Geography Working Paper Series
No. 2000-4

Introducing the Dissemination Area for the 2001 Census: an Update

by

Henry Puderer

Geography Division
Statistics Canada

June 2001

Catalogue no. 92F0138MIE, no. 2000-4 (updated edition)

ISSN 1481-174X
ISBN 0-662-30789-5

Aussi disponible en français.

Geography Working Paper Series

The Geography Working Paper Series is intended to stimulate discussion on a variety of topics covering conceptual, methodological or technical work to support the development and dissemination of the division's data, products and services. Readers of the series are encouraged to contact the Geography Division with comments and suggestions. A list of titles appears at the end of this paper.

A paper version, catalogue no. 92F0138MPE, is available for $10.00 per issue. There is a shipping charge per issue for delivery outside Canada: CDN $6.00 to the United States and CDN $10.00 to other countries. All prices exclude sales tax. There is no charge for downloading the document yourself on the Internet (*http://www.statcan.ca*).

For inquiries about the working paper series, please contact:

Geography Division
Statistics Canada
Jean Talon Building, 3rd floor
Ottawa, Ontario K1A 0T6

Telephone: (613) 951-3889
FAX: (613) 951-0569
Internet: geohelp@statcan.ca

TABLE OF CONTENTS

ABSTRACT ii

1. INTRODUCTION 1
2. DISSEMINATION AREA DESIGN CRITERIA 1
   2.1 Temporal Stability 2
   2.2 Minimum Population (Reduced Area Suppression) 3
   2.3 Maximum Population (Uniformity) 4
   2.4 Intuitive (Visible) Boundaries 5
   2.5 Compact Shape 6
   2.6 Homogeneity 6
3. OPERATIONAL FACTORS 7
   3.1 Dissemination Area Code Structure 7
   3.2 Maximum Two-digit Block Code 7
   3.3 Production Time-line 8
4. FEEDBACK 8

ACKNOWLEDGEMENTS 9
REFERENCES 9

Appendix 1 - The Block 10
   Map 1. Blocks in an urban area for the 2001 Census 11
   Map 2. Blocks in a rural area for the 2001 Census 11

ABSTRACT

Up to and including the 1996 Census, the enumeration area (EA) served as both the primary collection area and the basic dissemination area. Optimising both the collection and dissemination functions was a difficult undertaking and compromises were made. For the 2001 Census, the new digital cartographic file called the National Geographic Base (NGB) and the block program have made it possible to create separate enumeration areas for data collection and dissemination areas for data dissemination. This working paper describes the design criteria that have been specified for the implementation of the dissemination area (DA). These criteria address the improvements most frequently requested by users such as increased temporal stability, reduced area suppression, intuitive boundaries, compactness and homogeneity. In addition to integrating and reflecting the recommendations from the user community, the dissemination area design incorporates operational factors related to the DA code structure, maximum code size and production timeframes.

Note: This working paper was first released in March 2000 and described the design criteria and their parameters that were being considered for creating the dissemination area (DA) for the 2001 Census. Since that time, the design criteria have been finalised. New information is highlighted as updates in this version of the paper (June 2001). The remainder of the paper is the same as the version released in March 2000.

1. INTRODUCTION

Up to and including the 1996 Census, the **enumeration area** (EA) served as both the primary collection area and the basic dissemination area. Optimising both the collection and dissemination functions was a difficult undertaking and compromise eroded the optimum for both. As a consequence, there have been long-standing requests from the user community to make modifications to the enumeration area that would improve it as a dissemination area (Statistics Canada, 1994, 1999). The most frequently requested improvements were for increased temporal stability, reduced area suppression, intuitive boundaries, compactness and homogeneity. It was never possible to make these modifications to the EA without seriously compromising its function as a collection unit. Nor was it possible to produce a separate dissemination area at a cost that was not prohibitive. But the situation has changed as we prepare for the 2001 Census.

First, a **National Geographic Base** (NGB), which is a digital cartographic base for all of Canada, has been created.[1] The availability of a digital base means that software can be developed to economically automate the delineation of a dissemination area designed to address the above noted users' requests.

Second, the 2001 Census will see the introduction of the **block** program. Essentially, blocks are the polygons formed by the intersection of streets, although these morphological blocks are split when the boundaries of selected standard geographic areas (like municipalities) or collection requirements dictate. At the time of enumeration, all dwellings will be geographically referenced to specific blocks, created prior to Census day. The availability of the block is critical. The design of the collection geography can be largely separate from the dissemination geography because these blocks can be aggregated one way for collection and another way for dissemination.[2]

Because it is now both technically and economically possible, the decision has been made to create the **dissemination area (DA)**, a standard geographic area designed specifically for data output, to replace the enumeration area for dissemination. Dissemination area delineation will be accomplished using a system developed in-house called Generalised Area Delineation System (GArDS). GArDS aggregates small geographic areas (in this case, blocks) according to a set of delineating or design criteria to produce a set of the desired geographic areas (in this case, dissemination areas) in a larger geographic work area called the largest delineation zone (LDZ). The design criteria are assigned penalty weights. The solution that is accepted is the one that has the lowest total penalty weight, which is an aggregate of the penalty weights for all of the combined criteria for all geographic areas within the LDZ. The rate at which the penalty weight increases can also be controlled so the impact of the different criteria can be modified according to their assessed importance to the desired solution. This working paper describes the design criteria that have been specified for the implementation of the dissemination area.

Note: Since this working paper was first released in March 2000, the design criteria have been finalised. New information is highlighted as **updates** in this version of the paper (June 2001).

---

[1] The National Geographic Base (NGB) is a digital base being created in partnership with Elections Canada. Source files for the NGB include the 1996 Census Street Network Files (SNF) produced by the Geography Division, Statistics Canada; Natural Resources Canada's digital coverage (NTDB) and the Digital Chart of the World (DCW) coverage for the northern parts of Canada, plus digital files from Elections Canada (NTDB with names tagged to streets).

[2] The block is also a new geographic area for dissemination of the 2001 Census. Current plans envision releasing population and dwelling counts by block but no characteristic data. The smallest geographic area for which characteristic data will be released is the dissemination area. For more details about the block program see Appendix 1.

2. DISSEMINATION AREA DESIGN CRITERIA

On the basis of feedback from the user community, six criteria were considered in the design of the DA. In order of priority, these are:

- Temporal stability
- Reduced area suppression
- Uniformity
- Intuitive boundaries
- Compact shape
- Homogeneity.

This does not mean that the DA will be perfect in all respects because of the many trade-off conditions among the design criteria. However, what can be realised is an improvement in the characteristics of the DA relative to those of the EA, which in turn will improve the usability of the census data that are subsequently made available. What follows is a discussion of the rationale for the individual criteria and the trade-off process that was an inherent part of the selection.

2.1 Temporal Stability

The boundaries of enumeration areas change from one census to another because of changes to the boundaries of standard geographic areas, the amount of population growth and changes to field collection specifications.[3] The degree of change varies from census to census.

Users often see these changes as a frustrating inconvenience. They can confound longitudinal analysis completely or force difficult and sometimes costly adjustments to the area of study in order to get comparable data across censuses.

Because instability in the EA was in large measure a consequence of the requirement for the EA to respect changes to the boundaries of standard geographic areas, it would be desirable, from a temporal stability perspective, to have the DA totally independent of any changes to boundaries of standard areas. This option could be considered because the block could now provide the link between data collection areas and the standard geographic areas for data dissemination.

However, user feedback had indicated that for maximum utility, dissemination areas should not straddle municipal (CSD) boundaries. A DA that respects the boundaries of CSDs would be especially useful where there is no census tract (CT) program. A DA that respects the boundaries of census tracts increases the utility of the CT program by providing the capability to drill down further to analyse intra-CT distributions. Therefore, after considering various alternatives that represented trade-offs between DA stability and utility, the approach chosen will require the DA to respect the boundaries of census subdivisions (CSDs) and census tracts (CTs).

What is the impact of this approach on DA stability? Census tract boundary revisions rarely occur and only when essential. Therefore, respecting the boundaries of CTs has minimal impact on the stability of the DA. However, future CT splits and CT creation or changes to CT boundaries – usually on the periphery of a census metropolitan area (CMA) or census agglomeration (CA) – will result in DA changes. However, these changes are localised and once put into effect, stability returns to the local area as the zone of change pushes outward. Similarly, change is anticipated where a CSD is incorporated or where a CSD boundary changes as a consequence of an annexation. Again, the expectation is that the change

[3] In order to aggregate census data to the standard geographic areas for dissemination, the EA must respect the boundaries of the standard geographic areas. These boundaries are subject to change from census to census. For example, between censuses, municipal boundaries change as municipalities are created or dissolved or annex parts of adjacent municipalities (municipalities are a major component of the census subdivision (CSD) standard geographic area). Historically we could expect 20% of the 6,000 CSDs to change boundaries from one census to another. The federal electoral district (FED) is another standard geographic area that EAs must respect. When new FED Representation Orders come into effect, 20% of the previous census EAs are altered. Population growth and the associated dwelling growth, changes to the road network, and changes to field collection specifications, such as the maximum number of dwellings per EA, also force changes to the EA from one census to the next.

will be localised and a one-time change. In general, where growth occurs, DA splits are likely, as in the case of the CT program. If decline should occur, aggregations could be considered, especially with respect to the minimum population threshold criterion. Minor boundary changes to either a CT or a CSD would in most instances result in similar changes to the affected DA, whereas significant changes (for example, a census agglomeration entering the census tract program) could result in a complete redo for the area affected.

There are two aspects to the temporal stability criterion. The above has discussed the forward part – building a solid base for the DA starting with the 2001 Census and maximising stability for the 2006 Census and beyond. But what about the reverse aspect, namely, retaining comparability to the 1996 Census EA? During consultations, opinion on this particular item was more divided than on any of the other criteria. One group clearly wished to see comparability to the 1996 Census EA as a design criterion while another group favoured dropping this historic baggage. In the end, the latter view prevailed because preserving comparability also meant preserving sub-optimal DAs. It would also constrain the creation of new DAs, thereby creating an environment supporting the generation of more sub-optimal DAs.

The **temporal stability criteria** have been implemented as described above. The DA will be delineated respecting the limits of census tracts (CT) and census subdivisions (CSD). (CSDs for the most part comprise cities, towns, villages, rural municipalities and Indian reserves and settlements.) The 1996 enumeration areas (EA) will not directly influence the delineation of the dissemination area.

Given the decision regarding 1996 EAs, attention is now focusing on how best to bridge the expected gap for data dissemination between the 1996 EAs and the 2001 DAs. The level of EA to DA concordance will not be known until after June 2001 when the delineation of DAs is expected to be complete. At that time, the situation will be assessed to determine if it is reasonable to produce an EA to DA concordance file or whether an alternative to this product is required to meet data user needs. Dissemination proposals will be presented for public discussion by fall 2001.

2.2 Minimum Population (Reduced Area Suppression)

The criteria for delineating enumeration areas for census collection conflict with the techniques used to protect the confidentiality of individual responses for dissemination. One technique called area suppression results in the deletion of all census characteristic data for geographic areas with population counts fewer than 40 (fewer than 250 in the case of income data). [4]

The application of area suppression rules to the 49,361 enumeration areas (EAs) for the 1996 Census meant that income data were suppressed for 13,085 EAs (27%) with fewer than 250 people, while other characteristic data were suppressed for 4,765 EAs (10%) with fewer than 40 people. In fact, 2,278 (5%) of the 49,361 EAs in 1996 had zero population. The need to reduce such high levels of area suppression for the smallest standard geographic area/dissemination unit is axiomatic. In the past it was not attainable given the dual roles of the enumeration area. However, it is largely attainable as part of the dissemination area design.

A minimum population of 500 has been specified as the target count for the dissemination area. A penalty weight is applied to the automated solution if the population count for the DA falls below this target value. The value of the weight increases as the difference between the solution and the target count increases. The rate at which the penalty weight increases can also be controlled so the impact of the different criteria can be modified according to their assessed importance to the DA solution. The minimum population criterion has been assigned the highest penalty weights. Nonetheless, DAs with population counts below 500 will still be delineated. The DA solution that is accepted is the one that has the lowest total penalty weight, which is an aggregate of the penalty weights for all of the combined criteria for all DAs within the largest delineation zone (LDZ).[5, 6] However, given that the minimum population criterion has the largest

[4] The other technique used is random rounding. Under this method all figures, including totals, that are greater than 10 are randomly rounded either up or down to a multiple of "5" and figures that are 10 or less are randomly rounded to "0" or "10".

[5] The LDZ is the geographic work area for which a solution is being sought. For the DA, it is defined by the intersection of census subdivisions (CSDs) and census tracts (CTs).

penalty weight, the DA in most instances will not have population counts significantly below the target value and certainly DAs below 250 will be rare, subject to two caveats.

The first caveat is that DAs are required to respect the boundaries of CSDs and CTs (because of the “temporal stability” criterion described above). Therefore, the automated system will produce solutions for largest delineation zones defined by the intersection of CSDs and CTs. There will be a number of individual CSDs and CTs that fall below the target population. For instance, using 1996 Census results, there were 2,329 CSDs and 94 CTs with a population less than 500. The number of CSDs and CTs with fewer than 250 people is 1,449 and 69 respectively and the number with below 40 people are 637 and 40 respectively. As well, there are CTs that straddle CSDs, which will result in LDZs with population counts below the target 500. There are 256 CTs that straddle CSD boundaries. Assume that something like 50,000 DAs will be created for the 2001 Census and that all of the 256 straddling CTs create LDZs of fewer than 40 people. Then we can expect that at least 6% of the DAs will have fewer than 500 population, 4% will have fewer than 250 and 2% will have fewer than 40. Although not perfect, it will be a significant improvement compared to enumeration areas.

The second caveat relates to the source of the population data that will be used in the DA delineation. We would like to use the 2001 Census population counts by block. However, the date when the final set of DAs is required may preclude use of the 2001 Census data. This means the DA delineation will have to use population estimates by block. The methodology for this aspect of the DA creation has not been investigated sufficiently to describe either the process or the anticipated quality. Nonetheless, we envision distributing 1996 EA counts across the corresponding blocks with some adjustment for growth. As a consequence of the associated error with this process, we expect cases where the DA population count is lower than the target 500. However we also expect the estimates to be sufficiently robust to prevent the creation of DAs with a population count below 250.

2.3 Maximum Population (Uniformity)

The purpose of this criterion is to keep the DA population size uniform. This criterion will have the benefit of generating more DAs within the LDZ. This will help when trying to fit DAs to other geographic areas of interest. Without an explicit maximum population criterion, the delineation would tend more towards larger and fewer DAs within the LDZ.

The maximum population target is also 500. It is not as critical as the minimum criterion since it does not affect the availability of data. Therefore, the penalty scores assigned to the maximum population target are less than those assigned to the minimum target. Consequently, the overall tendency will be to produce DAs greater than the target of 500, but to constrain the solutions so that DAs that significantly exceed the target are not created. As with the minimum population criterion, the use of population estimates will affect the final results but this is not considered too serious in this instance.[7]

During the deliberations concerning the minimum and maximum population criteria, a related issue was brought into discussion. The issue was of the method of selecting population targets and data reliability considerations. In general, because data reliability decreases with cell size, the objective is to try to have cell sizes above some minimum size. Consequently, from the perspective of data reliability, the larger the minimum threshold, the better. However, increasing the minimum threshold beyond a certain level runs counter to the objectives of a basic standard geographic dissemination area, so some maximum limit is required. Table design is also a critical component when addressing data reliability. In the end, trying to generate uniformly sized DAs (in terms of population size) was seen as the goal since the resulting

---

[6] This approach does not directly take into account the spread of the individual scores for each DA within the LDZ. As a consequence, use of the overall solution with the lowest aggregate penalty score within the LDZ could result in a poor DA while generating optimal solutions for the other DAs within the LDZ. An alternative approach could ensure that individual DA scores are considered so that the creation of a poor DA is prevented, but this would cause slightly poorer quality solutions for the other DAs within the LDZ, which means the lowest aggregate penalty score solution would not be selected. Whether the second approach is better than the current one being programmed needs to be assessed during the testing phase.

[7] Low population DAs will result where the distribution of the 1996 EA population by block is inaccurate and high population DAs will result where the growth adjustment is not accurate.

consistency would benefit in the design of optimum output tables. The target of 500 was seen as a good all round choice with respect to both area suppression and data reliability concerns.

Minimum Population (Reduced Area Suppression) and Maximum Population (Uniformity)

The initial approach was to have two criteria—one for the minimum and one for the maximum population—both set at 500. Testing revealed that this approach was too restrictive and gave so much weight to the population variable that the other design criteria were almost excluded. Any deviation from the population target resulted immediately in the assignment of a penalty score. The penalty score increased with the distance from the 500 target and consequently the population variables soon became the over-riding design criteria.

The solution was to reformulate the two variables into one population variable in the form of a population range of 400 to 700. This flexibility in the population target resulted in far better results as the other design criteria began to make an impact on the outcomes. Subsequent testing demonstrated that an overall population average of about 550 was attained for the DA. Individual DAs are highly clustered in the 400 to 700 range (75%), but there are instances where DAs have values outside of this range (10% are below and 15% are above).

Sub-blocks

Since release of the March 2000 working paper, a number of questions have been posed about the treatment of high-rise or townhouse complexes that are internal to a morphological block—that is, they don't form a physical block and are often represented by a single address. These situations form logical blocks called **sub-blocks**. Where the population of a sub-block equals or exceeds 300, a separate DA is formed; otherwise, the sub-block is considered part of the morphological block for the purpose of DA creation.

2.4 Intuitive (Visible) Boundaries

Enumeration areas were required to respect all the boundaries of the standard geographic areas in order to aggregate the data to these areas for dissemination. On occasion these standard geographic areas had "invisible" boundaries, i.e., some or all of the boundary did not follow any visible feature on the ground (such as a street, railway, water feature, power transmission line, etc.). As a consequence, the EA boundary also had an invisible component. Furthermore, in the absence of suitable visible features, invisible EA boundaries were sometimes deliberately created in order to limit the size of the EA to reflect collection workload specifications. EA boundaries that follow invisible features can make it difficult for users to determine the geographic coverage of an EA and the associated uncertainty can raise questions about the quality of the data application. Consequently, the user community has requested that only visible and easily identifiable boundaries be used to delineate standard geographic areas – what generically have been called "intuitive" boundaries.

Because DAs are aggregations of blocks, the intuitive boundary criterion is in most instances assured because blocks respect roads. However, blocks are also required to follow the boundaries of selected standard geographic areas and any additional limits required by the collection operation to constrain workloads. As described above, some of these boundaries (or at least sections) are invisible. Therefore, the task for the DA creation is to apply significant penalty scores to block boundaries that are invisible. Again, as is the case with the other criteria, this does not prevent the creation of DAs with invisible boundaries (or at least partially invisible boundaries), but it significantly limits the likelihood of such outcomes. The exception condition to the preceding statement is the fact that the DA respects the boundaries of all census subdivisions (CSDs) and census tracts (CTs) and therefore will have invisible boundaries to the extent that the CSDs and CTs do.

For the most part, dissemination areas will have **intuitive (visible) boundaries** because they follow the roads that bound the blocks from which DAs are built. The roads include highways, regional roads, hard and loose surface roads, dry weather roads and bridges. In some instances where more substantive roads are not available for block determination, footbridges, cart tracks and trails may be used. Nationally, the road vintage is up-to-date at least to fall, 1997. Selected high growth census subdivisions may have road networks updated from fall 1999 to spring 2000.

Since the March 2000 release of this paper there have been a number of inquiries as to whether or not road type had a greater influence on the DA limits than outlined in the above paragraph. In the design implemented for 2001, there are no explicit controls on the type of roads that form a DA limit or that may be located internal to a DA. For example, there is nothing in the DA implementation to explicitly prevent a DA from straddling a limited access highway. On the other hand, given that the DA does respect census tract limits and local authorities have carefully selected these limits, implicit control is present. Testing indicates that this form of implicit control is producing acceptable results in this regard. This approach avoids the costs of ordering the road network and the additional complexity of programming for this type of control.

In general, other features such as railways, hydro lines and hydrographic features are not used for block determination and consequently do not form DA limits.

2.5 Compact Shape

As with intuitive boundaries, the demands placed on the EA did result in the generation of strange EAs in terms of shape. For example, if two standard geographic areas straddled each other producing a long narrow sliver where the two boundaries overlapped, an EA needed to be created. Similarly, strange EA shapes resulted where accessibility was a problem. Oddly shaped EAs can create problems when users are attempting to create a "best fit" to their area of interest via the aggregation of EAs. Therefore, users have identified the need to have the dissemination areas as compact as possible.

In order to address this requirement, a compactness index is being included in the DA design. The index is derived by comparing the area of the DA to its perimeter. A square has an index value of one (1.0) while a polygon that is twice as wide as it is long will have a compactness index of two (2.0). In general, if a polygon is stretched so that one side is “n” times longer than the other side, the compactness index will be equal to “n”. Values greater than one (1.0) reflect increasingly less compactness as the DA shape differs more and more from that of a perfect square.

Again, the inclusion of this criterion does not prevent what could be judged as inappropriate results. The inclusion of this criterion only makes certain types of DA delineations more probable.

Testing indicates that good results in terms of **compact shape** are achieved ninety percent of the time. The ten percent of the cases that are viewed as needing improvement predominantly exhibit three types of conditions—the horseshoe or U-shaped DA, the strip DA, and the cookie bit and donut DA.

The horseshoe or U-shaped DA is a result of street patterns designed to discourage through traffic in suburban neighbourhoods (for example, a large block formed around crescent shaped streets within a neighbourhood and bounded by arterial roads at the edge of the neighbourhood). This problem rarely presents itself in older parts of cities where grid road networks are often found.

The strip DA is formed where a strip of land is sandwiched between a road and a boundary feature (often following a hydrographic feature or highway) without interruption by other features. A strip DA also often forms along side of linear features where the rest of the road pattern takes on the pattern of a comb—one long block with many penetrating roads. Allocation of this type of block to a DA results in a tentacle like extension to the DA.

The cookie bit and donut DA is the result of high population concentrations (high-rise apartment buildings) which form just on the edge of an existing DA (cookie bit), or within a DA (donut). This latter situation is in many instances the result of a decision to permit high-rise apartment buildings to form a DA where a minimum population of 300 was present without bounding roads.

2.6 Homogeneity

On the basis of submissions from the user community, homogeneity is the last important factor suggested for inclusion in the delineation of a standard dissemination area. Dwelling type is most often suggested as the variable that should be used to achieve the desired homogeneity. However, of the six key factors recommended for inclusion in the dissemination area design, it is the only one that we have elected not to explicitly include in the design.

The reason for this decision is the lack of the appropriate data by block. Support of this criterion would require that estimates of dwelling counts, by type of dwelling, be generated by block from the 1996 EA level census data. It was concluded that the quality of dwelling type estimates that could be generated would simply be inadequate to the task, so inclusion of an explicit homogeneity criterion was dropped.[8]

3. OPERATIONAL FACTORS

In addition to the recommendations from the user community, there are also operational factors that need to be included in the dissemination area design. The most important of these is the structure of the DA code.

3.1 Dissemination Area Code Structure

Early in the design process the need was identified to have a hierarchical structure represented by an eight-digit code with embedded geographic proximity. These operational requirements were determined or previously satisfied by the province / federal electoral district / enumeration area (PR / FED / EA) structure. The code size (eight-digit) is a consequence of the PR / FED / EA code size namely, 2 + 3 + 3 = 8. This code size is required to avoid redesigning the many programs that need to access this code during census processing. The hierarchical structure with geographic proximity is required to support edit and imputation when a geographically close donor record is required.

All code requirements will be accomplished via a province / census division / dissemination area (PR / CD / DA) hierarchy. The eight-digit code maximum is satisfied with PR (2) + CD (2) + DA (4). Not only is the hierarchical structure requirement satisfied, the number of units at each level is approximately the same since there are 301 FEDs and 288 CDs. As with the EA, geographic proximity will be embedded in the code by assigning DA codes in a serpentine manner within census divisions.[9]

Some users have suggested adding additional geographic code levels to help better determine the set of DAs relative to the area of interest. Because there is not a limitation on the size of the DA code for dissemination purposes, these suggestions can be considered. For example, the CSD code could be added, as could the CT code where CT coverage exists. The dissemination program will review these suggestions.

3.2 Maximum Two-digit Block Code

A second operational requirement may limit the number of blocks that can be included in a DA to 99. This results from a database design that provides for a two-digit block code. Although this design limitation could be changed (e.g., a three-digit code could be created), it would be costly to implement. Consequently, the decision has been made to wait for the first DA test. The first test will not include this limitation. A final decision will be made after determining how many times DAs will be created exceeding the limit of 99 blocks and assessing the impact this limit would have on the creation of dissemination areas.[10, 11]

[8] When it became clear that the homogeneity criterion could not be supported directly, there was some thought that homogeneity could still be supported indirectly by keeping the DA small and compact. In other words, if DAs were comprised of from two to four blocks, then homogeneity would be obtained indirectly since dwelling type often tends to be consistent from block to block without sudden transitions. Although this view is unsubstantiated, it is likely that DAs of two to four blocks will be well below the norm in terms of population size.

[9] This code structure is an internal operational requirement. Alternative structures for dissemination purposes are possible and can be considered; i.e., suggestions are welcome.

[10] Given the maximum population target of 500, we believe the number of times an optimum DA would exceed 99 blocks will be low. In other words, we do not believe this limit on the number of blocks in a DA will adversely affect their formation.

[11] The limit of 99 blocks also applies to the number of blocks per enumeration area. Therefore, some collapsing of blocks will result but the incidence will be low.

3.3 Production Time-line

A three-month elapsed time production window is envisioned for the creation of the 2001 DA structure (from May 2001 to July 2001). Not only is this operational requirement forcing the use of block population estimates for the creation of DAs, but it may also further influence the DA specifications if the required production time exceeds that available. However, there are a number of potential solutions to be investigated before the specifications would need to be reduced to obtain the required time reduction. First, increased computing power is a possible solution. Streamlining the algorithms without reducing the specification requirements is also a possible solution. Limiting the number of solutions to be considered by limiting the amount of computational time for a LDZ is possible, although this approach does not ensure the best solution will be found. It is important to point out however that discussion of this requirement is highly speculative and without empirical production experience at this time. Nonetheless, it is an operational requirement with the potential to affect DA specifications and as such is included in this paper for completeness.

Operational requirements have necessitated the production of DAs prior to June 2001. Consequently, production has been advanced to the period of January 2001 to May 2001—not the May 2001 to July 2001 period previously anticipated. As stated before, this means that 1996 Census population counts must be used as the base for estimating block population counts. The estimate of block counts of sufficient accuracy to support DA delineation is possible only where block-face geocoding[12] is available for the 1996 Census—primarily in census metropolitan areas (CMAs) and census agglomerations (CAs) with census tracts. In all other areas, block estimates of sufficient quality could not be generated. Consequently, for the 2001 Census the delineation of DAs will be restricted to CMAs and census tracted CAs.[13] Outside of these areas, the DA boundaries will be the same as the 2001 enumeration area boundaries.

To replace the EAs-cum-DAs for the 2006 Census, the plan is that DAs will be generated automatically using the individual block counts from the 2001 Census.[14] The DAs generated for 2001 using 1996 data will also be reassessed using the 2001 population data and user reaction. Minor modifications to these DAs are expected.[15]

4. FEEDBACK

The specifications for generating the dissemination area for the 2001 Census are for the most part fixed. Nevertheless, we are always interested in receiving feedback. Comments, suggestions or inquires about any aspect covered in this working paper are most welcome and should be directed to Geography Division at the address provided at the beginning of this working paper.

---

[12] Block-face geocoding is the process by which individual dwellings are associated with a block-face given that the address of the dwelling falls within the address range for the block-face. This association is made to support the retrieval of data for user defined (custom) areas. Because block-faces aggregate to form blocks (usually four block-faces to one block), accurate block population counts can be generated via aggregation: block = sum of block-faces = sum of dwellings per block-face = sum of dwelling population count.

[13] This statement is adequate as a general summary for DA coverage. However, the actual implementation is somewhat more restrictive. Only census tracts within CMAs and CAs where the 1996 Census population geocoded to a block-face equalled or exceeded 80% were included in the DA creation. This accounts for about 63% of the 1996 Census population. In census tracts that had geocoding rates below 80%, the 2001 EAs are being used as DA equivalents for 2001 without any additional processing.

[14] A production window for this activity has not been determined. Consequently, no plans have been established for the release of 2001 Census data using this DA structure.

[15] Since the March 2000 release of this working paper, there have been a number of inquiries as to the potential for direct input (as with the CT programme) into the DA delineation process to reflect local anomalies. Others have expressed their disappointment with the use of 1996 Census data rather than 2001 Census counts. They would prefer DAs that reflect current conditions (e.g., high growth areas) even if that necessitated a delay in the release of data. For the 2001 Census, these concerns cannot be addressed. However for 2006, the DA process will be reviewed and where operationally feasible will be modified in order to reflect the collective needs of Canadians.

ACKNOWLEDGEMENTS

This working paper is a compilation of ideas and research conducted over the past five years and reflects the efforts of a number of staff from the Geography Division (Pessoa, Earl and Puderer, 1996). The contributions of Louise Earl, Peter Murphy, José Pessoa and the reviews of Carolyn Weiss and the Dissemination Area Working Team chaired by Rick Rideout are acknowledged. Thanks are extended to Karole Kidd who edited the final draft of this working paper and Réjeanne Loranger who reviewed the French translation.

REFERENCES

Pessoa, José, Earl, Louise and Puderer, Henry (1996). *The 2001 Logical Dissemination Unit Model: A Proposal for the Development of a New Small Area Geographic Unit for the Dissemination of Data*, unpublished report, Geography Division, Statistics Canada (Ottawa).

Statistics Canada (1994). *1996 Census Consultation Report*, 1996 Census of Canada, Report No. 2 (Ottawa: Statistics Canada).

Statistics Canada (1999). *2001 Census Consultation Geography Report*, 2001 Census of Canada, Catalogue No. 92-126-SPB (Ottawa: Industry Canada).

Appendix 1

THE BLOCK

Up to the 1996 Census, dwellings and their occupants (i.e., census data) were geographically referenced to the enumeration area at the time of collection. Starting with the 2001 Census, the data will be geographically referenced to the block at the time of collection.[16]

First and foremost, blocks are the polygons formed by the intersection of streets[17] (see Maps 1 and 2). These natural or morphological blocks will be split to form two blocks wherever a block is traversed by the boundaries of selected standard geographic areas, namely federal electoral districts (FED), census subdivisions (CSD), census tracts (CT) or designated places (DPL). This is required in order to be able to aggregate the data to all standard geographic areas for dissemination.

A block may also be formed to satisfy collection requirements. For example, workload limits require that blocks be formed for large apartment buildings or collective dwellings. In these cases, blocks within blocks will result. As well, where the road network is sparse or even non-existent, enumeration area boundaries may be established to constrain the coverage area for the census enumerator. In these situations, the block will respect the boundary of the enumeration area.

Blocks are aggregated to build enumeration areas and dissemination areas. Since blocks are primarily an artifact of the road network, the number of blocks and their creation is a function of how up-to-date the road network database will be prior to the census. It is not possible to have a road network reflecting exactly the situation on census day. This will be especially true in high growth areas.[18]

Population and dwelling counts will be disseminated by block. Characteristic data will not be available by block. The smallest standard geographic area for which characteristic data will be provided will be the DA. However, the block will replace the EA outside block-face coverage as the basic unit for constructing user-defined areas, thus providing increased precision in the delineation of these areas.

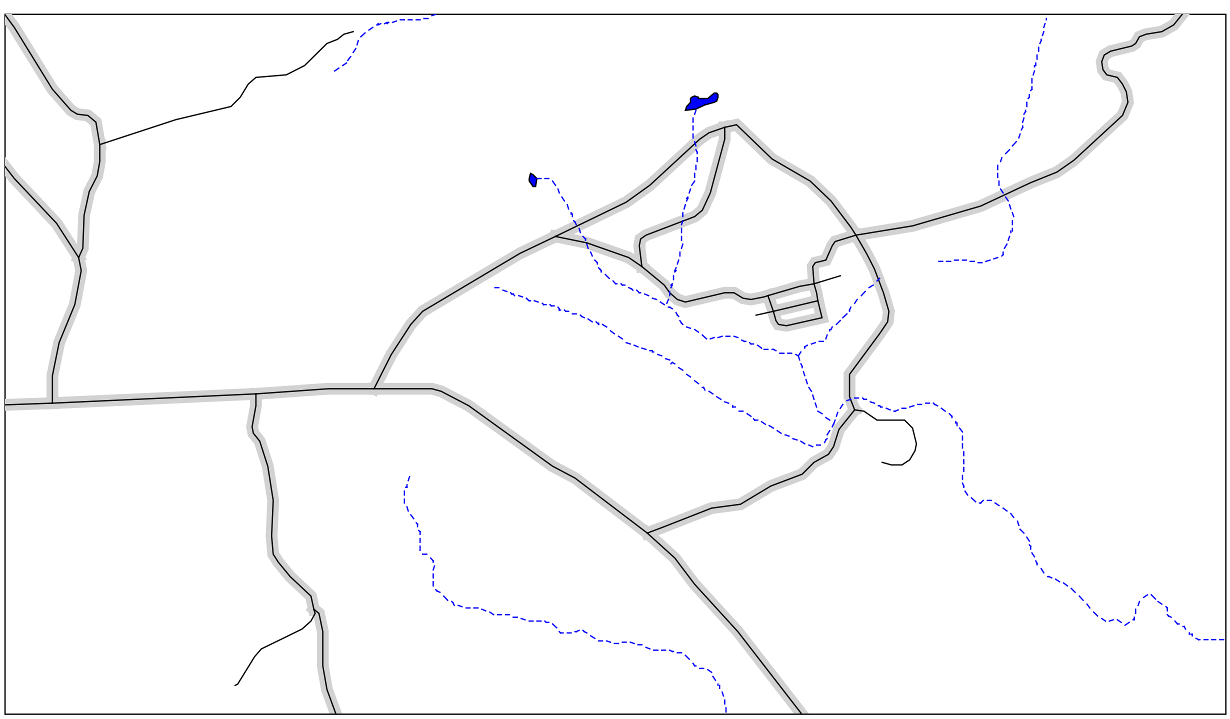
[16] Census data have been and will continue to be referenced to the block-face to support data retrieval for user defined areas. But this linkage to the block-face is done as a post-collection activity and only for selected areas.

[17] Certain morphological blocks (e.g., road medians, ramp areas) will be merged with adjacent blocks.

[18] At the writing of this working paper, the road network on the NGB was up-to-date to about 1995 and generated about 630,000 blocks. However, the road network continues to be updated. The latest round of updates currently in progress will bring the base up-to-date as of spring 1997.

Map 1. Blocks in an urban area for the 2001 Census

Map 2. Blocks in a rural area for the 2001 Census



Geography Working Paper Series

<u>Number</u>

1993-1 ***A Comparison of Census Geographic Areas of Canada and the United States***, Carolyn Weiss, Michael Ratcliffe and Nancy Torrieri (November 1993)

1996-1 ***Whither Land Area? A Case for Retaining the Status Quo for the 1996 Census***, Carolyn Weiss (December 1996)

1998-1 ***Automated Land Area Tests for the 2001 Census: Preliminary Results Using the 1996 Digital Cartographic Files***, Carolyn Weiss and Augustine Akuoko-Asibey (April 1998)

2000-1 ***Census Metropolitan Area and Census Agglomeration Influenced Zones (MIZ) with Census Data***, Sheila Rambeau and Kathleen Todd (January 2000)

2000-2 ***Census Metropolitan Area and Census Agglomeration Influence Zones (MIZ): A Description of the Methodology***, Chuck McNiven, Henry Puderer and Darryl Janes (January 2000)

2000-3 ***Delineation of Canada's North: An Examination of the North–South Relationship in Canada***, Chuck McNiven and Henry Puderer (January 2000)

2000-4 ***Introducing the Dissemination Area for the 2001 Census: an Update***, Henry Puderer (June 2001)

2001-1 ***Geographic Structures as Census Variables: Using Geography to Analyse Social and Economic Processes***, Robert Mendelson (March 2001)