

Work Session on Geographical Information Systems
(Brighton, United Kingdom, 22-25 September 1997)

Item (vi) of the provisional agenda

**PRICING OF GEOGRAPHIC PRODUCTS
FROM THE 1996 CENSUS**

Submitted by Statistics Canada ¹

¹ Prepared by Pamela Tallon.

I. INTRODUCTION

1. Statistics Canada conducts the Census of Population every 5 years, the last one being on May 14, 1996. The Census produces many different products and services, including a series of geographic products to help clients understand the spatial nature of data and to better analyse that data. There has been considerable interest in the geographic products, particularly over the past few years with the tremendous growth in desktop GIS and mapping software packages.

2. Discussions on the marketing of geographic products at Statistics Canada were included in the past two sessions of this Work Session. The first, *Statistics Canada's Geography Marketing Strategy for the 1996 Census*" prepared for the September 1994 session, was written 3 years prior to the release of the 1996 products, and described the strategy to improve on the 1991 experience and provided a benchmark upon which the 1996 marketing process could be evaluated.

3. The second, *Marketing Issues Facing Statistics Canada Relating to Geographic Products*" prepared for the April 1996 Work Session, described some of Statistics Canada's principles regarding the marketing of its data and related some of the issues faced in the formulation of its marketing policy, particularly in relation to the geographic products. This was written at the time just prior to the development of new marketing strategies for the 1996 products.

4. In both papers, the pricing of Statistics Canada's geographic products was discussed and how the high prices were considered a barrier to acquiring them. Since then, the prices have been reviewed and new policies established for geographic products from the 1996 Census.

5. This paper first provides an overview of Statistics Canada's cost recovery policy, which is the underlying context for pricing all of Statistics Canada's products and services. Subsequent sections provide the considerations under which the 1996 geographic product prices were established, a description of each of the 1996 geographic products, and the formulation of their 1996 prices. Comparisons with previous census prices are also provided.

II. COST RECOVERY

6. In its marketing efforts, Statistics Canada categorises two broad groups of clients. The first is the general public who acquires its statistical information through the media, its Regional Reference Centres across the country and a network of public libraries. The second group is comprised of those users of detailed and large-volume information who require it for their businesses, whether they be in government or the private sector. Statistics Canada charges these users realistic prices in order to cover the costs of making the product available to the market place.

7. Cost recovery has several benefits to Statistics Canada's marketing efforts:

- a) allows the product and service line to be expanded since they are self-funding;
- b) imposes a rigorous discipline on the development of products and services since managers must find a market of sufficient size to generate the offsetting revenues and ensure that the products meet client needs;
- c) minimises the inefficient use of resources to develop products for which market demand is marginal;
- d) encourages managers to look for more efficient ways of delivering products to keep their prices within market value.

III. PRICING CONSIDERATIONS FOR GEOGRAPHIC PRODUCTS

8. Market Research activities over the past few years have indicated that pricing of geographic products needed review. This review was carried out in terms of market value of similar products. As a result, the prices of most geographic products from the 1996 Census will be substantially reduced. The pricing for geographic products was based on "high prices/low volume" in the past. For 1996, the strategy will focus more on "low prices/high volume" with the stipulated objective to maintain current revenue levels.

9. The prices of the geographic products are always established to cover costs associated with product development, marketing and support over the entire product and services line and not on a product by product basis. These costs are expected to be recovered over the 5 year census cycle, rather than on a year-to-year basis, to coincide with the life span of the products. Market value is also considered, though in the past, there were very few similar products in the market place with which to compare. A market sensitivity study was conducted for the 1991 products. It concluded that high prices were appropriate at that time.

10. Many of the 1986 and 1991 prices were based on algorithms derived from a unit cost (cost per geographic unit or record in a particular file) plus a base cost per order. Thus, clients paid only for the amount of data they received. Prices were in exact dollars according to the established algorithms. Volume discounts were not implemented.

11. The 1996 strategy for pricing the geographic products followed a number of principles:

- a) the prices will be reduced to better reflect their market values;
- b) the prices will be defensible in terms of the amount of data the product contained, although a minimum price will be established (\$100 or \$200 depending on the product);
- c) volume discounts will be implemented for products covering the whole of Canada or a province or territory;
- d) prices will be rounded to the nearest \$50 or \$100;
- e) standard prices will be established for products for Canada and each of the provinces and territories, as well as for large urban centres where appropriate;
- f) standard Statistics Canada policies will apply to geographic products to be used in a networked or multi-site setting. These policies dictate that a factor will be applied to the price depending on the number of users. Details are provided in Appendix A.

12. In 1991, a base cost of \$100 per product line was added to each customer's order. If a customer purchased more than one file of the same type (e.g. a street network file for Toronto and Vancouver), the \$100 was added only once. This base price was eliminated for the 1996 products by establishing minimum prices per product.

IV. THE 1996 GEOGRAPHY PRODUCT LINE AND PRICING

13. Geography products emanating from any census are developed from the geographic infrastructure built and maintained mainly for the census collection activities. For 1996, this geographic infrastructure was composed of principally two layers of spatial data and two attribute databases.

14. The two layers of spatial data are:

- a) The digital street network database for the large urban centres. This digital layer is complemented with a paper map base of approximately 3,200 maps covering small urban centres and rural areas for the rest of Canada where no digital files exist.
- b) The Enumeration Area boundary layer. Digital boundaries were created for the almost 50,000 enumeration areas in Canada. All other census geostatistical boundaries are derived by aggregating enumeration area boundaries

An Enumeration Area (EA) is the area of responsibility for one census enumerator. EAs range in size from 125 dwellings in rural areas to 440 dwellings in urban areas and are created to respect the boundaries of all other standard geostatistical areas. EAs are the smallest geographic area for which census data are published.

15. The two attribute databases are:

- a) The Geographic Attribute Database (GADB), which contains the attributes of all of the geostatistical areas and the linkages among the various geostatistical units. Examples of these attributes include the geostatistical area name and code, and the population and dwelling counts. Information from the previous census is also maintained.
- b) The Postal Code Concordance Database, which contains the linkages between Canada's 6-character postal codes, as defined by Canada Post Corporation, and enumeration areas. In addition, for each postal code, one or more x,y point coordinates are assigned to each postal code to represent the approximate location of the postal code on a maps.

In general, Statistics Canada's geographic products are derived directly from these infrastructure databases. In some cases, additional information must be created specifically in order to make a product better meet our clients' needs.

16. The following sections describe each of the standard 1996 geography products and the considerations in determining their 1996 prices. Comparisons with the 1986 and 1991 pricing are also provided. Statistics Canada's digital geographic product line consists of two basic types of data: spatial data products and reference (i.e., tabular) data products.

V. SPATIAL DATA PRODUCTS

17. Spatial products are digital files that provide the shape and location of geographic features such as boundaries or streets. Two (2) types of products will be discussed: boundary files and street files.

18. All of Statistics Canada spatial products are provided in both ARC/INFO® Export and MapInfo® formats, in latitude/longitude coordinates in the reference datum NAD27. Since the spatial data products also include attribute information in a database format, all of Statistics Canada's spatial products are provided in English and French to accommodate the attribute field names in the appropriate language. Diskette and CD-ROM are the main mode of delivery for most of the geographic products. CD-ROM will be offered when the product spans more than two or three diskettes.

BOUNDARY FILES

19. Boundary files are computer files of geographic data which, when used with the appropriate GIS or mapping software, depict the limits of standard geostatistical areas. Each boundary file represents one specific geostatistical level (e.g. counties, municipalities, federal ridings, etc.). A total of twelve geostatistical levels are available.

Statistics Canada produces two types of digital boundary file products: Digital Boundary Files (DBF) and Digital Cartographic Files (DCF)

1996 Digital Boundary Files (DBF)

20. Digital Boundary Files (DBF) portray the official boundaries used for the collection activities of the 1996 Census and, therefore, often extend as straight lines into bodies of water. As such, boundaries you would expect to follow a shoreline or coastline will not. The boundaries cut through lakes, jut into oceans, or follow the approximate centre of a river. This ensures that all land and islands are included for the enumeration of the census and that official limits are followed.

21. The 1996 DBFs are similar to the 1991 DBFs in terms of general description and features, though the boundaries have been updated to 1996 and the attribute information has been enhanced. The 1996 DBFs may not match the corresponding 1991 products even in areas where no changes to the boundaries occurred. This results from the impact of various cartographic adjustments, updates in the street network files and some re-digitising of boundaries. DBFs are useful for reference mapping and sophisticated GIS applications employing topological manipulations, but they aren't appropriate for mapping where realistic shorelines are required along Canada's perimeter.

1996 Digital Cartographic Files (DCF)

22. In the 1996 Digital Cartographic Files (DCF), boundaries were modified to follow the coastlines and shorelines on the perimeter of Canada's land mass, including major islands. The 1996 DCFs include a separate map layer of lakes and some rivers. This "water" layer can be used for additional reference purposes when mapping or displaying the boundaries.

23. The 1996 DCFs differ from the 1991 DCFs with the provision of the separate water layer and detail of the shoreline increased. Bodies of waters such as lakes and large rivers were "cut into" the 1991 boundary layer. Although the shoreline used in areas covered by street network files is the same in the 1991 and 1996 files, the 1996 shoreline in the rest of Canada is more detailed than that used for the 1991 files. This more detailed shoreline allowed, for the first time, the creation of a 1996 DCF at the EA level.

24. The EA DCF is created from the EA DBF by "clipping" a suitable shoreline file depicting Canada's perimeter onto the DBF. For 1996, the shoreline was based on a file from Natural Resource Canada (Canada's national topographic mapping agency). Some generalisation to reduce the number of points was required to ensure that the boundary polygons met the constraints of most desktop mapping packages.

25. The DCFs for all of the other geostatistical areas were created by aggregating the EA DCFs. The number of points per boundary polygon was checked at each level of aggregation to ensure the point limit was not exceeded. If it were, further generalisation was applied at the EA level, and the aggregation redone. This ensured that all of the DCF products are vertically consistent.

DBF and DCF Pricing

26. In 1991, the DBFs were priced at \$1 per geostatistical area plus \$100 per order. The DBF for all EAs in Canada (approximately 46,000) was the most expensive product at \$46,095. If a client purchased the EA DBF for Canada or any of the provinces or territories, they also received the DBFs for all other standard geostatistical areas at no additional cost.

27. The 1991 DCFs were released into the market place about a year after the DBFs. The market had changed in the intervening year, and given the perceived demand for boundary files depicting a more realistic shoreline, the DCFs were set at a base price of \$0.50 a geostatistical area plus \$100 per order. 1991 DCFs were not available at the EA level.

28. The prices for the 1996 DBFs and DCFs were established by calculating a base price of \$.25 per geostatistical area with a minimum value of \$100. Thus for 1996, the DBFs and DCFs are set at the same price and range from 4100 to \$9,800. In most cases, clients purchase the DCFs (for mapping purposes). The extra water layer is provided with the DCF in its basic price.

29. The table below depicts some typical prices for the DCFs.

Product	1986 DCF	1991 DBF	1991 DCF	Private Sector 1991 DCF	1996
EAs for Canada (50,000 areas)	not available	\$46,095	not available	\$18,000	\$9,800
Municipalities for Canada (6,000 areas)	\$4,2001 \$5,4602	\$6,106	\$3,103	\$1,325	\$1,200
Census tracts for Toronto (800 areas)	\$700	\$912	\$506	\$400	\$200
Census tracts for Regina (50 areas)	\$200	\$149	\$125	\$125	\$100

The 1986 digital boundary files were available in a number of different formats.

1GIMMS, ARC/INFO and SAS/Graph formats

2Atlas*Graphics, MapInfo and MapAnalyst formats

STREET FILES

30. The Street Network Files (SNF) are computer maps that represent a street network and other features such as rivers and railways. Attribute information includes the type ("class") of feature (such as a street, highway, river, etc.), names (such as the street name), and address ranges at street intersections. SNFs are available for most of the larger urban centres across Canada, covering more than 62% of the population but less than 1% of Canada's total land area.

31. The SNFs are derived from Statistics Canada's Street Network Database (SNDB) that maintains both polygon and arc information, however, the essential features in the market-based product are the arcs and their attributes. Most polygon information in the SNDB, maintained for internal operations, is dropped. For 1996, two additional layers of information will be included with the SNF product - a layer of water features (e.g. lakes and large rivers) and a layer of municipal boundaries that comprise the urban centre. The addition of the water layer will make it easier for clients to shade water polygons blue. Since most SNFs do not extend to the limits of the urban centre, the municipal boundaries layered on the SNF will make it easier for clients to determine which municipalities in an urban centre have SNF coverage and which do not.

32. The major enhancement to the 1996 SNFs is the updated streets and address ranges to Census Day, May 14, 1996. This will be a marked improvement from the 1991 version, currently in the market, containing vintages as early as 1986 for address ranges and up to about 1990 for street information.

33. A new product was introduced in 1991 offering a slimmed down version of the SNF. This product, called the **Skeletal Street Network File (SSNF)**, is designed to provide some basic reference information for thematic mapping where the detailed street networks are not required. The SSNFs are actually a subset of the SNF. Street names are included, but not address ranges.

34. Major changes from the 1991 file structures will not be done, although efforts will be made to configure the attribute tables associated with each file to facilitate their use in ArcView® and MapInfo® (e.g. address geocoding capabilities within the particular software).

SNF and SSNF Pricing

35. SNFs and SSNFs are packaged and sold by urban centre. In most cases, the digital coverage does not extend to the limits of the urban centres, but does include the most populated municipalities.

36. For 1991, the SNF prices were based on a base value of \$350 plus \$0.45 per polygon (formed by the intersection of all features within the SNF). The prices were retained in exact dollars (not rounded).

37. For 1996, the SNF prices were calculated at \$0.15 per polygon with a minimum of \$200, and then rounded to the nearest \$50 or \$100. The resulting numbers were then compared to prices used by Statistics Canada distributors (as an example of market value) to ensure they were within reasonable limits. Some adjustments were made where necessary. A 20% discount applies to a full Canada or provincial set of SNFs.

38. The SSNFs are priced the same as in 1991, that is at \$100 or \$200 depending on the size of the urban centre. The 1991 SSNFs were provided free with the corresponding DCFs for census tracts since they were designed to match only the census tract boundaries. With the enhancements made to the 1996 DCFs, SSNFs will match all geostatistical area boundaries and customers will have to purchase them separately. In some cases, the 1996 product may be the same as the 1991 product.

39. The following table provides examples for some large and small urban centres for both SNFs and SSNFs.

Urban Centre	1986 SNF	1991 SNF	Private sector 1991 SNF	1996 SNF	1991 & 1996 SSNF
C a n a d a (2 0 0 , 0 0 0 polygons)	\$75,202.85	\$88,345	\$15,000	\$20,000	\$2,900
T o r o n t o (2 7 , 0 0 0 polygons)	\$9,896.20	\$12,144	\$2,000	\$2,900	\$200
Ottawa-Hull (1 2 , 0 0 0 polygons)	\$6,219.95	\$5,668	\$1,600	\$1,450	\$200
Regina (3,000 polygons)	\$1,063.55	\$1,379	\$350	\$400	\$100

VI. GEO REFERENCE PRODUCTS

41. In addition to spatial products, there will be four 1996 digital geographic reference products. These digital reference products provide descriptive information about geographic features that can enhance GIS applications. Included in this set of products is the new 1996 Postal Code Counts CD-ROM which contains the 1996 Census population and dwelling counts for all postal codes in Canada as reported by Census respondents. The four products are GeoRef, Postal Code Conversion File (PCCF), Block-Face Data File (BFDF), Postal Code Counts (**New for 1996**).

42. Each reference product provides information related to a particular geographic feature:

- The most detailed level of information in GeoRef provides data for one enumeration area (EA.).
- Each record of the Postal Code Conversion File (PCCF) provides information on a postal code.

Each record of the Block-Face Data File provides information on one block-face, which is one side of a city street between consecutive intersections defined in SNF areas only.

43. Information related to a feature includes fields such as linkages to standard geostatistical areas through codes and names, as well as 1996 population and dwelling counts, depending on the product.

44. Important for GIS users, however, is that each file (except the Postal Code Counts CD-ROM) provides an x,y point coordinate which is useful in mapping the featured geographic unit. Each reference product can be formatted and added into a desktop GIS or mapping package as a table (or equivalent). The x,y coordinates can then be converted to a point layer for displaying on a map.

GEOREF

45. The **1996 GeoRef CD-ROM** is a user friendly query and access tool providing access to a wide selection of geographic information. It includes software and data for all of Canada on CD-ROM. It allows the user to explore the links between geographic units and to obtain information on those units. 1996 Census population and dwelling counts are included for all standard geostatistical areas (except postal code), including the almost 50,000 EAs. Information can be exported from GeoRef in a number of different formats.

46. **The price for a 1996 GeoRef** is substantially reduced from the equivalent product in 1991 - from \$500 to \$60. This is a result of Statistics Canada's strategy to make basic geographic information more accessible to a greater number of people. The equivalent 1986 product was \$850.

47. The 1986 and 1996 equivalent products were flat ASCII files packaged for Canada and each of the provinces and territories. The 1996 GeoRef is available at the Canada level only. Clients can purchase custom extractions from GeoRef if they do not have CD-ROM capabilities or if they only want a subset of the information. Extractions are priced to cover the work required.

POSTAL CODE CONVERSION FILE (PCCF)

48. The Postal Code Conversion File (PCCF) links the six-character postal codes (as defined by Canada Post) to Statistics Canada's standard geostatistical areas and provides point coordinates for mapping approximate locations of postal codes. There are currently over 700,000 postal codes in Canada.

49. The PCCF enables clients to relate postal codes to census geography or vice versa for tabulating client lists by a particular area (such as municipalities or census tracts), linking census demographics to postal codes (generally representing client addresses), or obtaining the postal codes for target geographic areas (for mailing). The PCCF is provided as a flat ASCII file. Updated files are provided semi-annually.

50. **The 1991 PCCF was priced at** approximately \$0.03 per record. This reflected the basic information provided in the file (the postal code, the geostatistical codes of areas in which the postal code was located, and the x,y point coordinate). Postal codes were updated every 6 months and after a client had purchased the product at the base price, clients could purchase semi-annual updated files at reduced rates. In addition, clients received a 25% discount if they subscribed to a minimum number of updates.

51. The price for the 1996 PCCF was based as one-half the 1991 price (or \$0.015 per record) with a minimum of \$200. The table below provides some examples. Updated file prices and subscriptions are the same as for 1991, ranging from \$100 to \$1,500 (or \$75 to \$1,125 with the 25% discount).

52. The table below provides some typical examples.

Urban Centre	1986 PCCF	1991 PCCF	Private sector 1991 PCCF	1996 PCCF
Canada (700,000 records)	\$9,500	\$19,000	\$7,500	\$9,000
Ontario	\$3,520	\$7,000	\$3,375	\$3,500
Saskatchewan	\$605	\$770	\$1,500	\$500

BLOCK-FACE DATA FILE (BFDF)

53. The BFDF provides information on the smallest geographical unit, called the block-face, available for urban centres where an SNF is maintained. A block-face is one side of a city street between two consecutive intersections. Information provided includes street names and address ranges, geostatistical codes of the areas in which the block-face is located, and the 1996 Census population and dwelling counts for the households living on the block-face. An x,y point coordinate is calculated from the SNF to represent the block-face and provided in the BFDF for mapping purposes. The BFDF is provided as a flat ASCII file by urban centre.

54. The BFDF was not available as a standard product in 1986. The prices for the 1991 BFDFs were based on the number of records for an urban centre and ranged from \$108 to \$683. The cost for all files available for Canada was \$3,914. The 1996 prices are based on the size of the urban centre - \$100, \$200 or \$300. The total price for all files is \$2,000.

POSTAL CODE COUNTS CD-ROM

55. The Postal Code Counts CD-ROM is a new product for 1996. It includes the 1996 population and dwelling counts for all postal codes reported by every census respondent in Canada. The data are provided with easy-to-use Windows-based software that enables users to search, query and retrieve information from the database.

56. The price for the Postal Code Counts CD-ROM for all of Canada was established at \$60 following the same strategy as that for GeoRef. Feed-back received from Statistics Canada's regional sales force indicates that \$60 may be too low given the value of postal code information to the business community.

VII. CUSTOM SERVICES

57. If the standard geographic products do not satisfy a user's need, Statistics Canada provides a Custom Service to produce the product as a special request. Examples include special data retrievals or merges using any of the geographic files (postal codes, attribute files, boundary files and street network files), special population compilations using basic census collection records and digitising user-defined boundaries. The minimum cost for this service is \$100 (reduced from \$350 in 1991). Cost estimates are provided to customers based on the complexity of the request. Custom mapping is also available at a minimum of \$350.

VIII. IN CONCLUSION

58. Prices for Statistics Canada geographic products from the 1996 Census were established to make them more competitive in the market place. Price reductions range up to 75%, depending on the product. As of July 1997, only the 1996 Digital Boundary Files and Digital Cartographic Files are available; the remaining products will be released in the fall of 1997.

59. Multi-use and multi-site pricing factors are still being discussed. Most of the 1991 geographic products were considered "corporately" priced where multi-use and multi-site factors were considered to be built into the established price. The application of the factors for the 1996 products may negate the price reductions that have been made.

60. New licensing policies for the redistribution of Statistics Canada's products by the private sector are almost finalised. These new policies, in combination with the price reductions, should make the geographic products more accessible to small firms who could not afford the high 1991 prices.

61. Statistics Canada has a practice of giving privileged status to the educational sector. In the early 1990's, it developed E-STAT, a product combining user-friendly mapping software with census and other Statistics Canada data. E-STAT is only available to the education market and sold at a subsidised price.

62. In addition, Statistics Canada has entered into agreements with consortia representing some 50 universities across Canada. These agreements give the universities access to all Statistics Canada's electronic products (including the geographic products) at affordable prices. This "Data Liberation Initiative", a five-year pilot project, will ensure that the Canadian academic community has access to Statistics Canada's data for teaching and social and economic research.

Appendix A - Multi-site and Multi-use Licensing

Excerpts from Statistics Canada's Policy on Licensing (November 12, 1992)

1. End-user - Pricing will differentiate between the stand-alone and networked end-use, and between single site and multi-site end-use within the same organisation.
2. Non-standard end-user licence agreement - Situations which may involve a nonstandard agreement include requirements for local area network (LAN), site or corporate end-use of Agency electronic information products.
3. Guidelines for Licensing - Introduction - End-User licences will in most cases be for a single user at a single location. Product pricing is based upon this type of use and the standard "shrink wrap" type of licence can be used. Increasingly, however, with electronic based products, it is essential to further determine whether the product is to be shared amongst many users at one location or perhaps at many locations. For such use, additional charges will apply, and depending on the situation, a "nonstandard" licence document may be required.
4. LAN Pricing - The LAN surcharge will be applied when more than three (3) users access the data on a network or if the network serves more than one building or civic address. Three (3) surcharge levels will apply. For up to ten (10) users, a factor of 1.75 will be applied to the regular end-user price. For up to twenty users, the factor 2.50 will apply. For over 20 users, a factor of 3.0 will apply. (Note these factors are different for custom data tables).
5. Site Licence Pricing - is considered to be equivalent to a LAN and the same price factors will be applied to the end-user price.
6. Corporate/Multi-site Pricing - Corporate/Multi-site Licences will provide unlimited use within a corporation at one or more locations. A factor of 3.0 will apply to standard retail product prices. When the retail price is over \$10,000, a volume discount may be applicable and Licence Services staff should be contacted to review the case and to provide a quotation. The complete data package details and user environment will be required for this analysis.