapped students receiving special services, and so forth.

Ultimately, some MRR become obso-

lete—a vacancy is filled, a welfare recipient makes good, a taxpayer is absolved of liability by the statute of limitations. What happens then? Alternatively, information on the MRR must be altered—a taxpayer's liability is increased by audit. What becomes of the superseded information on the original MRR?

The administrative needs have increasingly been met by the creation of data base management systems (DBMS). These systems update MRR and provide for output and retrieval needed for efficient administration. The systems are often proprietary, specialized in purpose and output, and sometimes dependent on particular hardware configurations. The software for these systems, often carefully structured to optimize day-to-day processing. may create an efficient flow of information for administration but may neglect larger data assembly tasks that are required for evaluation, policy research, and comparative studies. Unfortunately, these characteristics imply that an unanticipated use of MRR or an unanticipated question may be prohibitively expensive to answer. An historical record may be difficult to retrieve and link when data points from different eras relate to different DBMS or data require a link to information that is not stored as part of the DBMS.[2]

Another trend is that the public has become increasingly concerned about the surveillance capabilities of the computer and the use of MRR as a surveillance mechanism. Damage to individuals arising from breach of confidentiality has been another concern. These concerns have given rise to more stringent legislation concerning privacy and protection of information at the same time that they have generated policies and legislation that open MRR to the individual who may be the subject of the record, as under the Freedom of Information Act.[3]

Concern about confidentiality, legislative mandates from privacy acts, and agency interest in secondary analysis of data have combined to place the greatest emphasis on data release through public use samples (PUS) in which information is not identifiable and the capacity to infer identity is limited by censoring underlying data sources in some dimensions, particularly geographic detail. Such files are expensive to create and limit the capacity to assemble relevant data. Why should access be so limited? Restriction of access to records should contain a time limitation. The deceased suffer little damage as a result of disclosure. Statutes of limitation and the minimal damage inherent in most disclosures of confidential data should be considered in permitting early research use of restricted and identifiable data by a widely qualified group of scholars.

Lastly, the high rate of change in administrative data processing has resulted in a phenomenon that could be called input without throughput. Delays in the implementation of DBMS, complications in electronic data entry systems, and pressures to maintain routine administration in the face of high staff turnover in data processing units create serious bottlenecks in routine administration. Administrative staff are often in the position where any demands beyond the most immediate cannot be considered and where even manual systems for collecting statistics on the operation are impaired or terminated.

An event that is not yet a trend, but that will have enormous implications for research use of administrative archives, is that the U.S. Office of Management and Budget has been mandated to control the "burden" of federal paperwork.[4] Every demand on the public that results in the collection of data is being monitored. The OMB intends to reduce this "burden" by 25% in the next two fiscal years. The control and the target reduction together imply that less information will be collected from the public in the future, unless productivity of present collections can be increased or burden can be reduced by automation.

These trends pose a severe threat to the retention of administratively collected MRR for secondary analysis. The disappearance of

paper records makes inadvertent destruction of records more likely. These trends also act to preempt archival research because the preservation of historical series is not needed for the narrow administrative mission of the agency collecting the data. The pressure to operationalize electronic administration, coupled with a desire to minimize cost may well lead to the design of DBMS that makes it difficult to undertake longitudinal analysis of caseloads, to link incoming applicants to past recipients, or to match the present and past of a particular individual. Confidentiality provisions of the law may be so strict as to inhibit information exchanges that link separate administrative records. This stricture has, for example, reduced the capacity of the U.S. Treasury to double-check earnings reported to the Social Security Administration (for credit to the retirement account).[5]

These trends pose a threat for the future use of MRR for secondary research. Examples of aborted or near abortions of research access to significant bodies of administratively collected MRR give an indication of the precarious nature of access to major bodies of MRR.

ABORTED DELIVERY OF DATA BASES FOR SECONDARY ANALYSIS

These examples also illustrate the variety of problems that beset the creation of adequate archival files of MRR for secondary research. Seven examples describe a range of serious delivery problems or failures.

1. Salary Data on Wisconsin School Teachers

One of the files of MRR that the SHSW-DPLS archival project attempted to accession included information on the salaries and Social Security account numbers of schoolteachers in the state. Release of the data was initially refused. The reason given was confidentiality of the information. Review of the statutes and administrative code revealed that the documents from which the MRR were drawn were open records. A clear legal mandate to permit public inspection of the paper records did not inhibit the Department of Public Instruction from attempting to restrict access to these records in machine readable form (Robbin, 1981a). Had Wisconsin law not been clear that MRR are one of the several media on which public documents are stored, information may well have been lost to the research community.[6]

2. Wisconsin Inheritance Tax Records

Paul Menchik and Martin David are conducting research on the relation of lifetime income to bequests (Menchik and David, 1980). They requested that the Department of Revenue release inheritance tax records to facilitate the research. David's request lies unanswered after 4 years. Successful release of those data would require (1) definition of a procedure to be followed by those desiring data access and (2) legislative stipulation of the obligations of the receiver of such restricted data. Neither is clear at the present time. (Fortunately for the research, the data were largely duplicated by public probate records. Failure to access MRR was an inconvenience, not a disaster.)

3. Survey of Consumer Expenditures, 1950

As part of an intensive program, the U.S. Bureau of Labor Statistics disseminated punch card copies of its 1950 Survey of Consumer Expenditures for a massive academic research effort that resulted in two famous volumes of papers (Friend and Jones, 1981). In 1980, no copies of the MRR could be located. Twenty or more working copies of the MRR had vanished. The problem here is that none of the 20 people took archival responsibility. The BLS had no control over its MRR. It destroyed its own copies in a fit of housecleaning or moving. It had no institutionalized memory that assigned value to the outdated survey information.

4. 1964–65 Wisconsin Tax Model

A similar story can be told about the sample of tax returns that was drawn to study the impact of "Federalizing" the definition of Wisconsin income, beginning in 1965. Discussions of tax reform in 1978 circled vainly around plans for simplification and the impact of deviations from federal definitions on taxpayers. No review of prior changes could be called forth because all documentation for the MRR has disappeared. (The sole copy of the documentation was destroyed in a fire several years ago.) The MRR are still available in a tape library. This case demonstrates that retention of physical files has little value when the effort required to interpret data is superhuman, if not impossible.

5. 1974 Wisconsin Tax Model

More recently, the State of Wisconsin invested a great deal of resources in preparing a data base for the study of tax incidence in the state. The primary information collected consisted of matched samples drawn from the income tax records, the file of Medicaid eligible persons, and records of applications for student aids. Considerable documentation and a simulation system for studying changes in tax law were erected around this data base. Today this material sits unexploited, while the Department of Revenue pursues other, more recent, policy problems. No provision has been made to preserve this body of information, no staff resources are available to manage it; and the research community is excluded from using the investment in these data by state statutory confidentiality restrictions on tax data. Catch-22!

This instance indicates the desperate need to provide for PUS. Resources must be provided to create PUS, and an administrative sensitivity to the need to produce such files must be nurtured. The agency developing an administrative data resource must also invest resources in the transfer of data to archival custody.

6. The 1960 Statistics of Income Sample

Tax data again are the subject of discussion. In this case, the public use sample was created, but the only known copy is not in archival custody. The Brookings Institution, which has had a major program of research on tax data, holds the information. However, it does not foresee continued funding for this research. Its staff has no experience in archival data preservation. The knowledgeable programming and research staff are being dismissed. How will these data be preserved?

7. Household Budget Study–Nyanza, Kenya (1971–72)

In this instance, all of the problems already suggested came to the fore. The collecting agency had no policy with respect to release of microdata. Access to the data by researchers has been capricious. The collecting agency had no physical control over the MRR, so it can not assure that the data have not been destroyed. No documentation exists, so that research uses depend on face-to-face communication with programmers who have long since taken on other duties.

The tragedy of this case is that Kenya has few resources for research and few individuals who can make use of microdata. Yet it is clear that the potential of the data was never exploited beyond a trivial level and that orderly retention and release of the data for academic research would be invaluable.

A summary of these seven cases is given in Table 1, which emphasizes the primary barrier frustrating access to the data.

DISTANCE OF ADMINISTRATIVE GOALS FROM SECONDARY RESEARCH

Discrepancies in the world view of the administrator and the secondary analyst make it clear why the demands for release of data and secondary analysis may seem peculiar even to the conscientious administrator. Table 2 displays discrepancies in the goals of agency

Table 1. Mortality of MRR by Cause

Data body	Pathology					
	Access		Control		Preservation	
	Media	Legal restriction	Physical	Documentation	Archive ?	PUS
1. Wisconsin teacher salaries	X					
2. Wisconsin inheritance tax		X				
3. 1950 consumer expenditures			X			
4. 1964-65 Wisconsin income tax				X		
5. 1974 tax model—Wisconsin					X	X
6. 1960 U.S. income tax—SOI					X	
7. Household budget, Nyanza 1970-72		X	X	X	X	X

Table 2. Distance of Administrative Goals from Secondary Analysis

	Agency	Analyst
1. Mission	Legislative definition; limited jurisdictional comparisons	Problem-solving; transcendental, comparative
2. Time horizon	Fiscal year	Indefinite
3. Product	Specific/defined	Serendipitous/evolutionary

data collectors and problem-solving researchers that lead to an unsatisfied demand for archival data.

1. Mission

Administrators have a mission that is legislatively defined and circumscribed by administrative code. No incentives exist to look beyond the limits that are so defined and, indeed, serious problems of "turf" may arise from broad interpretations of the mission and mandate. When studying programs and data from outside their jurisdiction, administrators' valuation of the data is still limited by the legislative mandate.

The research community, on the other hand, has an interest in problem-solving. Its activity cuts across legislatively defined mandates and often requires information of a much broader institutional character. The researcher may investigate commonalities and functionally determined reasons for differences in the missions defined for similar agencies in different jurisdictions or at different times. Research use thus may require linkage of materials gathered in several agencies and from several record series. (Legal,

procedural, and operational mechanisms to provide linkage of data collections in those agencies are few and far between.)

2. Time Horizon

The time horizon of administrators is typically myopic as compared to the researcher. Immediate demands for completing the administrative function, a budgetary horizon of 1–2 years, and legislative demands for information for modifying policy tend to stifle attempts to look at research with a long gestation period. At the other extreme, historical scholars look to centuries of records to ascertain the effect of policies on social change and the performance of the economy.

3. Product

The administrator is required to deliver specified material and reports. The administrator describes and accounts for activities of his agency. The researcher may engage in discovery and delivery of unspecified outputs.

The defined mission, limited time horizons, and specific products of the administrative agency need not conflict with research access. Access can be assured by incorporating a mandate for secondary analysis and research access to administrative data in the legislative mission defined for the agency. As Robbin (1981a) points out, defining a role for secondary analysis of data entails five aspects: (1) scope of protection for research access, (2) persons protected, (3) materials protected, (4) procedures for gaining access, and (5) institutionalization of the authority for retaining, disposing, and accessing data[7] [Robbin, 1981a:20]. These points will be elaborated in the final section of this article. First, we consider the delivery of MRR in the absence of a legal definition of the right to research access, the case that has been typical in the past.[8]

THE DELIVERY OF PUBLIC USE DATA FROM MRR

In the absence of institutions that organize access to MRR, possibilities for loss of valued information are great. Our research in three Wisconsin agencies uncovered problems of four kinds that stand in the way of orderly retention and delivery of MRR (See Table 3).

1. Discovery

The agencies we investigated had no comprehensive inventories of MRR (State Historical Society of Wisconsin). MRR were not integrated into records management practices. Centralized sources of information about agencies' MRR were either unavailable or inadequate. Documentation on MRR was in most cases nonexistent or scattered among various agency personnel responsible for the different aspects of MRR production, processing, and reporting. Tape libraries controlled physical reels without regard to content. Records and data centers saw themselves as repositories for magnetic tape, with responsi-

Table 3. Stages in the Delivery of MRR for Secondary Analysis

Stage	Problem
Discovery	What data on MRR?
Control	Can data be obtained?
Appraisal	Do data have value? Content Methodology Extent Linkage
Access	Is there a custodian? Is there an archive? Does research use proposed meet minimal standards of consistency?

bility for decisions about tape maintenance left in the hands of the agency. Records managers left decisions about retention to those with programmatic responsibility and concerned themselves with managing paper and microfilm records. As in California and Georgia, most records managers were unaware that documentation was essential to interpreting the computerized record. Individual analysts retained information on the content of various files for which they had programmatic responsibility. Data processors were often the only persons knowledgeable as to format and physical attributes of the information. Inject a substantial rate of turnover among staff in each of these three capacities and the potential for irretrievable loss of information is large. Moreover, computerized records, removed from casual inspection by the storage medium, could be totally hidden from public view.

2. Control

Following discovery, it is necessary to establish control. Magnetic tape containing potentially valuable data were regularly erased and reused when tape shortages occurred, without systematic review of the tape's contents by records management. A way must be found to preempt the withholding of information and the destruction of information by workers in the agency who have no interest in research access. This stage has usually involved informal and cooperative relationships between researchers and staff members in the agency. Control also implies that it is possible to obtain documentation of the information desired.

3. Appraisal

Assessing information is the next step. In the case of files that the project accessioned, this information was not always available, even after a time-consuming search of files in analysts' and data processors' offices and libraries. An additional problem that was uncovered was that computerized records did not always reflect encoding of the entire source document or were created from multiple source documents. The source documents and methodology of the measurements and procedures to encode information in machine-readable form must be available and exactingly reviewed. Content, validity, and sample design must be considered.

4. Delivery

Actual delivery of the data can then take place. The researcher must request the items of information; clearance to access the data must be given by the agency; and physical transfer can then take place.[9] These stages in the delivery of MRR cannot be short-circuited. The same stages must be followed in archival preservation and must be institutionalized to assure an effective right to research access.

CONCLUSIONS AND RECOMMENDATIONS

Institutionalizing Research Access

The scope of protection for research access must include both a mandate that society's interest in knowledge requires access to MRR and a definition of the obligations of the researcher who gains access. Society has a right to know because the society can benefit from the distillation of relationships out of data on individuals and transactions. Knowledge that can be extracted is generalized and consists of relationships, abstracted from idiosyncrasies of individuals. The potential of this knowledge can be realized with relatively small inputs of additional resources at the time of secondary analysis. Boruch and Reis add that "secondary analysis . . . increases the number of laboratories for research and . . . generate[s] ideas and critical assessments" (1981:60).

Data already embody a substantial invest-

ment that has been made in the interest of administrative efficiency. Access magnifies the return on the original capital investment. For example, Kleiner (1980) argues that administrative records, used in conjunction with survey data can provide more efficient labor market information. (See also Cartwright and Armknecht [1980] for the multiple uses that can potentially be made of administrative records.) The Privacy Protection Study Commission suggested that administrative records are a cost-saving device for a variety of research and statistical purposes (1977:489). Hulett believes that transfer of identifiable data for a statistical or research purpose can ensure the accuracy, timeliness, and consistency of major statistical or research reports (1975:203). Dobson's (1980) discussion of a National Health Institute data base is additional evidence of ways in which existing administrative records systems linked to new data collection activities provide a means of creating efficiencies and accountability in government and of responding to new policy questions.

The obligation of the researcher in dealing with administrative data is clear. No disclosure of individual data should be permitted, since knowledge is *general* relationships. However, the researcher may need to deal with restricted, confidential information in order to generate that knowledge. The researcher should therefore be willing to sacrifice a bond of liability in exchange for the privilege of accessing the data.[10]

The personae must be defined: Who exercises decision-making power in the agency to release the data? Who are qualified as researchers? The answers to these questions cannot be given in an ad hoc fashion. The researcher becomes understandably frustrated when it is difficult to establish who has the power to release data. On the other hand, society has the right to ascertain that the research being undertaken is feasible, given the data requested, and that the researcher is technically qualified to undertake the project.

Specification of the materials protected must include provisions for identifiable data and confidential data. Specification of procedures makes certain that the analyst who is denied access understands possible appeals when access to data is denied and does not suffer from a poorly defined review process for his request. Robbin indicates that administrative review is superior to judicial procedures for gaining access on appeal (1981a:180). The reason is that researcher and administrator have more in common than in opposition. The higher-level administrator generally stands to gain from knowledge that can be generated from MRR. Thus a good deal of commonality is lost by relying on an adversarial judicial procedure to gain access. Moreover, resources available for adversary proceedings are heavily biased in favor of the administrator.

The last element in the legal specification of access is now clear. Process without substance is worthless, so that a clear archival responsibility must be established. A custodian must be defined; others may not destroy MRR without permission. Resources for retention must be forthcoming to support the custodian in his duties

Archival Preservation

Resources are not being systematically allocated for preservation. The most logical procedure to finance the archive is a tax on the resources involved in the creation of MRR. After all, failure to set aside archival resources at the time records are created amounts to frustrating one of the purposes of such information—to assure a more satisfactory functioning of the society. Such functioning might be attained by the extraction of knowledge from the complete data body.

Reviewing Table 3 gives insight into the project's recommendations. The project's first recommendation is that state archives establish prearchival control over MRR. This forces an inventory and centralization of information on physical location and documentation. Prearchival control also entails a process of

scheduling recurring records in a system that controls preservation and destruction. The recommendation thus implies that secondary analysts have some hope of an answer to the question: What data are captured as MRR?

The corollary of prearchival control is that agencies must establish procedures to assure that MRR are not inadvertently erased, physically to maintain storage media (magnetic tapes) to prevent deterioration, and to control the reformatting of information to prevent the loss of data collected at an earlier time. Agency custodial responsibility rests with record managers. But they have little competence in dealing with computer systems and have little management authority to enforce adequate control over MRR.[11]

A second recommendation is that administrative agencies adopt standards of documentation developed within the social research community (see Roistacher et al., 1980; Robbin, 1981b). Documentation needs to be augmented by statutory references that mandate the collection of MRR and delimit access to nonagency users. It also needs to include descriptions of hardware and software dependencies when the data are not preserved in a standard data transfer format. Agencies' inventories of MRR can be an invaluable source of reference information, not only to the researcher and the public at large, but also for agency personnel throughout state government. An inventory, or guide to government resources, makes it possible for all concerned persons to gain faster access to data required for efficient social decision-making.

As Table 3 makes clear, once discovery is systematized and control of MRR is effected, the next intellectual problem is assessing the value of MRR for secondary analysis. Because the agency has interest in MRR that pertain to current operations and it is likely that historical analyses, long-term longitudinal studies, and research on specific subtopics will not be undertaken in the present, the research community must anticipate an intermediate stage in which the potential of MRR for future research must be appraised according to some general criteria. Not only the nominal material included on MRR, but the methodology must also be considered. Coverage of the universe, sampling, and bias in measures should be examined. The temporal extent of MRR may be important to their value, and exhaustive enumeration of populations may be important for future sampling. Lastly, identification that permits linkage to other MRR must be considered an essential attribute in assessing the value of a particular data body. (One should note that the existence of paper records identical to MRR are a minimal factor in the assessment. For the foreseeable future, the costs of encoding paper records and manually sampling them are so great that a decision not to preserve MRR is almost equivalent to destroying the capability of future statistical analyses. Efforts now underway to create PUS from the microfilmed versions of the 1940 and 1950 Census are to cost $8,000,000. Much of that information was on punch cards at one time.)

The decision to preserve data implies an institution to take archival custody. Few archival institutions with a capability to administer MRR now exist. In Wisconsin, the public records statutes invest the SHSW with archival custody over all public records, restricted or unrestricted. This makes it clear that the SHSW has the responsibility for appraising MRR and preserving those that have future value for secondary analysis. The problem that the SHSW faces is that it has little expertise in the management of MRR. Thus the project's fourth primary recommendation is that archivists throughout the country acquire the capability of managing MRR.

The last stage in the delivery of MRR to secondary analysts consists of assuring access. This entails two separate phases: the procedure for gaining access (described previously) and a testing algorithm in the archival program that can assure that documentation is complete and correct and that data are stored in an accessible format.

The research community and the archvist need to be present at the time that DBMS are

designed to assure that it is possible to take a census of the information at regular intervals, link the resulting MRR over time, and output data in relatively standard transfer formats. The concerned social scientist should now be clear that expertise is required from the research community far in advance of the delivery of MRR or PUS. As a group, the research community needs to be represented in the appraisal process where vision on styles of research, comparative data, and linkages of interest are extremely valuable.

Finally, and indeed the highest priority, the research community must convince legislators that knowledge generated from MRR is socially valuable. Knowledge must be guaranteed to the society by intelligent systems for assuring a right to access identifiable records on the part of bona fide researchers. We must seek ways to impress on legislatures that millions of dollars are already committed to creating MRR. For society to blind itself to this information in making policy decisions is to create waste in government. Once created, the cost of maintaining MRR and documentation is so small that potential benefits, albeit rare, warrant institutions for preservation.

NOTES

1. "A Pilot Program to Accession Machine Readable Public Records of Wisconsin State Agencies: A Cooperative Project between the Archives Division of the SHSW and the DPLS of the University of Wisconsin-Madison." The project was supported in part by the National Historical Publications and Records Commission and the Graduate School of the University of Wisconsin–Madison (November 1, 1979–April 30, 1981). The context for the SHSW-DPLS project was extremely favorable. The SHSW is recognized as the State Archives for all public documents. The public records statutes are sufficiently broad to include machine readable documents as public records. The bureaucracy is extremely favorable to research uses of administrative records and is predisposed to research access by University staff because of a long and fruitful exchange between the government and the University that has led to the effective solving of social problems. Some of the work carried out on this project is being continued by Alice Robbin with Volkswagen Foundation funding. Preliminary research into the information management practices in California and Georgia state governments indicates that the problems we identified in Wisconsin are shared by other state governments.
2. Administrative demands for information are quite different from demands for statistical analysis. Locating all the information on a few pathological cases that demand administrative review is a different task from locating a limited subset of information about all individuals, the typical statistical requirement. The consequences of this difference in design objectives for the DBMS is that a system friendly to administrative goals may not be friendly to statistical inquiries (see Anderson and Sim, 1977; Robbin and Hedstrom, 1981).
3. Even the Freedom of Information Act (FOIA) has come under scrutiny by the current administration. Under a proposed change, federal agencies will be permitted greater discretion in releasing government documents. Agencies have argued that the FOIA places an administrative burden on them (Taubman, 1981). Horton (1980) argues that the FOIA may be called a "failure" in the sense that it never addressed access as an information problem. He may well be right. What the Wisconsin, Georgia, and California data show is that the access question is very much an information management problem.
4. Paperwork Reduction Act of 1980, Public Law 96-511.

5. However, the Paperwork Reduction Act does suggest that interagency data exchanges need to be considered as a means of reducing reporting burden and improving governmental efficiency. In addition, the Senate in an amendment to taxpayer privacy statutes, has approved legislation encouraging the IRS to share information with other law enforcement agencies investigating nontax crimes (Pear, 1981). California Government Code provides for exchanges of certain administrative records for "compatible" purposes. Nevertheless, there is no evidence to suggest that researchers who need to link administrative records would be permitted the same sorts of access as members of a government agency. (See, for example, Alexander and Jabine, 1978.) As the American Statistical Association's Ad Hoc Committee on Privacy and Confidentiality noted, the IRS has discontinued, on Privacy Act grounds, address locator services it had been providing to qualified researchers conducting follow-up studies. And the Social Security Administration has "suspended a procedure whereby files containing survey data collected by University researchers merged with SSA earnings data and bearing case numbers (not individual identifiers) were made available to the university research group" (pp. 65,66).

6. A similar case involves the property tax rolls. The public records laws declare property tax assessments to be open documents at the county level. Failure to anticipate this situation in legislating restricted access to administrative data of the Department of Revenue implied that the MRR created by the department from these open records are restricted and unavailable for research access.

7. This typology was originally developed for evaluating the extent of federal and state statutory protection for access to confidential records for research and statistical purposes. See Nijelski and Peyser (1975). These criteria were then elaborated on by Boruch and Cecil (1979).

8. We do not intend to imply that statutory or legal access is not permitted. For example, the Privacy Act of 1973 does provide access to nonidentifiable data and identifiable records can be accessed "if an agency was to define 'research and statistical analysis' as a routine use of an agency records system" (Boruch and Cecil, 1979:247). In addition, the U.S. Public Health Services Act, Alcohol Abuse Research Law, Crime Control Act of 1973, and Drug Abuse and Treatment Act also provide access for research purposes. The Georgia Public Records Act and California Information Practices Act of 1977 permit access to individually identifiable records for legitimate research activities. However, research access is not a right, but is subject to administrative discretion.

9. See David (1980) for an idea of what secondary analysts would like in gaining access to data.

10. Miller (1971) has advocated the idea of bonding or stipulated judgments against researchers releasing data.

11. Lack of training underscores one of the important ingredients of a strategy for research access. A wide variety of personnel in agencies need training in a variety of skills to understand the value of secondary analysis for intelligent policy making and creation of valuable knowledge about society. This need is an opportunity for the research community to provide skills and training while at the same time establishing better communications to attack the problem of archival preservation and research access.

REFERENCES

Alexander, L. A., and T. B. Jabine
1978 "Access to Social Security Microdata Files for Research and Statistical Purposes: An Overview" Presented at the NBER Workshop on Policy Analysis with Social Security Research Files, March 15, 1978. (Published in Policy Analysis with Social Security Research Files, U.S. Department of Health, Education and Welfare, Social Security Administration, Office of Policy, Office of Research and Statistics, 1978, 3–32; and in Review of Public Data Use 8(1):13–26.)

American Statistical Association. Ad Hoc Committee on Privacy and Confidentiality
1977 Report. The American Statistician 31(2):59–78.

Anderson, R. E., and F. M. Sim
1977 Data management and statistical analysis in social science computing. American Behavioral Scientist 20(3):367–410.

Boruch R. F., and J. S. Cecil
1979 Assuring the Confidentiality of Social Research Data. Philadelphia: University of Pennsylvania Press.

Boruch, R. F., and J. Reis
1981 The student, evaluative data, and secondary analysis. New Directions for Program Evaluation 8:59–72.

Cartwright, D. W., and P. A. Armknecht
1980 Statistical uses of administrative records. Review of Public Data Use 8(1):13–26.

David, M. H.
1980 Access to data: the frustration and utopia of the researcher. Review of Public Data Use, 8(4):327–337.

Dobson A.
1980 Towards an NHI data base: the linking of administrative record systems to national health care survey results. Review of Public Data Use 8(2):167–173.

Friend, I., and R. D. Jones eds.
1981 Conference on Consumption and Savings. Philadelphia: University of Pennsylvania Press.

Horton, Jr., F. W.
1980 The Freedom of Information Act and the problem of access. Review of Public Data Use 8(2):117–120.

Hulett, D. T.
1975 Confidentiality of statistical and research data and the Privacy Act of 1974. Statistical Reporter June:197–209.

Kleiner, M. M.
1980 Statistical uses of administrative records for labor market information. Review of Public Data Use 8(1):43–50.

Menchik, P., and M. David
1980 The Effect of Income Distribution and Redistribution of Lifetime Savings and Bequests. Madison, WI: University of Wisconsin Institute for Research on Poverty, IRP DP-637-80.

Miller, R.
1971 "Computers and Privacy: What Price Analytic Power?" Proceedings of the Association for Computing Machinery.

Nijelski, Paul, and Howard Peyser
1975 Appendix B. A Researcher's Shield Statute: Guarding Against Compulsory Disclosure of Research Data. In *Protecting Individual Privacy in Evaluation Research.* Ed. Alice M. Rivlin, pp. B-1–B-36. Washington, D.C.: National Academy of Sciences.

Pear, R.
1981 Bill allows I.R.S. to share evidence. The New York Times July 29:6.

President's Reorganization Project for the Federal Statistical System
1981 Improving the federal statistical system: issues and options. Statistical Reporter, 81–5:133–221.

Privacy Protection Study Commission
1977 Personal Privacy in an Information Society. Washington, DC: U.S. Government Printing Office.

Robbin, A.
1981a "Technical Report Number 4. The Social Utility of Personal Information. An Examination of and Recommendations for Statutory Protection and Agency Policies and Practices Regarding Research Access to Confidential Records." Madison, WI: University of Wisconsin and State Historical Society of Wisconsin.
1982b Technical Guidelines for Preparing and Documenting Data. In R. F. Boruch, P. Wortman, and D. Cordray, eds. Reanalyzing Program Evaluations: Policies and Practices for Secondary Analysis of Social and Educational Programs. San Francisco: Jossey-Bass.

Robbin, A., and M. Hedstrom
1981 Technical Report Number 3. A Report on Data Processing and Machine Readable Records in the Department of Health and Social Services. Madison, WI: University of Wisconsin and State Historical Society of Wisconsin.

Roistacher, R. C., et. al.
1980 "A Style Manual for Machine-Readable Data Files and Their Documentation." Washington, DC: National Criminal Justice Reference Service, Report No. SD-T-3, NCJ-62766.

State Historical Society of Wisconsin
1981 "Final Report of The Wisconsin Survey of Machine Readable Records Project." Madison, WI: State Historical Society.

Taubman, P.
1981 Justice Department hopes to toughen rules on access to data. The New York Times May 4:1, 13,