

Counting Archives In: The Appraisal of the 1991 Census of Canada

JEAN-STÉPHEN PICHÉ and SHEILA POWELL*

Résumé Cet article est une étude détaillée des circonstances, de la méthodologie et des conclusions de l'évaluation archivistique des données du Recensement du Canada de 1991 réalisée par les Archives nationales du Canada. Les auteurs analysent les données du recensement – sous forme anonyme et nominative – sur papier, microfilm et support informatique. Grâce à leurs recherches sur la création et l'utilisation des données démographiques, sur la collecte des données de recensement auprès des Canadiens et sur la manipulation des informations par Statistique Canada en vue de les rendre valides à des fins statistiques, les auteurs ont déterminé qu'il était nécessaire de préserver à la fois les données nominatives fournies par chaque Canadien sur les questionnaires de recensement, la version anonyme des données que Statistique Canada rend publique pour fins de recherche ainsi que la consignation des modifications apportées aux données. Les auteurs présentent, en conclusion, les recommandations finales de leur évaluation archivistique et l'entente intervenue entre les Archives nationales du Canada et Statistique Canada pour assurer la sauvegarde de ces données.

Abstract This article is a detailed study of the circumstances, methodology, and findings of the National Archives of Canada's archival appraisal of the data resulting from the 1991 Census of Canada. The authors provide an analysis of the census data – in both anonymized and name identified form – in paper, microfilm, and electronic media. Through their research into the creation and uses of demographic data, the collection of census data from Canadians, and the manipulation of the data by Statistics Canada to render it valid for statistical use, the authors determined that it was necessary to preserve the name identified data provided on census forms by individual Canadians, the anonymized data that is made widely-available for research by Statistics Canada, and evidence of the changes made to the data by Statistics Canada in order to render it statistically "clean." The authors conclude with a discussion of their final appraisal recommendations and the agreement reached between the National Archives of Canada and Statistics Canada to preserve data.

Introduction

Canadian archivists are familiar with these words of Arthur Doughty, the second Dominion Archivist: "[archives] are the gift of one generation to another and the extent of our care of them marks the extent of our civilization."¹ The care and attention given to census records by national archives could be said

to mark the extent of a national archival “civilization.” Few classes of records are viewed as more valuable or lead to more discussion about the proper role of archives in preserving and making personal information available on individuals. In recent decades, a number of countries have publicly debated whether census forms containing information – such as names and addresses – that allows individuals to be identified should be made available to researchers, whether the forms should be preserved after the anonymized census data has been captured in automated format, and even if censuses should continue to be taken at all. Australia, which destroys all individual census forms after tabulation has been completed, is re-examining this practice. In Canada, conflicting views on the permanent preservation of census records for research use have offered a unique opportunity to analyze the processes that result in the creation of statistically valid census data and to examine the decisions which determined what data should be preserved for future use.

In early 1995, the authors – archivists in the Records Disposition Division of the National Archives of Canada (NA) – conducted an archival appraisal of the data collected, in paper, microfilm, and electronic media, by the 1991 Census of Canada, including, most importantly, name identified data. It occurred in response to a disagreement between the National Archives of Canada, which was seeking to preserve name identified census records, and Statistics Canada and the federal Privacy Commissioner, who were opposed to long-term preservation.

This article will examine the circumstances, methodology, and findings of the appraisal of the 1991 census data. We must emphasize at the outset that the main subject of both the appraisal and this article was and is census records that contain information which identifies individuals, specifically names and addresses of respondents. While the appraisal took into account all forms of data created as a result of the 1991 census, including anonymized data, the primary focus was on determining how to preserve the name identified data.

Moreover, there are a number of important archival issues relating to the census that are not addressed in this article: there is no detailed discussion of the legal issues surrounding the protection of personal information, of the ethics of its preservation in archives, or of its release to researchers.² Nor do we enter into the debate over whether creators of electronic records should be allowed to maintain custody of such records for archival purposes beyond the end of their administrative use or discuss related decisions pertaining to the 1996 census, which was appraised at the same time.

The Genesis of the Archival Appraisal

Prior to Confederation in 1867, the French and British colonial administrations both took censuses; the first census was taken in 1666. A decennial census was established by law in 1852, with the next census taken in 1861, and

every ten years ever since. The quinquennial, or mid-decade census, was first held nation-wide in 1956. The 1871 census, the first to be taken after Confederation, was also the first employed to determine representation by population in the federal House of Commons. Canadian census data is now used to set the boundaries of federal electoral districts, to calculate transfer payments from the federal government to provincial governments, and to provide federal, provincial, and municipal governments, academics, and businesses with data used in a vast array of programs.³

The first evidence of the acquisition of a Census of Population of Canada by an archives is found in the acquisition register of the Public (now National) Archives of Canada for the year 1917. The register contains the brief notation "1851, 1861 Census rolls, from the Dept. of Agriculture, Jan 11, 1917."⁴ This acquisition coincided with plans to create the Bureau of Statistics in 1918; the parent Department of Agriculture likely decided to transfer pre-Confederation censuses to the Public Archives because they no longer had an operational need to retain them and did not wish to store them indefinitely. Despite the confidentiality clauses in the *Statistics Act*, also of 1918, the 1871 census returns were transferred to the archives in 1941 after an accord was reached between the Chief Statistician and the Dominion Archivist, who agreed that the returns would be available only for scientific research in the social sciences. This agreement was short-lived, because subsequent Chief Statisticians did not support the availability of census returns for any kind of research outside the purview of Statistics Canada. In the late 1970s, pressure from the genealogical and academic communities for the release of additional census returns mounted, and there was the threat of a private member's bill in the House of Commons to force Statistics Canada to transfer the returns to the Public Archives. In 1985, Statistics Canada bowed to what was called the "Roots Syndrome" – the explosion of interest in family history that resulted from the American television mini-series *Roots* – and transferred returns from the 1881, 1891, and 1901 censuses to the NA.⁵

Two contradictory pieces of federal legislation govern access to census records and their transfer to the National Archives. The current *Statistics Act* restricts the use of census records which contain nominal information to Statistics Canada employees only, which technically prevents the transfer of such records to any other agency of the government, including the National Archives. Statistics Canada applies this legislation retroactively to all census information back to 1906.⁶ This practice is in direct conflict with the intentions of section 6(1) of the *National Archives of Canada Act*, which states that all government records appraised as having archival value must be transferred to the NA.⁷

This disagreement between the *National Archives of Canada Act* and the *Statistics Act* is at the root of a complex relationship between Statistics Canada and the National Archives regarding the disposition of census records.

It has manifested itself throughout the twentieth century as various acquisition and custodial issues have repeatedly surfaced and resurfaced. Debate over the conflicting authority of the two acts has occurred over research use, privacy of personal information, transfer of records, and the broader cultural issue of public interest and historical relevance.

In 1993, Bruce Phillips, the Canadian Privacy Commissioner, publicly expressed concern over the intrusive nature of the 1991 census, and the potential threat the census posed to the privacy of Canadians. The Privacy Commissioner is an ombudsman, appointed by the federal Parliament, who acts as a watchdog in matters of the collection, use, and disclosure of personal information by the federal government as these relate to federal employees and participants in federal government programs. Phillips' involvement in the debate over the census and the protection of privacy was prompted by thirty-three complaints received by his office relating to the 1991 census, most of which involved protests against the intrusive nature of some of the questions, particularly those on religion and fertility, and against the use of neighbours as census takers.⁸ Phillips responded to these complaints by taking the stand that any census is, by its very nature, a threat to personal privacy.⁹

In May 1994, the National Archives received a letter from Phillips, informing the NA that the issue of the preservation of census questionnaires in the National Archives had come to his attention as a result of investigations by his office into complaints against the 1991 census. Phillips wrote that he believed that the transfer of census questionnaires to the National Archives and their preservation therein was an inappropriate use of census data.¹⁰ Phillips went on to say that the destruction of name identified data would eliminate any risk of that data being used against individuals or groups, and intimated that he was seeking the concurrence of the National Archivist in the destruction of the census records.

As a result of this letter, senior officials of the National Archives met with representatives of Statistics Canada and the Privacy Commissioner's Office to discuss the situation. Michael Swift, Assistant National Archivist, then prepared a draft report that set out the viewpoints of the three parties on the preservation and destruction of the census returns, outlined the legal issues involved, and proposed three possible options for action.¹¹ The first option was to do nothing, with Statistics Canada continuing, as it did, to store the microfilm for the 1911 to 1986 censuses and the microfilm and original census returns for the 1991 census. The second option was for the National Archivist to agree to the destruction of all microfilmed and original census returns. The third option (and the one recommended by Swift for adoption by the three parties) was to transfer all name identified census records from 1911 to 1991 to the National Archives, where they would be preserved, but not made available for research.

Statistics Canada rejected the National Archives proposal, specifically, that

all microfilmed census returns dating from the 1911 to 1991 censuses be transferred to the NA, stating that the only feasible option open to them was to seek authority for the destruction of the 1991 census records. This constituted Statistics Canada's formal request for authority to dispose of the records, and in late March 1995 we embarked on the appraisal of the name identified census returns in paper and microfilm format, as well as the databases containing anonymized census data. Our instructions were to report back to the National Archivist by the end of April 1995.¹²

The terms of reference for the appraisal were established in collaboration with Statistics Canada. It was agreed that the appraisal would concern only the data on individual Canadians and households collected by the 1991 Census of Canada. The records consisted of the census forms completed in all Canadian households (and maintained by Statistics Canada in paper format for all forms and on microfilm for a portion of the forms), the visitation records completed on paper by Census Representatives (better known as census takers), and the electronic databases created for purposes of processing and anonymizing the census data. Excluded from the appraisal would be all policy and operational records, and any other Statistics Canada records relating to the census in general, to census-taking procedures, to the financing of the census, to decisions on changes to the questionnaires, and to the publicizing and use of the census; these records would be the subject of later appraisals. It was also decided that the report would provide not one, but a series of options for the long-term preservation of the 1991 data, outlining the advantages and disadvantages of each, including completeness of information and resource implications.

The use of the word *preservation* rather than *acquisition* is deliberate. Knowing Statistics Canada's reluctance to relinquish any of its sensitive records to the control of another agency and the Privacy Commissioner's crusade to destroy all name identified census records, we refrained from broaching the subject of transfer in discussions with staff of Statistics Canada and in our appraisal report to the National Archivist. By dealing with the issue of preservation rather than transfer, the NA avoided becoming bogged down in issues of security and access, and left itself open to possibilities of distributed custody of the archivally-valuable census records.

The Appraisal Hypothesis

Our research began with two assumptions that were unshaken during the course of the appraisal: that the data resulting from the 1991 census was of archival value, and that the inclusion of personal identifiers in the census records was integral to its value. Our conclusions regarding the census' value stemmed in part from our knowledge of the extensive use, by many varieties of researchers, of the 1851 to 1901 censuses already in the custody of the NA. It was assumed that the assumption of archival value could be applied auto-

matically to the 1991 census records. Since the National Archives had already informed Statistics Canada and the Privacy Commissioner of its conclusion that the records were of unquestionably high archival value, it was unnecessary to expound any further on the value of the 1991 data. We also accepted the recommendation of Terry Cook's 1991 RAMP study, *The Archival Appraisal of Records Containing Personal Information*. This study states that "the national census is the single most essential personal information record in terms both of research for many disciplines and for genealogists, and of providing the core demographic information vital to the design, delivery, and modification by the government of its own major programs."¹³ This conclusion was affirmed by a seven-country group of experts from the International Council on Archives, which clearly suggests that Canada's perspective on the value of census records is shared by the international archival community.

From the outset, we and our managers were certain that, regardless of their form or format, the census data to be preserved had to include identifying information in the form of names and addresses in order to permit future use that requires the identification of the providers of information. Current experience with users of census records in the custody of the National Archives demonstrates that many users require identifying information in order to conduct a wide variety of research. Genealogical research clearly requires that one be able to link the names and addresses of respondents to the data that they provided; many other areas of research, such as studies of communities, health trends, and social migrations, also require address information in order to group relevant data together. Apart from pointing to known research methods, we noted that archivists are not able to predict what new research types and needs may emerge in the future, or whether that research will require personal identifiers.

Our final assumption, that the most usable census records are in electronic form, resulted from the sheer number of individual census forms – over ten million forms in 1991. It is no longer feasible, given the huge number of records, to sort and analyze data manually. It is only in electronic form and through electronic processing that the records can be used in a meaningful way and to their highest potential by researchers.

Macro-appraisal analysis of other data collected by the federal and provincial governments led us to determine that the census was the single most complete and uniform body of demographic data in Canada. The provinces are responsible for maintaining records of births, marriages, deaths, adoptions, divorces, and changes of names. These records contain much of the data on individuals that has been traditionally sought by genealogical researchers: date of birth, date of death, names of parents, occupation of parents, residence, place of birth, cause of death, religious denomination, and date and place of marriage.

The crucial difference between provincial vital statistics and the census

records is that the provincial data is event-driven and thus recorded only at certain points in an individual's life when these events occur, while the census collects data at regular intervals throughout the course of a person's life. For example, provincial vital statistics on an individual who never married and who had no children would be limited to those collected during registration of their birth and death. On the other hand, census questionnaire forms would provide information at regular five year intervals on other aspects of a person's life, such as address, marital status, language, and the identity of the person who pays the rent or mortgage in the family. This information is collected on all individuals, and even more is collected on twenty per cent of the population through the long census form (Form 2B). This data is extensive, including information on ethnic origin and immigration data, aboriginal status, education, religion, labour force participation, income, housing, and disabilities.

Data is also collected by a number of other federal government programs. Taxation records and records maintained for the purposes of administering federal Income Security Programs, such as the Canada Pension Plan and Old Age Security, contain information on date of birth, place of residence, income, marital status, and other individual characteristics, depending on the type of program. There is, however, no federal government system that contains all the types of data that are captured through the census. For the departmental systems, specific data elements are collected for the purposes of administering and delivering specific programs within a limited period of time. The data is relevant only to those programs and the more limited needs of those citizens interacting with them, and it is maintained only as long as is necessary to the delivery of these programs. The census, by definition, covers all Canadians.

It became evident as we began our census research that our focus was not on each single census record or on the one big database of census records. Rather, we realized that taking a census constitutes a business function that involves several processes ranging from data gathering and processing to dissemination. These processes result in the creation of a number of different types of records in various forms and media. It also became clear that in order to preserve census records that will meet the future needs of government and of researchers, it was crucial to capture data at each step of the census activities and, *as well*, preserve evidence of the many ways in which the data changes throughout the entire census process. Above all, we feared that if nothing but the final product of the census, the Dissemination Database, was permanently preserved (though this, in itself, is significant as the primary source of data used by government, business, and researchers to conduct analysis), the use of editing techniques to alter attributes of records or even to create entirely new records in order to make the data clean for statistical research would mean loss of information. The data available to future researchers would not necessarily reflect *real* individuals and their households accurately. With this acquired perspective, we changed our initial appraisal hypothesis,

which had focused on how to “save” the paper census returns, to one of preserving, first, the data that is initially provided by Canadians on their census forms, including name identified information, and second, the final “clean” census database. We also believed it was necessary to preserve, in electronic form, the data originally supplied by respondents and loaded onto computer databases; that is, in cases where Statistics Canada had modified records in order to render them statistically valid, we wanted to ensure that the original data supplied by respondents was still preserved.

Our desire to preserve the data provided on the original census forms along with the personal identifiers was consistent with the traditional conception of the census as a treasure trove of discrete records on individual Canadian households. We also saw the census as a single, vital store of information that informed, through its use by all levels of government in Canada, as well as the business, academic, and other communities, the most important social and economic policy decisions taken in Canada in the 1990s. We believed that it was crucial to preserve the data in the same format and with the same interfaces that were present when the data was used in research leading to government and private sector policy decisions; users in the future, whether governmental or not, must be able to have access to the same 1991 census data that was used make policy decisions in the 1990s.

It became clear, as the appraisal proceeded, that it was also important to preserve evidence of the changes that the data undergoes at the hands of the statisticians as they create statistically valid data. These complex processes result in the final, anonymized, statistically valid data that is used by governments, business, research bodies, and eventually the general public, through statistical analyses and micro and macro socio-economic studies, to create an image of our country, a collective impression of who we are, and where we are as a society. Statistics is a social science subject to error and bias. Canadians eventually must have original evidence available in order to be able to challenge and critique conclusions that were based on the statistical analysis of “polished” or “edited” census data. We believe that this is best achieved through the ability to gain access to evidence of the changes the data underwent at the hands of the statisticians.¹⁴

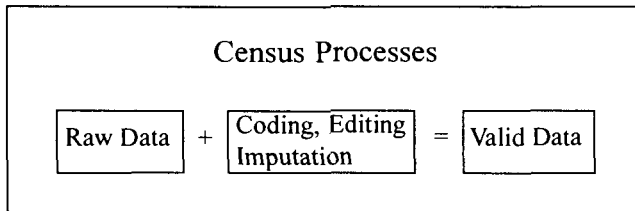
With our revised goal to preserve the data provided by Canadians, evidence of the processing stages, and the final census database, we jumped into a brief, but intensive period of research into the census records. The following section provides an overview of the processes of collecting, processing, and disseminating census data.

The Census Records

There are significant differences between the raw data that is *collected* from individuals via the census questionnaire forms, and the statistical data that is

disseminated to users as the final, “clean” product. A number of steps are taken between the collection and dissemination phases that render the data accurate and usable as statistics; we refer to these steps as *processing* (See **Figure One**).

Figure One



Collection

Collection consists of the steps involved in acquiring the raw data on individuals and households on census questionnaires. At this stage, personal identifiers are included in the records. (Personal identifiers, such as names and addresses, are removed during the processing stage, described below.) The most important records resulting from data collection are the completed paper census returns, mainly forms 2A and 2B. The 2A forms, filled out by eighty per cent of the population, contain strictly demographic information such as age, sex, marital status, and housing unit. The 2B forms collect additional socio-economic information for twenty per cent of the population, such as place of birth, citizenship, ethnic origin, religion, languages, labour force activity, income, fertility, and population mobility. The physical extent of the forms for 1991 is enormous: over ten million forms, the physical extent of which is approximately 19,000 linear metres, or 60,000 one foot boxes. In addition to the census returns, a “visitation record” is created, which contains the names and addresses of citizens living in one particular Enumeration Area (the primary unit of collection of forms). There were 46,000 such areas across the country for the 1991 census. Statistics Canada, in conformity with its practice for previous censuses, had started to microfilm the completed census forms, but had halted the project for cost reasons. Statistics Canada completed filming of forms, and verifying the quality of the film, from the provinces of Newfoundland, Nova Scotia, New Brunswick, and Prince Edward Island. Forms from Québec and parts of Ontario were filmed, but no verification was carried out. The forms from the remaining provinces and the two territories were not microfilmed.

Census questionnaires are distributed, filled out by citizens, and mailed

back to a Census Enumeration Area unit where they are verified against a visitation record. Census enumerators use the visitation record to help contact citizens who have not completed their census forms. Once this operation is completed, records are shipped to a regional processing centre.

Processing

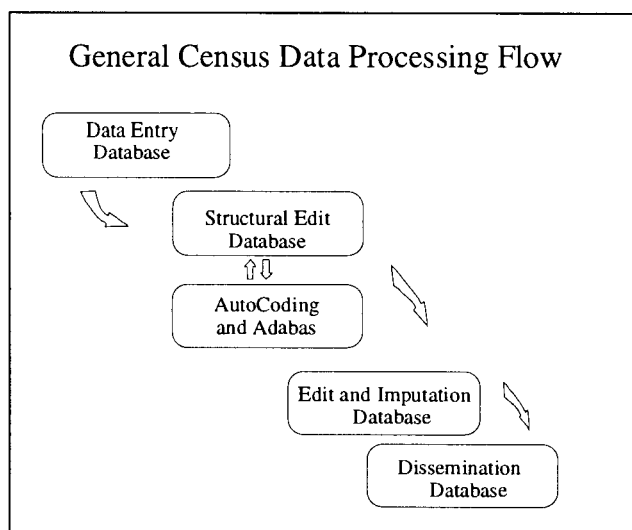
The processing stage consists of the measures, including coding, editing, and imputation, that are taken to transform the raw data into statistically valid data. The data at this stage is generally not considered ready for use in generating statistical reports or research. *Coding* is the process of transforming the raw data into numeric codes in a succession of processing databases. It is performed either manually or by an automated process. *Editing* applies in a general sense to a number of steps that are taken to ensure that the data is free of errors and other incongruities and inaccuracies that could render it statistically invalid. *Imputation* is a complex stage of editing that entails the creation of one or more values in a record where no values existed *or* replacing a statistically unacceptable value in a census record with an acceptable value derived from one or more other similar records. (Examples follow later.) If the imputation software detects anomalies in the values of certain responses on a questionnaire, imputed values are substituted.

Initial processing, which is the transformation of data from the paper census forms into electronic format, in the Data Entry Database, is performed on the premises of the Revenue department by contract employees working for Revenue Canada under the auspices of Statistics Canada.¹⁵ It is important to note that the names and addresses of individuals whose data is recorded on the paper census forms are not captured electronically, but are replaced by an anonymous personal identifier code, called the "PROV-FED-EA-HOUSEHOLD-PERSON" number (this number is drawn from each paper census form). This anonymous personal identifier allows for the protection of the privacy of individuals when census information is available to the public.

Once regional processing is completed, the electronic data is forwarded to headquarters along with all the paper census returns. The electronic data is entered into the Structural Edit Database, where the first level of verification and editing of data values is performed. Various verification routines are run, including the detection of potential errors during the original enumeration of the population, e.g., in the relation between number of individuals and number of dwellings. All records containing errors are then edited manually, and the incorrect values are replaced with values from the paper forms or with values that are structurally valid.

Next, all textual attributes of records maintained in the Structural Edit Database are downloaded into the Autocoding Database. This database performs a routine to convert all textual fields into numeric fields. The values are merged

Figure Two



back into the Structural Edit Database, and the records are then ready to be processed in the Edit and Imputation Database.

The Edit and Imputation Database runs a series of in-house statistical programs against the data in order to determine if any values in individual records should be imputed. One program (SPIDER) analyzes records and replaces values that do not meet criteria set in the various decision tables, as we have noted above. If, for example, the content of an attribute/field is empty or is outside the parameters of the decision tables, the program will identify a similar record elsewhere in the database that is called a "donor" and will borrow the required value and copy it back to the attribute/field of the record where an error was detected. As a hypothetical example, if a census form indicates that an individual's income was one million dollars, that he or she had ten children, and in the rest of the person's neighbourhood the average income was \$40,000 a year, and the average number of children per family 1.5, then the software would calculate a more likely income and number of children, based on the responses of the other "average" households in the neighbourhood, and substitute it in the census database for the original response, even if the original response was completely correct. If there was no completed census questionnaire at all for an individual household, *all* of the data on that household would be imputed at this stage.

Each record can contain both imputed and unimputed data. The rate of imputation in the census varies from one to thirty per cent, depending on the variable. The records from the Edit and Imputation Database are downloaded

into the Retrieval Database, which performs querying routines to generate subsets used for thematic statistical research. The key characteristic of the Edit and Imputation Database is that it not only contains all the census data, but all the imputed *and unimputed* values of variables. When the data is transferred into the Retrieval Database, the unimputed values are dropped where the changes have been made; as a result, if imputation has been performed on any variables in a record, only the imputed values will appear in the Retrieval Database. The data has now been cleansed of errors and omissions that could cause inaccuracies when used for statistical purposes. This “clean” data is made widely available, at the dissemination stage, to all of Statistics Canada’s users within the public and private sectors.

Dissemination

In order to optimize the querying and speed capabilities of research on census data, as well as to ensure the protection of the data through technological change, Statistics Canada has opted for the migration of census data from the Retrieval Database into a Sybase relational database resting on a Unix platform, called the Dissemination Database. This database is a client-server application that allows users to formulate requests for data and obtain quick responses. The Dissemination Database also incorporates data dictionaries, access rules, and transaction logs, and a library database, which is used to manage requests. This database is also attached to another database containing all census geographic variables, which allows researchers to specify requests based on geographical location.

Appraisal Recommendations

Our revised appraisal hypothesis was to capture, first, the data that is initially provided by Canadians on their census forms, including personally identifying information, and second, the final “clean” census database. And, in cases where Statistics Canada had modified database records in order to render them statistically valid, we wanted to ensure that the original data supplied by respondents was still preserved in electronic form. Once we had a thorough understanding of the processes that resulted in the collection and alteration of data, we were ready to determine what data should be preserved.

Our terms of reference required us to present to the National Archivist a series of preservation options for further deliberation within the National Archives and to use in negotiations with Statistics Canada. We provided three distinct preservation options, each of which recommended the preservation of the Dissemination Database along with one of the three forms of the census documentation providing name identified information from the collection stage: the original questionnaires, the microfilmed questionnaires, and the vis-

itation records. A fourth option proposed both the preservation of the Dissemination Database and the creation by Statistics Canada of an electronic database of the names and addresses of all individuals represented in the Dissemination Database. The creation of this database would obviate the need for preservation of the paper or microfilmed questionnaires, and would render the records searchable by name.

We knew, however, that this option was not likely to be adopted: Statistics Canada had promised Canadians that it would not create an electronic database of the names and addresses of census respondents, and the Privacy Commissioner would almost certainly have opposed the creation of such a database.

In order to satisfy our goal to preserve evidence of the changes that the data underwent as it was processed, we also recommended for all options that the unimputed values of variables that had been altered as a result of the imputation process – located in the Edit and Imputation Database – be preserved, and that these values be added and linked to the records to which they relate within the Dissemination Database. This addition will not prevent future users from viewing the Dissemination Database in its original form. We also recommended that Statistics Canada be required to provide the National Archives with the algorithm to convert the identifying code for each record in the Dissemination Database to the original identifying code used on the census questionnaire and other databases.¹⁶

The National Archivist considered these recommendations, and determined that the preservation of all of the paper census forms and the “enhanced” Dissemination Database (that is, the Dissemination Database with the addition of the unimputed values from the Edit and Imputation Database) was the best option for capturing as complete a record of the census as possible. Once the National Archivist decided on the preservation option, he was obliged to negotiate the physical transfer of the paper forms to the NA, and the preservation, by Statistics Canada, of the enhanced Dissemination Database on their own premises.

The Agreement with Statistics Canada

The National Archives of Canada and Statistics Canada concluded their agreement on the disposition of the records of the 1991 Census of Canada in December 1995, and the records disposition authority was issued by the NA the same month.¹⁷ Because the National Archives did not press for legal control of the census records, but instead asked for continued preservation by Statistics Canada on NA premises, Statistics Canada was able to concur in the agreement. As a result, the records disposition authority allowed Statistics Canada to destroy the paper forms that had been microfilmed, on the condition that the quality of the microfilm had been verified and that the master

version of the microfilm be stored on National Archives premises. The legal control of these records, however, remains with Statistics Canada, as does the application of access and privacy legislation. In essence, the National Archives is the custodian of the 1991 census paper or microfilmed forms, but cannot make them available to the public, even after the expiry of the period of ninety-two years prescribed by the *Privacy Act* regulations; the records are still the property of Statistics Canada, and for now are subject to the restrictive "never release" clauses of the Statistics Act.

Statistics Canada agreed to preserve the Dissemination Database on its own premises and at its own expense. It also agreed to preserve all unimputed values of modified attributes in the Edit and Imputation Database, as well as the unique PROV-FED-EA-HOUSEHOLD-PERSON number, which could eventually serve to link names and addresses to records within the anonymized Dissemination Database. The National Archives will monitor Statistics Canada's adherence to these requirements on a regular basis.

The National Archives is now faced with the dilemma of storing millions of census forms that it does not legally own, and that it will not be able to provide to researchers for the foreseeable future. It is currently investigating means of reducing the extent of the forms that must be maintained, while at the same time preserving the name identified information from the forms. One possible solution is the microfilming or digitizing of the two pages of the form that contain the names and addresses of the members and the identifying code of each household. This is also an expensive venture, but could be the only means of releasing prime archival storage space to other archival records while at the same time preserving the vital data from the paper forms.

Conclusion

The appraisal of the 1991 Census of Canada and the agreement for preservation that was reached with Statistics Canada has opened up a new understanding of the census and of the census-taking and census-creating processes and has created new roles for both the National Archives and Statistics Canada in preserving and making available census records. We believe that the agreement with Statistics Canada will enable us to capture for future users some of the transactionality of census data gathering and processing, which is necessary to reconstruct the context surrounding the creation of the single most important source of demographic information on the Canadian population. The groundwork has been laid for pursuing a continuing joint initiative to perform macroappraisal on the entire function of statistics gathering by the federal government.

Our relationship with Statistics Canada will no doubt lead the National Archives into further explorations of non-custodial or distributed approaches to the preservation of statistical data, and provision of access, in cases where

acquisition by the NA is not possible, and a new role for the NA in the dissemination and description of statistical records. To some, this means the death of archives; to others, it means a brave new archival world. We feel that our main accomplishment in conducting this research was that we took an approach to the census that saw it as more than a store-house of data on individuals and families. We saw the processes relating to the census as a complex, on-going relationship between the Canadian government and the Canadian population, a relationship that is becoming increasingly problematic as the state's legitimate need for accurate demographic information collides with a growing fear of the erosion of the rights to privacy of the individual and the protection of personal information. The National Archives and Statistics Canada share a common mandate to foster research into Canadian society and to provide the raw materials for that research; at the same time, these two institutions must also ensure that personal information on Canadians is not used in an inappropriate manner. Our challenge now within the National Archives is to increase our understanding of the remaining conflicting issues of privacy protection and access to data, and to serve in a sensitive manner as the brokers between these two competing interests. Finally, we believe that, regardless of the mechanics of how the records of the 1991 census are preserved and made available, we have contributed a new, archival interpretation to our understanding of census records, and fulfilled our role of ensuring that future researchers eventually have access to the only body of records in which all Canadians are named and numbered.

Notes

* This article is a much-revised version of a paper given at the 1996 Annual Conference of the Society of American Archivists in San Diego, California. We would like to thank Terry Cook, formerly of the National Archives of Canada and now of the University of Manitoba, for his comments on this article and his constant encouragement. Opinions expressed by the authors are not necessarily those of the National Archives of Canada.

1 Introduction to Arthur G. Doughty, *The Canadian Archives and Its Activities* (Ottawa, 1924), p. 5. Doughty immediately went on to write:

Each day that passes is a triumph for an archive, for each day some mere scrap of paper permits justice to prevail. And yet while so much of our happiness and safety as individuals, and as a people, is inseparable from archives, the average man bestows little thought either upon their commercial or their historical value. Indeed those who are deeply concerned about these matters are not infrequently accorded scant sympathy. Serious minded people often question the wisdom of preserving records, and the fiat is issued for their destruction. Nor is this altogether surprising. For twenty years or more, perchance, no one has required them. Why should they be kept?

Decades later, Doughty's words remain relevant to archivists, especially those who advocate the preservation of records whose continued survival is deemed needless or inappropriate by some members of society.

2 For a comprehensive discussion of access issues relating to the census see David H. Flaherty,

- "Access to Historic Census Data in Canada: a Comparative Analysis," *Canadian Public Administration* (Winter 1978), pp. 481-98.
- 3 This information on the history of the Canadian census is taken from Statistics Canada, *Census Handbook* (Ottawa, 1992), pp. 7-12.
 - 4 National Archives of Canada, Manuscript Division, *Acquisition Register 1* (1908-1937), p. 64.
 - 5 National Archives of Canada, RG 37, vol. 29, file 60-3-DBS and National Archives of Canada, file 9430-50/S5, vol. 2.
 - 6 A confidentiality directive was passed as an Order-in-Council on 31 March 1911 (P.C. 646). Because the 1906 statistics legislation stipulates that all rules and instructions related to the census set out by the Governor-in-Council have force of law, Statistics Canada interprets the 1911 confidentiality provision to apply to the 1906 census and all subsequent censuses.
 - 7 Access to census records is governed by the *Privacy Act*. Depending on who is the custodian of the records, the regulations to the *Privacy Act*, which provide guidelines for the application of the *Act*, govern access to census records differently. If census records are in the custody and under the control of the National Archives, they can be released after 92 years (*Regulations to the Privacy Act*, Section 6d). Because no post-1901 census records have been transferred to the National Archives, this means that all census returns currently at the National Archives are open to researchers. As for post-1901 nominal census records held at Statistics Canada, there are no provisions in the regulations to the *Privacy Act* for making these records available to the public.
 - 8 The Privacy Commissioner of Canada, *Annual Report, 1994-1995* (Ottawa, 1995), pp. 33-38. Of the thirty-three complaints received by the Privacy Commissioner, seven were "not well-founded" or were discontinued.
 - 9 Speech by Bruce Phillips, Privacy Commission of Canada, to the National Statistics Council, Ottawa, 26 November 1993. Transcript provided by the Office of the Privacy Commissioner.
 - 10 Bruce Phillips, Privacy Commissioner, to Jean-Pierre Wallot, National Archivist, 29 April 1994, File EX6237-50/S5, Vol. One.
 - 11 "Census Records, Privacy Issues and the Preservation of Canadian Heritage," A Draft Discussion Paper Prepared by Michael Swift, Assistant National Archivist, June 1994, File EX6237-50/S5, Vol. One.
 - 12 Under normal circumstances, the National Archives would have entered into an agreement, known as a Multi-Year Disposition Plan, with Statistics Canada. This agreement would have set out joint priorities for the disposition of Statistics Canada's records; each priority would be the subject of a formal request by Statistics Canada for authority to dispose of records. Following a request for authority to dispose of a specific body of records, the National Archives performs an archival appraisal, which is informed by an analysis of the records by a National Archives Information Systems Analyst. In the case of the 1991 census appraisal, the appraisal took place outside of the context of Statistics Canada's Multi-Year Disposition Plan, and was required to be completed in such a short time that two people were assigned to it: an archivist, Jean-Stéphane Piché, and his supervisor at the time, Sheila Powell. Technical advice was also provided by Rick Schnarr of the Electronic Special Projects Division of the NA. For more information on the National Archives' approach to the disposition of federal government records, see Bruce Wilson, "Systematic Appraisal of the Records of the Government of Canada at the National Archives of Canada," *Archivaria* 38 (Fall 1994), pp. 218-31.
 - 13 Terry Cook, *The Archival Appraisal of Records Containing Personal Information: A RAMP Study With Guidelines* (Paris, 1991), p. 22.
 - 14 The 1840 US census serves as an example of why it is valuable to have a first hand look at how the answers to the census questions are recorded, or more significantly, misrecorded and misinterpreted. The final results of the 1840 census showed an alarming rate of insanity and idiocy among the African-American population of the northern states, which proponents of

slavery held up as proof that freedom led to insanity among former slaves. The statistics, however, were wildly inaccurate due to a problem with the format of the census form that led information to be recorded in the wrong category. For a discussion of the problems with the 1840 and other American censuses, see Margo J. Anderson, *The American Census: A Social History* (New Haven and London, 1988).

- 15 Revenue Canada employees involved in census processing are sworn in as employees of Statistics Canada, and are therefore allowed to view census records. The authors were also sworn in as Statistics Canada employees in order to conduct their appraisal.
- 16 While the Dissemination Database does not contain the names and addresses of census respondents, it does contain a code that can be converted into the unique identifying code (the FED-PROV-EA-HOUSEHOLD-PERSON number) used on the original census forms and in the other processing databases. In this way, the names and addresses of respondents from the forms can eventually be linked to the corresponding census data in the Dissemination Database.
- 17 Records Disposition Authority Number 95/024.