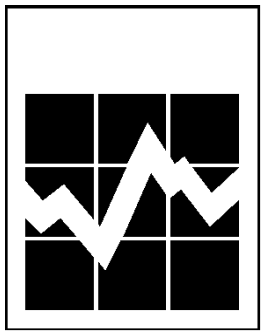

Survey of Information Technology Occupations, 2000: Employer Survey

Methodology Report

September 26, 2000

Prepared by:
Business Survey Methods Division
and
Small Business and Special Surveys Division
Statistics Canada
R.H. Coats Building 3-C
Ottawa, Ontario
K1A 0T6

Survey of Information Technology Occupations, 2000: Employer Survey

Methodology Report

INTRODUCTION

The Pilot Survey of Information Technology (IT) Occupations, 2000: Employer Survey was conducted by Statistics Canada, in February, March and April, 2000, on behalf of Human Resources Development Canada. The main objective of this survey of IT employers was to study labour market conditions for IT occupations and to produce statistical information on required skills and employment-related issues specific to information technology occupations within the business service sector in Canada.

One of the goals of this pilot survey was to assess the feasibility of conducting a national survey covering a variety of industries. The results of this pilot will help to design an effective sampling strategy that will allow reliable estimates to be produced from a national survey.

POPULATION

The target population consisted of locations, with at least six employees, coded to three specific industry categories (classified to the North American Industrial Classification System (NAICS) at the four-digit level) and found in specific geographical regions (see chart below).

NAICS	DESCRIPTION	REGIONS
5241	Insurance carriers	Ontario only
5413	Architectural, engineering and related services	Quebec only
5415	Computer systems design and related services	All of Canada

These three industry categories were thought to employ significant numbers of employees in the occupations of interest, based upon the results of the 1996 Census. The sampling frame was approximated by a list of 5,760 locations on the Statistics Canada Business Register as at December 6, 1999. A SAS program run on the mainframe computer was used to create the sampling frame according to the size and industry criteria. The inclusion of locations with six or more employees was intended to avoid the problems inherent in selecting smaller locations. There has tended to be a high turnover rate for small businesses and the sample might have been quickly depleted with locations that are “out-of-business” if locations with fewer than six employees were included. Also, the nature of the work in small businesses often makes it difficult to associate an employee to one specific occupation and the diversification of occupations is often very limited.

CHOICE OF STRATA

Twenty-one strata were used to ensure adequate representation of locations by region, industry and size. Locations in each region-industry combination were divided into three size categories to improve the representation of the sample. The region-industry combinations were: the Atlantic provinces, Quebec industry 5413, Quebec industry 5415, Ontario industry 5241, Ontario industry 5415, the Prairie provinces (including the Northwest Territories and Nunavut), and British Columbia (and Yukon). The size categories were 6-25 employees, 26 to 50 employees, 51 or more employees.

COLLECTION METHOD

Computer-assisted telephone interviews (CATIs) were used to survey businesses at the location level to collect information on the numbers of employees, hiring and recruitment practices, employee retention, and training and development of IT workers.

SAMPLE SIZE

A sample of 3500 locations was drawn, with a further 315 drawn during the collection period in anticipation of low response rates in certain strata. While efforts were made to achieve a 5% bound on error for estimates of percentages, some locations were excluded from the sample selection process to minimize overlap with other concurrent surveys, most notably the Workplace Employee Survey (WES). See the document "2000 Pilot Survey of Information Technology Occupations: Control of Overlap with Other Surveys" for more information on sample selection. To reduce the response burden of one enterprise, of their 32 locations that had been selected, only six of its locations were asked to respond.

		POPULATION SIZE				**SAMPLE SIZE**			
		Number of Employees				Number of Employees			
NAICS	Region	6 - 25	26 - 50	> 50	Total	6 - 25	26 - 50	> 50	Total
5415	Atl	103	31	18	152	96	29	15	140
	Que	641	121	112	874	392	114	97	603
	Ont	1740	256	217	2213	474	254	183	911
	Pra	511	87	65	663	451	82	49	582
	BCY	394	62	50	506	370	57	43	470
	Total	3389	557	462	4408	1783	536	387	2706
5241	Ont	252	89	151	492	217	61	109	387
5413	Que	640	118	102	860	523	111	88	722
Quebec Total		1281	239	214	1734	915	225	185	1325
Ontario Total		1992	345	368	2705	610	315	292	1217
Overall Total		4281	764	715	5760	2523	708	584	3815

RESPONSE RATE CONSIDERATIONS

Prior to collection, a number of steps were taken to help increase the response rate. Telephone numbers for all locations were verified using various sources, such as Canada 411, CD Prophone, Infodirect, Canadian Business Directory and Bell Canada. As well, for locations that

did not have an address specified on the Business Register, research was conducted to find this missing information.

There was some concern that large locations (those with 100 or more employees) may have an organizational structure that is not well suited to this survey. A pre-contact of all of these locations was conducted to identify the most appropriate person(s) to respond to the survey. In a few cases, more than one respondent provided the information for a particular location.

Each location included in the sample was sent an introductory letter describing the purpose of the survey, a list of the questions that would be asked and definitions of the occupations that were being targeted.

COLLECTION

Collection was carried out by the Operations and Integration Division using a computer-assisted telephone interview (CATI) method. The Operations Research and Development Division developed the CATI application in BLAISE and maintained the system during the collection process. Collection began on February 14, 2000 and ended on May 1, 2000.

In the course of an interview, CATI interviewers asked for the number of employees in each of 21 occupation categories of interest according to a revised version of the National Occupational Classification (NOC), then the CATI system randomly selected two occupation categories for which there were one or more employees.

The response status by location was as follows:

STATUS	Agreed to share data with HRDC	Did not agree to share data with HRDC	No IT employees or contract workers	TOTAL
Responded	**1168**	**140**	**368**	**1676**
with no employees in any of the 21 occupation categories	*37*	*65*	*368*	*470*
with at least one employee in any of the 21 occupation categories	*1131*	*75*	*0*	*1206*
Non-respondents	**0**	**576**	**0**	**576**
Out-of-scope (including duplicate)	**136**	**691**	**71**	**898**
Out-of-business (including unable-to-locate)	**0**	**665**	**0**	**665**
TOTAL	**1304**	**2072**	**439**	**3815**

The majority of locations classified as out-of-business were not able to be located despite repeated attempts to make contact. Nearly all of the locations classified as out-of-scope had fewer than six employees, although the frame indicated otherwise. The response rate, calculated as the number of respondents divided by the sum of respondents and non-respondents, was 74%. However, the number of responses featuring occupation data out of the total sample size was

34%. As a consequence, many results were suppressed due to low numbers of observations or high coefficients of variation or standard errors.

The Business Register keeps track of businesses' births, deaths and numbers of employees. It is never absolutely current, so it is inevitable that out-of-scope and out-of-business locations are included in the sample. The maintenance of accurate information on IT companies is especially difficult as IT companies comprise a relatively volatile industry.

THE QUESTIONNAIRE

The questionnaire was divided into seven sections. Three sections (A, B and G) collected data on the location and/or the respondent, while the other four sections (C through F) collected data on employees in one or two occupations of interest. The occupation categories were chosen from those in the following table, provided that there were one or more employees in the chosen occupations. The final survey questionnaire can be found in Appendix A.

NOC Code	Description of occupation category
0115	IT Training managers
0611.5	Web managers
0213	Computer and information systems managers
2171.1	Information systems business analysts and consultants
2171.2	Systems security analysts
2171.3	Information systems quality assurance analysts
2171.4	Systems auditors
2172.1	Database administrators
2172.2	Data administration analysts
2175	Network systems and data communications specialists
2173	Software engineers
2147	Computer engineers, except software
2133	Electrical and electronics engineers, except computer engineers
2174.1	Computer programmers
2174.2	Interactive media developers
2281.1	Computer and network operators
2281.2	Web technicians
2282	Technical support analysts
2283	Systems testing technicians
5121.2	Technical writers
5241	Graphic designers and illustrating artists

WEIGHTING

Each location was assigned a stratum-specific location weight calculated as the number of locations in the population divided by the number of locations in the sample. The location weights were adjusted for non-response. Totals of IT employees and contract workers by occupation were estimated at the location level.

Results for questions in sections C through F of the questionnaire were estimated at the location-occupation level. Location-occupation weights were calculated in two stages. One, the location

weights were multiplied by a factor to account for the random selection of two occupation categories from among those with at least one employee. Two, the weights were adjusted so that the total numbers of employees estimated by question C2 (number of permanent and temporary full- and part-time employees) agreed with the totals estimated at the location level from section B of the questionnaire.

EDIT & IMPUTATION

Four types of edits were used in the data editing process.

Edit #1	Check that an answer is within a range of allowable answers according to the answer key.
Approach	Every answer was compared to the set of possible coded answers for that question, adjusted for skip patterns.
Edit #2	Check that a response to a question concerning the numbers of employees (e.g. vacancies) was of a magnitude consistent with the number of employees in that occupation.
Approach	Answers having to do with numbers of employees failed this edit check if the answer was more than 300% of the number of employees in that occupation.
Edit #3	Verification of derived values (consistency edits).
Approach	Answers to quantitative questions were checked against minima and maxima calculated from other answers to the questionnaire.
Edit #4	Identification of extreme values
Approach	Frequency tables provided summaries of data values. Unexpected values were investigated individually.

The project manager reviewed all occurrences of data edit failures in conjunction with the methodologists. No formal outlier detection and imputation method was used.

Imputation of the numbers of employees and contract workers was done at the location level using ratio imputation. An imputed number of employees in an occupation was calculated as the product of the location's number of IT employees and the stratum-specific average proportion of IT employees in a specific occupation.

Imputation of data from sections C through F of the questionnaire was done after the data file was converted from one line per location to one line per location-occupation. Where possible, imputation of these data was done by occupation and stratum. In strata with no observations of particular occupations, data were imputed by occupation and size, or if not possible, by occupation only. Values for question C2 (number of permanent and temporary full-time and part-time employees) were imputed using ratio imputation whereas mean imputation was used to impute values for all other quantitative questions. Although mean imputation has the advantages

of simplicity of implementation and of interpretation, this method introduces bias and causes the coefficients of variation to be understated by an unknown amount.

The number of permanent full-time employees in an occupation at a particular location was imputed using the following formula, where the ratio on the right is calculated from the sums of all non-imputed ("good") observations within a stratum:

$$no.\ of\ employees\ at\ that\ location \times \frac{total\ no.\ of\ \ permanent\ full-time\ employees}{total\ no.\ of\ employees}$$

A similar approach was used for the other parts of question C2.

There were 18 quantitative variables that could have been imputed. The percentage of variables that were imputed by occupation and region were:

NOC_Code	Atlantic 5415	BC/Yukon 5415	Ontario 5241	Ontario 5415	Prairies 5415	Quebec 5413	Quebec 5415	All
0115	0%	1%	1%	11%	2%	30%	16%	10%
0611.5	25%	1%	0%	12%	7%	13%	7%	8%
0213	5%	4%	5%	8%	3%	4%	8%	6%
2171.1	0%	3%	11%	8%	8%	9%	5%	6%
2171.2	0%	6%	3%	18%	25%	3%	4%	12%
2171.3	0%	3%	14%	6%	0%	0%	10%	6%
2171.4		0%	0%	1%	4%	0%	0%	1%
2172.1	1%	5%	0%	8%	15%	1%	6%	6%
2172.2	7%	3%	0%	17%	0%	0%	16%	8%
2175	1%	1%	6%	1%	0%	17%	8%	4%
2173	1%	5%	7%	7%	4%	1%	2%	4%
2147		16%	0%	6%	2%	25%	13%	9%
2133	0%	14%		11%	7%	2%	0%	5%
2174.1	1%	2%	2%	5%	3%	0%	6%	4%
2174.2	0%	1%	0%	0%	0%	0%	0%	0%
2281.1	0%	0%	0%	5%	17%	11%	11%	7%
2281.2		0%	0%	0%	6%	50%	0%	4%
2282	1%	0%	1%	2%	2%	3%	12%	4%
2283	0%	2%	0%	0%	0%	0%	0%	1%
5121.2	14%	0%	0%	5%	0%	0%	0%	2%
5241	0%	1%		1%	1%	1%	0%	1%

ESTIMATION

Estimates were generated using the Statistics Canada Generalized Estimation System (GES). GES is a SAS based application for producing estimates by domain (e.g. occupation and

province category). Coefficients of variation were provided for quantitative point estimates. Standard errors were provided for estimates of proportions and ratios.

These estimates have been put into Excel tables for ease of use. The content of these output tables appears in Appendix B.

DATA QUALITY

There were no observations available for three occupation categories in the Atlantic provinces and two occupation categories in Ontario for industry 5241. Due to skip patterns, some observations were unavailable for other occupation and province-industry categories.

Since all the estimates produced from this survey were based on sample results, they were subject to sampling error. Sampling error of a total can be expressed as a coefficient of variation (CV). The CV is a percentage that expresses the size of the standard error as a proportion of the estimate to which it is related. For example, a CV of 10% means that the standard error is 10% of the estimate. The following table provides a guideline of the quality of an estimate of a total:

Value of CV	**Rating**
0 to 5%	Very Good
6% to 15%	Good
15% to 33%	Good to Poor-- use with caution
33% and over	Very poor -- may not be acceptable

Many of the estimates fell into the "Good to Poor" or "Very poor" categories. The main reason for this is that the variables used to create these estimates varied widely from unit to unit in the population. Highly variable observations made precise estimation of population values more difficult.

Estimates were suppressed for reliability and/or confidentiality purposes according to the following criteria:

Type of estimate	**Measure of dispersion**	**Result suppressed if**
Total	Coefficient of variation (CV)	CV > 33.5%
Proportions	Standard error (SE)	SE > 15.5%
Ratio	Standard error (SE)	SE > 15.5%

Some estimates of proportions were also suppressed to prevent residual disclosure.

All estimates created with three or fewer sample units have been suppressed for confidentiality purposes. Estimates of aggregate totals for industry 5415 were suppressed if it were possible to derive a suppressed estimate at the region by industry level.

The limited sample size of the pilot survey, as well as the use of a non-standard sampling strategy, resulted in suppressed estimates. The relatively low quality of estimates was expected for the pilot survey but indicates that a different sampling strategy might be appropriate for the full production survey. The use of mean imputation would also have led to underestimates of

dispersion.

MICRO-DATA FILE

Two micro-data files were created for this survey and are only available to Human Resources Development Canada as a result of the Data Sharing Agreement. One file was produced at the occupation-level, which means that for each location, the 21 IT occupations are listed on separate lines. A second file was produced at the location-level. This file includes responses to sections A, B and G of the questionnaire. A complete data dictionary was included with this file to help identify the variables and give the possible values they may assume. Data from respondents who said "no" to the data sharing agreement were removed from this file.

Of the responding locations that reported having at least one IT employee or contract worker in one of the 21 occupation categories, 89.2% agreed to share their data with Human Resources Development Canada.

Appendix A: The Survey Questionnaire