In the 1960s and 1970s, the American Statistical Association (ASA), the federal statistical community, the U.S. Congress, and the general public debated the expanded data collection activities and analysis made possible by advances in statistical methodology and computer technology. A key element of that debate involved public concerns about threats of intrusive governance surveillance of individuals and their claim to privacy. Within ASA, two temporary committees addressed these matters, first in the early 1960s, and then in the mid-1970s. The 1970s committee recommended the association create a permanent committee, which it did, as the Committee on Privacy and Confidentiality in 1977. This paper provides an overview of the technical, political and administrative issues that arose from protecting the privacy and data confidentiality interests for data collection, research, and public release of information in the mid to late 20th century.

Introduction

In 1977, ASA President Leslie Kish and the Board of the ASA established the Committee on Privacy and Confidentiality (P&C) as a permanent ASA committee. The action to establish a permanent committee followed from the 1976 ASA report and recommendations of an Ad Hoc Committee on Privacy and Confidentiality appointed by ASA President Lester Frankel. The initial charge to the permanent committee was to continue the work of the ad hoc committee, namely to "continue the review of new legislation and of legislative and policy guidelines [on privacy and confidentiality]; monitor the application of privacy and confidentiality laws and regulations; communicate important developments in this area to the ASA members; provide an early warning system to the board to take official positions when appropriate; provide appropriate liaison with congressional committees and federal agencies; and, at the request of the president or board, represent ASA in hearings on privacy and confidentiality."1

---

The establishment of the permanent committee signaled that ASA would affirm and strengthen its longstanding interest in and contributions to the development of statistical confidentiality standards for data collection, processing and management. Those contributions date from the origins of ASA in the nineteenth century down to the present. A look back at the earlier policy development and the committee’s origins in the controversies that arose during the 1960s and 1970s provides insights for the P&C committee’s charter, the ASA and persons working in the field of privacy and data confidentiality.

To understand this history, we searched and examined the voluminous analytic and evidentiary record on issues of statistical confidentiality and privacy in government and academic sources, and the internal records of ASA, namely the archived materials of the committee and related ASA officials available at the Park Library Collection at Iowa State University.

Definitions

The legal, administrative, and political issues surrounding the terms “privacy” and “confidentiality” in statistical policy are not the same. Definitions were developed during the last century and a half, but they were not linked as we discuss them today in the statistical literature until roughly the 1970s.

Following Justice Brandeis’ famous phrase that the right to privacy is “the right to be let alone,” privacy law defines a ‘right to privacy’ as an individual’s (person or organization) right “to determine what sort of information about themselves is collected, and how that information is used.” A violation of this right, a breach of a person’s privacy interest, implies an unlawful intrusion of an individual’s privacy rights and redress under the law. The right to privacy also includes an individual’s refusal to provide information. Exceptions to the right of refusal include information that is required by law or is transmitted as part of a contractual relationship between the individual and the recipient of the information, as in a lawyer/client relationship.

Confidentiality refers to a quality or condition accorded to information as an obligation not to transmit that information to an unauthorized party.  

---


Confidentiality involves the recipient’s responsibility for the protection of the individual’s information once provided, and is guided by contract law, professional ethics, and administrative practices designed to guarantee an individual’s private information is not divulged without consent.4

The organization or individual requesting the information assures the respondent that the information will be held in confidence and that the requestor guarantees protection of the information from unlawful or unethical use. Or from the data supplier or respondent’s perspective, confidentiality means that the information will be reported to the data collector and will not be shared with others. “Statistical confidentiality” in this sense is a special set of restrictions relating to the collection, use, and disclosure of information.

Former census director Kenneth Prewitt has recently emphasized the statistical community’s need to understand the distinction, and made a “plea” that research and analysis “start from a fresh conceptual foundation—one that considers rather than blurs the distinction between privacy and confidentiality.”

Prewitt argued that the “citizen saying ‘I won’t answer that question because it is none of your business’ is sending a different message from the citizen who is saying ‘I won’t answer that question because you may share it without my knowledge or permission,’ and neither response is the same as ‘I won’t answer because I don’t trust the government not to use my answers against me.’”

He emphasized that “privacy, confidentiality, trust, and fear of misuse mingle in complicated ways, but this does not relieve us of the responsibility to sort them out as best we can. To treat these public concerns in a largely undifferentiated fashion, as much of the research literature has, makes for a poor guide to nuanced policy and practice responses by statistical agencies.”5

In other words, respondents and data collectors have differing perspectives on the propriety and legitimacy of the data transaction. Privacy is fundamentally an issue of the individual data provider, a willingness to provide information, or a financial, legal, or other responsibility to do so. Confidentiality is fundamentally a


responsibility of the data collector, hence the use of the concept of data stewardship to encompass the collector’s role.

The Development of Privacy and Confidentiality Protection in Statistical Practice

The statistical community broadly defined, as represented in the American Statistical Association, has long been concerned with the development of statistical confidentiality practices, and has primarily addressed issues of “privacy” in the context of efficient, ethical, and informative data collection and analysis. The statistical community historically has been concerned with respondent privacy concerns insofar as they interfere with the work of data collection and analysis.

Statisticians and data collectors developed their understanding of methods and standards for managing and protecting the information they collected through a process of trial and error. For example, there were “privacy” challenges to census questions in the U.S. dating back to the first half of the 19th century. There were public complaints over the 1840 census and it was the subject of cartoons in the press and other media as early as the 1850s. A Saturday Evening Post cartoon from 1860, for example, mocked a boorish enumerator with the caption, “I jist [sic] want to know how many of yez [sic] is deaf, dumb, blind, insane and idiotic—likewise how many convicts there is in the family—what all your ages are, especially the old woman and the young ladies—and how many dollars the old gentleman is worth!”

One of the more serious concerns was that census enumerators would use the information they collected for private gain, or to harass their neighbors. During the 19th century, the data values reported on census forms were considered public documents. Completed census forms were initially posted in public places so people could check accuracy of the responses. Over time, objections from the public and the expense of hand copying duplicate returns ended these practices. In the context of the difficulty of collecting and publishing the census, the officials who managed the census in the nineteenth century showed little concern about these matters beyond the embarrassment to the reputation of the enterprise.

Data collection efforts increased dramatically in the late nineteenth century to provide more information on the economic and social issues of the new industrial economy. The Bureau of Labor (now BLS), established in 1884 quickly confronted the concerns of employers and businessmen to provide information that they feared might be accessible to competitors. Commissioner Carroll Wright went to great pains to assure economic respondents that their information would be held in confidence, though he had no statutory authority to support his claims.

Nevertheless, in an era when data management and analysis was done on paper, by hand counting methods, the sheer unwieldiness of the data provided significant protection. The census office until 1902 was a temporary agency, literally mothballed each decade with paper files put in storage.  

Debate about non statistical uses of information collected by the federal government began to impact the statistical system as the regulatory powers of the federal government grew. During the first decade of the twentieth century, the permanent Census Bureau and the Bureau of Labor Statistics moved into the new Department of Commerce and Labor and became sister agencies to others with regulatory enforcement powers. This new administrative location prompted the clarification of the status of information collected by the statistical offices, establishing a firewall of confidentiality for data collected in censuses and surveys, and introducing confidentiality provisions for economic data into the Census Act in 1909. By 1929, those protections were extended to population data.  

In other words, from approximately 1880 to 1930, government survey statisticians worked out rules and procedures for guaranteeing respondent confidentiality, and codified them in administrative practice and law. Cartoons, like the one in the Saturday Evening Post, objected to questions about personal wealth and disability. Congress, which until 1930 legislated the actual questions on the census each decade, responded by removing such sensitive questions from the decennial census. Throughout this period, ASA members contributed to working through the issues, as

---


government officials, academics, or corporate leaders. The pages of JASA and the records of ASA government advisory committees, detail the debates.9

Thus when social science data collection and government survey work expanded during the 1930s to assess patterns of unemployment, migration, build the system of national accounts, and the new social security system, they relied upon the legal and administrative practices for data stewardship that was set up from the mid 19th century. Those practices assumed a centralized system of record keeping often with literally one copy of key documents, secured and preserved by the collecting agency, primarily on paper or index cards, supplemented perhaps with punch cards for machine tabulation.

The invention of the digital computer began to change the technical underpinnings of data stewardship in the 1950s. First used to tabulate the 1950 census, main frame computing capacity spread rapidly within large public and private institutions in the 1950s and 1960s, and the advent of computer tape files made it possible to convert mechanically counted punch cards to a much more concise digital record which could be shared among researchers. That in turn led to discussions of wider research and policy use of existing datasets, the potential for matching records from different databases, and the possibilities of data sharing among researchers beyond the original collectors. The Census Bureau produced the first Public Use Microdata Sample (PUMS) files from the 1960 census, on both tape and punch card media. They reported that only 85 users bought the new product, still reflecting the scarcity of large scale computing power outside of the federal government.10

Throughout the first half of the twentieth century, there were also occasional challenges to the confidentiality of statistical data collected by government. For example, during World War II when the Second War Powers Act of 1942 lifted the confidentiality restrictions of Title 13 of the Census Act for both population and economic data so the data could be used for administrative and regulatory purposes to support the war effort. The confidentiality provision was restored in 1947.11

9 The American Statistical Association and American Economic Association established an advisory committee to the Census Bureau at the end of World War I. Anderson, The American Census and Duncan and Shelton document the role of ASA in statistical policy development broadly defined. Stuart Rice, for example, 1933 President of ASA, took key positions in the federal statistical system from the 1930s to the 1950s. He served as the acting chair of the Committee on Government Statistics and Information Services (COGSIS), Assistant Director of the Bureau of the Census, and from 1940, the first head of the Office of Statistical Standards in the Bureau of the Budget (now OMB). See the Amstat News piece on his contributions, “Stuart Arthur Rice: Student Activist to Statistical Statesman,” [http://magazine.amstat.org/blog/2010/09/01/rice9_10/](http://magazine.amstat.org/blog/2010/09/01/rice9_10/)


the 1950s, the Department of Justice challenged the Bureau of Mines’ claim to data confidentiality of its information on the oil industry.\footnote{12} And the Federal Trade Commission and the Antitrust Division of the Department of Justice litigated an antitrust case which forced the St. Regis Paper Company to disclose the information on its own file copy of the economic census form.\footnote{13} The Supreme Court upheld the Department of Justice position. In 1962, the statistical community and agency officials successfully lobbied Congress to amend Title XIII to prevent such a disclosure in the future.\footnote{14}

The statistical community was alert to these challenges, and though in 1962 they did not initially generate huge consternation, they did prompt a discussion within the ASA Board. This led to a recommendation to create a committee to “develop and to recommend a confidentiality position for the Association.”\footnote{15} In 1962, President Philip Hauser appointed a “Committee on the Confidentiality of Statistical Returns,” with Frederick Mosteller as chair, and committee members William H. Shaw, Martin R. Gainsbrugh, Harold F. Dorn, and Ernest J. Engquist and charged the committee to report and make recommendations to the Council.\footnote{16} The committee convened for the next two years and quickly confronted the complexity of the issues. Their discussions reflected differing positions on the issues of protecting statistical confidentiality. Harold Dorn, in a December 28, 1962 memorandum to the committee, acknowledged there were serious issues with protecting the confidentiality of the data, but was dubious of the committee’s ability to resolve the issues from the outset: “I can think of no aspect of the problem of confidentiality in this area to which the American Statistical Association can make a useful contribution.” Shaw, in a January 4, 1963 letter to the committee disagreed: “...if the statistical profession cannot resolve at least some of the more nagging confidentiality questions, it would seem to be avoiding an area which could vitiate a good part of the efforts connected with techniques of data collecting.” Raymond Bowman, ASA President and Director of the Office of Statistical Standards in the Bureau of the Budget identified some ten issues for the committee, including issues of data sharing of sampling frames and directory information; matching; confidentiality policies of trade associations vs. government. For the next two years

\begin{flushleft}
\end{flushleft}

\begin{flushleft}
\footnote{12} Id.
\footnote{13} Id.
\footnote{14} Anderson and Seltzer, “Challenges;” Anderson and Seltzer, “Federal Statistical Confidentiality.”
\end{flushleft}

\begin{flushleft}
\footnote{15} William Shaw to Donald C. Riley, August 23, 1962; ASA MS349, Box 34, Folder 2.
\footnote{16} Hauser to Mosteller, October 1, 1962, Ibid.
\end{flushleft}
the committee wrestled with these questions. In January 1965, it was “discontinued for the time being.” It is not clear that the committee ever filed a formal report.

Meanwhile, myriad economists, statisticians and social scientists, recognized the potential for secondary data analysis of the expanding information infrastructure of the federal government overall and the statistical data collections in particular. They began to articulate bold proposals for archiving, coordinating and sharing data. Those efforts led to recommendations in various professional associations and the creation of a committee of the Social Science Research Council on the “Preservation and Use of Economic Data.” Their 1965 report recommended the creation of a national data center to preserve, archive and disseminate the data across the federal statistical system to agencies and private researchers. Raymond Bowman, Director of the Office of Statistical Standards in the Bureau of the Budget, and ASA President in 1963, supported the proposal, and commissioned a report from Edgar Dunn in late 1965 on next steps, including any required legislation. A second commission appointed by the Bureau of the Budget recommended (Kaysen Report) in October 1966 the creation of a new position of Director of the Federal Statistical System in the Executive Office of the President, to manage the proposed national data center.

The social scientists and officials from the federal statistical system who proposed the national data center were concerned to this point with administrative structures, technical questions of data harmonization and computer capacity, duplicative data collection, and the potential for increased social research. When Congress began to focus on the proposals in 1967, however, they interpreted the propriety of the proposed national data center in the context of other legislative concerns, most notably an emerging debate about personal privacy, the power of large organizations to abuse the information in their computerized data collections, and the concerns of ordinary citizens and businesses about misuse of their information. In the media, the national data center was quickly translated into an image of an Orwellian “data bank” which would collect impertinent information on individuals without their knowledge and preserve them for potentially embarrassing or inappropriate revelations to government bureaucrats or

17 Dorn to the Committee; Shaw to the Committee; Bowman to the Committee, Ibid.; Donald Riley to Mosteller, Gainsbrugh, Shaw and Bowman [Engquist had died], January 28, 1965, Ibid.
anonymous private researchers. The words “privacy” and “confidentiality” began to be linked rhetorically, though from the ASA perspective, “privacy and confidentiality” still meant the array of practices and issues surrounding data confidentiality and stewardship, not the different issues of monitoring and surveillance that were on Congress’ legislative agenda.

The social scientists and statisticians struggled to explain their good intent, including the confidentiality protections embedded in the proposal. Congressional critics responded that the protections were irrelevant if the entire enterprise was perceived as an invasion of privacy. The Johnson and Nixon administrations, slowed planning and ultimately quietly shelved the national data center plan. Congress began work on a number of bills to acknowledge the twin issues of information privacy and freedom of information, passing the first Freedom of Information Act in 1966, the Privacy Act in 1974 and a second freedom of information act in 1976. These laws focused on restricting government agency authority to collect and store information on individuals that related to their official use, and on the public’s right of access to government information.

Thus by the early 1970s, the controversy over a federal “data bank” had added new aspects to the issue of the statistical confidentiality. Congressional legislation on an individual’s right to privacy implicitly also led to the question of what information the government should be permitted to collect and possess on individuals, and on whether information from different files could or should be linked. What had been a somewhat arcane, quiet discussion among social scientists and data professionals became a political hot potato that consumed time and energy and potentially threatened further data development and coordination. The Nixon administration responded to the “data bank” controversy and other emerging concerns with the statistical system by appointing a Commission on Federal Statistics. The Commission produced a two volume report in 1971, and included a chapter with recommendations on “privacy and confidentiality.” The Commission also recommended the creation of a statistical policy committee in the National Academy of Sciences. The Committee on National Statistics (CNSTAT) was founded in 1972, “to improve the statistical methods and information on which public policy decisions are based.”

Looking back, one can see why the issues of privacy and confidentiality became public concerns in the 1960s. Despite the fact that data professionals had recognized these issues for many years, and had developed a policy framework for protecting micro data, the rapid expansion of main frame computer use, the expansion of government activities affecting individuals and firms, and a number of scandals about research misconduct, made the public suspicious of assurances that

their data were protected from misuse and unauthorized access and release. For example, once individuals recognized that their paychecks or savings bonds were actually IBM punch cards, they realized their daily activities could be digitized and how the computerized world would affect their lives. In the 1960s, only large organizations, corporations, or the state, could afford these new technologies. It would be another 10-15 years before innovative developers in Silicon Valley designed the microcomputers that would democratize data use. By the 1990s, the Internet developed, and with personal computers and cell phone technology, the digital world became omnipresent in personal lives and the operations of small businesses as well.

The Association Takes Up the Issues

During the 2nd half of the 20th century, statisticians and researchers continued to press for more and better data, and thus also began to confront how their professional expertise could be brought to bear on both the concerns about the newly redefined questions of “privacy and confidentiality” as well as data access. Innovation occurred in statistical methodology during the 1970s and in the organizational work of the American Statistical Association.

On the statistical front, for example, Ivan Fellegi published a seminal article in the Journal of the American Statistical Association (JASA) in 1972 addressing the technical issues of disclosure in data tabulations.20 He pointed out that the increasing demand for detailed tabulations had raised a new problem of disclosure for agencies. “[I]t is also incumbent on the statistical offices to maintain a continuous scrutiny of their own publications to insure that from their publications no information can be deduced concerning particular respondents” (pp 8-9). He proposed a theory and data sensitivity tests for identifying and thus preventing “direct” or “residual disclosure” in publications, special tabulations, summary tapes, and the emerging practice of creating PUMS files. By the 1980s, these analyses led to a new field of “disclosure limitation methods,” which took up these technical insights to develop tools for identifying tabulations or data files at risk of identifying individual respondents, and then for developing statistical tools for protecting data.

On the organizational front, discussion of a committee or forum for considering issues of privacy and confidentiality resumed as well. ASA’s Social Statistics Section, which was a major section for government statisticians, recognized the need for association wide consideration of the issues once again. As Ed Goldfield reported in his historical review of the section for the ASA Sesquicentennial, “Another cross-cutting issue that was assuming larger proportions [of the energy of the section] was that of privacy and confidentiality, and the section gave attention to it. An example was the session cosponsored by the section at the 1975 ASA annual meeting -- Privacy, Freedom of Information, and Federal Statistical Programs. In

---

response to a request from the ASA national office, the section spearheaded the
developmental activity that led to the formation of an ASA ad hoc committee on
privacy and confidentiality in 1975. The committee produced a 20-page report that
was published in *The American Statistician* in May 1977 and was forwarded to the
Privacy Protection Study Commission. The ad hoc committee recommended that it
be succeeded by a continuing committee on privacy and confidentiality.”21

The Ad Hoc Committee was chaired by Joseph Gastwirth and met in 1975 and
1976.22 It produced a detailed report of the complex issues of confidentiality facing
statisticians and data analysts in the new and publicly sensitive environment of
social, economic, health, and educational research. New comprehensive federal
statutes in the area of privacy protection and freedom of information required the
statistical profession to look at existing practices and see to what extent these new
laws affected data collection, tabulation and dissemination. The committee
surveyed federal agency practices and discovered uneven patterns in applying data
protection methods. Experienced agencies like Census and BLS with long
established protocols and statutory or administrative protections had different legal
obligations and higher levels of statistical expertise than the smaller statistical
offices embedded in regulatory agencies where administrative data also served
statistical analyses. The Ad Hoc Committee reported on long standing data sharing
arrangements among some agencies, e.g., IRS to Census, and barriers to data sharing
in other venues such as IRS to Agriculture. They debated the advisability of using
survey or administrative data from another file to supply information a respondent
left blank. A minority of members (including Chairman Gastwirth) opposed such
“borrowing” without the respondent’s knowledge. The committee suggested areas
for clarification and changes to the new Privacy Act of 1974 and FOIA procedures,
recognizing that as they wrote, the implementation of these new statutes was very
much a work in progress.

Above all, the work of the Ad Hoc Committee demonstrated the need for ongoing
attention to the issues of privacy and confidentiality by the Association. In 1977.
The ASA leadership agreed and established the Committee on Privacy and
Confidentiality. As the initial charge to the committee indicates, the Association
recognized that it could no longer stand aside from the larger political and policy
debates about confidentiality, or speak quietly and authoritatively in the
administrative offices of government or the private conversations among

21 Edwin D. Goldfield, “History of the Social Statistics Section.” The American
22 American Statistical Association. “Report of Ad Hoc Committee on Privacy and
Confidentiality.” The American Statistician 31, no. 2 (May, 1977): 59-78. Other
members were John H. Aiken, A. Ross Eckler, Edwin D. Goldfield, Donald Guthrie,
Simmons, and Rita Zemalch.
researchers. The Association needed to monitor developments in Congress, testify and submit position papers, and prepare ASA leadership for public statements on issues of privacy and data confidentiality, and draw upon the expertise and commitment of members to support best practices in research and data development. The new permanent Committee on Privacy and Confidentiality was established in 1977 to implement these mandates.

Appendix

ASA Archival Materials in the Parks Library, Iowa State University, Ames, IA.


1. The records of the 1962-1965 Committee on the Confidentiality of Statistical Returns are in Box 34, Folder 2.

2. The records of the Ad Hoc Committee on Privacy and Confidentiality, 1974-76, are in Box 77, Folders 2-4.

3. The records of the ASA Advisory Committee to the Bureau of the Budget on Statistical Policy, 1947-1967, are in Box 15, Folders 11-12; the 1968-71 records are in Box 49, Folder 10.