

CENSUS CONFIDENTIALITY AND PRIVACY: 1790 - 2002

TABLE OF CONTENTS

INTRODUCTION	1
THE GROWING NEED FOR CONFIDENTIALITY SAFEGUARDS	3
1790-1840: Shouldn't This Information Be Confidential?	3
Census results publically posted	4
First concerns about economic data	4
1850-1950: The Long Journey to Title 13	5
Secretary of the Interior's statement	6
Superintendent of the Census' declaration	6
Local supervisors replace U.S. marshals	7
The 1880 and 1890 Census Acts	7
Confidentiality strengthened and extended to individuals	9
Access to individual census records	11
DISCLOSURE LIMITATION AT THE CENSUS BUREAU	20
The Disclosure Review Board	34
RESTRICTED ACCESS TO CONFIDENTIAL DATA	35
Choosing and Setting up New RDCs	36
Securing the office	37
Security functions of personnel at the RDC	39
Protecting Confidentiality for the Life Cycle of Research Projects	39
Initial inquiry	40
Project selection	41
Evaluating Confidentiality Safeguards at the RDCs	42
PRIVACY AT THE CENSUS BUREAU: 1974 - 2002 AND BEYOND	43
SUMMARY	51

INTRODUCTION

If the framers of the U.S. Constitution thought that the census might be viewed as an intrusion on personal privacy or foresaw any need to keep census data confidential, their misgivings were not evident when they approved Article 1, section 2, providing for a decennial census. Confidentiality and privacy may not have been an issue then since the first enumeration in 1790 collected minimal information and only produced statistics needed by the Federal Government for a few specific purposes, such as—

- ! Reapportioning seats in the House of Representatives
- ! Levying direct taxes on each state based on its population
- ! Determining the country's military potential in case of war

Only later, as the amount of data collected became more extensive, would census officials gradually become aware of the public's privacy concerns and the need to establish confidentiality safeguards that today are an integral part of census-taking operations.

By 1850 however, the need for information had expanded and Federal Government officials, statisticians, and others saw the census as a means of gathering more information on a growing number of demographic (people) and economic (business) topics.¹ Consequently, as the number (and sensitivity) of questions in the census increased, so did the potential for abusing privacy (an individual's or business' interest in personal or proprietary information weighed against the Government's need to know) and

¹Margo J. Anderson, *The American Census: A Social History*, Yale University Press, New Haven, CT, 1988.

confidentiality (the Government's responsibility not to disclose individual census information to anyone else).

Although the concepts of privacy and confidentiality are difficult to separate, most of this monograph's focus is on the confidentiality of census information and its historical evolution between the 1790 and 2002 censuses.² Since the American public's privacy concerns are of more recent origin, dating back to the events in the 1960s and early 1970s that lead to the Privacy Act of 1974, the Census Bureau's historical response to these concerns is briefly dealt with in the last section.

The monograph includes the following four sections—

1. The Growing Need for Confidentiality Safeguards. This section traces (more or less chronologically) the evolution of confidentiality safeguards between the 1790 and 1980 censuses—starting with the absence of these safeguards, continuing through the enactment of Title 13, United States Code, the protection of individual census records, and concluding with the U.S. Supreme Court's decision stating that even address listings are confidential.
2. Disclosure Limitation at the Census Bureau. In addition to assuring the confidentiality of individual census records, the Census Bureau uses disclosure limitation to prevent the identification or harm of any person or establishment from its published consolidated data. This section discusses how disclosure limitation works, its history at the agency, its most recent applications during the 1990 and

²This monograph was concluded shortly before the data-collection process began for the 2002 Economic Census.

2000 censuses, and concludes with the oversight provided by the Disclosure Review Board.

3. Restricted Access to Confidential Data. This section describes the confidentiality measures put in place by the Census Bureau's Center for Economic Studies (CES) since 1994 to enable researchers to use economic microdata at seven offsite Research Data Centers (RDCs) located around the country.
4. Privacy at the Census Bureau: 1974 - 2002. This section briefly describes what has transpired at the Census Bureau since the passage of the Privacy Act of 1974, the agency's internal structure for establishing privacy policy, and privacy research already conducted and now underway.

THE GROWING NEED FOR CONFIDENTIALITY SAFEGUARDS

1790-1840: Shouldn't This Information Be Confidential?

The first 50 years of census taking is most notable for the absence of concern about confidentiality. A few individuals did oppose the first census in 1790, and enumerations in the early 1800s on religious grounds. They cited the Bible (II Samuel 24: 1-15), where it is said that King David's taking of the census of Israel and Judah resulted in an epidemic that killed 70,000, as a reason to refuse giving any information. They also pointed out that the unwelcome results of other biblical censuses were military service and taxes. Apart from this opposition though, little if any evidence can be found that Americans were afraid of what their government would do with this personal information once they provided it to census takers.

Census results publically posted

Perhaps due to this lack of public opposition, the laws authorizing the taking of the census did not provide for the confidentiality of the information collected. In fact, the laws governing census taking between 1790 and 1840 required the assistant U.S. marshals (who were responsible for data collection between 1790 and 1870) to post the returns in “two of the most important places” in their enumeration districts.³ The idea was that anyone incorrectly enumerated, or not enumerated at all, could come forward and correct the mistakes or be added to the list. Most Americans cooperated with the first census (there were penalties for refusing) and apparently raised few objections when the assistant marshals posted the returns in their enumeration districts.⁴ In the population censuses, this procedure of posting census results remained in effect until the 1850 census.

First concerns about economic data

The conduct of the census of manufactures, which began in 1810, was somewhat different from that of the population census. First, response was voluntary and generally remained so until 1880. Congress believed that persons engaged in manufacturing were likely to cooperate and would share the members’ favorable view of the need for manufacturing data.⁵

³Carroll D. Wright and William C. Hunt, *History and Growth of the United States Census: 1790-1890*, Government Printing Office, Washington, DC, 1900, pp. 131-47.

⁴Ibid.

⁵Ibid., p. 136.

The marshals did not post the returns from the 1810 census of manufactures but did file copies with the U.S. District Court clerks. Response to the 1820 census of manufactures (both in coverage and quality) was so poor that Congress opted not to include any questions on manufacturing in the 1830 census act.⁶ Congress restored the questions on manufacturing in the 1840 census but, bowing to confidentiality concerns, instructed the assistant marshals to assure respondents to these questions that no individual or company names would appear in the statistical tables. Further, the assistant was to “consider all communications made to him in the performance of this duty, relative to the business of the people, as strictly confidential.”⁷

1850-1950: The Long Journey to Title 13

From the 1850 census on, no returns of any sort were posted in public places. The 1850 census act required each assistant marshal to prepare an original and two copies of his returns. The original was to be deposited with the appropriate county clerk. The marshals then sent one set of the copies to the U.S. Secretary of the Interior for processing and transmitted the second set to the secretary of the state government or territory. These documents most likely came under whatever access rules these local officials specified but the law itself made no reference to confidentiality.

⁶The 1790 - 1950 decennial censuses were authorized by acts passed by Congress prior to each census. The 1960 and subsequent censuses are authorized by Title 13 passed in 1954.

⁷Ibid., p. 145.

Secretary of the Interior's statement

Census procedures during the 1850 census, though, were in a sense ahead of the law at this time. The U.S. Secretary of the Interior, Thomas McKennan, newly charged with responsibility for the enumeration, officially reminded the marshals and their assistants about current policyS

Information has been received at this office that in some cases unnecessary exposure has been made by the assistant marshals with reference to the business and pursuits, and other facts relating to individuals, merely to gratify curiosity, or the facts applied to the private use or pecuniary advantage of the assistant, to the injury of others. Such a use of the returns was neither contemplated by the act itself nor justified by the intentions and designs of those who enacted the law. No individual employed under sanction of the Government to obtain these facts has a right to promulgate or expose them without authority.

...all marshals and assistants are expected to consider the facts intrusted to them as if obtained exclusively for the use of the Government, and not to be used in any way to the gratification of curiosity, the exposure of any man's business or pursuits, or for the private emolument of the marshals or assistants, who, while employed in this service, act as the agents of the Government in the most confidential capacity. When your original copies are filed with the clerks of the courts and secretary of your state, they will be under the control of those officers and subject to the usual regulations of the respective offices, and you can enjoy the same access to them which can be had by every citizen. To the publication of the mere aggregate number of persons in your district there can be no objection.⁸

Superintendent of the Census' declaration

Twenty years later the superintendent of the 1870 census, Francis Amasa Walker, reinforced Secretary McKennan's statement by informing the marshals and their assistants that "strict and literal compliance [to the 1850 Census Act, under which they still operated]

⁸Ibid. p. 150, Circular to the United States Marshals and Assistants.

in every particular will be enforced,” and in instructions to the assistant marshals [the actual enumerators] added—

No graver offense can be committed by assistant marshals than to divulge information acquired in the discharge of their duty. All disclosures should be treated as strictly confidential, with the exception hereafter to be noted in the case of the mortality schedule [where professional review by a local physician was authorized]. Information will be solicited of any breach of confidence on the part of assistant marshals. The [Department of Interior] is determined to protect the citizen in all his rights in the present census.⁹

Local supervisors replace U.S. marshals

Some proposed procedural changes for the 1870 census were not implemented until 1880. These changes, once in place, increased the U.S. Census Office’s control over the enumeration itself, over how the returns could be made more accurate than in the past, and over how these returns would be protected from abuse. A major change was that the U.S. marshals and their assistants were relieved of census responsibilities and replaced by local supervisors appointed by the President with the advice and consent of the Senate. These supervisors were authorized to select indigenous enumerators “solely with reference to their fitness, and without reference to their political party or party affiliations.”¹⁰

The 1880 and 1890 Census Acts

The 1880 and 1890 census acts indicate that the principle of census confidentiality was still evolving, and appeared to be more focused on information about property or business than with personal characteristics. For example, if members of the general public resisted answering questions, on any grounds, about whether they were paupers or

⁹Ibid., p. 156.

¹⁰Ibid., pp. 936-43.

convicts, as they had been asked in 1850 and 1860, the census either dropped the items entirely (as in 1870) or moved them to special institutional questionnaires (as in 1880 and later years).

The 1880 Census Act required the enumerators on oath not to disclose any information they collected to anyone except their supervisors. In Section 12 of the act confidentiality on both the supervisors' and enumerators' parts appeared to be limited to "statistics of property or business." Section 13 of the 1890 act removed that stricture and prohibited disclosing "any information." For both acts, violation could lead to a \$500 fine.

Beginning in 1880, the completed schedules were no longer to be filed with local officials, but were to be sent by the local census supervisors directly to the Interior Department in Washington. An amendment to the original 1880 act called for these supervisors to—

- ! Prepare lists of the names of the people they had canvassed, with age, sex, and color, and file these lists with the county clerk.
- ! Advertise in at least three public places where they (the enumerators) would be to make corrections or additions.
- ! "Make known to the bystanders, ...the results of such inquiry for correction and the whole number of persons by him enumerated..."¹¹

¹¹Ibid., pp. 942-3.

The 1890 act eliminated the filing of name lists with county clerks, but it allowed them and other local officials to buy the lists, which would include name, sex, age, birthplace, and color or race, at the rate of \$0.25 per 100 names.¹²

The same permission appeared in subsequent census acts, varying only in the fee to be charged, and also was extended to individuals to cover “such data from the population schedules as may be desired for genealogical and other proper purposes.” No limitations on those data were spelled out, such as who might or might not be eligible to receive them. This was left to the Director’s discretion. The long-standing permission to furnish individual data as described above to governors of states and territories and to courts of record as well was not removed from the census law (Title 13, United States Code, Section 8) until 1976.

Confidentiality strengthened and extended to individuals

In the first Presidential proclamation on the census, President Taft in 1910 unequivocally promised confidentiality for all the information collected (the same words appeared in subsequent decennial proclamations)—

The sole purpose of the census is to secure general statistical information regarding the population and resources of the country, and replies are required from individuals only in order to permit the compilation of such general statistics. The census has nothing to do with taxation, with army or jury service, with the compulsion of school attendance, with the regulation of immigration, or with the enforcement of any national, state, or local law or ordinance, nor can any person be harmed in any way by furnishing the information required. There need be no fear that any disclosure will be made regarding any individual person or his affairs. For the due protection of the rights and interests of the persons furnishing information, every

¹²Ibid., p. 948.

employee of the Census Bureau is prohibited, under heavy penalty, from disclosing any information which may thus come to his knowledge.¹³

The 1910 census act continued the request for enumerators to keep all census information confidential. It increased the fine (\$500) for violating confidentiality to a maximum of \$1,000 and/or 2 years in prison. Further, the 1910 law for the first time specifically prohibited the Census Bureau from publishing any data in which an economic establishment might be identified.¹⁴ (The 1930 act [Section 28] extended this prohibition to identifying individuals as well.)¹⁵

Even with these changes, in 1910 confidentiality remained questionable. For example there were numerous cases of over- and under-counting that had to be investigated and resolved. During this investigation the Census Bureau found that in some areas unauthorized third parties had collected (or compiled) census data about individuals and handed these over to enumerators, who simply transcribed them to their official schedules and collected (and probably split) the fees. The agency called for legislation to prohibit such practices, as indeed was done in the 1920 act (Section 29).¹⁶

¹³"Proclamation for the Thirteenth Decennial Census," March 15, 1910.

¹⁴U.S. Bureau of the Census, *Census Bureau Legislation: Department of Commerce and Other Executive Departments*. (Robert H. Holley, comp.), Government Printing Office, Washington, DC, 1936, p. 48ff.

¹⁵*Ibid.*, p. 36.

¹⁶*Ibid.*, p. 44.

Access to individual census records

As indicated by the previously discussed census acts, the extent to which the original census records were kept confidential depended on the confidentiality requirements imposed by the laws authorizing the various censuses. Before the Census Bureau became a permanent agency in 1902, the public had been permitted unrestricted access to the census records from 1790 through 1880, since the laws at that time did not require these records to be kept confidential.

While the public did not have access to the 1890 and 1900 census schedules, apparently the only reason for this was that there were about 12.6 million family schedules¹⁷ for 1890\$so many that they could not be bound into volumes (of which an estimated 30,000 would have been required), and the 1900 schedules were still in the process of being bound into some 2,800 volumes. Again, the laws authorizing those censuses did not prohibit disclosure.

In 1904, with many of the older volumes in bad condition due to frequent handling, the Census Bureau decided to stop all public access to the original records and offer instead census transcripts, certified if required, for a small fee. (Aside from the usual genealogical interest in the individual census records, more and more Civil War veterans needed transcripts to prove their ages for pension purposes.)

Government requests for access to individual census records increased during World War I. Since the 1910 census law did not prohibit disclosure, the Census Bureau

¹⁷In 1921, before the age of microfilming, virtually all the 1890 population schedules were lost to fire and subsequent water damage.

furnished many transcripts to the U.S. Department of Justice, local draft boards, and individuals, especially in connection with cases where the individuals had been arrested for draft evasion. Men who were in doubt about their ages, and therefore their duty to register, also obtained transcripts as needed, as did both sides in legal cases related to the draft.

The Census Bureau also provided the U.S. Provost Marshal General with national estimates (using the 1910 census as a benchmark) of the number of men in various age groups for comparison with the registration figures. The U.S. Solicitor General, in an opinion dated June 26, 1917, held that, as the census law then stood, the “Director of the Census might, in the exercise of his discretion, furnish to the officials in charge of the execution of the Selective Service Law, information in regard to the names and ages of individuals, as it did not appear that any person would be harmed by the furnishing of such information for the purpose for which it was desired.”¹⁸

In a similar situation in 1920 requiring access to individual census records, the U.S. Department of Justice asked census officials in Toledo, OH to provide information about individuals’ citizenship from the 1920 Census of Population (then underway) for use in deportation cases instituted by the U.S. Department of Labor. Here, the Solicitor General followed a line of guidance similar to the one cited above, but noted that the Director also might take into consideration whether the request would interfere with the progress of the census. In addition, the opinion pointed out that the 1920 Census Act

¹⁸Cited in letter, E.R. Magie, Acting Solicitor, to the United States Secretary of Commerce, January 15, 1920.

prohibited the Director from disclosing information from private business concerns, but did not restrict his discretionary disclosure of individual information (by the Director, but not by any other Census Bureau employee without the Director's permission) from the population and agriculture censuses^S

Under the law, the Director's relation to the information gathered by the Government for census purposes seems to be in the nature of that of a custodian or guardian, who is to see that it is used for the purposes for which it was gathered and not for private purposes to the harm or detriment of the person or persons from whom it was obtained under the implied promises that it would be considered confidential.¹⁹

In 1921, a number of the states and local institutions started campaigns to reduce illiteracy in this country. To this end they sought and received the 1920 census on the subject, and realized that not only did the Census Bureau have the general data, but also names and addresses. A number of states and organizations found funds, and the Census Bureau put clerks to work compiling lists for them from the census records.

Prior to the 1930 census, the agency began interpreting confidentiality much more strictly. It established as policy that anyone applying for a census transcript is entitled to his or her own record and that of his/her minor child. But for anyone else the requestor must have a signed authorization. For a deceased person, a death certificate or similar evidence must be presented, as well as proof that the applicant is either a direct blood-line descendant or an heir. Thus, for example, census transcripts are not available to collateral descendants (such as the niece of a maiden aunt) or to someone simply trying to find out who was living in a particular household in the past 72 years. Even the release of a name

¹⁹Ibid.

requires an authorization, although normally the transcript will show the householder's name in addition to that of the applicant and his/her relationship to that householder.

In 1930, the Women's Bureau (a Federal agency) asked the Census Bureau for a list of the names, addresses, occupations, and employment status of women living in Rochester, NY. In light of the 1929 Census Act imposing strict confidentiality, the Census Bureau referred the request to the U.S. Attorney General. The Attorney General decided such information could not be released under the new statutory provisions.²⁰ Similar requests from law-enforcement or security agencies subsequently were routinely turned down. These included identification of foreign-born persons in a particular Washington, DC neighborhood where an official residence was being considered and a confidential verification—without authorization from the suspects—for a census search of addresses claimed as alibis in drug or immigration cases.

One of the best known cases involved the 1942 request for census information on Japanese Americans living on the west coast. Until recently, the Census Bureau maintained that it did not provide the U.S. War Department with names or addresses from the 1940 census in order to facilitate the round-up of Japanese Americans for relocation in internment camps during World War II.

By law, no one—neither the census takers nor any other Census Bureau employee—is permitted to reveal identifiable information about any person, household, or business. Thus, when the United States entered World War II and the War Department wanted to relocate Japanese Americans living on

²⁰Attorney General, William D. Mitchell, to the Secretary of Commerce, 36 *Op. Atty. Gen.* 362, Sept. 29, 1930.

the West Coast in 1942, it could not obtain their names and addresses from the 1940 census.²¹

The Census Bureau has never denied that it responded to the War Department's request by providing 1940 census data on Japanese Americans for small geographic areas down to the census tract and block levels. A memorandum written in 1980 by the then Director, Vincent Barabba, states

A Government report, "Japanese Evacuation from the West Coast," notes that the 1940 census data were the single most important source of information used for evacuation and resettlement purposes, and the Bureau prepared a duplicate set of punch cards on Japanese Americans to assist in this effort. Shortly after the United States entered the war late in 1941, the war agencies asked that a Census Bureau statistician be assigned to assist them. Early in 1942 a statistician was transferred to California for this work. He requested and obtained a duplicate set of punch cards from which aggregate data were tabulated on an expedited basis. Virtually all of the data tabulated at this time was eventually published in the 1940 census reports issued in 1943. The 1940 census publications showed separate statistics on Japanese Americans by counties within States. Similar data on racial or ethnic origin had been published from the 1930 census, though in less detail. The extra set of punch cards permitted the 1940 aggregated information to be compiled on a more rapid basis than the normal schedule for published census reports.²²

Although the Census Bureau concedes that its staff provided census tabulations that were used in the internment effort, what is unclear is whether the then-in-effect legal prohibitions against revealing individual census records were violated. In response to this uncertainty, the recent Director, Kenneth Prewitt, made the following statement on March 24, 2000.

²¹ U.S. Bureau of the Census, *Factfinder for the Nation*, "History and Organization, May 1988, p.3.

²² Vincent P. Barrabba, unpublished memo, July 1980.

The historical record is clear that senior Census Bureau staff proactively cooperated with the internment, and that census tabulations were directly implicated in the denial of civil rights to citizens of the United States who happened also to be of Japanese ancestry.

The record is less clear whether the then in effect legal prohibitions against revealing individual data records were violated. On this question, the judicial principle of innocent until proven otherwise should be honored. However, even were it to be conclusively documented that no such violation did occur, this would not and could not excuse the abuse of human rights that resulted from the rapid provision of tabulations designed to identify where Japanese Americans lived and therefore to facilitate and accelerate the forced relocation and denial of civil rights.

I would also like to state clearly that for many years the Census Bureau was less than forthcoming in publicly acknowledging its role in the internment process. Silence was not the worst offense, for there is ample evidence that at various times the Census Bureau has described its role in such manner as to obfuscate its role in internment. Worst yet, some Census Bureau documents would lead the reader to believe that the Census Bureau behaved in a manner as to have actually protected the civil rights of Japanese Americans. This distortion of the historical record is being corrected.

The internment of Japanese Americans was a sad, shameful event in American history, for which President Clinton, on behalf of the entire federal government, has forthrightly apologized. The Census Bureau joins in that apology and acknowledges its role in the internment.

In the post-war period, important safeguards to protect against the misuse of census tabulations have been instituted, notably stronger legal provisions to protect data confidentiality and the Census Bureau's introduction of disclosure avoidance techniques. Adherence to these safeguards preclude a repeat of the 1941/42 behavior.

Over the past half-century, and especially following the Civil Rights Act of 1964 and the Voting Rights Act of 1965, the critical importance of summary data for enforcement of voting rights and civil rights in this age is an important contrast to the misuse of information in the early 1940s.²³

²³Kenneth Prewitt, e-mail to Bureau of the Census staff, March 24, 2000.

From the 1790 through the 1920 censuses, Congress enacted legislation for each decennial census and for the other censuses and surveys occurring in the intervening years. In 1929, Congress enacted the law which formed the basis for what was codified in 1954 as Title 13, United States Code, and the Census Bureau has operated under this law ever since. While title 13 assures confidentiality for all records in the Census Bureau's custody (except those for the census of governments, which are taken from public documents), it nowhere states how long that confidentiality shall last. Virtually all Census Bureau records, including those from the population and housing and the economic censuses, are required by law (Title 44, United States Code) to be sent to the U.S. National Archives and Records Administration. Once in National Archives custody, such records are subject to the provisions of title 44.

Pursuant to the requirements of title 44, the Census Bureau and the National Archives entered into an agreement in 1952 providing that microfilm copies of the population census records—those containing information about individuals—were released for public use after 72 years. As a result of the 1952 agreement, a person has access, for example, to as many pages of the 1930 Federal population census (and prior censuses) as he or she wants on microfilm furnished by the Archives under title 44. Records from the 1930 census were made available on April 1, 2002. On the other hand, existing paper or microfilmed records from much later economic censuses, which identify the responding firms, can be used at the Archives without any restriction after 30 years.²⁴

²⁴By law (Title 44, United States Code, Section 2107) most other agency records more than 30 years old and not in current use must be transferred to the Archives, where they are made available for public use as soon as practicable unless some further restriction is placed on them.

Confidentiality also became an issue in the economic censuses in 1958. Not long after the 1958 census, the Federal Trade Commission (FTC) asked the St. Regis Paper Company for its file copies of the census of manufactures reports it had completed in recent years. The company refused, maintaining that the file copies came under the same guarantee of confidentiality against “purposes of taxation, investigation, or regulation” as the originals. At issue was the fact that the Census Bureau had routinely given respondents file copies of the economic censuses and surveys taken by mail for many years, a practice that made it much easier to discuss and reconcile reported figures if questions arose. Because of the timing or the nature of the specific inquiry, the returns often contained estimates that would not appear in the respondents’ accounting records, but were of interest to the FTC. St. Regis Paper Company apparently did not want to release these estimates.

The FTC pursued the issue through the courts until in December 1961 the U.S. Supreme Court ruled that the census law did not protect the copies that respondents had retained.²⁵ The implications for the economic censuses and surveys were serious. It appeared that respondents generally would be reluctant to furnish information, especially in the form of estimates or approximations. As a result of this experience, Congress amended title 13 specifically to extend census confidentiality to file copies.

The Census Bureau’s firm commitment to maintaining the confidentiality of completed personal census forms was dramatically conveyed to census employees in the

²⁵368 U.S. 208, 1961.

Colorado Springs, CO, District Office and the Denver Regional Census Center (RCC) on Wednesday, August 13, 1980. Late that afternoon, four Federal Bureau of Investigation (FBI) agents arrived at the District Office in Colorado Springs armed with a search warrant authorizing them to seize census documents, including completed questionnaires, in the course of their investigation of a case involving alleged questionnaire falsification and payroll fraud.

A census employee recognized the seriousness of this challenge to census confidentiality and alerted her superiors at the Denver RCC. She also tried to persuade the FBI agents that their warrant did not supersede the section of the U.S. Code that prohibits disclosure of confidential information to those who are not sworn census employees. About 90 minutes later, a census supervisor arrived in Colorado Springs, assessed the situation, and reported back to her boss at the RCC, who contacted Census Bureau headquarters in Suitland, MD.

The Census Bureau's Director, Vincent Barabba, recognized the potentially disastrous situation into which both the FBI and the Census Bureau had stumbled and immediately began tracking down the Director of the FBI, Judge William Webster. Barabba finally reached him at a Washington, DC, restaurant, where Webster was eating dinner. After a brief flurry of telephone calls to their subordinates in Colorado, the two men agreed that a mutually satisfactory conclusion could be reached while the disputed questionnaires remained in the custody of the Census Bureau. Barabba dispatched a senior official to Colorado Springs to negotiate the details of the agreement.

Ultimately, the documents were placed in a secure room protected by two locks, with one key held by the FBI and the other by a local census official. Under this arrangement, only sworn census employees were allowed to enter the room, but an FBI agent had to be present when the door was opened. While the door was unlocked, an agent was stationed outside the room to monitor the activities of the census personnel. The Census Bureau brought in experienced enumerators from outside the Denver area to reinterview the respondents in the area where the alleged fraud had taken place and compare the original questionnaires with those from the recanvass. Census Bureau officials prepared a report that described all significant discrepancies uncovered but did not reveal any confidential information.²⁶

Under title 13, even the address lists used in the census are confidential. In the litigation that followed the 1980 census, the City of New York obtained a district court order requiring the Census Bureau to turn over its address registers (not the household registers) for the city so that local officials might compare the listings with their own records. The agency's director at the time, Vincent Barabba, refused. In 1982, the case (*Baldrige et al. V. Shapiro*) ultimately reached the U.S. Supreme Court, which decided in the Census Bureau's favor: Address lists could not be disclosed, either through civil discovery or under the Freedom of Information Act.²⁷

²⁶Associated Press 1980; Barabba 1980; Clemence 1986; Colorado Springs, *Gazette Telegraph*, 1980

²⁷455 U.S. 208, 82 S. Ct. 289.

DISCLOSURE LIMITATION AT THE CENSUS BUREAU

Disclosure avoidance, disclosure analysis, or disclosure limitation refers to the statistical methods the Census Bureau uses to hinder anyone from identifying an individual respondent or establishment by analyzing published census or survey data, especially, by manipulating the arithmetical relationships among the data. At the same time, the agency has the responsibility of releasing data for the purpose of statistical analysis. The desire then is to release as much statistically valid and useful data as possible without violating the confidentiality of the data as required by title 13. Disclosure limitation techniques are applied to the data prior to their release in an effort to protect confidentiality.²⁸

The three most commonly used forms of data release from the Census Bureau and other statistical agencies are microdata, frequency count data, and magnitude data.²⁹

A microdata file consists of records at the individual respondent level, be it a person or establishment. Each record consists of characteristics associated with that respondent, in which no specific identifiers are disclosed. Typical variables for a demographic microdata file include age, occupation, and income of a responding individual. Variables for an economic microdata file might include employment size and value of shipments of an establishment. Because economic data are highly skewed and establishments often easily identified by just a few characteristics, the disclosure risk of

²⁸Laura Zayatz, Paul Massell, and Phil Steel, "Disclosure Limitation Practices and Research at the U.S. Census Bureau," U.S. Bureau of the Census, 1999, p. 1.

²⁹*Ibid.*, p. 1.

economic microdata files can be quite high. For this reason, the Census Bureau releases very few economic microdata files, and those that are released have extremely little detail.

To assess disclosure risk, microdata files must be reviewed by disclosure experts at the Census Bureau. If it is determined that the potential for disclosure exists, the review may result in the—

- ! Removal or reduction in detail of any variable considered likely to identify an especially small and visible population such as persons with high incomes.
- ! Introduction of “noise” (small amounts of variation) into selected data items.
- ! Use of data swapping (i.e., locating pairs of matching households in the database and swapping those households across geographic areas to add uncertainty for households with unique characteristics.
- ! Replacement of a reported value by an average in which the average associated with a particular group may be assigned to all members of a group, or to the “middle” member (as in a moving average).

Frequency count data are summary or tabular data. Tables of frequency count data present the number of units of analysis (persons, households, establishments) in a table cell (an intersection of a table row and column). An example of this is a table where the columns represent categories of race and the rows represent categories of sex, and the table cells show the counts of the number of people having these characteristics. The Census Bureau does not consider frequency count data for establishments to be sensitive because general information about an establishment, particularly classifications that would

be used in frequency count tables, is publicly available. But when the tables of counts are based directly on decennial census data, disclosure limitation procedures must be applied.

Magnitude data also are summary or tabular data, but instead of counts magnitude data present the aggregate of a “quantity of interest” over all units of analysis in the cell. The quantity of interest must measure something other than membership in the cell. For example, tables presenting the total value of shipments within the manufacturing sector by Standard Industrial Classification group by county-within-state are tables of magnitude data. Magnitude data are generally nonnegative quantities reported in Census Bureau surveys or censuses of business establishments.

As stated previously, the distribution of these reported values is likely to be skewed, with a few entities having very large values. Disclosure limitation in this case concentrates on making sure that the published data cannot be used to obtain an individual establishment’s response, based on a knowledge of its operational characteristics.

It is less likely that sampling alone will provide magnitude data disclosure protection because most sample designs for economic surveys include a stratum of the larger volume entities, which are selected with certainty. Thus the units that are most visible because of their size receive no protection from sampling. Maintaining the confidentiality of these units is accomplished by employing one or more of the following techniques.³⁰

³⁰Ibid., pp. 1-2.

Data suppression consists of both cell and table suppression, the latter of which is rarely utilized by the Census Bureau. In cell suppression, a data item that could lead to disclosure may be suppressed, i.e., the figure is omitted and replaced by an asterisk or another symbol that indicates that the figure is being omitted to maintain confidentiality for the subjects of the table. Care must be taken, however, to ensure that the disclosing figure then may not be deduced by arithmetic means. If it can, another figure in the same row and another in the same column must also be suppressed, assuming it is desired that no changes be made in the row and column totals. It is necessary to suppress at least four figures in a two-way distribution to avoid a disclosure.³¹

Data confidentiality can frequently be accomplished by changing the structure of tables in such a way that the disclosure possibility is eliminated. Thus, rows or columns can be combined into larger intervals or new groupings of characteristics. This may be a simpler solution than the suppression of individual items, but it tends to reduce the descriptive and analytical value of the table. An indirect, but common example of rolling-up exists in databases in which the Standard Industrial Classification system is used. That hierarchical system has 2-, 3-, and 4-digit levels providing successively greater detail. When data are suppressed at the 4-digit level, the 3-digit level summary provides the benefits of intermediate rolling-up.

Other examples of rolling up data include topcoding, bottom-coding, and recoding the data into intervals. For an income data table, large income values could be topcoded

³¹Federal Committee on Statistical Methodology (1994), *Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology*, Washington, DC: U.S. Office of Management and Budget, p. 12.

to show incomes that are greater than \$100,000 per year as \$100,000 per year. Small income values could be bottom-coded to show incomes that are less than \$40,000 per year as \$40,000 per year, and the categories in between could be recoded into \$10,000 intervals.

This process involves changing the figures of a tabulation in some systematic fashion, with the result that the figures are insufficiently exact to disclose information about individual cases but are not distorted enough to impair the informative value of the table.

Ordinarily, rounding is the simplest example. For example, figures in a table could be rounded to the nearest multiple of five. Where the figures are very large, this will have little or no effect on the informative value of the tables. If all cells in a table are rounded by the same rules, totals will not always agree with the sums of the detailed cells. If this is considered undesirable, the most detailed cells can be rounded and then added to obtain totals at various levels. Other examples of data perturbation include—

- ! Adding random noise. One method of disguising high visibility variables, such as income, is to add or multiply the data by random numbers.
- ! Swapping. This technique involves selecting a sample of the records, finding a match in the database on a set of predetermined variables and swapping all other variables.

! Blanking and imputing. The blank and impute method involves selecting a few records from the microdata file, blanking out selected variables and replacing them with imputed values.³²

As early as the 1920s the Census Bureau prevented disclosure in the economic area by presenting the data in broad or grouped (“collapsed”) categories, by withholding other figures, or by suppressing certain cells (deleting and so marking the table entries). In general, the process was one where subject-matter specialists “eyeballed” the tables prior to publication and manually censored any suspicious number.³³

Disclosure was no particular problem in the 1930 population census publications; the addition of the words “or individuals” to the 1930 census act applied most immediately to the issue of releasing personal data to third parties, as discussed earlier. There were no published tabulations for 1930 below the level of census tracts, wards, or similar areas of 5,000 to 12,000 people, and for these, cross-classification was limited.

The Census Bureau tabulated—but did not publish—housing data, including the occupants’ race and number of persons per room, by census block for 1940 for 191 cities; the Works Progress Administration prepared analytic maps from these data in 1942-44. Beginning with the 1940 census the agency started withholding summary data from its demographic census reports when the risk of disclosure appeared high.

³²Ibid., pp. 20-24.

³³Lawrence H. Cox, Sarah-Kathryn McDonald, and Dawn Nelson, “Confidentiality Issues at the United States Bureau of the Census,” *Journal of Official Statistics*, Vol. 2, No. 2, 1986, pp. 135-160; U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, Statistical Working Paper 2, *Report on Statistical Disclosure and Disclosure-Avoidance Techniques*, Government Printing Office, Washington, DC, 1978.

By 1960, the Census Bureau had developed statistical routines for avoiding indirect disclosure. In such cases, simple counts of population or housing units for the small area in question were published, but information about individuals or housing were suppressed.³⁴

The Census Bureau began using an electronic computer (UNIVAC I) in 1951, which dramatically increased both users' and the agency's ability to cross-tabulate data. Further, it became possible to build disclosure analysis into the tabulation process by prescribing, for example, the frequency with which particular characteristics would have to appear in order to be published. By establishing a so-called "threshold rule," a known person's income could not be discovered by looking at county-level income data cross-classified by sex and occupation.

Disclosure analysis requirements increased when the agency began issuing public-use microdata files. This occurred first with punchcards containing selected statistics from the 1950 Censuses of Population and Housing, and then, for the 1960 and later censuses, with either summary tape files or samples of the raw data with personal and geographic identification removed. In all of these areas, the Census Bureau's disclosure analysis has become increasingly sophisticated. It has been extended to data furnished by or to other Federal agencies.

For the 1990 census, the 100-percent data were published in the form of tables, usually at the block level. The average block contained 36 people. Some of the more

³⁴Sherry Courtland, "Census Confidentiality: Then and Now," *Government Information Quarterly*, Vol 2, No. 4, 1985, pp. 407-418; Paul T. Zeisset, "Making Decennial Census Data Available," *ibid.*, pp. 1-23.

detailed tables were published at the block group level, which contained 400 people. The procedure used to protect the short form (100%) data was the confidentiality edit.

For the above edit procedure, a small sample of census households from the internal census data files was selected. The data from these households were swapped with data from other households that had identical characteristics on a certain set of key variables but were from different geographic locations. Which households were swapped was not public information. The key variables were number of people in the household of each race by Hispanic/NonHispanic by age group (under18, 18 and above), number of units in the building, rent/value, and tenure (own or rent).

All tables were produced from this altered file. Thus, census counts for total number of people, totals by race by Hispanic/NonHispanic by age 18 and above³⁵ as well as housing counts by tenure were not affected. A greater percentage of records was swapped for small blocks because those records possessed a greater disclosure risk. All data from the chosen households were swapped except for Indian tribe. It was felt that it did not make sense to move a member of one tribe into a location inhabited by another tribe.³⁶

The 1990 census sample data also were published in the form of tables. Some of the tables were published at the block group level—about 400 people. The fact that it was a sample provided protection for all areas for which sample data were published except

³⁵Public Law 94-171 counts, also known as redistricting counts.

³⁶Laura Zayatz, Philip Steel, and Sandra Rowland, *Disclosure Limitation for Census 2000*, U.S. Bureau of the Census, 1999, pp. 1-2.

for small block groups. In small block groups, some values from one housing unit's record on the internal file were blanked and imputed using the 1990 census imputation methodology. This altered file was used to create all tables. Which values were altered was not public information.

The 1990 census microdata file contained records from 5 percent of all households in the nation. The microdata were created from the internal file after the blanking and imputation described earlier, so some protection was provided by that procedure. All identified geographic areas (PUMAs--public use microdata areas) contained at least 100,000 people. Income values and some other continuous values such as age and rent were topcoded. Some very detailed categories from items such as ethnicity and Indian tribe were collapsed into broader categories. And, of course, all identifying information such as name and address was stripped from the file.³⁷

The 1990 census permitted single-race identification only. That is, respondents were to check one box only on the questionnaire in response to the race question. The 2000 census asked respondents to check the applicable boxes, giving a total of 63 possible answers to the race question. This led to changes in disclosure risk as well as processing procedures because of the additional detail in the tables.

The American FactFinder (AFF) (accessed through the Census Bureau's website, www.census.gov) was developed to allow for broader and easier access to the data.

There are three ways in which users can obtain data from the AFF—

³⁷From the 1990 census, the Census Bureau published a 5-percent state file, a 1-percent metropolitan area file, and a 3-percent elderly file. All three files were mutually exclusive. Laura Zayatz and Paul Mackun, *Disclosure Limitation for the Census 2000 Public Use Microdata Files*, U.S. Bureau of the Census, April 2001, section 4.

- ! Through Tiers 1 and 2, users can obtain predefined tables free from disclosure problems.
- ! Through Tier 3, or the Advanced Query System (AQS), users are allowed to create their own data products.

The goal is to allow users to submit requests for tabular data electronically. A request to the AQS passes through a firewall to an internal Census Bureau server with a previously swapped, recoded, and topcoded microdata file. The table is created and electronically reviewed for disclosure problems. If it is judged to have no disclosure problem, the table is sent back electronically. This does not affect the disclosure limitation procedure for the public use microdata. Those disclosure limitation techniques will have been applied to these data before they are made available on the AFF or any other Census Bureau website. However, this is a new way of publishing tabular data, so the agency needed to develop new disclosure limitation practices for the AFF.³⁸

For 100 percent Census 2000 data, as in 1990, the agency swapped a set of selected records. Unlike 1990, the selection process in Census 2000 swaps records with the most risk of disclosure, and census blocks with high imputation rates are avoided. Only records which are unique in their block, based on key variables, are swapped. The probability of a unique record being swapped has an inverse relationship to block size. In addition, records representing households containing members of a race category which appears in no other household in that block will have a higher probability of selection. All

³⁸Op cit., *Disclosure Limitation for Census 2000*, p. 3.

data products are created from the swapped file. The Census Bureau tested and evaluated various parameters using data from the 1995 and 1996 census tests and the 1998 dress rehearsal.

For Census 2000 sample data, the agency uses swapping (rather than blanking and imputation) to protect the data. This increases the amount of distortion (giving the data more protection). Swapping has the additional quality of removing any 100-percent assurance that a given record belongs to a given household. It is consistent with the 100 percent procedure. And, it retains relationships among the variables for each household.

For these sample data, the Census Bureau uses two different sets of key variables - one to identify the unique sets of data, the other to find the swapping partners. The Census Bureau also holds some variables fixed (unswapped). For example, travel time to work and place of work for a household may not make sense if swapped with a household geographically far away.

The procedure for producing the masked file then is very similar to the procedure for the 100 percent data. Block groups replace blocks because the block group is the lowest level of geography for publishing sample data. The threshold value for not swapping in block groups with a high imputation rate differs, and the probability of a unique record being swapped has an inverse relationship with block group size; it also has a direct relationship with block group sampling rate. The lower the sampling rate, the more likely that the sample unique is not unique in the entire block group population. So a smaller sampling rate should lead to a lower chance of being swapped.

The disclosure limitation procedures for the Census 2000 microdata files are similar to the procedures used for the 1990 data. Because of additional concerns this decade about advances in technology and the abundance of databases in the private sector, the disclosure limitation techniques are a bit more conservative. For example, some variables such as different income types and travel time to work are rounded. The microdata are created from the internal file after the swapping described earlier. All PUMAs contain at least 100,000 people.³⁹

Income values and some other continuous values such as age and rent are topcoded. Topcodes for variables that apply to the total universe include at least ½ percent of all cases. Topcodes for variables that apply to subpopulations include either 3 percent of the appropriate cases or ½ of 1 percent of all cases, whichever is the higher topcode. Some very detailed categories from items such as ethnicity and Indian tribe are collapsed into broader categories. And, of course, all identifying information such as name and address are stripped from the file.

The American FactFinder does not provide an open-ended or unconstrained opportunity to construct any or all possible tabulations from the full microdata files. As stated previously, a query for a table through the AFF's Tier 3 passes through a firewall to an internal Census Bureau server with a previously swapped, recoded, and topcoded

³⁹The Census Bureau has decided on two (mutually exclusive) files. There will be a 5-percent state file where PUMAs must contain at least 100,000 persons and follow state boundaries. This file will contain less detail for several variables than in 1990. There will be a 1-percent characteristics file with "super-PUMAs" which must contain at least 400,000 persons and follow state boundaries. This file will contain approximately the level of detail for most variables as in 1990. Op. cit., Zayatz and Mackun, section 4.

microdata file. All tables generated from the sample data are weighted. The query and the resulting table must each pass through a filter.

If a user requests a tabulation for more than one geographic area or for a combination of areas, each area must individually pass the query filter.

The external user is advised in the user interface that the blockgroup is the lowest level of geography permitted for 100 percent data and the tract is the lowest level of geography permitted for sample data for an external user. Requests for split blockgroups or split tracts are not permitted. A minimum population requirement is also imposed for each area. The user interface permits no more than three dimensions (page column, and row) and one universe not including geography.

The query filter also delimits the use of variables such as race, Hispanic origin, group quarters, cost of electricity, gas, water, fuel, property taxes, property insurance cost, mortgage payments, condo fees/mobile home costs, gross rent, selected monthly owner cost, household/family income and individual income types. External users may obtain only predefined categories or recoded values of these variables. Most variables have several sets of recodes that the user can choose from. So if the user is requesting a table from a large geographic area, he can choose a very detailed list of recodes. If a user is requesting a table from a small geographic area, he or she can choose a short list of recodes to try to ensure that the table will pass the results filter.

If the query passes the query filter rules, the query is sent from the external server outside the firewall to the internal server inside the firewall to the full microdata files. The

full microdata files contain all of the predefined categories for race, Hispanic origin, group quarters, etc.

If the query passes the query filter rules, the query is sent from the external server outside the firewall to the internal server inside the firewall to the full microdata files. The full microdata files contain all of the predefined categories for race, Hispanic origin, group quarters and modified sensitive variables.⁴⁰

Each resulting tabulation selected from the full microdata files obtained through AFF must meet certain criteria or it will not provide the user with the tabulation. If a user requests a tabulation for more than one area or for a combination of areas, each area must individually pass the results filter. The filter is designed to prevent the release of sparse tabulations which can lead to disclosure. If a tabulation does not meet the criteria, the user will receive a message stating that the tabulation cannot be released for confidentiality reasons. The rules, their parameters and population threshold values were tested in 1999 and finalized for Census 2000. The U.S. Census Bureau's disclosure limitation rules are designed to prevent the release of sparse tables. They do not guarantee that there will be no cell values of one.⁴¹

The Disclosure Review Board

Before data are publicly released, they have to be approved by the Census Bureau's 9-member Disclosure Review Board (DRB). The DRB consists of six members representing the economic, demographic, and decennial program areas that serve 6-year

⁴⁰Op cit., *Disclosure Limitation for Census 2000*, pp. 5-6.

⁴¹Ibid., p. 6.

terms. In addition, the Board has three permanent members representing the research and policy areas. DRB members review proposed data releases against a check list. It asks for information about file contents and disclosure limitation procedures applied to the file.

After reviewing a request, the DRB may approve or deny the request. If the request is denied, the Board will state the reasons for denial and may offer suggestions on changes that could be made in order to obtain approval. If Census Bureau staff members are not satisfied with the DRB's decision, they may appeal to a steering committee consisting of several Census Bureau Associate Directors. Thus far, there have been few appeals, and the Steering Committee has never reversed a decision made by the Board.⁴²

RESTRICTED ACCESS TO CONFIDENTIAL DATA

The Census Bureau's Center for Economic Studies (CES) was established in 1982 to provide restricted access to the agency's economic microdata for the manufacturing sector. The CES provides an opportunity for unique research that generally advances science and informs policy makers while providing basic measurement research and feedback that can advance Census Bureau data programs.⁴³

Researchers in economic fields can submit proposals to the CES. The proposal must specify the research they want to conduct, the economic microdata they will need,

⁴²Op cit., *Disclosure Limitation Practices and Research at the U.S. Bureau of the Census*, 1999, p. 2.

⁴³"Introduction to the RDC Program," U.S. Bureau of the Census, Center for Economic Studies website, <http://www.ces.census.gov/ces.php/rdc>.

and the nature of the results they expect to publish. If the proposed research is deemed beneficial to the Census Bureau's programs, it may be approved by the agency. If their proposal is accepted the researchers become eligible for Special Sworn Status to work with data that cannot be publicly released due to confidentiality concerns. Researchers with Special Sworn Status are bound by law to maintain confidentiality, just like any Census Bureau employee.

To obtain Special Sworn Status, applicants must take the title 13 oath of non-disclosure and receive training in title 13 requests. If the Special Sworn Status employee plans to access Federal tax returns and return information (FTI), the employee must complete title 26 training to ensure the protection of these data.

It is often too expensive for researchers to relocate to the CES, located at the agency's headquarters in Suitland, MD. To remedy this situation, the Census Bureau has established seven regional Research Data Centers (RDCs).⁴⁴ These include—

- ! Boston Research Data Center, which opened in 1994
- ! Carnegie Mellon Census Research Data Center, which opened in 1997
- ! California Census Research Data Center, which opened in 1999
- ! Triangle Census Research Data Center, which opened in 2000
- ! Chicago Census Research Data Center, which opened in 2002
- ! Michigan Census Research Data Center, which opened in 2002

⁴⁴Ibid., Currently Operating Research Data Centers.

Choosing and Setting up New RDCs

The RDCs are joint ventures involving the Census Bureau and the National Science Foundation (NSF), and the sponsoring institutional organizations. The Census Bureau and the NSF sought to place the RDCs in areas that promised the best combination of high quality research and benefits to the Census Bureau's mission. The two agencies select host sites based on proposals submitted to them. The NSF provides some funding for the RDCs.⁴⁵ Successful RDC partners demonstrate

! An understanding of the constraints placed on a restricted access site, including the need to

- Work with the Census Bureau to protect the confidentiality of the underlying microdata
- Ensure that researchers use the data only for statistical purposes
- Maintain the Census Bureau's longstanding reputation for providing objective information that is not influenced by political considerations⁴⁶

In choosing RDC partners, the Census Bureau emphasizes the need to protect confidentiality of its data covered by Title 13, United States Code and FTI, obtained under Title 26, United States Code. An RDC prospectus explains to potential partners the

⁴⁵These costs are covered by funds raised by the local research community. During the first 3 years of operation, the NSF provides up to \$100,000 per year as "seed money" to partially cover these costs. The RDC's recover these costs through laboratory fees on research projects.

⁴⁶Arnold P. Reznick, *Increasing Access to Longitudinal Business Survey Microdata: The Census Bureau's Research Data Center Program*, Census Bureau, May 2000, pp. 4-5.

application process and how the Census Bureau expects the RDCs to operate. The prospectus places heavy emphasis on maintaining security and protecting confidentiality.⁴⁷

Securing the office

Setting up and maintaining a secure office and computing environment for an RDC has three aspects—physical (office) security, computer security, and data security. For physical and computer security purposes, the Census Bureau’s partner in the RDC endeavor is considered a contractor, and the RDC office is a contractor site.⁴⁸ The main security requirements for these sites are

- ! Before the RDC facility may open, a security plan must be developed and approved (and is updated periodically), according to specific Census Bureau procedures. The RDC office is subject to unannounced inspection by the Census Bureau’s ADP Security Branch.
- ! The RDC office must be in a secure (locked) room that has a security system that meets Census Bureau specifications. Researchers enter the facility through use of a combination of identification cards, key codes, combination (cypher) locks, and door keys; the exact setup varies across RDC offices.
- ! The RDC computer network must be isolated from any other networks, including the Internet. It is not possible to access confidential data from outside the RDC (e.g., by remote terminal access).

⁴⁷Arnold P. Reznick and Alfred R. Nucci, *Protecting Confidential Data at Restricted Access Sites: Census Bureau Research Data Centers*, *Of Significance: A Topical Journal of the Association of Public Data Users*, Vol. 2 Number 1, 2000, p. 44.

⁴⁸*Ibid.*, pp. 44-45.

- ! Secured storage containers (e.g., safes or lockable file cabinets) are required to store confidential data. The RDC keeps all confidential data, including computer backups, physically secure.
- ! Only Census Bureau employees or persons having Special Sworn Status are allowed into the RDC facility.
- ! Researchers are accountable for their computer use, through the use of passwords and system logs.
- ! Researchers are not allowed to bring into the RDC facility laptop computers, zip drives or other portable mass storage devices, or devices with wireless modems, such as some personal digital assistants.
- ! The RDC computers are set up to prevent copying of data to removable storage media.
- ! Approved procedures exist for transferring sensitive information between the RDC and Census Bureau headquarters or other secure sites.
- ! Approved procedures (shredders, burn bags) are required for disposing of sensitive information that is no longer needed. Procedures also must be in place for clearing magnetic media that have held sensitive information.

Security functions of personnel at the RDC

To ensure security, the agency stations at least one Census Bureau employee who is trained in confidentiality and security policies and procedures at each RDC. The person called the RDC administrator ensures that researchers learn the Census Bureau's "culture

of confidentiality,” provides guidance to the researchers regarding confidentiality, and carries out disclosure procedures on output researchers wish to remove from the RDCs.

Protecting Confidentiality for the Life Cycle of Research Projects

Protecting confidentiality requires continual effort during all phases of research projects. The following sections describe how the agency accomplishes this.

Initial inquiry

Just as they must for projects at the CES, Researchers must submit proposals to carry out projects at RDCs. These proposals must follow a set of guidelines. The Census Bureau has found that it is desirable for prospective researchers to work informally with RDC personnel while writing their proposals, particularly in assessing disclosure risks and in deciding on the types of output that can be removed from the RDC. As researchers explore the possibility of carrying out research at RDCs, the agency informs them of the requirement to protect confidentiality, and indicates the types of projects are likely to have significant disclosure risks. The Census Bureau emphasizes that RDCs are reserved for researchers doing analytical projects that primarily involve statistical modeling with potential benefit for agency programs. The RDCs are not designed to supplement the Census Bureau’s existing data program operations by producing large scale special tabulations from microdata.

In summary, the Census Bureau has learned to ask that researchers specify as much as possible the research samples they wish to use and the kinds of research outputs they wish to remove. If their projects are approved, the agency holds the researchers to these outputs, and makes clear that it cannot guarantee the release of any particular

output. The Census Bureau recognizes that this policy restricts the flexibility of inquiry, but researchers are not limited in the analyses they can make within the RDC. (One exception—the agency does not allow casual “browsing” of the data.) At the very least, requiring researchers to specify the output they wish to produce for release calls their attention to the problem and forces them to focus their projects.⁴⁹

Project selection

The Census Bureau currently chooses new research projects every 2 months according to a proposal review cycle. Several general factors are used in evaluating proposals:

- ! Whether the project requires access to confidential Census Bureau data.
- ! The potential to benefit Census Bureau data programs—for example, by adding to or improving the databases at the Census Bureau, or by suggesting improvements to the Census Bureau’s data programs. This condition is required by law if the project is to be approved.
- ! Scientific integrity (endorsement by scientific granting agencies is an indicator to the agency that the project is sound).
- ! Feasibility (whether the project can be done with the available data in the proposed amount of time for a reasonable cost).
- ! Disclosure risk.

⁴⁹Ibid., p. 45.

In evaluating disclosure risks, it is important to be aware of complementary disclosure risks across projects. Tracking this is always difficult, but it is made more difficult across the RDCs. The agency now handles this largely at headquarters, which reviews every project proposal, but it has learned that the communication across the RDCs is important too, since headquarters is not always involved in consultation with researchers as they specify their projects.

Risks of disclosure remain even after the Census Bureau has informed researchers about the types of projects that present significant disclosure risks, and if the researchers have worked with the agency to develop their project proposals. Researchers, especially those with little or no experience handling confidential data, often do not completely understand the steps needed to minimize disclosure risks. Moreover, in many cases researchers find it difficult to specify expected research output before beginning their projects⁵⁰this is true especially for researchers who are proposing to use data sets they have not used previously. For these reasons, the agency receives a number of proposals that pose significant disclosure risks or unclear risks, and it often must ask researchers to revise and resubmit or, short of that, to provide more information about their proposed research samples or research output.⁵⁰

Evaluating Confidentiality Safeguards at the RDCs

RDCs pose unique problems because many of the underlying data sets and the research output at the RDCs differs from traditional Census Bureau data products. The

⁵⁰Ibid., pp. 45-46.

more they differ, the less applicable are the Census Bureau's traditional disclosure criteria. In these situations, the agency has adopted very conservative criteria for release of research output to assure against the inadvertent release of identifiable data.⁵¹

The Census Bureau has learned that protecting confidentiality requires constant effort at all stages of choosing, establishing, and operating the RDCs, and at all stages of the life cycle of the research projects carried out at the RDCs. The process is perhaps best described as one of learning by doing.

PRIVACY AT THE CENSUS BUREAU: 1974 - 2002 AND BEYOND

Within the Federal statistical community, confidentiality and privacy are inextricably linked by legislated requirements to protect statistical data. Confidentiality is largely a data dissemination issue that involves adequate protection of data from unauthorized use. Privacy is largely a collection issue that is a subjective sense of what respondents want to reveal and the limits of intrusion.⁵² Data protection refers to policies and procedures that ensure data confidentiality and minimal intrusion.⁵³ Both confidentiality and privacy at the Census Bureau are deserving of thorough treatment but, for the sake of clarity, this paper has focused on the evolution of confidentiality practices at the Census Bureau while making only tangential reference to privacy. However, to ignore the privacy issues

⁵¹Ibid., p. 48.

⁵² National Academy of Sciences, *Private Lives and Public Policies*, National Academy Press, Washington, D.C., 1993, p. 22.

⁵³ Ibid, p. 23

affecting the Bureau and the responses to them would leave an impression that privacy is not very important, but the reality is quite the contrary. Thus a brief overview of the Bureau's privacy program follows.

The Federal government's privacy principles are encoded in the Privacy Act of 1974.⁵⁴ Signed by President Gerald Ford on January 1, 1975, the Act went into effect on September 27, 1975.⁵⁵

The Act primarily is concerned with the confidentiality and accuracy of personal information and seeks to restrict both the amount of personal information collected by federal agencies and the amount that is transferred or disseminated to other agencies and third parties. It gives individuals the right to see records about themselves held in systems of records, to obtain copies of their records, to have their records corrected or amended with agency approval, and to have a statement of disagreement filed in their record if the agency doesn't allow the correction or amendment. Record systems are a group of records from which information is retrieved by name, a number, such as Social Security number, or other personal identifier. The grouping normally has some administrative or statistical purpose.⁵⁶

⁵⁴ 5 U.S.C. , Section 552a (amended 1997, 5 U.S.C.A. Section 552a).

⁵⁵ General Services Administration, *Public Papers of the Presidents, Gerald R. Ford, 1975*, Book 1, Government Printing Office, Washington, D.C., 1977, p. 1; *Federal Register*, October 2, 1975, page 45606.

⁵⁶The Privacy Act defines a statistical record to be “. . . a record in a system of records maintained for statistical research or reporting purposes only and not used in whole or in part in making any determination about an identifiable individual . . .” Because statistical records are not used to make decisions about individuals, individuals do not have a right to have them changed.

All but one of the Bureau's eight record systems contain individually identifiable information collected from the public solely for statistical purposes.⁵⁷ Examples are the "Individual and Household Statistical Survey Records and Special Studies Records" and the "Population and Housing Records of the 2000 Census" systems of records. These records are from personal interviews with and questionnaires completed by individuals designated for statistical sample surveys and the decennial census.

The Census Bureau also obtains administrative records from program agencies such as the Social Security Administration. Administrative records are records collected and/or maintained by government agencies for the purpose of administering programs, such as determining Social Security benefits or housing subsidies. They may contain information on social, economic, and other characteristics of individuals. These records may be used by the Census Bureau to supplement and evaluate information collected directly from the public for such statistical methodological purposes as performing quality control studies of Census Bureau collected data and improving population estimates. Depending on the specific uses made of them, they may enter a Census Bureau systems of records, such as the "Statistical Administrative Records System." All administrative records are protected by the Privacy Act because they are from systems of records at supplying agencies.

The Census Bureau is in full compliance with the Privacy Act, but compliance alone does not allay all public concern about government intrusion. This unease may

⁵⁷ The one exception is a system of records pertaining to Census Bureau employee productivity measurement records.

compromise the Census Bureau's best efforts to collect data because individuals who believe their privacy is at risk may be less likely than those without such concerns to respond to census questions and surveys. In fact, the decline in the 1990 decennial census response rate has been attributed partly to privacy issues.⁵⁸

As a result of the decline in 1990 decennial census response rates compared to 1980, the Census Bureau initiated research to understand census-related privacy concerns better. Beginning in 1992, the Census Bureau sponsored qualitative and quantitative research to improve its understanding of public reaction to its surveys and the census. From 1995 through April 2000, just after Census 2000, the Bureau sponsored four public opinion surveys concerning privacy and confidentiality issues. Supplementing this quantitative research were Bureau-sponsored conferences with privacy experts and case studies of privacy practices followed by federal statistical agencies and private organizations to identify best practices.

The survey research consisted of three major comparisons: trends in privacy attitudes; the effect of the census information environment on beliefs, attitudes, and privacy concerns; and the relationship between privacy attitudes and response behavior. Results related to trends in privacy concerns showed small, yet statistically significant, increases between 1995 and 2000 in the percentage who were very worried about their personal

⁵⁸There has been much research investigating the relationship between privacy concerns and response rates. For a review, see Thomas S. Mayer "Privacy and Confidentiality Research and the U.S. Census Bureau, Recommendations Based on a Review of the Literature", U.S. Census Bureau, February 7, 2002. Specifically in regard to the 1990 census, see Fay, R.E., Bates, N., and Moore, J. "Lower Mail Response in the 1990 Census: A Preliminary Interpretation", paper presented at the 1991 Annual Research Conference of the U.S. Census Bureau, Arlington, VA, March 17-20, 1991.

privacy and the loss of control over personal information. Census 2000 publicity enhanced the public's knowledge and endorsement to cooperate with the census. For Census 2000, three attitudes were related to lower census returns (although the overall return rate exceeded that for 1990): concerns about privacy, concerns about census misuse, and support of data sharing in order to eliminate the census. Full results of this research are available.⁵⁹

A series of five cross-sectional opinion tracking surveys also were conducted between March and the middle of April, 2000. These surveys tracked opinions about the census as an invasion of privacy, whether the Census Bureau's promise of confidentiality could be trusted, whether answers could be used against people, and whether the law requires the census form to be filled out. The data indicated that in some ways privacy reactions to Census 2000 differed from the 1990 census. Census 2000 evidently produced more sensitivity and a more diffuse privacy reaction. For example, during Census 2000 people increasingly came to believe that their answers could be used against them. This reaction did not occur during the 1990 census.⁶⁰ All of the research described above will help guide planning for the 2010 census by identifying privacy and confidentiality concerns that may influence census participation.

During the most recent census, privacy concerns were prominently and forcefully expressed about the Census 2000 "long form". Delivered to about 1-in-6 American

⁵⁹U.S. Census Bureau, *Consolidated Report for: The Social Security Number, Privacy Attitudes, and Notification (SPAN) Experiment*, November 7, 2002.

⁶⁰ Martin, Elizabeth, "Changes in Public Opinion During the Census", paper presented at the Census Advisory Committee of Professional Associations, October 19, 2000, p. 31.

households, the long form asked 52 questions about 33 topics compared to only six on the “short form”, which was delivered to the remainder of the households. Many respondents and members of Congress complained about the perceived intrusiveness of the schedule’s length and respondents’ associated right to privacy.⁶¹ The News Hour on PBS television even devoted an entire segment to privacy and the long form.⁶² While complaints about specific questions were received, the length of the form generated the primary concerns. Seven bills were introduced in the Congress to limit the number of questions or the penalties incurred for not answering all questions. None of the bills received legislative action.

Currently, the Data Stewardship Executive Policy Committee (DSEP) serves as the Census Bureau Executive Staff focal point for decision-making and communication on policy issues related to privacy. Data stewardship includes elements of privacy, confidentiality, security, record linkage, data access, and data dissemination. The mission of the DSEP is to assure that the Census Bureau can effectively collect and produce data about the nation’s people and economy while fully meeting the Census Bureau’s legal and ethical obligations to respondents to respect privacy and protect confidentiality.

Three high-level corporate teams report to the DSEP and serve as the focal point for issue identification, research coordination, and policy development. These are the Disclosure Review Board, formed to ensure individuals cannot be identified by the data

⁶¹ Kenneth Prewitt, “Congressional Testimony before the Subcommittee on the Census, Committee on Government Reform”, U.S. House of Representatives, April 5, 2000, pgs. 7-8.

⁶² Public Broadcast System, March 30, 2000, “Nosy Census”, <http://www.pbs.org/newshour>

released to the public; the Committee on Administrative Records Policies and Procedures, which examines policy issues related to administrative records including policies for protecting the privacy of administrative records; and the Privacy Policy and Research Committee formed to help the Census Bureau protect respondent privacy and data confidentiality. This committee examines such privacy and confidentiality issues as internal control of personal identifiers on survey and census computer files to reduce the opportunity for employee browsing, developing privacy principles, discussed below, and improving informed consent messages provided to respondents. Informed consent messages are provided to survey and decennial census respondents to ensure that they are informed about the purpose and planned statistical uses of the information collection so that they can make an informed choice about participating in the data collection.

Developing privacy principles is yet another effort to ensure information collection practices that secure full privacy and confidentiality protection at the Census Bureau. The privacy principles are meant to address the increasing difficulties in collecting and disseminating quality data linked to heightened public concern about data collection by government agencies as well as by commercial entities. The principles are not intended to repeat legal requirements under the Privacy Act and Title 13, but supplement them by communicating the ethical issues involved in asking the public to provide personal information to us. There currently are four privacy principles. They are:

- (1) Principle of mission necessity – *The Census Bureau will only collect personal, sensitive information that is necessary for meeting the Census Bureau’s mission and legal requirements.*

(2) Principle of informed consent – *The Census Bureau will ensure that participants in data collection activities are informed about the purpose and planned statistical uses of the information collection.*

(3) Principle of protecting respondents from unwarranted intrusion – *The Census Bureau will respect respondents' rights to decide the conditions of their participation in censuses and surveys and will respect their rights as research participants.*

(4) Principle of confidentiality – *The Census Bureau will ensure that confidentiality protections are included in its procedures to collect, process, and release data.*

The principles inform the public of the Bureau's ethical standards thereby reassuring respondents. New policies and controls tied to the principles ensure appropriate and consistent practices across the agency and guide employees in performing their work.

The 21st century will offer new, technological, data collection, and data dissemination challenges to protecting privacy and confidentiality for statistical agencies. Technology has brought into our lives amazing new capabilities for data collection and use, such as data collection by the Internet and linkage of discrete data sets, that also carry worrisome opportunities for unwanted intrusion and disclosure. The Census Bureau will continue to solve these challenges as it has across the centuries documented in this paper. Respondent privacy and data confidentiality also will continue to be primary concerns for the Census Bureau in accomplishing its mission to be the preeminent collector

and provider of timely, relevant, and quality data about the people and economy of the United States.

SUMMARY

The Census Bureau (along with its predecessors) is well into its third century of census taking. Today's privacy and confidentiality concerns, both within the agency and among the public at large, are the result of a long evolutionary process. From the beginning in 1790 to the early 19th Century, there appeared to be no process at all. For the remainder of the 19th Century, the pace of these slowly evolving concerns about confidentiality and privacy issues seems glacial by today's standards. At the turn of the 20th Century, the census-taking process had become sufficiently complex to warrant the establishment of a permanent agency, which is today's Census Bureau. These complexities resulted in a growing number of concerns, mainly at the Census Bureau and within the statistical community, about safeguarding the confidentiality of the data that individual respondents were willing to provide to the Census Bureau. Yet each time the agency's decision makers were faced with new challenges and the need to make changes to ensure that its respondents' data were kept confidential, the basic questions remained: "What did we do in the past and what are we doing now that needs to be changed, or what else do we need to do make sure our confidentiality safeguards do not fail us."

Today, the Census Bureau no longer can concentrate its resources just to ensure data confidentiality. Even though the agency must continue to improve its confidentiality safeguards, the American public now is just as concerned about threats to its privacy as it

is by the confidentiality of the data that are collected. As a result, the Census Bureau must convince its data providers that it is not only in their country's best interest, but also in their individual best interest to entrust their personal information to an agency of the Federal Government.

Furthermore, technological advances continue at an increasingly rapid rate changing the privacy and confidentiality landscape confronting the Census Bureau. Probably the most significant technological change between the 1990 census and Census 2000 was the emergence of the Internet. Now that Census 2000 data can be accessed on the agency's American FactFinder site, the safeguards that ensured the confidentiality of the 1990 census data are no longer adequate to protect the Census 2000 information.

It is probably a safe prediction that the evolutionary pace of confidentiality and privacy issues confronting the Census Bureau will continue to accelerate. What may be less certain, however, is how the American public will respond to the agency's intensified efforts to create a greater general awareness of "the census" and to convince its data providers that nothing bad, and in fact something good, will happen if they provide this particular Government agency with some of their personal data. Census 2000 witnessed a reversal, albeit slight, of the declining response rates since the 1970 census, so it is conceivable that an upward trend in response rates could develop. Also, the basic questions of how to protect the data and how to convince people to provide the data are likely to remain the same, although there may be additional basic questions in the future.

REFERENCES

Anderson, Margo J., *The American Census: A Social History*, Yale University Press, New Haven, CT, 1988.

Associated Press, "FBI investigating census workers," August 14, 1980.

Barabba, Vincent P., "Chronology of recent developments," unpublished memorandum, August 15, 1980.

Barabba, Vincent P., "Confidentiality of Census Records During World War II," unpublished memorandum, July 1980.

Clemence, Ted, "Census Summer in Colorado Springs [CO]," unpublished memorandum, July 1986.

Colorado Springs [CO], *Gazette Telegraph*, August 14-15, 1980.

Courtland, Sherry, "Census Confidentiality: Then and Now," *Government Information Quarterly*, Vol. 2, No. 4, 1985.

Cox, Lawrence H., et al., "Confidentiality Issues at the United States Bureau of the Census," *Journal of Official Statistics*, Vol. 2, 1986.

Gates, Gerald W., *Privacy Research Involving Expanded Statistical Uses of Administrative Records, Processing of Government Statistics and Social Statistics Systems of the American Statistical Association*, 1999.

General Services Administration, *Public Papers of the Presidents, Gerald R. Ford*, 1975, Book 1, Government Printing Office, Washington, DC, 1977.

Magie, E.R. Letter from the Acting Solicitor General to the Secretary of Commerce, January 15, 1920.

Martin, Elizabeth, "Changes in Public Opinion During the Census," paper presented at the Census Advisory Committee of Professional Associations, October 19, 2000.

Mitchell, William D. Memorandum from the Attorney General to the Secretary of Commerce, 36 *Op. Atty. Gen.* 362, September 29, 1930.

National Academy of Sciences, *Private Lives and Public Policies*, National Academy Press, Washington, DC, 1993.

Prewitt, Kenneth, "Additional Comments by Census Bureau Director Kenneth Prewitt on the Role of the Census Bureau and the Use of Census Tabulations in the Internment of Japanese Americans in World War II," e-mail to U.S. Bureau of Census staff, March 24, 2000.

Prewitt, Kenneth, "Congressional Testimony before the Subcommittee on the Census, Committee on Government Reform," U.S. House of Representatives, April 5, 2000.

"Proclamation for the Thirteenth Decennial Census," March 15, 1910.

Public Broadcast System, "Nosy Census," March 30, 2000, <http://www.pbs.org/newshour>.

Reznek, Arnold P., "Increasing Access to Longitudinal Business Survey Microdata: The Census Bureau's Research Data Center Program, May 2000.

Reznek, Arnold P. and Alfred R. Nucci, "Protecting Confidential Data at Restricted Access Sites: Census Bureau Research Data Centers," *Of Significance: A Topical Journal of the Association of Public Data Users*, Vol. 2, No. 1, 2000.

U.S. Bureau of the Census, *Factfinder for the Nation*, "History and Organization," May 1988.

U.S. Bureau of the Census, *200 Years of Census Taking: Population and Housing Questions*, Government Printing Office, Washington, DC, 1989.

U.S. Bureau of the Census, 1970 Census of Population and Housing, *Procedural History*, series PHC(R)-1, Government Printing Office, Washington, DC, 1976.

U.S. Bureau of the Census, "Introduction to the RDC Program," [On-line] available: <http://www.ces.census.gov/ces.php/rdc>, June 2001.

U.S. Bureau of the Census, *Census Bureau Legislation: Department of Commerce and Other Executive Departments*, (Robert H. Holley, comp.), Government Printing Office, Washington, DC, 1936.

U.S. Bureau of the Census, "TXE Tests & Experimentation Program," [On-line] available: <http://cww.census.gov/msdir/pred/prednet/txe/experiment.htm>, June 2001.

U.S.C., Section 552a (amended 1997, U.S.C.A. Section 552a.

U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, Statistical Working Paper 2, *Report on Statistical Disclosure and Disclosure-Avoidance Techniques*, Government Printing Office, Washington, DC, 1978.

U.S. Department of the Treasury, Internal Revenue Service, *Tax Information Security Guidelines for Federal, State, and Local Agencies*, OMB No. 1545-0962, Publication 1075, June 2000.

U.S. House of Representatives, *Confidentiality of Census Information*. Report to Accompany H.R. 12884, 91st Congress, 1st Session, Report No. 91-407.

U.S. Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Working Paper 22, *Report on Statistical Disclosure Limitation Methodology*, Washington, DC, 1994.

Wright, Carroll D. and William C. Hunt, *History and Growth of the United States Census: 1790-1890*, Government Printing Office, Washington, DC, 1900.

Zayatz, Laura and Paul Mackun, *Disclosure Limitation for the Census 2000 Public Use Microdata Files*, U.S. Bureau of the Census, 2001.

Zayatz, Laura, et al., *Disclosure Limitation for Census 2000*, U.S. Bureau of the Census, 1999.

Zayatz, Laura, et al., *Disclosure Limitation Practices and Research at the U.S. Census Bureau*, U.S. Bureau of the Census, 1999.

Zeisset, Paul T., "Making Decennial Census Data Available," *Government Information Quarterly*, Vol. 2, No. 4, 1985.