
Data Quality and Confidentiality Standards and Guidelines (Public)

2001 Census Dissemination



Prepared by: 2001 Census Dissemination Project

Date: November 1, 2002

Revised: July 16, 2003

Table of Contents

	Page
1. INTRODUCTION.....	4
1.1 Background.....	4
1.2 Executive Summary.....	4
1.3 Census Release Criteria.....	5
2. CONFIDENTIALITY (NON-DISCLOSURE) RULES	6
2.1 Area Suppression for Standard and Non-standard Geographic Areas.....	6
2.2 Random Rounding	7
2.3 Confidentiality Adjustment for Place of Work Counts.....	8
2.4 Disclosure Avoidance for Statistics	8
2.4.1 Statistic Suppression	9
2.4.2 Special Statistic Calculations	9
3. CONFIDENTIALITY PRACTICES.....	10
3.1 Confidentiality Adjustment for Population and Dwelling Counts.....	10
3.2 Preventing Disclosure Through Content Planning	10
3.3 Census of Agriculture and Census of Population Linkage.....	10
3.4 Public Use Microdata Files	11
3.5 Access to Census Databases	11
4. DATA QUALITY RULES	12
4.1 Incompletely Enumerated Areas	12
4.2 Partially Enumerated Areas	12
4.3 Data Quality Suppression for Standard Geographic Areas	12
4.3.1 Global Response Rates.....	12
4.3.2 1996 Adjusted Population Flag	13
4.4 Area Suppression for Income Characteristic Data	13

DATA QUALITY AND CONFIDENTIALITY STANDARDS AND GUIDELINES
2001 CENSUS DISSEMINATION

4.5	Data Quality Indicators	13
4.6	Other Methods of Data Quality Suppression	14
4.6.1	Distribution Suppression.....	15
5.	DATA SUPPRESSION – OTHER	16
5.1	Indian Reserves – Citizenship and Immigration Data	16
5.2	Incidence Reporting	16
5.3	Zero Suppress	16
5.4	Place of Work Flow Suppression	16
6.	APPENDIX A – Summary Table of Confidentiality Rules (Historical)	17
7.	APPENDIX B – Summary Table of Data Quality Flags (Historical).....	19

1. INTRODUCTION

1.1 Background

Data disseminated by the census are subjected to a variety of automated and manual processes to determine whether the data need to be suppressed. This is done primarily for two reasons, (1) to ensure non-disclosure of individual respondent identity and characteristics and (2) to limit the dissemination of data of unacceptable quality.

Additionally, suppression of data may be applied for product specific reasons due, typically, to formatting issues. The term **product** refers, primarily, to tabular output. Data may be either modified in the product or removed from the product altogether to effect the suppression rules required.

This document is intended to summarise the rules applied by census for the above purposes.

1.2 Executive Summary

The Data Quality and Confidentiality Standards were developed for application in the 2001 Census Dissemination Project. They include a summary of the rules and practices that were carried over from 1996, along with new rules and practices adopted for 2001.

New for 2001 are the following:

- introduction of new geographical hierarchies, specifically the dissemination area (DA), with the resulting expected decrease in the amount of suppression;
- change to the random rounding routine to further reduce the risk of disclosure of respondents or their characteristics;
- change to suppression rules for income characteristics to include two threshold criteria, population size greater than or equal to 250 and number of private households greater than or equal to 40;
- the number of same-sex common-law couples will be made available for all geographic areas with population of 5,000 or more;
- change to suppression rules for geo-coded and custom area files built from block, block-face or six-digit postal code levels to have a minimum suppression threshold of 100;
- more detailed breakdown of data quality flags identifying non-response rates;
- confidentiality adjustment for Place of Work data;
- clarification of the statistic suppression rules and practices;
- enhancements to the statistic calculation methods; and
- confidentiality adjustment for population counts, whereby all census block population counts less than 15 will be rounded to base 5.

The following information is not necessarily new for 2001 but has been included with the intention of having one comprehensive reference document related to data quality, confidentiality and security practices employed in the dissemination of the 2001 census of population data. These sections include:

- practices for avoiding disclosure in table content planning;
- security measures surrounding the access to census databases;
- confidentiality practices used in the census of population/census of agriculture linkage; and
- practices for producing public use microdata files (PUMF).

1.3 Census Release Criteria

The following questions were given consideration when reviewing and assessing the current release criteria and established policy for census data tabulations:

- Has the release criteria been relaxed over the last three or four census cycles. And, if so, why?
- What additional data outside the agency, if any, might give cause for concern with respect to current population release thresholds?
- What new data mining tools or software being used today might impact on census data release thresholds?
- Are our release criteria for tabular data more or less stringent than other statistical agencies?
- Does the census random rounding policy still work as an effective disclosure avoidance technique, and should we be looking at others?

The current release criteria and established policy for census weighted tables has proven to be successful at ensuring non-disclosure. Evolving client demands, data mining, the ever-increasing data combination possibilities and software developments give reason to perform regular assessments, research new methods and make changes where necessary.

The census release criteria have been modified slightly over time since the 1981 census, adapted to meet changing needs and conditions. In general, although some population thresholds for geographic area suppression were adjusted downwards from 1981 to 1986 and remained constant since then, there was no evidence of any negative impact resulting from this change. In other ways the criteria have actually been strengthened, such as the additional protection added to the random rounding rules, the addition of the higher suppression threshold for postal code information, and suppression of certain statistics (e.g. standard deviation = 0). The population universes upon which the population thresholds are based were also modified in 1991 and 1996 to add protection against disclosure. In addition, a new suppression method has been implemented in 2001 that will prevent the dissemination of tables in which the number of units (individuals, families or households) is below a given threshold. The suppression rules and random rounding practices specific to the dissemination of 2001 data are provided in **Section 2** of this document. A summary table of historical confidentiality rules can be found in **Appendix A** at the end of this document.

2. CONFIDENTIALITY (NON-DISCLOSURE) RULES

The following describes the various suppression rules used to ensure confidentiality (or non-disclosure) of individual respondent identity and characteristics. All census data are subject to confidentiality suppression rules. New for 2001 is a confidentiality adjustment performed on the total population and dwelling count data (refer to Section 3.1).

One of the changes to the 2001 census involves the move from enumeration areas (EAs) to dissemination areas (DAs) as the low level geographic dissemination unit. The objective behind implementing the DA was to reduce the amount of suppression (from 28% for EAs in 1996 to 5% when DAs are fully implemented for 2006) and make the dissemination unit more uniform and consistent than the EA was. There were approximately 50,000 EAs in 1996. The population thresholds for the DA range between 400 and 700 with the average being 540 in order to maintain the same number of areas for 2001. For rural areas, DAs will be equivalent to EAs. For the 2001 census, the delineation of dissemination areas (DAs) will be restricted to census metropolitan areas (CMAs) and census tracted census agglomerations (CAs). Outside these areas, the DA boundaries will be the same as the 2001 enumeration area boundaries.

2.1 Area Suppression for Standard and Non-standard Geographic Areas

Area suppression is used to remove all characteristic data for geographic areas below a specified population size¹.

The specified size for all standard areas or aggregations of standard areas, except blocks, block-faces or postal codes is **40**. No census data are released for areas below a population size of 40. The population under consideration for all 100% data tabulations is the total population. For all other tabulations except Place of Work data and same-sex-common-law-couples data, the population under consideration is the non-institutional population. For Place of Work data, the population under consideration is the employed labour force. For same-sex-common-law-couples data, the population under consideration is the non-institutional population in private households.

The specified size for six-character (FSA-LDU) postal codes, geo-coded areas and custom areas built from the block, block-face or LDU levels is **100**. Consequently, no census data are released for these area types below a population size of 100.

Additional area suppression is applied for data quality reasons if the census tabulation contains any data showing individual, family or household **income** characteristics. See Section 4.4 for further information.

New for 2001, the number of same-sex common-law couples will be made available for all geographic areas with a population of 5,000 or more. Data on same-sex couples will not be available **under any circumstances** including custom requests for geographic areas with population under 5,000. The 5,000 threshold is based on population in private households from the 2B census questionnaire. Studies are ongoing to evaluate the quality of the data on same-sex common-law couples for populations below 5,000. These studies take into consideration the low prevalence of same-sex common-law couples in Canada and the fact that respondents were asked in 2001 to respond to this question for the first time in a census.

¹ The population size is defined as the number of persons in 100% data tabulations. For the 20% sample data tabulations, it is the lower of the actual or estimated number.

The standard geographic areas include FSA-only postal codes (first three characters of the postal code). Generally, blocks and individual urban block-faces (one side of the street between two intersections) will be too small to meet the above thresholds. Where an aggregation of blocks or block-faces fall above the threshold, data can be retrieved through a custom tabulation.

For census tabulations that include **Place of Work** data, the 40, 100 and income size limits are applied to the **Employed Labour Force** counts for these areas, rather than the population of the areas. Tabulations containing both places of residence and places of work as geographic areas have the 40, 100 and income size limits applied to both place of residence (population) and place of work (employed labour force).

A table summarizing the history of area suppression rules for the census can be found in **Appendix A** at the end of this document.

2.2 Random Rounding

It is useful here to review the random rounding technique used by the Canadian population census. All counts in census tabulations are subjected to a process called random rounding. Random rounding transforms all raw counts to random rounded counts. This reduces the possibility of identifying individuals within the tabulations.

For 2A (100%) data (short form), all counts are rounded to a base of 5. This means that all 2A counts will end in either 0 or 5. The random rounding algorithm employed controls the results and rounds the unit value of the count according to a pre-determined frequency. The table below shows those frequencies. Note that counts ending in 0 or 5 are not changed and remain as 0 or 5.

Unit Values of	Will round to count ending in 0	Will round to count ending in 5
1	4 times out of 5	1 times out of 5
2	3 times out of 5	2 times out of 5
3	2 times out of 5	3 times out of 5
4	1 times out of 5	4 times out of 5
5	never	Always
6	1 times out of 5	4 times out of 5
7	2 times out of 5	3 times out of 5
8	3 times out of 5	2 times out of 5
9	4 times out of 5	1 times out of 5
0	always	Never

2B (20%) data (long form, completed by 1 in 5 households) require a slightly different random rounding algorithm. All counts greater than 10 are rounded to base 5, as is done for 2A data. Counts less than 10 are rounded to base 10. This means that any 2B counts less than 10 will always be changed to 0 or 10. The table below shows the effect of rounding on 2B counts with a value less than 10.

Count of	Will round to 0	Will round to 10
1	9 times out of 10	1 times out of 10
2	8 times out of 10	2 times out of 10
3	7 times out of 10	3 times out of 10
4	6 times out of 10	4 times out of 10
5	5 times out of 10	5 times out of 10
6	4 times out of 10	6 times out of 10
7	3 times out of 10	7 times out of 10
8	2 times out of 10	8 times out of 10
9	1 times out of 10	9 times out of 10
0	always	Never

The random rounding algorithm uses a random seed value to initiate the rounding pattern for tables. This method used to seed the pattern can result in the same count in the same table being rounded up in one execution and rounded down in the next.

The introduction of noise into the data is necessary to make a data table more secure. The effectiveness of rounding compared with other confidentiality techniques such as cell suppression is due to the fact that it largely “suppresses” small cells (by rounding them down to 0) without disrupting medium and large-sized cells. This trick is possible because rounding makes it impossible to determine the presence of units in cells rounded to 0. This is sufficient to make the great majority of rounded tables secure.

A table summarizing the history of random rounding practices for the census can be found in **Appendix A** at the end of this document.

2.3 Confidentiality Adjustment for Place of Work Counts

The place of work counts for census blocks are available on a custom basis. These counts will be adjusted to reinforce the confidential nature of the data. In fact, all census block counts for Employed Labour Force will be rounded to a base of 5. This adjustment, however, will be controlled. That is, aggregates of the adjusted population counts for dissemination areas will always be within 5 of the actual values.

The area suppression rule with a population size threshold of 100 does not apply for these counts.

2.4 Disclosure Avoidance for Statistics

Statistics (such as mean, standard error, sum, median, percentile, ratio or percentage) **are not** subject to random rounding. However, when shown in tabulations accompanying the counts used to calculate the statistic, their presence can result in disclosure of individuals. To prevent this, we use statistic suppression methods or special statistic calculations.

2.4.1 Statistic Suppression

The following three conditions will result in the suppression of statistics:

- (1) It is possible (mainly for cells with small counts) that quantitative values were imputed from a single donor record. For example, an income cell with three individual records may in fact be only one actual response to income and the other two income amounts were imputed from the first record. When this occurs, the income characteristics of a single individual could be disclosed if the mean and standard error statistics are produced. To prevent this, and more generally to prevent disseminating statistics based on a narrow range of values, all statistics of a cell are suppressed if the minimum and maximum of the income amounts in the cell are equal.
- (2) For all quantitative variables, all statistics are suppressed if the number of actual records used in the calculation (not rounded or weighted) is less than 4.

Note: The number of records used in the calculation is not necessarily the number of records in the cell but, rather, the number of records that are applicable or available to the calculation of the statistic in the cell.

- (3) For all quantitative variables, all statistics are suppressed if the sum of the weights is less than 10.

2.4.2 Special Statistic Calculations

- (1) The statistic value is never rounded, except for frequencies.
- (2) All statistics based on ranks (medians, percentiles) are calculated the usual way, that is, never rounded.
- (3) All dispersion statistics (standard error) are calculated the usual way, that is, never rounded.
- (4) When a sum is specified, if the program sums a **quantitative variable** such as a dollar amount, a number of weeks, a number of hours or an age, then the program multiplies the unrounded average by the rounded, weighted frequency. Otherwise, the program rounds the actual weighted sum.

When a division is specified (averages, percentages, ratios, etc...), the program must apply the point (4) to both numerator and denominator before it proceeds with the division.

Note: Statistics based on ranks like median and percentiles are always calculated via linear interpolations. That means that, for low count cells, these statistics are not reliable. That is the reason why no confidentiality measure is applied to them.

3. CONFIDENTIALITY PRACTICES

3.1 Confidentiality Adjustment for Population and Dwelling Counts

The population counts of small census blocks may be adjusted to reinforce the confidential nature of the data. In fact, all census blocks with a population count less than 15 will be rounded to a base of 5. This adjustment, however, will be controlled. That is, aggregates (or roll-ups) of the adjusted population counts for dissemination areas will always be within 5 of the actual values, and large census sub-divisions will not be affected. Total control is achieved at the census division level, which means that the population counts for all census divisions will remain unchanged.

3.2 Preventing Disclosure Through Content Planning

Prevention of direct or residual disclosure must also be addressed when determining product content. When assessing the potential for disclosure, a number of factors must be considered. The detail of individual variables, cross-classification of variables and the geographic level of the data will all contribute to the risk. For example, there may be no risk in producing a table containing information on households by number of rooms in the dwelling and detailed groupings of dwelling value showing various characteristics of the household members for large geographic areas. However, the risk of disclosure would increase for the lower levels of geography.

The most common method used for preventing disclosure is defining content that is appropriate for a given geographic level. Increasing population thresholds or applying manual suppression as needed are other methods that can be employed. Since these are typically product-specific requirements, they are not part of the automated suppression systems.

3.3 Census of Agriculture and Census of Population Linkage

Census of agriculture and census of population 2B (long form) data are matched using geographic information and the age and sex of farm operators. Match rates are about 95% and weighting is performed to account for non-matches. Data are available for all members of households where a farm operator resides.

Census of agriculture data include farm type, farm sales, area of crops and numbers of livestock while the census of population provides socioeconomic data, including education, income and occupation of families and household members. Pre-planned standard products are produced at the province level only. Some multi-dimensional tables are available in pre-planned products with the usual being 2-3 dimensions.

Custom products are available at sub-provincial levels based on aggregations of weighting areas, where 5,000 persons is the usual population threshold, although it can go as low as 2,500. The data are randomly rounded and low-bounded to ensure confidentiality. Suppressions are done manually if cells are below a specified size.

All verification of tabulations is done internally by census of agriculture staff. There is a group responsible for reviewing all AG-Pop data before release.

3.4 Public Use Microdata Files

The 2001 Public Use Microdata File (PUMF) product consists of a series of files based on different foci: individuals, families and households. Each file contains approximately 3% of the Canadian population.

Microdata files are unique among census products in that they give users access to non-aggregated (or non-totaled) data. This makes PUMF a powerful research tool. The files contain a large number of variables. The user can group and manipulate these variables to suit his/her own data and research requirements. Tabulations not included in other census products can be created, or relationships between variables can be analyzed using various analytical tools.

The Census Public Use Microdata Files (PUMF) provide quick access to a comprehensive social and economic database about Canada and its people. They consist of samples of anonymous responses to the completed questionnaire. The PUMF files contain statistical information about Canadians, the families and households to which they belong and the dwellings in which they live.

Statistics Canada has to protect the confidential information that it collects. Owing to the very nature of a microdata file, various measures are taken to fulfil this commitment. Data for small geographic areas are not available in these files. The user will find information only for selected census metropolitan areas, selected census sub-divisions, the provinces and the territories. The breakdown of some sensitive variables was reduced for the Atlantic region. Some of the values of sensitive variables were suppressed because their combination could have been used to identify a person, a family or a household. Also, the income variables were subjected to reduced low, and high, income limits.

3.5 Access to Census Databases

All census of population data reside in several databases on different computer platforms within Statistics Canada. The access to this information is controlled through the use of computer file access permissions that are administered and maintained by representatives from Census Operations Division. Access permissions and security procedures are regularly reviewed and monitored to maintain maximum confidentiality.

4. DATA QUALITY RULES

The following section describes the methods used to restrict the dissemination of census data of unacceptable quality.

4.1 Incompletely Enumerated Areas

Any dissemination area with a high level of non-enumeration due to non-response is removed from census tabulations. If all of the dissemination areas for a specific higher-level geography have high levels of non-enumeration, then the higher level area is also removed from the tabulation. This occurs most frequently at the census subdivision (CSD) level, where an entire Indian Reserve CSD does not respond. This may also occur at the census tract (CT) level.

Although there is no census data collected for non-response areas, the areas themselves are included as part of the standard geographic hierarchies on the census databases. Retrieval and tabulation software will retrieve these areas but with no data.

In the 2001 and previous censuses, some dissemination areas for Indian Reserves were not enumerated due to non-participation/co-operation. Non-response rules require these non-enumerated areas to be identified and removed from products. As well, higher-level geographic areas containing non-enumerated areas must be identified in the products.

4.2 Partially Enumerated Areas

Any geographic area that contains an incompletely enumerated area is considered a partially enumerated area. Partially enumerated areas are flagged to end users as containing non-response (or incompletely enumerated) areas. In print publications, this is done through the usage of symbols attached to the geographic name. In electronic products, a numeric flag is provided with the area identifier indicating inclusion of non-response areas.

4.3 Data Quality Suppression for Standard Geographic Areas

4.3.1 Global Response Rates

Global response rates are determined for each of the census geographic areas. These areas are flagged on the database according to the non-response rate. Geographic areas with a non-response rate higher than or equal to 25% are suppressed from tabulations. Geographic areas with a global non-response rate higher than or equal to 5% and lower than 25% have been broken into 2 categories for 2001 census and will be flagged according to the following ranges: falling between 5% and 10% and falling between 10% and 25%. These geographic areas are identified in tabulations but not suppressed.

Place of Work data is not released for areas where the global non-response rate is greater than or equal to 25% and the majority of the employed labour force live and work in the same area.

4.3.2 1996 Adjusted Population Flag

Users wishing to compare 2001 census data with those of other censuses should take into account that the boundaries of geographic areas may change from one census to another. In order to facilitate comparison, the 1996 counts are adjusted as needed to take into account boundary changes between the 1996 and 2001 censuses. The flag has also been used to refer to corrections to the 1996 counts. The flag is also used to identify areas that have been created since 1996, such as newly incorporated municipalities (census sub-divisions) and new designated places. However, most of these flags are the result of boundary changes.

4.4 Area Suppression for Income Characteristic Data

Area suppression, when applied for data quality purposes is used to replace all income characteristic data with zeroes for geographic areas with populations and number of private households below a specific threshold.

If a census tabulation contains any data showing individual, family or household income characteristics, then the following rule applies. The specified population size for all standard, custom and geo-coded geographic areas is **250**, and the specified size for number of private households is **40** (except for Place of Work tabulations). The threshold of 40 private households is based upon the fact that weighted data are being used. With the weighting factor for each household being 5, setting a threshold of 40 ensures that there will be a minimum of 8 households used in the calculation. Income characteristic data are zeroed out for areas below these thresholds.

4.5 Data Quality Indicators

Data quality indicators (commonly referred to as data quality flags) are attached to each standard geographic area disseminated. In the census database environments, the data quality indicators consist of a 5-byte numeric field.

The following table describes the data quality indicator field and its' contents for 2001. Note that a zero in any of the 5 bytes is the default for the respective indicator and means that no data quality action is required.

Byte	Byte Description	Byte Value	Value Description
1	Incomplete Enumeration	0	<default> – Fully enumerated area.
		1	Incompletely enumerated Indian Reserve or Indian Settlement (suppressed).
		2	Excludes census data for one or more incompletely enumerated Indian Reserves or Indian Settlements.
2	2A Data Quality	0	<default>

**DATA QUALITY AND CONFIDENTIALITY STANDARDS AND GUIDELINES
2001 CENSUS DISSEMINATION**

Byte	Byte Description	Byte Value	Value Description
		1	Data quality indexes showing a global 2A non-response rate higher than or equal to 5% and lower than 10%.
		2	Data quality indexes showing a global 2A non-response rate higher than or equal to 10% and lower than 25%.
		3	Data quality index showing a global 2A non-response rate higher than or equal to 25% (suppressed).
3	Population and Dwelling Counts Error Flag	0	<default>
		1	An error exists in the 2001 population and dwelling counts for this area.
4	2B Data Quality	0	<default>
		1	Data quality indexes showing a global 2B non-response rate higher than or equal to 5% and lower than 10%.
		2	Data quality indexes showing a global 2B non-response rate higher than or equal to 10% and lower than 25%.
		3	Data quality indexes showing a global 2B or 2A non-response rate higher than or equal to 25% (suppressed).
5	1996 Adjusted Population Flag	0	<default>
		1	1996 adjusted count; most of these are the result of boundary changes.

Note: The 2A and 2B data quality flags do not apply to the population and dwelling counts. Some of the flag descriptions have been modified for 2001 and a new range has been introduced for the non-response rates in the 2A and 2B data quality flags. The flag legend for historical census years can be found in **Appendix B** at the end of this document.

4.6 Other Methods of Data Quality Suppression

The methods of suppression mentioned to this point provide sufficient data quality suppression and identification for most census data products. However, in some products, the specifying area or production area may require that additional data quality suppression be performed. Examples of additional suppression could include increasing population thresholds or applying distribution or cell suppression. These are typically product-specific requirements and therefore are not part of the automated suppression systems. In all cases, some form of manual process is required.

4.6.1 Distribution Suppression

The most common example of other methods of data quality suppression is distribution suppression. This occurs in selected **standard income products** where income distributions are suppressed when the **total number of units** (persons, families, households) within the income distribution is less than **250**.

5. DATA SUPPRESSION – OTHER

Suppression of data also occurs if, and when certain census questions are not asked of all respondents. Additionally, suppression of data may be applied for product-specific reasons due, typically, to the size of the product and/or the constraints of the output format on which the product is being disseminated.

5.1 Indian Reserves – Citizenship and Immigration Data

Citizenship and immigration questions are not asked on Indian Reserves. Consequently, no data for these variables are collected or disseminated for Indian Reserves.

5.2 Incidence Reporting

Incidence reporting is a process used to order or rank characteristic data by size within products. It can be used as a method to select only the n highest categories of a characteristic for inclusion in a product.

5.3 Zero Suppress

Zero suppress refers to the removal of records in which all of the counts are equal to zero. This method is used to reduce the size of an output product by removing any rows of the output matrix where all data are equal to zero.

5.4 Place of Work Flow Suppression

Flow Suppression is a process used to remove records with very low counts from place of work flow tables. It can be used as a method to select only those commuting flows with a count greater than a threshold value for inclusion in a product. The default value is 20. These are typically product-specific requirements and therefore are not part of the automated suppression systems.

**DATA QUALITY AND CONFIDENTIALITY STANDARDS AND GUIDELINES
2001 CENSUS DISSEMINATION**

6. APPENDIX A – Summary Table of Confidentiality Rules (Historical)

1981-1996

Census Years ▶ Confidentiality Rules ▼	1981	1986	1991	1996
Random rounding	Rounding up or down to base 5	Rounding up or down to base 5.	For 100%, rounding to base 5. For 20%, rounding to base 10 for weighted frequencies less than 10. Weighted frequencies over 10, rounding to base 5, (custom tabulations only).	For 100%, rounding to base 5. For 20%, rounding to base 10 for weighted frequencies less than 10. Weighted frequencies over 10, rounding to base 5, (all tabulations).
Area suppression	Geographic areas < 50 for self-enumeration areas were suppressed. Geographic areas < 25 for canvasser areas were suppressed. For products containing income characteristics, distributions or quantitative statistics, the threshold was 250.	Standard geographic areas with population count < 40 were suppressed. Non-standard geographic areas with population count < 100 were suppressed. For products containing income characteristics, distributions or quantitative statistics, the threshold was 250.	All standard and non-standard geographic areas with population count < 40 were suppressed. For 100% data, the decision was based on population count. For 20% data, the decision was based on non-institutional population estimates. For products containing income characteristics, distributions or quantitative statistics, the threshold was 250.	All standard and non-standard geographic areas with population count < 40 were suppressed. For 100% data, the decision was based on population count. For 20% data, the number used was the lower of the 2A non-institutional population count or the 2B non-institutional population estimates. For products containing POW geographic areas, population counts < 40 were suppressed. For POW flow tables, both residence and work suppression rules applied. For products containing 6-digit postal code the threshold was 100.

**DATA QUALITY AND CONFIDENTIALITY STANDARDS AND GUIDELINES
2001 CENSUS DISSEMINATION**

Census Years ▶ Confidentiality Rules ▼	1981	1986	1991	1996
Area suppression	<p>Native's Package, self-enumerated areas with native population count < 500 were suppressed; canvasser areas with native population count < 100 were suppressed.</p> <p>All suppressed data were included in higher geographic levels.</p>	<p>All suppressed data were included in higher geographic levels.</p>	<p>All suppressed data were included in higher geographic levels.</p>	<p>For products containing income characteristics, distributions or quantitative statistics, the threshold was 250.</p> <p>Suppression of data also occurs if and when certain questions are not asked of all respondents, such as on Indian Reserves.</p> <p>All suppressed data were included in higher geographic levels.</p>
Cell suppression	<p>Custom tables involving Industry, Occupation or Income, all cells < 25 were suppressed (set to 0).</p>	<p>Custom tables involving Industry, Occupation or Income, all cells < 10 were suppressed (set to 0).</p>	<p>Income statistics based on cells < 10 were suppressed.</p>	<p>The standard deviation rule was established. If the standard deviation was equal to 0 indicating that all quantitative values were equal, then the statistic was suppressed.</p> <p>For all quantitative variables, statistics are suppressed if the associated unrounded count used in calculating the statistic is less than the random rounding threshold which is 10 for all 2B products.</p>

**DATA QUALITY AND CONFIDENTIALITY STANDARDS AND GUIDELINES
2001 CENSUS DISSEMINATION**

7. APPENDIX B – Summary Table of Data Quality Flags (Historical)

The following table describes the data quality indicator fields and their contents as they apply to census years prior to 2001. Note that a zero in any of the 5 bytes is the default for the respective indicator and means that no data quality action is required.

Byte	Byte Description	Byte Value	Value Description
1	Incomplete Enumeration	0	<default> – Fully enumerated area.
		1	Incompletely enumerated Indian Reserve or Settlement (suppressed).
		2	Excludes census data for one or more incompletely enumerated Indian Reserves or Indian Settlements.
2	2A Data Quality	0	<default>
		1	Data quality indexes showing a global non-response rate higher than or equal to 5% and lower than 25%.
		2	Data quality indexes showing a global non-response rate higher than or equal to 25% (suppressed).
3	Boundary Error	0	<default>
		1	Areas where boundaries were found to be in error subsequent to collection.
4	2B Data Quality	0	<default>
		1	Data quality indexes showing a global non-response rate higher than or equal to 5% and lower than 25%.
		2	Data quality indexes showing global non-response rate higher than or equal to 25% (suppressed).
		3	Employee resident error rate is higher than or equal to 5% and lower than 25% and global non-response rate is less than 5%.
		4	Both the employee resident error rate and global non-response rate are higher than or equal to 5% and lower than 25%.
5	1996 Adjusted Land Area	0	No <default>
		1	Yes