

1986 - 1991 ENUMERATION AREA CORRESPONDENCE FILE

DATA QUALITY STATEMENT

AND

RECORD LAYOUT

April 1992

Table of Contents

1.0 Introduction	2
2.0 Data Quality	2
2.1 Lineage	2
2.1.1 Large Urban Centres.....	3
2.1.2 Areas Not Covered by the Street Network Files (SNF).....	3
2.2 Attribute Accuracy.....	5
2.3 Logical Consistency	5
2.4 Completeness.....	5
3.0 File Record Layout	6
Appendix 1.....	9

1.0 Introduction

A digital file containing the relationship of the 1986 enumeration area structure to the 1991 enumeration area structure has been created as a result of the 1991 Census.

The relationship between the 1986 enumeration area structure and 1991 enumeration area structure is determined by the spatial coverage of the two structures. The 1986 enumeration areas that cover the same area as a 1991 enumeration area creates a relationship between the two structures.

The 1991 Enumeration Area Correspondence File was used to create the *Equivalent Enumeration Areas, 1991 and 1986* of the Enumeration Area Reference Lists series (Statistics Canada Catalogue No. 92-318). Appendix 1 describes the verification methodology used for the tables found in that publication.

2.0 Data Quality

The purpose of a data quality statement is to provide detailed information to users to evaluate the suitability of the data for a particular use. This report documents the data quality of the 1986 - 1991 Enumeration Area Correspondence File according to four fundamental components: lineage; logical consistency; completeness; and attribute accuracy.

This report is based on the best information available at the time of its release. It in no way constitutes a warranty of the data in the event that characteristics are observed that deviate from those stated here.

2.1 Lineage

Lineage includes descriptions of the source material from which the data were derived and the methods of derivation, including the dates of the source material and all transformations involved in producing the final digital files.

The 1986 - 1991 Enumeration Area Correspondence File was created over a period of three years. Preliminary 1991 enumeration areas were created and their relationship to the 1986 enumeration area structure were determined. These data were subsequently updated as changes to the boundaries occurred prior to the final 1991 enumeration area structure.

The 1986 - 1991 Enumeration Area Correspondence File was created using the two different methodologies described below.

2.1.1 Large Urban Centres

In most large urban centres and a few small urban centres in Canada, Statistics Canada maintains digital street network files. These files contain streets, railways, hydrography, and other physical features, as well as municipal boundaries and other relevant information such as feature names and address ranges. The street network files were originally digitized from maps at various scales ranging from 1:2400 to 1:50000. Street Network Files (SNFs) cover more than 60% of the population, but less than 1% of the total land area.

Of the 45,995 enumeration areas in 1991, 24,399 (53.05%) of these are found in Street Network File areas. Of these 24,399 enumeration areas, approximately 10,000 were in Street Network File areas where the 1991 enumeration areas were created using an ARC/INFO® based Computer-Assisted Districting software package developed within the Geography Division. The preliminary creation of the enumeration areas covering these areas was by computer assistance. Following their initial creation, enumeration areas in these areas underwent manual changes as modifications to higher-level boundaries occurred.

In areas that underwent enumeration area creation through Computer Assisted Districting, the initial 1986 - 1991 enumeration area correspondence was created by the program. A Geography Division version of 1986 enumeration area digital boundaries was utilized in this process. As subsequent changes to the enumeration area structure occurred, their impact on the 1986 - 1991 enumeration area correspondence was evaluated manually. This evaluation was performed using the 1986 enumeration area collection maps.

Approximately 16,000 of the 1991 enumeration areas in street network file areas not created using computer assistance had their 1986 - 1991 enumeration area correspondence determined manually. This manual process was identical to the manual evaluation mentioned above.

2.1.2 Areas Not Covered by the Street Network Files (SNF)

In those areas not covered by the digital street network files, the 1986 - 1991 enumeration area correspondence was determined manually. As enumeration areas were delineated, the 1986 - 1991 enumeration area correspondence was determined by examination of the 1991 enumeration area limits and the 1986 enumeration area limits. The 1991 enumeration areas in these areas were delineated on compilation maps that were copies of the 1986 enumeration area reference maps. Therefore, the 1986 enumeration area limits were underlying the 1991 enumeration area limits.

2.2 Attribute Accuracy

Attribute Accuracy refers to the accuracy of the relationship of the 1986 to 1991 Enumeration Area structure as provided in the file.

To determine the accuracy of the 1986 - 1991 enumeration area correspondence on this file, a random stratified sample of 1991 enumeration areas was drawn from the file. The sample contained a total of 1,121 1991 enumeration areas and all 1986 enumeration areas corresponding to them on the file. The three strata analyzed and the number of 1991 enumeration areas in each were:

- 1) Enumeration areas in areas covered by the SNF where initial Enumeration Area creation utilized the Computer Assisted Districting Program (371);
- 2) Enumeration areas in areas covered by the SNF where enumeration area creation was performed manually (373); and
- 3) Enumeration areas not covered by SNF areas (377).

The sample was drawn such that the distribution of enumeration areas in each of the strata reflected the distribution of all enumeration areas by regions of the country.¹

The relationship between a 1991 enumeration area and the 1986 enumeration area structure was deemed to be in error if any of the 1986 enumeration area(s) covering the 1991 enumeration area were not identified correctly on the file. This incorrect identification included a 1986 enumeration area missing from the relationship, a 1986 enumeration area erroneously linked to the 1991 enumeration area, or the flag² indicating a part or full 1986 enumeration area relation improperly set.

Upon completion of the sample, all errors identified were corrected on the file.

Table 1 presents the accuracy estimates based on the sample.

Table 1: Estimated Accuracy Rates, Variance, and Standard Error by Stratum

Stratum	Sample Size	Accuracy Estimate	Standard Error
Non-SNF Area	377	98.939%	0.524%

¹For this study the country was divided into six regions: Maritimes; Quebec; Ontario; Prairies; British Columbia; and the Territories.

²The file contains a 1986 Enumeration Area Part Flag. This flag indicates whether the whole of the 1986 Enumeration Area is related to the 1991 Enumeration Area, or if only part of the 1986 Enumeration Area is related.

SNF Areas, Manual Enumeration Area Creation	373	93.298%	1.278%
SNF Areas, Computer-Assisted Districting	371	97.035%	0.867%
Canada		96.854%	0.004%

The table indicates that in 1000 correspondences between 1986 and 1991 enumeration areas, the number of expected errors is 32. The standard error is used in determining the confidence of the expected error rate. With 95% confidence, the errors in 1000 correspondences is between 24 and 39.

2.3 Logical Consistency

Logical Consistency describes the fidelity of relationships encoded in the data structure of the digital spatial data.

The file underwent two automated logical consistency checks to further ensure its data quality. The first check ensured that all 1986 and 1991 enumeration areas resided on the file. The second check ensured that the 1986 enumeration area part flag was set correctly based on the data within the file.

These two checks were performed on the file prior to the sample being drawn to measure the attribute accuracy. Following the correction of errors found during the sample, these two checks were performed again. The checks indicated that corrections resulting from the errors found in the sample caused logical inconsistencies to develop. These inconsistencies were due to errors in other records. These new errors were manually resolved through a review of the 1986 and 1991 enumeration area structures for the records in question. The errors were corrected, the two checks were performed again, and no logical inconsistencies were found within the file.

2.4 Completeness

Completeness contains information about selection criteria and definitions used.

The 1986 - 1991 Enumeration Area Correspondence File contains all 1986 and 1991 enumeration areas.

3.0 File Record Layout

The record layout of the 1986 - 1991 Enumeration Area Correspondence File is found in Table 2.

Table 2: Record Layout of 1986 - 1991 Enumeration Area Correspondence File

67,417 Records			CESP.CGDB91.EA8691.FINAL.DATA			
Field	Size	Start	End	Type	Description	
1	2	1	2	Numeric	Province Code	
2	3	3	5	Numeric	1991 Federal Electoral District Code	
3	3	6	8	Numeric	1991 Enumeration Area Code	
4	3	9	11	Numeric	1986 Federal Electoral District Code	
5	3	12	14	Numeric	1986 Enumeration Area Code	
6	1	15		Character	1986 Part Flag	
					Value	Definition
					*	part of 1986 EA
					blank	complete 1986 EA

Figure 1 is presented to provide an example of how to read the file.

Figure 1: Example of File Layout

Record	Province Code	1991 Federal Electoral District Code	1991 Enumeration Area Code	1986 Federal Electoral District Code	1986 Enumeration Area Code	1986 Part Flag
1	10	001	001	006	005	*
2	10	001	002	001	003	*
3	10	001	002	001	004	*
4	10	001	003	001	021	*
5	10	001	004	001	010	*
6	10	001	005	001	009	
7	10	001	006	001	007	*
8	10	001	007	001	008	
9	10	001	008	001	005	*
10	10	001	008	001	006	
11	10	001	008	001	007	*
12	10	001	008	001	012	*
13	10	001	008	001	013	*

Record 1 indicates that the 1991 enumeration area 10-001-001 corresponds to the 1986 enumeration area 10-006-005, and that no other correspondence exists. The asterisk in the last field indicates that only part of the 1986 enumeration area corresponds to the 1991 enumeration area. Therefore, the 1991 enumeration area 10-001-001 corresponds only to a part of the 1986 enumeration area 10-006-005.

Record 2 and 3 indicate that the 1991 enumeration area 10-001-002 corresponds to 1986 enumeration areas 10-001-003 and 10-001-004. Again, the asterisks in field 6 of these records indicate that only a part of the 1986 enumeration areas correspond to the 1991 enumeration area. Therefore, the 1991 enumeration area 10-001-002 corresponds to parts of the 1986 enumeration areas 10-001-003 and 10-001-004.

Record 6 indicates that the 1991 enumeration area 10-001-005 corresponds to the 1986 enumeration area 10-001-009. Since this record does not have an asterisk in field 6, the complete 1986 enumeration

area is within the 1991 enumeration area. The file does not have any other record for the 1991 enumeration area 10-001-005, therefore this 1991 enumeration area is equal to the 1986 enumeration area 10-001-009.

Records 7 and 11 both concern the 1986 enumeration area 10-001-007. Both records indicate that this 1986 enumeration area is only partly related to the 1991 enumeration area on the records.

APPENDIX 1

Verification methodology used for the tables in the *Equivalent Enumeration Areas, 1991 and 1986* (Catalogue No. 92-318)

Tables 1 and 2 were verified using one-to-one 1986 and 1991 listings taken from the EACORRES file of the Census Geographic Data Base. All of the one-to-one corresponding enumeration areas (1991 to 1986 and 1986 to 1991) in Prince Edward Island, the Northwest Territories and the Yukon Territory were verified. The beginning and end of every column in tables 1 and 2 for the remaining provinces were also verified. This verification confirmed that all one-to-one enumeration areas were present in the tables.

Table 3 was verified using manually-created verification tables. Listings for each province and territory that flagged enumeration areas with one-to-many and many-to-many correspondences were provided from the EACORRES file. From these listings, sorted by 1986 and 1991 enumeration area code, the manual verification tables were developed. All groups of enumeration areas for Prince Edward Island, the Northwest Territories and the Yukon Territory were verified. A table showing 138 groups for New Brunswick was also created. For the remaining provinces, verification tables of between 15 and 28 groups of corresponding enumeration areas were created.

The manual tables were developed so that, at least, a group of equivalent areas represented the following concordances: one 1991 enumeration area-to-many 1986 enumeration areas; many 1991 enumeration areas-to-one 1986 enumeration area; and many 1991 enumeration areas-to-many 1986 enumeration areas. The only exception to this was Manitoba, where no one 1986 enumeration area-to-many 1991 enumeration areas was found in the first 25 groups of equivalent enumeration areas.

The reference numbering system developed to index 1991 and 1986 enumeration areas was verified by spot-checking groups of enumeration areas to ensure that all the enumeration areas were listed with the correct reference number. The reverse procedure of reference number by enumeration areas was also performed.

The tables were approved after the final verification had been completed. The verification process itself for table 3 was done in stages for Prince Edward Island, until the programming was proven correct. After the final assembly of table 3, the entire verification process described above was followed. No errors were found during the final verification process to the target equivalent groups of enumeration areas.