

68N0003XPB

no. 61

c. 2

**Project to Improve Provincial Economic
Statistics**

**Projet d'amélioration des statistiques
économiques provinciales**

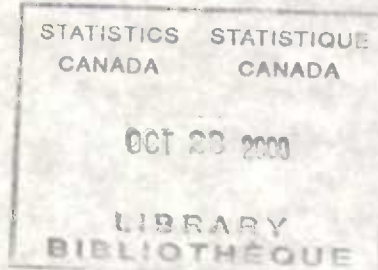
An Overview of the 1998 GIFI Database

Aperçu de la base de données de l'IGRF 1998

Available also on the StatCan INTRANET site /
Disponible aussi sur le site INTRANET de StatCan

Technical Series

Number 61



Série technique

Numéro 61



Internet: www.statcan.ca
Intranet: <http://pipes>



Statistics
Canada

Statistique
Canada

Canada

An Overview of the 1998 GIFI Database

The purpose of this paper is to present an overview of GIFI at a very basic level and to encourage new users to take advantage of the breadth and depth of this data set. The paper is in four sections. The first is the introduction and describes the background of GIFI and what the benefits are to Statistics Canada. The next section is the background, giving the user an idea of how financial data were collected in the past, the present situation and the future of GIFI. The third section includes tabulations from the entire data set and describes the Metadata. The last section contains GIFI-based estimates for a particular industry chosen as an example, the Personal and Laundry Services industry.

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

For further information on the materials covered in this paper, please contact Bonnie Bercik (613) 951-6790 or Diane Proulx (613) 951-7192
Fax: (613) 951-0411

An Overview of the 1998 GIFI Database

Jennifer Sarah Jones
July 26, 2000

Table of Contents

Introduction	3
Motivation of the paper.....	3
The Harmonized Sales Tax (HST)	3
The role of Statistics Canada and PIPES.....	4
The Unified Enterprise Statistics Program (UESP).....	5
What is the General Index of Financial Information (GIFI)?	6
Why introduce GIFI?.....	7
How was GIFI developed?.....	7
Who will benefit from GIFI?	8
Which businesses are included and not included in GIFI?.....	8
What is the structure of GIFI?	9
What is GIFI-Short?.....	12
How are GIFI data collected by CCRA?	12
What are the benefits of using GIFI data?.....	13
What are the principal challenges facing Statistics Canada when using GIFI?.....	13
How to gain access to GIFI?	15
Background	16
The Business Register (BR).....	16
The Legal and Statistical Structures of Businesses.....	17
Consolidation and allocation of Legal Entity data to the establishment level.....	18
How was financial information data collected in the past?	19
How can GIFI do what we have done in the past only better?.....	20
What is the present situation with GIFI?	21
What is the expectation for the future?	23
Tabulations	24
Metadata.....	24
Which industries are more accurately portrayed using GIFI?	28
Questions which help to show what is currently in the database.....	32
Personal and Laundry Services Industry	35
An example of the statistical use of GIFI.....	35
NAICS Sub-Sector 812 – Personal and Laundry Services.....	35
Personal and Laundry Services – Statistical estimation using 1998 GIFI data	39
BR information about the Personal and Laundry Services sub-sector	41

Comparing the BR and GIFI data for Personal and Laundry Services	43
Possible solutions in comparing GIFI data with the BR.....	45
Problems encountered using GIFI data with respect to the Personal Laundry Services Industries.....	46
Conclusion	46
References.....	47

Introduction

Motivation of the paper

Canada Customs and Revenue Agency (CCRA), previously known as Revenue Canada, is undergoing many changes in their methods of collecting corporate income tax. The redesign of their systems will allow CCRA to receive corporate financial statements electronically in a secure and timely manner. To support their new system, CCRA has created an index of financial items that will standardize the collection process while allowing accountants flexibility in reporting. This index is called the General Index of Financial Information (GIFI) and will be mandatory for corporations to complete by the end of the year 2000. This means that for the 2000 tax year, all incorporated businesses will submit their financial statements using GIFI codes. GIFI will eventually be made mandatory for unincorporated businesses too, and this will probably occur within the next 3 to 5 years.

Statistics Canada has legal access to financial statements for statistical purposes. In the past, gathering financial information from CCRA has been both costly and slow. In addition, the data received either contained little information about many corporations or a lot of information about a small sample of corporations. GIFI will provide not only a very detailed set of statistics, but will also be a census of 1.3 million corporations. This data set will create an opportunity to greatly improve the quality of statistics produced at Statistics Canada, in a timely and inexpensive manner.

There are many benefits for Statistics Canada programs. The level of detail and accuracy in these statistics is high. While GIFI can be used to improve statistical quality, the survey response burden placed on business can also be reduced. The data will provide potential benefits for both old and new statistical programs, notably those generating the much-needed provincial statistics for the Harmonized Sales Tax (HST) allocation. There are also many technical and methodological challenges Statistics Canada will have to face in order to make good use of these administrative data. These include the process of consolidation and allocation of legal entity data to the establishment level. There is the issue of confidentiality, the challenge of mapping GIFI information directly or indirectly to that on survey questionnaires, altering systems to deal with large amounts of data and dealing with an increased number of users.

The purpose of this paper is to present an overview of GIFI at a very basic level and to encourage new users to take advantage of the breadth and depth of this data set. The paper is in four sections. The first is the introduction and describes the background of GIFI and what the benefits are to Statistics Canada. The next section is the background, giving the user an idea of how financial data were collected in the past, the present situation and the future of GIFI. The third section includes tabulations from the entire data set and describes the Metadata. The last section contains GIFI-based estimates for a particular industry chosen as an example, the Personal and Laundry Services industry. Reading this paper from start to finish will give the reader a synopsis of GIFI. Alternatively, the reader can choose particular sections based on her/his needs.

The Harmonized Sales Tax (HST)

The Harmonized Sales Tax (HST) was implemented on April 1, 1997 in Nova Scotia, Newfoundland, New Brunswick and by the Government of Canada. The HST is the combination of the Goods and Services Tax (GST) and the Provincial Sales Tax (PST) for each province. In most provinces, corporations are expected to fill out a tax form, not only for each province they do business in, but also for the Government of Canada. This has placed a large burden on companies doing business in more than one province. Consequently, the federal government and three provincial governments introduced the HST. The HST is collected at the same rate of 15%

in each province and on the same types of goods and services. In addition, only one tax form needs to be filled out, one group of administrators oversees everything and all of the structures and rules are the same.

The HST is a value-added tax. This means producers are credited the HST they pay on their intermediate goods. Intermediate goods are goods that are used to produce final goods or services. If another producer then uses these goods or services as their intermediate goods, thereby producing another good or service, they will also be credited with the tax. In effect, the tax is being pushed down the line to the final product. Therefore, the final consumer of the product pays the value-added tax on the entire production process.

CCRA is the agency in charge of collecting all of the HST. Officials from the treasury departments administer the formula for allocating HST revenues among the governments. The HST form that businesses fill out includes the total HST they have paid for their purchases and their entitled tax credits for their inputs in production. This net amount is then sent to CCRA where all of the HST is put into a common revenue pool. The revenue pool is currently made up of the HST and the GST from all non-harmonized provinces. CCRA receives these revenues without sufficient data to indicate how the portions are to be split among the four governments. To tackle this problem, the HST revenue allocation formula was created. This formula requires detailed and reliable statistics for each province to determine the breakdown of revenue for each government.

The role of Statistics Canada and PIPES

The HST revenue pool could have been allocated among the four governments by requesting detailed information from all HST registrants about their taxable transactions. However, that would have imposed a heavy compliance burden on those businesses. Hence, the four governments decided to consider using macroeconomic statistics instead. The focus of Statistics Canada before this time was on national rather than provincial data, although some provincial information was being produced. Also, the input-output tables for each province were not being produced, the detail and reliability of data between provinces were not very consistent and there was not enough commodity and industry information available for the allocation formula. Statistics Canada had a long way to go to produce high quality statistics for use in the revenue allocation formula.

The System of Provincial Accounts provides a great deal of detail and covers a large range of pertinent economic information about each province. The data describe goods or raw materials used in production, what goods are being produced, who buys what products, sources of income in the province and the flow of money in the provincial economy. The System of Provincial Accounts is made up of three components: The Provincial Economic Accounts, the Provincial Production Accounts and the Provincial Input-Output tables. The Department of Finance uses these statistics in the revenue allocation formula to calculate shares needed for the HST allocation. Using these data within the HST formula will be more efficient compared to collecting the information directly from all businesses.

A project was set up at Statistics Canada to bring the quality of provincial statistics up to the required level. This project is called the Project to Improve Provincial Economic Statistics (PIPES). The project's goals are to upgrade the Provincial Economic Accounts, to produce annual Input-Output tables for each province and to ensure there is an equal level of detailed and accurate statistics for each province. This is quite a challenge given that no similar project has ever been carried out at Statistics Canada or for that matter any other statistical agency in the world. PIPES was set up to improve the quality of data and the divisions that have been involved in PIPES now have the job of maintaining and further improving these provincial economic statistics in the future. As a result of the project, there has been an increase in the quality of surveys and administrative data. In turn, this will produce better data needed for the HST.

PIPES targets both survey and administrative data sources. The most important type of administrative source is the corporate income tax file. In the past, Statistics Canada received these financial statements from CCRA in the form of an 'admin' universe file with a few variables and a sample file with more detailed information. In effect, the admin file contains little information about many corporations whereas the sample contains a lot of information about a few corporations. The General Index of Financial Information (GIFI) will be replacing both the admin file and the sample. GIFI plays a very important role in PIPES because it contains a lot of information about many corporations. In fact, it includes every piece of financial information reported by the corporation to CCRA for nearly the entire corporate population. GIFI is such a valuable resource because it is a census comprised of detailed, accurate, provincial data vital to the building of the provincial economic accounts that in turn are crucial for the HST allocation process. In addition, GIFI contains sufficient detail on costs to permit the calculation of gross revenue, intermediate expense and value-added. Using this valuable resource, PIPES is able to keep respondent burden down and limit costs while improving the quality of the statistics produced.

The Unified Enterprise Statistics Program (UESP)

To implement the HST allocation formula, there need to be provincial statistics that are consistent and coherent with substantial breadth and depth. With this in mind, the Business and Trade Statistics field conceived the Unified Enterprise Statistics Program (UESP). The purpose of this program is to increase the quality, consistency and detail of business statistics while minimizing the burden on respondents.

One major purpose of the UESP is to improve respondent relations through the unification of surveys. UESP encourages integrated questionnaires, sampling, collection and post-collection processing. To do this, all questionnaires can be sent to the enterprise headquarters, rather than separately to their individual establishments, if the business so prefers. This also helps to increase the coherence of the data. Surveying at the enterprise level helps to verify the profile of each enterprise, which in turn contributes to a better understanding of complex enterprises. By integrating surveys, the agency increases the consistency of the data. Also, standardizing the terminology and concepts of the questionnaires helps in attaining this. This ensures that an enterprise that receives many questionnaires receives them in a coherent and reasonable format without duplication of questions. Other parts of the program include the creation of one survey frame. This survey frame is the business register (BR). There is also a main contact management system to measure the response burden on each enterprise. Last but certainly not least, the program is designed to increase the use of administrative data.

How are administrative data used?

By increasing the use of administrative data, it is possible to produce more data with more breadth and depth, as well as a more complete coverage of the value-added statistics that are needed for the provincial accounts. The breadth of the economic statistics is improved by closing data gaps and the depth is improved because far more data records are available, thus permitting estimates to be produced in more detail by industry and province.

The administrative data can be used directly for enterprises with one legal entity (LE) and with activities in only one industry and one province – in other words, for 'simple' enterprises. The data can be used directly to generate production statistics. When there is one LE but the activities of the business are in more than one industry or province, then administrative data can be used for enterprise statistics. However, they cannot directly provide provincial or industrial information about production unless there is a profile of the enterprise and there is information about the allocation of variables from the enterprise level to the establishment level. For enterprises with

more than one LE and covering more than one industry or province, administrative data cannot be used directly for either establishment or enterprise statistics without a profile of the enterprise and information for use in the consolidation and allocation process.

GIFI data can also be used to stratify a sample, which is to allocate sample size over industries, provinces and size groups. This is because Gross Business Income (GBI), a key stratification variable, can be obtained directly from GIFI. Also, GIFI is being used alongside the CCRA Business Number (BN) to choose suitable donors for imputation in cases of non-response. GIFI combined with other tax data provides accurate information for modeling techniques and statistical estimation. This helps greatly in minimizing respondent burden, lowering survey costs and increasing reliability of detailed survey estimates. Consequently, surveys are gradually moving to this unified framework.

What is the General Index of Financial Information (GIFI)?

The General Index of Financial Information database, otherwise known as the GIFI database, is a census of corporate financial statements available from the Canada Customs and Revenue Agency (CCRA). It is an index, based on a unique coding system that allows corporations to prepare their financial statements as they please and report them to CCRA in a more standardized, efficient and timely manner. The index is made up of items most commonly used by corporations on income statements, balance sheets and statements of retained earnings. Each item has a unique four digit code. These codes range from 0000 to 9999. The balance sheet statement has codes ranging from 1000 to 3849 and the income statement codes range from 8000 to 9999. All together, there are currently 685 codes being used on GIFI, this means of the 10,000 codes available there are plenty of codes left for possible future expansion. Below is a chart with a few examples, giving an idea of what variable codes are available in GIFI:

BALANCE SHEET		INCOME STATEMENT	
1000	Cash and Deposits	8000	Trade sale of goods and services
1120	Inventories	8160	Fishing revenue
2360	Long-term loans	8910	Rental
2700	Short-term debt	9570	Rebates
3849	Retained earnings / deficit - end	9999	Net income/loss

The purpose of the coded system is to allow businesses the flexibility to structure their financial information the way that they choose. When a business prepares their financial statements to be attached to their T2 or T1 tax form, each item will receive a GIFI code; this code along with the Canadian dollar amount is then sent electronically to CCRA, with the aid of suitable computer software. The first year of implementation is the 1998 tax year. For both 1998 and 1999, all corporations fill out their tax forms as usual and CCRA transcribes these items to a GIFI code. For tax year 2000, it will be mandatory for corporations to complete their tax returns electronically. This new data set will provide Statistics Canada with much needed detailed, accurate and timely corporate financial information never before had by the agency.

Why introduce GIFI?

CCRA is modernizing its systems for collecting corporate income tax. By taking advantage of new technology, the agency is developing a means to transmit corporate tax information electronically. Therefore, their corporate EFILE system will be overhauled to transmit tax information and their processing system will be redesigned to handle tax files sent electronically. CCRA has developed GIFI to support the new electronic system. As noted, this index is a way of identifying financial information electronically.

The motivation behind CCRA's systems redesign is to provide a faster, more accurate, secure and efficient way to process corporate income tax information. When corporations send in their tax forms electronically there is no need to transcribe items from paper to computer. This increases the accuracy and efficiency of CCRA. In addition, verification of information will be done electronically instead of manually. The former labour-intensive method was slow, now computers are doing the verifying almost instantly, thereby making the procedure more precise and efficient. For businesses this will prove to be a significantly faster way of processing tax returns, with the lag decreasing from a previous 60-day return to an expected return of only a few days. CCRA audits will also be facilitated by the more complete database. There will also be savings realized through less paper handling, reduced storage costs and lower operating costs. Finally, an encrypted code will allow financial information to remain secure.

This will be more efficient not only for CCRA but for businesses, tax professionals, accountants and government agencies.

How was GIFI developed?

GIFI was developed by officials not only from CCRA, but also from Statistics Canada, the Department of Finance and the Alberta and Ontario provincial governments. These organizations consulted such associations as the Canadian Institute of Chartered Accountants, the Financial Executives Institute, the Certified General Accountants' Association of Canada, the Tax Executives Institute, the Society of Management Accountants of Canada and the Canadian Federation of Agriculture. They also consulted accountants and software developers and looked at hundreds of corporate income tax returns to create the GIFI system.

In the past, CCRA has received financial statements sent by a business in a style unique to the accountant of that business. This allows accountants to make their product better tuned to the needs of their corporate clients. On the other hand, CCRA wants to take advantage of the opportunities for increased efficiency that are afforded by using electronically sent tax forms and the only way to do this is with standardized forms. If corporations continued to send in their statements in their own customized style, this would defeat the purpose of an electronic system. Initially CCRA decided on a completely standardized form. Accountants did not like this idea because it did not leave them with a lot of flexibility. CCRA then came up with the idea of an index. The purpose of an index is to allow flexibility for accountants while keeping the tax form relatively standard. This allows CCRA to deal with a select number of variables while providing accountants with the opportunity to choose which variables they would like to use. Of course, this index includes generic variables for any items that cannot be routinely slotted. These codes have been set up so the addition of new variables will not require changing the previous codes. For example, 8299 will always be total revenue.

There are several ways an accountant might view a corporation's financial flows. This is important for the individuals developing GIFI to understand and take into consideration. One is according to the nature of the expense. This involves such elements as wages, utilities and so on. Another way is to divide the organization by product line. This involves separating each product and having all of its expenses and transactions in one group. A third classification is the process line.

This divides the accounts according to separate production processes, as in a vertically integrated business. Yet another way is to look at the different functions of the organization, for example, training, administration, retailing and so on, and structure the accounts on that basis. Accountants choose an approach along these lines when they are setting up financial statements. GIFI combines all of these approaches to make the index a flexible one.

Another part of the development process is that of the tax preparation software. In order for corporations to file their tax forms electronically, they must prepare their statements in a tax software package. CCRA has encouraged accounting software package companies such as ACCPAC and BEDFORD to redesign their accounting software to support GIFI. Links have to be developed between choices in the accounting software and GIFI items. The accountant fills out the tax form using the software and then clicks a button. The software then automatically assigns a GIFI code to each item and sends it to CCRA. This accounting software is presently available, but the system is not yet in place at CCRA to receive the information electronically so corporations must mail in their GIFI codes and amounts for the time being until the system at CCRA is ready to receive forms on-line.

Who will benefit from GIFI?

Many individuals and organizations stand to benefit from GIFI. Accountants will receive quick and detailed replies to their questions from CCRA because the information will be quickly and easily accessed and processed by CCRA. The response time of their tax returns will be greatly improved. Auditors can also use GIFI for compliance reviews.

The Statistics Division within CCRA uses GIFI to produce statistics and feed data to the Department of Finance for tax simulation models. These simulation models help Finance develop tax policies and legislation, and with GIFI they will be able to do so in a more timely manner. Statistics Canada will also benefit from the introduction of GIFI. This database will help to improve the statistics used in the system of national accounts, industry statistics, financial statistics and other statistical programs.

Which businesses are included and not included in GIFI?

GIFI is made up of nearly the entire T2 and T1 business population. Incorporated businesses fill out T2 tax forms. An incorporated business is one that is a legally separate entity from the owner. A T1 tax filer (or unincorporated business) is a business that is not legally separate from the owner. This is usually a smaller business and can be something as simple as a computer consultant working from home or an individual owning a hobby farm. Presently, all T2 corporations are in GIFI and in the future T1 businesses will be included also. Joint ventures or partnerships are also covered by GIFI. They are defined as two corporations that have joined for a point in time to perform a task together, but which are not considered one corporation. Because of this, each corporation fills out a T2 separately for its own activities and one jointly for the combined task.

All of the financial statements included in the GIFI database are at the legal entity level. A legal entity (LE) is an incorporated or unincorporated business or company, technically defined as an entity granted by law the power to own assets, borrow resources and transact business (PIPES Technical Series 1, 1997). The basic GIFI data are at the legal entity level and have not been consolidated to the enterprise level and allocated to the establishment level.¹

¹ *The allocation and consolidation processes are explained more fully in the section entitled "Consolidation and allocation of Legal Entity data to the establishment level".*

T2 and T1 filers include both simple and complex enterprises. A simple enterprise is a business that operates in only one province and its activities are only in one industry. A complex enterprise, on the other hand, is defined as a business that has functions in more than one province or in more than one industry. An industry is defined by the North American Industrial Classification System (NAICS), a classification system that Statistics Canada uses to group businesses into different 'industries'.

All filers will be required to fill out GIFI reports except insurance companies and non-resident corporations. GIFI is not structured to fit the needs of insurance companies that have activities in underwriting, for example, life insurers, deposit insurers or property and casualty insurers. Presently, the Industrial Organization and Finance Division (IOFD) is running a survey to include these insurance companies. Eventually, insurance companies will be added to GIFI once financial items relevant to the industry are agreed upon.

Inactive corporations also use GIFI. Also, the first year after a business becomes incorporated they fill out two balance sheets, both an opening and a closing balance sheet. However, corporations do not use GIFI when they fill out a deferred expense statement because of their participation in development activities that have not yet produced significant sales revenue and have deferred expenditures.

What is the structure of GIFI?

There are 685 financial items currently available on the GIFI database. Of these items, most businesses report 40 to 50. A few, about 5% of all businesses, report only the mandatory variables, while others report even more than the average 40-50 items. The index contains items that are found on both a Balance Sheet and an Income Statement. The Balance Sheet accounts for 48% of all GIFI items and is divided into three sections as listed below. Each percentage is the proportion of total items.

Total Assets	34%
Liabilities	13%
Retained Earnings	2%

The Income Statement makes up the remaining 52% of GIFI categories and is divided up into two sections:

Revenue and Expenses	34%
Income and Expenses	17%

Within these sections are sub-sections. Total Assets are comprised of Current Assets, Capital Assets and Long-Term Assets. Total Liabilities are made up of Current Liabilities, Long-Term Liabilities and Shareholder Equity. Retained Earnings Information only has one sub-sector known as Retained Earnings Information. The Revenue and Expenses section from the Income Statement contains three sub-sections: Non-Farming Revenue, Non-Farming Cost of Sales and Non-Farming Operating Expenses. The Income and Expenses section is broken down into Farming Revenue, Farming Expenses and Extraordinary Items and Income taxes. A more detailed breakdown is available in Table 1 below.

Table 1 – The Structure of the GIFl Balance Sheet and Income Statement Items

Structure of Balance Sheet Items in GIFl			
Section	Sub-Section	Block Totals	Total Items
Assets	Current Assets	12	61
	Capital Assets	14	124
	Long-Term Assets	10	47
	Total	36	232
Liabilities	Current Liabilities	11	39
	Long-Term Liabilities	12	38
	Shareholder Equity	5	10
	Total	28	87
Retained Earnings	Retained Earnings	6	12
	Total	6	12
Total Balance Sheet Items:		Total	70
			331

Structure of Income Statement Items in GIFl			
Section	Sub-Section	Block Totals	Total Items
Revenue and Expense	Non-Farming Revenue	10	67
	Non-Farming Cost of Sales	10	47
	Non-Farming Operating Expenses	25	121
	Total	45	235
Income and Expense	Farming Revenue	7	52
	Farming Expenses	5	59
	Extraordinary Items and Income Taxes	7	8
	Total	19	119
Total Income Statement Items:		Total	64
			354

Total GIFl Items:	Total	134	685
--------------------------	--------------	------------	------------

Within these sub-sections, block totals have been defined and calculated by the Tax Data Division (TDD). There are 70 block totals in the balance sheet and 64 block totals in the income statement. Altogether there are 134. These block totals are not part of the 685 total GIFl variables. A block total is defined as the sum of dollar items. They can be the sum of many items or just one item. Which items are included in a block total is not always obvious.² A block total is easy to recognize because it is either highlighted on a list of GIFl variables³ or it is a four-digit unique code preceded by the letters BT in GIFl tables⁴. Block totals are not to be confused with generic items, although both have four-digit codes. The difference is a generic item is a variable on the GIFl. It is not a sum like the block totals. Rather, a generic item is an item the accountant will choose if she or he is not sure which item best fits. This allows flexibility for the accountant. A generic item has a four-digit code just like a block total, but the code is preceded by an L. In summary, the block total is a sum of many items including the generic variable, but is a variable

² See the "Conceptual and Physical Data Model" under "Metadata" on the TDD web site.

³ A complete list of GIFl variables can be found on CCRA's web site under the Corporate income tax section.

⁴ Tables are explained in more detail in the Metadata section of this paper.

created by Statistics Canada. The chart below includes examples of both block totals and generic variables, and how these amounts are derived.

Code	Variable Name	Block Total (Derived by SC)	Generic Item (Available on the GIFI)
1000	Cash and Deposits	This field is the sum of items 1000 to 1007	Includes any amount of cash and deposits, which cannot be broken out into fields 1001 to 1007
2700	Short-Term Loans and Debt	This field is the sum of items 2700 to 2706	If it is not possible to determine the breakdown of short-term debt, then enter the entire amount in this field.*
8520	Advertising and Promotion Expense	This field is the sum of items 8520 to 8524	Include amounts reported as advertising or promotion expenses which are not included in fields 8521 to 8524.
9180	Property Taxes Expense	This field is just item 9180	**

* This field includes exclusions that have not been mentioned in this chart.

** There are no items similar to this field.

The GIFI filer must ensure that certain GIFI variables balance. There are four rules for balancing. These rules are as follows:

Total Assets = Total Liabilities + Total Shareholder Equity
 Total Revenue – Total Expenses = Net Non-Farming Income
 Total Farm Revenue – Total Farm Expenses = Net Farm Income
 Net Income / Loss = Net Income / Loss after taxes and Extraordinary items

It is also required that each corporation include all mandatory variables requested by CCRA. There are nine mandatory variables. Of these nine variables, there are either six or seven mandatory variables that each corporation must report depending on the activities and situation of that corporation. These variables are as follows:

Mandatory:

Item 2599 total assets
 Item 3499 total liabilities
 Item 3620 total shareholder equity
 Item 9999 net income/loss after taxes and extraordinary items

Mandatory if corporation has non-farming activity:

Item 8299 total revenue
 Item 9368 total expenses

Mandatory if corporation has farming activity:

Item 9659 total farm revenue
 Item 9898 total farm expenses

Mandatory if there are any amounts recorded in items 3660 to 3744 (Retained Earnings)

Item 3849 retained earnings/deficit-end

In the future, new items will be introduced. In fact, there are already some codes set aside for this purpose. Some codes will be exclusively used by T1 or unincorporated businesses and others will be used for the Net Income Stabilization Account (NISA) program. NISA is a safety net program for farmers, offered by both the federal and provincial governments for income stability.

What is GIFI-Short?

GIFI-Short was designed for corporations that do not use tax preparation software when filing their tax forms. The corporation must have both gross revenue and gross assets less than three million dollars. GIFI-Short is a paper version of GIFI and contains 100 of the most commonly used balance sheet and income statement items. Included in this form is a separate section that contains items essential to farming corporations. This is only a short-term solution. It will be mandatory for all corporations to fill out their financial statements using GIFI software in the near future, but there is no firm date for this yet.

How are GIFI data collected by CCRA?

It is mandatory for each corporation to send in its financial statements each year to CCRA. These financial statements are divided into four parts. The first is the Balance Sheet. This is made up of details for assets, liabilities and the statement of retained earnings. The second part is the Income Statement. The Income Statement includes details of revenue and expenses for both farming and non-farming corporations. The 'Notes Checklist' is the third form and contains a few questions to help CCRA determine who completed the tax form and the extent of their involvement. It also helps determine the type of information in the Notes to the Financial Statements. These Notes, the final item included with GIFI, are also mandatory. They explain anything within the financial statements that the corporation feels needs to be explained. For example, they could explain what makes up an item or how items are broken down. When the electronic version of GIFI becomes mandatory, in October 2000, these notes will have to be sent in a text format.

CCRA is presently in a transition mode from their old, manual system to the new, electronic system. Corporations send their financial statements and T2 tax returns to their local CCRA branch. The local branches then assess each tax return as they normally would. Currently, tax items are transcribed to GIFI items by CCRA. The individual capturing the data enters in the code followed by the dollar amount of each code, followed by an asterisk. For example, 1000 10000* 1060 25000* and so on. These data then go through a series of very limited edits. If the file passes the edit stage, it goes straight to the Taxation Centre (TC). There are six centres; this is where the data are stored. If the data do not pass the edit stage, they go through a re-key and, whether they pass or not, are sent to one of the TCs. Each file that fails an edit receives an invalid flag. From the output centre, the files are then sent to the Business Returns Processing Division (BRPD) at CCRA headquarters. From there, the tax information is sent to the Statistics Division at CCRA, the final stage before being sent to Statistics Canada. The Tax Data Division (TDD) at Statistics Canada receives eight SAS tax files. One being the invalid data file. The rest are the files which passed the edit stage. Each file contains a Business Number (BN), a GIFI code and the associated dollar amount for each code. TDD transfers all of the data from these files into an Oracle database for use by subject matter analysts at Statistics Canada.

The difference in the future is that there will be no need to transcribe tax items to GIFI codes and no need to have data processors entering data from paper to computer. This will all be done electronically using accounting software packages. In the first year of GIFI, which was the 1998 tax year, GIFI software was available for corporations to make use of. Although the software was not available until around June 1999, businesses that had not yet filled out their 1998 tax forms were encouraged to use the software. For 1998, around 3% of corporations submitted their tax return in GIFI format and approximately 97% tax forms had to be transcribed by CCRA. It is expected in the 1999 tax year that the number of filers sending in their returns in GIFI format will increase, while those files needing to be transcribed will decrease. All filers covering tax years

ending after December 31, 1999 must file in the GIFI format. However, financial statements cannot be filed electronically until October 2000, when the new CORTAX system goes into operation at CCRA. After this time, it will be mandatory for corporations to file their statements electronically.

What are the benefits of using GIFI data?

Statistics Canada can reap many benefits from the introduction of GIFI. Overall, the quality of statistics produced by the agency will greatly improve. This is because GIFI is a census of financial data that is both detailed and accurate. The process will be more efficient because almost the entire population of T2 and eventually T1 businesses' financial statements will be available on GIFI. The database will also provide very detailed statistics due to the increased industrial, regional and size-of-business statistics available. The accuracy will greatly improve because there will be no need for manual re-keying so there is less potential for human error. In addition, all verification of data will be done electronically. With these electronic verification checks the data can be received by Statistics Canada in a more timely manner, therefore increasing the production timeliness of the statistics the agency produces. There is also the potential to facilitate monitoring of deaths, births and restructuring of businesses. A business is presumed inactive or dead if it has not filed a tax return. Businesses in this category receive a flag on TDD's GIFI database. Births are noted in a function called the Opening Balance Sheet Indicator (OBSI) which can be accessed in the Balance Sheet table of GIFI.

There are other benefits from GIFI. The main one is the reduction of burden placed on survey respondents. This can be done in one of three ways. Linking GIFI items to questions on the surveys can shorten surveys. Another way is to replace surveys biennially, triennially or entirely by making use of GIFI for the required data in these off years. In addition, GIFI can be used to cover the very small enterprises in an industry. These enterprises, defined as those with less than \$30,000 in annual revenue, are never surveyed. Finally, it increases the power of a survey sample for the small to medium size enterprises. There is also a potential for costs to be lowered. This can occur through the elimination of transcription and the reduction in paper handling costs. The reduction in paper also means faster results because there is no more searching for financial statements, no more photocopying and no more transmitting all of this information physically to Statistics Canada. Another important issue is confidentiality. Financial statements are sent from the corporation to CCRA and from CCRA to Statistics Canada using an encrypted code to ensure the security of the information.

What are the principal challenges facing Statistics Canada when using GIFI?

There are many challenges that Statistics Canada will have to overcome to fully exploit the GIFI database. One of the first obstacles is coping with a huge volume of raw data from CCRA. A new system had to be designed and developed to deal with the situation. So far, there is enough storage capacity in the system, but there are still more records to come from CCRA. There will be more divisions trying to access GIFI so the number of technical support questions will increase. There is also the issue of accuracy. For reference years 1998 and 1999, CCRA is transcribing the data manually. This means there is potential for human coding errors and there is the concern that CCRA has not had sufficient resources to read all the notes to the financial statements. Even in reference year 2000, when the data are collected and verified electronically, there are still the concerns that GIFI codes automatically being assigned within the accounting software may not always be accurate for each business since there are likely to be transition problems.

Another challenge involves businesses choosing a tax-filing period. Chances are they choose their own fiscal period as their reporting period. This varies from business to business. By law, this period must be any consecutive 12-month period. Of course, this is not the case for new firms, dead firms or merging firms, which can report less than twelve months. Ideally, Statistics Canada would prefer each business to file at the end of the calendar year. To deal with this problem, the agency must 'calendarize' the data. If a form is sent in, but its fiscal period is not January to December, then the data must be shifted to fit that calendar reference year. In the past, this problem was often ignored and fiscal periods were added together as if they were all calendar year periods. That approach was sometimes an acceptable approximation when the resulting statistics applied only at high levels of aggregation, but it is no longer adequate when estimates are produced for a large number of province-industry cells. Another problem occurs when LEs have more than one reporting period. This can happen either because a LE splits into two LEs, or two LEs merge. Presently, the agency is dropping all 'extra' records with the shorter fiscal period to avoid this double counting. In the future, we would like to merge all records together by adding all the flow variables together, taking the stock variables from the last record, and re-balancing the final record.

Keeping data confidential is difficult but absolutely essential. The *Income Tax Act* and the *Statistics Act* protect corporations from disclosure of their financial statements. CCRA imports and exports financial statements using a very sophisticated, encrypted computer system. Once the data reach Statistics Canada, there are two main strategies. One is that employees using the data must be specifically authorized by management to do so. The user of these data must sign a form stating that they understand the data are confidential and must be treated appropriately. In addition, the Duffet rules are used to avoid inadvertent residual disclosure of confidential information.⁵

GIFI data are received from CCRA at the legal entity level. In general, legal entities may have more than one establishment, engaged in more than one industry. This makes it difficult to assign a unique NAICS code in some cases. If a business has one and only one establishment associated with each LE, known as a one-to-one relationship, then assigning NAICS codes is straightforward. If there are several LEs, and/or one LE with several establishments in different industries, this can cause problems. There are about 1500 enterprises in Canada that have more than one LE and many of these are among the largest businesses in the country. Today, to assign a NAICS code to one of these LEs, TDD goes to the Business Register (BR). On the BR, each LE has a corresponding enterprise and each of these enterprises are assigned a NAICS code based on the dominant activity of the enterprise. The LE is assigned the same NAICS code as the enterprise. But this is problematic. For example, a LE may have activities in retail, but because the dominant activity of the enterprise is manufacturing, if a summation of GIFI records is done the data for the retail industry will end up in the manufacturing industry. This means that industry data are subject to being either under- or over-estimated. A task force is presently looking at the 1500 multi-LE enterprises and considering what the best approach might be to deal with this problem. Fortunately, for most GIFI records (covering the small businesses) this problem is not relevant.

Linking GIFI records to the Business Register via the CCRA Business Number and then assigning the BR's NAICS code to the GIFI record obviously does not work if the links are not there. It is important that these links exist to maximize the number of industrially coded records available. If there are missing links then essentially records on GIFI will be lost. There are many reasons why these links might be missing. There are BNs on GIFI that cannot be found on the BR. Even if there are no BN linkage problems, it could be that a NAICS code has not been assigned on the BR. Another task force is presently addressing the BR linkage problem.

⁵ The Duffet rules are referred to again in the tabulation section of this paper under the analysis outlined in Table 6 - NAICS 452991 - Home and Auto Supplies Stores.

In general, administrative data do not cover all of the information needed by Statistics Canada. Sometimes the statistical units that we would like to look at are just not available. This is because GIFI is only a financial data set, so non-financial information such as commodity inputs and outputs is not available. To fill in these gaps, Statistics Canada will always need surveys. Another problem is linking particular administrative variables to survey questions. Sometimes GIFI variables do not conform perfectly to our target variables and need to be extracted indirectly, for example, by adding a few items together. Finally, even if a GIFI item is a direct match to a target variable, there is no guarantee that the detail required will be available on GIFI. Just because there is a code for the item does not mean that all businesses will choose to fill it out.

How to gain access to GIFI?

Corporate financial information is highly confidential information. The federal government has policies protecting the confidentiality of a corporation's financial statements. For any individual to gain access, they must receive a security clearance. Authorization must also be received from a supervisor and the user must sign the data access form indicating he/she understands the confidentiality issue behind the use of the GIFI data set. Every GIFI user must abide by a number of important security procedures:

1. The user cannot have an A/B switch on their computer. This is because the B switch is external and has the potential for unauthorized external individuals to hack into the system and retrieve sensitive information.
2. Summary tables, on a personal computer, must include the warning "This Document contains semi-aggregated data that has not been checked for confidentiality. MUST BE TREATED AS 'CONFIDENTIAL'".
3. There are special procedures when printing from the mainframe.⁶
4. Users must ensure their workstation is always locked when leaving it unattended. This will deter unauthorized internal individuals from misusing the data.

Once the user understands the importance of security, he/she can go to the Tax Data Division (TDD) web site, under Data Access, to get more information about accessing the data. There are two documents. The first explains the procedures for gaining access and the second is an electronic copy of the data access authorization form. Forms are also available from TDD. The user will have to give a reason for wanting access to the data, and will have to give a description of which data they would like (in this case it is GIFI data). The Tax Data Division also needs to know who will be using the data and when data access will be terminated. Both the user's director and the director of the tax data division must grant authorization. How long it takes to get access once the request form has been completed depends on the availability of staff at that time. Usually the wait is a few days to just over a week.

Once access has been granted, the Tax Data Division will let the user know. The user will then have to choose a password that will able him/her to access the data. The password will need to be entered each time a program is run. There are two ways to enter your password. One is by connecting to ODBC and creating a PROMPT. This prompt is more secure because it asks the users to enter their password twice, while the program is running. On the other hand, it is time consuming. The second method involves connecting directly to Oracle. In this case, the user's login, password and the path they wish to access are written right into the code. This is the easiest method because the password only needs to be entered once, but then extra care must be taken to ensure the secrecy of the password. The two commands in SAS, within the PROC

⁶These special procedures and other information can be found in a document entitled "Procedures for gaining access to Tax Data". This document is on TDD web-site under Data Access.

SQL prompt, look like this:

```
Connect to ODBC as mydb (PROMPT);  
Connect to ORACLE as mydb (user=lastfir password=XXX path="...");
```

Another handout will be given to the user when access is granted. This explains how to hook up their personal computer to the Oracle database. Of course, the user's LAN Administrator can help them if they have any questions or difficulties.

To get started, there are a few skills the user will need to have. The data are in an Oracle database, but the data are not directly manipulated with Oracle tools. To manipulate the data, the user will need to learn either SAS or MS-Access. To access the data, the user will also need to learn the basics of SQL. MS-Access is a fairly easy software tool to use. It has a user-friendly interface and produces nicer tables, while SAS is often the preferred choice because it is more flexible and efficient when manipulating large data sets. For both the SAS and MS-Access user, running small programs that do not call for a lot of variables or a lot of manipulation of these variables is fine. If many variables are needed then programs will have to be either run at night or, more commonly, off the server.

The user will also need to learn the metadata of GIFI.⁷ This is not simple. The user will have to understand how GIFI codes are set up (both block totals and variables) and what the different commands are when accessing the most appropriate record. This includes anything from choosing a particular NAICS code to eliminating all records with multiple fiscal year endings. Finally, which pieces of data are stored in which tables and how these tables are linked in Oracle are both very important to understand.⁸ Getting started is probably the hardest part. There are sample programs, or templates, available from TDD. These can be copied into SAS or MS-Access to ensure GIFI is hooked up and running correctly.

Background

The Business Register (BR)

The entire business universe, consisting of every business in Canada, is three to four million enterprises in size. The vast majority of these businesses are very small and Statistics Canada does not have the resources to keep all of their records on file. By definition, a very small business is one that has total revenue below \$30K per year. If the business has total revenue below \$30K and is *incorporated*, then it is included on the BR. The BR contains information about all incorporated businesses and some unincorporated businesses in Canada. This file was created initially in the 1980's from the payroll deduction system. Only the companies having employees were on file. The BR had information about births, deaths and size of business. By the mid-nineties, CCRA had created the single Business Number (BN) system. A BN is a number that is assigned to every legal entity in Canada and is used to identify a business for tax purposes. This BN system, combined with the use of GST files that cover non-employers, permitted the BR Division to increase the size of the BR to 2 million. Of these two million enterprises, approximately one million are incorporated enterprises and one million are unincorporated enterprises. Currently, GIFI Data are available for all of the incorporated businesses, but none of the unincorporated ones.

⁷ Refer to the section of this paper entitled *Metadata*. For more information, refer to footnote two.

⁸ A copy of the tables in Oracle and how they are linked can be found on TDD's shared drive in a file called *Simple GIFI Diagrams*. This file is located under the tax dev. external directory, which is located under the external directory.

Enterprises are also broken down by complex and simple. A complex enterprise is an enterprise that contains establishments in more than one province or in more than one industry. A simple enterprise is the exact opposite. It operates in one province and has activities in only one industry. There are approximately 10,000 complex enterprises, which comprise around 70,000 establishments. Of these 10,000 only 1500 are multi-legal enterprises. Although complex enterprises are relatively few in number, they account for about half of all business output and most of them are incorporated. Simple enterprises on the other hand account for the approximately 1,990,000 businesses remaining on the BR. About half of all simple enterprises on the register are unincorporated and the other half are incorporated.

The Legal and Statistical Structures of Businesses

Statistics Canada receives business financial information from CCRA's tax data. Tax forms filled out by businesses are the responsibility of the LEs of that business. The legal level corresponds to either a T1 or T2 tax form (see Diagram 1 - Legal Structure). The Business and Trade Statistics Field is interested in collecting statistics based on industry activity. However, the legal structure does not necessarily split perfectly by industry. A complex business with establishments in more than one industry may have one or more than one LE, and if it has more than one then there may not be a one-to-one mapping between the LEs and the establishments. Therefore, to the extent that complex businesses play a significant role in an industry, it is difficult to produce statistical estimates based solely on LE-based tax data.

There is also an operating structure on the BR, which is the way that a business thinks of itself and its activity. Using both the legal and operating structure of a business, Statistics Canada creates a statistical breakdown of an enterprise known to us as the statistical structure. (See Diagram 2 - Statistical Structure).

Diagram 1 – Legal Structure of an Enterprise

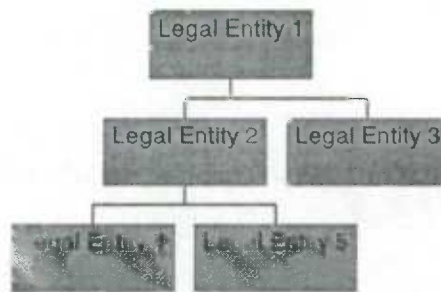
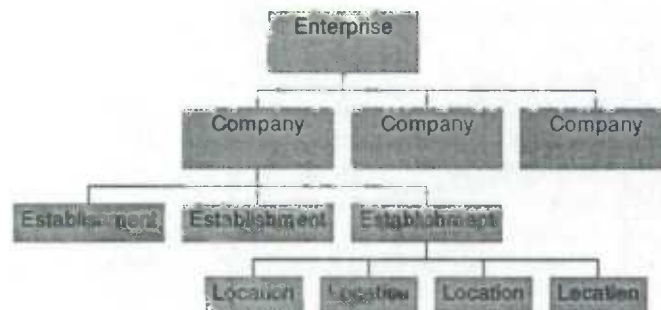


Diagram 2 – Statistical Structure of an Enterprise



The statistical structure is broken down into four parts. The easiest way to look at it is from the bottom up. The location is the physical place where business takes place. Usually a manager is allowed to make decisions for that location exclusively, but is not allowed to make decisions about such things as the accounts or hiring. The establishment is made up of a collection of locations that are in the same industry. The manager is allowed to make decisions about day to day affairs like hiring, selling and buying, and the accounts. This manager still has a boss and cannot make any long-term decisions. The company is comprised of several establishments. Decisions at this level are made on investment, financing, and any other long-term decisions or decisions that involve growth of the company. The enterprise is made up of all the companies, or essentially the entire business. It oversees everything and owns everything. In the case of a simple enterprise all of these levels coincide at the enterprise level. Statistics Canada is mostly interested in the establishment because this is where an industry code is assigned. The establishment units of a large complex enterprise are comparable with the small production units of simple businesses. Which industry an establishment is involved in dictates which survey each establishment belongs to. The reason why we need establishment level data are to keep all data consistent. If we want to add up any variables across industry, province or nation or if we want to compare variables then data must be consistent. It is also important to note that only income statement information is allocated down to the establishment level and balance sheet information is not. This is because establishments, by definition, are not in charge of such things as assets and liabilities. The enterprises and companies are in charge of balance sheet information. Therefore, balance sheet information is consolidated to the company or enterprise level, while income statement information is allocated down to the establishment level.

Previously the 1980 Standard Industrial Classification (SIC) was used to assign an industry code. The SIC was produced in the late 1970s, but the codes are now out of date because industries are not as they were 20-25 years ago. Statistics Canada now uses the North American Industrial Classification System (NAICS). Statistics Canada, the United States Office of Management and Budget, and the Instituto Nacional de Estadística, Geografía e Informática, Mexico, designed this coding system and published it in 1997. Not only are the codes standard for all of North America but the industries themselves are more up to date and, therefore, more applicable. In order to assign the codes we must go from LE level data to establishment level data. However, this is not always one step. Data must be consolidated up to the enterprise level and then allocated back down to the establishment level using survey data and the BR. Then an industry code can be assigned.

Consolidation and allocation of Legal Entity data to the establishment level

One challenge with GIFI is that the data are at the LE level. In order to get the data to the establishment level, they must be consolidated to the enterprise level (if the enterprise is one of 1500 or so multi-LE businesses) and then allocated back down to the establishment level (if the enterprise is "complex"). For simple enterprises with one LE and one establishment, no consolidation or allocation process is needed, so this is not an issue. These enterprises make up 99% of all businesses. For the remaining 1%, which account for about half of all business activity, this is not always an easy task. The process is outlined below.

Consolidation from the LE level to the enterprise level is done in two different ways. The first is using information from IOFD's Quarterly Financial Survey. This survey asks multi-legal enterprises specific questions about the enterprise as a whole. When this survey information is available, it can be used as the consolidated data for the enterprise, in place of the data from the LEs of that enterprise. The second way is to approximate and involves simply adding up the data from all of the LEs of the enterprise. Which LEs belong to which enterprise can be determined from the information that is collected via IOFD's *Corporations Returns Act* program.

The allocation of enterprise level data to the establishment level is a little more complicated because the enterprise does not think of itself in terms of the statistical structure. There are three ways to get establishment level data. The first is from part 2 of a Unified Enterprise Survey questionnaire. This part of the survey contains questions directed to an establishment of an enterprise. This is the preferred method because the information is accurate and detailed. The second way of allocating involves what is called an allocation worksheet. This questionnaire is similar to the part 2 but is more general. This is for businesses that have many establishments and require more flexibility due to time and other constraints. Data are collected from a collection entity (CE) which might, for example, consist of all establishments within a single province. One CE corresponds to many establishments and is defined for collection purposes. Of course, allocation must still occur from the CE level to the establishment, and this is done using data from the allocation worksheet. Finally, a third way to allocate is to go to the BR and find quantitative stratification information about each establishment such as, for example, the number of employees. From this, the relative size of the business can be determined and other variables can be split up between each establishment based on this size indicator. Of course, these variables are only rough estimates and are not always up to date, so this type of breakdown is not always accurate.

One major problem occurs when a LE files twice in the same year. This could be because the LE is trying to change its fiscal year or maybe a merger took place. The best way to deal with these multi-year filers would be to merge all records by adding all flow variables together, taking the stock variables from the last record, and re-balancing the final record.

There are also opportunities that Statistics Canada needs to take advantage of. For example, when a LE corresponds exactly to an establishment it would be advantageous for the BR to recognize this. This would minimize the amount of work involved in consolidation and allocation just to get the same breakdown. The problem is that it is time consuming to go through every complex enterprise to investigate this.

How was financial information data collected in the past?

T2 Survey Universe File (SUF)

IOFD's T2 SUF was created by CCRA for STC instead of from the BR, as was the procedure for the Tax Record Access group (TRA). CCRA produced a file called the CORPAC file, which contained a few pieces of information about a corporation as well as five financial variables. These variables are total assets, total equity, total revenue, total profit before tax and taxable income. This file was sent to IOFD by tape and contained about 1.6 million businesses. After IOFD cleaned and imputed the missing data it was renamed the ADMIN file. This ADMIN file is essentially IOFD's version of the BR. This process will continue until October 2000, at which time the CORPAC file will be replaced with the CORTAX file.

The Quarterly Financial Survey (QFS)

The Quarterly Financial Survey (QFS) is a questionnaire that IOFD sends out to a sample of around 6,000 incorporated businesses in order to gather quarterly financial information. Up to and including 1998, the sample was chosen from the T2 SUF whereas from 1999 on the sample will be chosen from the BR based on the UES SUF. The largest corporations are surveyed as well as one third of medium sized corporations but none of the smaller corporations. The questionnaire is sent to all multi-legal businesses where they are asked to give their consolidated enterprise level information. Estimates for the small businesses are developed by projecting information from tax data in previous years, using the survey information for medium size enterprises as the projectors. The estimates are revised subsequently when the annual tax data for small

businesses become available. After these data are collected, IOFD then uses weights to estimate for the entire population.

On the questionnaire, businesses are asked questions about their corporate income statements and balance sheets. This is because the SNA needs quarterly financial statistics for such accounts as the Quarterly income and expenditure accounts, Financial flow accounts, National balance sheet accounts and Balance of international payments. These accounts provide information about investment activity, financing patterns, rates of return and other financial performance ratios. There are other users of these data, for example, other government departments, private sector and the academic sector. These surveys will be ongoing because the SNA needs quarterly financial statistics. In addition, these surveys have been modified to include questions about value-added data, which is information crucial to the input-output tables.

QAFS (Quarterly and Annual Financial Statements)

Annual estimates, for the QAFS, were produced in the following manner. Quarterly data, which has been collected at the enterprise level, were annualized. Also, legal entity level tax data were 'rolled up' to the enterprise level. The profit variable, from these two sources, was compared at the enterprise level to check for any outstanding frame or data problems. Any taxation variables not collected on the quarterly survey were merged with annualized quarterly data. The final sample, therefore, consists of 'good'⁹ annualized quarterly survey financial data, supplemented by tax information for larger enterprises. For the rest of the incorporated enterprises, data was received from the three samples determined by the former Tax Record Access Section of Business Register Division. Weights are recalculated and applied to produce estimates.

How can GIFI do what we have done in the past only better?

The process of collecting financial data has changed so much from the past to the present. With the introduction of GIFI's unique coding system and electronic data system, things are expected to improve immensely. Financial statements will now be collected in a more timely, comprehensive, less costly and efficient manner. The data will be both detailed and accurate and enable the production of value-added statistics at the provincial level. In addition, the reliability of the data will improve because GIFI will contain almost the entire population of corporations. Finally, the financial statement information will be in a relatively standardized format making it easier for users to group statistics together for their own needs.

Information about simple enterprises will be available directly from GIFI. There will also be information about complex enterprises; it is just more problematic. It will be possible to reduce the survey burden on respondents. Surveys can now be done biennially or triennially instead of annually, by using tax estimates in the off years. The agency will be able to use tax data instead of survey data for smaller industries like the Personal and Laundry Services industries. In addition, GIFI can be used to cover the very small enterprises in an industry. These enterprises have less than \$30K in revenue and, therefore, are never surveyed. Finally, it increases the power of a survey sample for the small and medium sized enterprises.

Another development is the upcoming change in the corporate and personal income tax laws that allow provincial statistical focal points access to corporate income tax information. Under current laws, Statistics Canada is allowed access to financial statements but is not allowed to share statistics produced using these data with provincial focal points. Formerly, the statistics that the agency produced were mainly created from survey data. With the introduction of GIFI, statistics produced will be a blend of both survey and admin data. Therefore, this change in federal law

⁹ 'Good' annualized quarterly survey financial data means that there exists at least two quarters of actual figures, one of which must be the last quarter of the year.

facilitates a much greater exploitation of GIFI in increasing the quality of statistics while reducing response burden and maintaining confidentiality issues.

What is the present situation with GIFI?

Statistics Canada employees first had access to GIFI (in the Oracle format) in October of 1999. The Tax Data Division is now centrally maintaining this database in the Tax Microdata System (TMS). Presently, there are only incorporated and joint venture businesses available on GIFI at the Legal Entity level. There are 134 block totals and 685 variable items, while the number of records depends on which Oracle table the users are accessing. As of March 1, 2000, there were 881,298 rows in the T2 Tax and Financial Declaration table each representing a business for the 1998 tax year. There are 889,106 rows in the 1998 T2 Declaration table, one for each financial declaration. There are more rows in the T2 Declaration compared to the T2 Tax and Financial Declaration because some businesses report more than once in a fiscal period. There are 924,947 rows in the 1998 balance sheet, one for each opening and closing balance sheet. There are 36,880 opening and 888,067 closing balance sheets for 1998. For the 1998 income statement, there are 883,930 rows, each representing an income statement.

In the 1998 tax year, 97% to 99% of data were transcribed for GIFI by CCRA. Conversely, fewer than 3% submitted their information in GIFI format. The focus of TDD is on balancing the data to make sure they are consistent. In 1999, the plan is the same but the number of tax forms needing to be transcribed will begin to decrease and the number of forms sent in with GIFI codes already assigned will increase. In addition, for 1999, the division is hoping to concentrate on performing stronger edits, historical edits, and partial imputations, looking at deaths, and revisiting matching criteria and multiple fiscal periods. The pessimist will look at the data for 1998, and to a lesser extent 1999, and will dwell on its problems. The optimist will look forward to a stream of big improvements year by year.

Available on the GIFI database is what is known as GIFI views. These views contain the records mentioned above but available at each processing stage. There are presently three views: GIFI-RAW, GIFI-CLEAN and GIFI-E&I (Edit and Imputation). Thus, if a GIFI database user sees unusual results in the GIFI-E&I database, then he/she can trace back to see where the data points originated.

GIFI-RAW contains GIFI records sent directly from CCRA. They are the raw data that has not yet been touched other than to exclude record duplications and those with multiple fiscal periods. The purpose of this view is to retain the original files. GIFI-CLEAN contains records that have gone through the initial cleaning and checking for obvious errors. GIFI-E&I view is comprised of records that have been edited and imputed. The purpose of this view is to fill in missing LE information that is in the UES survey universe file (SUF) but not in The GIFI database. Presently the plan for 1998 data is to ensure there are estimates for block totals but there are no plans to impute below the block total level. Finally, there are no historical edits being done because there is no historical data at this point.

The General Edit and Imputation System (GEIS) is used by Tax Data Division to perform edits and imputations on GIFI data. It selects donors from other tax records to impute for missing records within the same 6-digit NAICS. In 1998, GIFI transcription was done by hand so this process will be greatly reduced once the entire GIFI process is in place. Post-GEIS is the edit and imputation system designed to pro-rate the variables to ensure they balance exactly. It is also designed to calculate specified sub-totals.

The strategy for the 1998 tax year edit and imputation is mainly to balance the data. Balancing ensures that the following equations hold.

Total Assets = Total Liabilities + Total Shareholder Equity
Total Revenue - Total Expenses = Net Non-Farming Income
Total Farm Revenue - Total Farm Expenses = Net Farm Income
Net Income / Loss = Net Income / Loss after taxes and Extraordinary items

First off, it cannot be expected that these equations always hold exactly, so a threshold needs to be set. As long as the equations balance within a \$100 threshold then it is accepted that the equations balance. Below is the strategy for balancing the balance sheet, income statement and retained earnings.

Balancing the Balance Sheet and Income Statement:

- Step 1: Certain items must be positive and certain items must be negative. These items are automatically changed and flagged.
- Step 2: Derive variables for total assets, total liabilities and total shareholder equity, total revenue, total expenses, total farm revenue, total farm expenses, total income adjustments and net income/loss after taxes and extraordinary items from the items provided. Do this even if totals exist.
- Step 3: Ensure that both the balance sheet and income statement equations balance within the tolerance, then move on to step 4, otherwise to step 5.
- Step 4: If pro-rating can be used to balance the equations exactly then proceed. No flag is needed. If any of the variables in the equations are derived then indicate this with a flag, then drop the derived values in step 2. Pro-rate any variables within the "balancing" equations so they sum exactly, do not flag them. Then, pro-rate the items that make up the equation variables if they violate the threshold. Indicate these changes with a flag. If there are no items reported but the equation variables are not equal to zero then the item needs to be flagged for donor imputation (GEIS) for the missing items in that section.
- Step 5: If there is no way to balance the equation then flag the record for donor imputation of the entire balance sheet or income statement. For the income statement, changes can sometimes be limited to just the income adjustment section and the revenue and expense information can be kept. Then the record can be flagged for donor imputation of the income adjustment section.

For the retained earnings information section, item 3600 must equal item 3849 and item 3680 must equal item 9999. If they are not equal a flag is set.

A partial imputation is an imputation performed on a GIFL record when the record is missing all details within a section. These partial imputations are done using GEIS. Presently, TDD is deciding whether to invest the resources in performing partial imputations on the 1998 data. In any case, partial imputations will eventually be performed. To do this, all records are examined for outliers to exclude them from becoming a recipient or a donor in GEIS. If one of these excluded records needs a donor then it will be examined manually. If an entire balance sheet or income statement needs to be imputed, then it receives a mass imputation flag. Depending on the variable that is missing the appropriate mandatory variables are chosen by GEIS to be used for imputation. Finally, if a total is missing it will be donated.

The job of Post-GEIS is to perform an imputation where a missing item is imputed for but the original total is retained. This imputed item then must be pro-rated. Also, sub-totals are derived for specified variables and block totals.

Mass imputation is done on records that are missing entirely. It could be that the record is within the UES SUF and has no tax link. For records like this, information about them has to be received from the BR in order to select a suitable donor. If there are no GIFL records available or the record received has no usable data then the record also receives a mass imputation flag and information about the record will be received from the T2 TY98 ARF in order to find a donor. Finally, items not

in the UES sample and missing in GIFI require mass imputation because there is a need to estimate non-UES industries.

Once a suitable donor is found, pro-rating is done to ensure that the newly imputed information balances with the rest of the record, within the threshold. Pro-rating is calculated by dividing the record variable on the ARF by the same donor variable on GIFI. This ratio is then multiplied by the donor variable to get the imputed value. For example, if the missing variable is total expenses then total revenue from the ARF is divided by total revenue from the donor unit. This ratio is then multiplied by the donor variable, total expenses, to get the imputed value.

Presently calendarization is not being performed on GIFI data. There is also the problem of businesses with multiple fiscal periods. There are Legal Entities with more than one reporting period, either because another LE was created from that existing LE or two LEs have merged. This becomes a problem because items will be double counted. The present solution is to keep the longest fiscal period and discontinue processing the shorter record.

A journal table is kept to record the history of all updates to every record.¹⁰

With these changes taking place, alterations must be made to the UES survey design. Part I of the UES was designed to collect annual financial information for the 225 variables needed for the UES Chart of Accounts. This was sent out for reference year 1997 but was not needed for RY1998. This is because of the introduction of GIFI. It is important to note that in 1998 this was not the case for multi-legal entity enterprises. A smaller Part I was sent out to gather consolidated data. In 1999, the Part I survey was not sent out because the Quarterly Financial Survey is now linked to the BR and its sample includes all multi-LE enterprises of significant size.

What is the expectation for the future?

Eventually, it will be mandatory for every T2 and T1 business to file their tax return electronically with CCRA. This means that Statistics Canada will receive this census of financial information electronically. As was mentioned in the previous section, Statistics Canada has received nearly 900,000 records for the 1998 tax year. These records were transcribed manually at CCRA from financial statements that were supplied on paper by corporations as part of their 1998 income tax filing. This is because GIFI software is not widely used and CCRA's new 'CORTAX' computer system is not in place as of yet. The situation will be very similar for the 1999 tax year. However, it is expected that more corporations will use GIFI software, so fewer records will need to be transcribed. There should be just over one million records available in the 1999 GIFI file. By the end of the year 2000, it will be mandatory for all T2 corporations to fill out their tax form using GIFI software. This means there will be over a million records available for the 2000 tax year that were coded using GIFI software and not manually transcribed. This will eliminate the cost for transcription, while increasing both the timeliness and accuracy of the data. By October 2000, it is expected that CORTAX will be up and running, so corporations will be able to file their financial statements electronically. This means that for the 2000 tax year, nearly all T2 tax returns will be received electronically. Additionally, it is expected that within 3 to 5 years GIFI will be modified to handle the requirements for T1 tax filers. Then Statistics Canada will receive financial information electronically about every business in Canada. This will provide the agency with detailed statistics about such variables as sales, expenses, value-added and profits, by both region and industry.

As the GIFI systems at CCRA are developed and improved, Statistics Canada will not need to focus as much effort on balancing and editing the administrative data. The focus will be more on historical imputation and advanced processing steps, such as calendarization. When donor imputation is used, historical continuity of micro-records is lost. The plan is to use donor imputation on small LEs where it should not cause significant problems, but for medium and large

¹⁰ Refer to the paragraph entitled *Journal tables in the Metadata section*

LEs it will be better to impute using a cell-average growth rate applied to the previous year's micro-record. This will avoid the loss of historical continuity.

The strategy for imputation in the future is as follows. Based on the UES SUF, we expect approximately one million LEs. However, there will always be some of the expected GIFI records that do not appear, due to amalgamations, business closures or other reasons. There will also be some unexpected GIFI records that do appear, due to the formation of new legal entities that had not yet been detected in the BR. The BR will be notified of all unidentified LEs in order to permit timely updates for purposes of the UES post-SUF. At the same time, TDD and BSMD will add imputed records to the E&I view for each of the expected but missing records. This will be done twice a year, once early in the process, after 75% of the records are received, and a second time later in the process, after almost all records have been received.

Presently, there are three GIFI views available to the user. These are sometimes referred to as GIFI-RAW, GIFI-CLEAN and GIFI-E&I¹¹. Eventually another view will be added, called GIFI-ENTERPRISE. This will include all single-legal records and all consolidated multi-legal records. Therefore, this view will contain all consolidated enterprise level GIFI records. The purpose is to enable users to link a GIFI record to an enterprise on the BR, using the BN. In turn, each LE associated with an enterprise on the UES SUF should be able to be linked to either a single or consolidated GIFI record.

The strategy also includes a calendarization processing step. In addition, if a record has multiple fiscal periods, all records will be merged by adding flow variables together, taking the stock variables from the last record, and re-balancing the final record.

Statistics Canada has assembled five working groups to address a number of specific issues that have arisen with respect to the use of the new GIFI database.¹² Task group A has been set up to tackle the issue of consolidation. The group will create procedures for consolidation and will recommend an appropriate allocation of responsibilities for divisions. Task group B's job is to define a new UES Chart of Accounts (COA). The group will do this by taking into consideration IOFD's COA, the UES model questionnaire, the implicit GIFI COA and the SNA's accounting framework. Task group C is addressing the BN linkage problem. If the BR cannot provide accurate and comprehensive linkages, then GIFI data cannot be effectively utilized for statistical purposes. Task group D is developing a strategy for estimating as many as possible of the variables that are required for the UES COA by drawing upon the information in the GIFI database. This will help reduce the survey response burden imposed on businesses. Finally, task group E will find common software approaches to help users access and analyze GIFI variables.

Tabulations

Metadata

When creating a view of GIFI data in either SAS or MS-Access it is very important to understand how to retrieve the correct data. GIFI data are not set up in simple spreadsheets, but rather are housed in a much more flexible relational database. The GIFI Oracle database is comprised of many tables that contain many fields and variables. The tables can all be linked together. For example, the BN number can be used to link records in the T2 declaration table to the corresponding records in the balance sheet or income statement tables. The variables include such items as assets, liabilities, retained earnings, revenue and expenses. These fields or

¹¹ *These GIFI views are explained more thoroughly in the section entitled 'What is the present situation with GIFI?'*

¹² *The purposes of these task forces are outlined in the 49th PIPES Technical Series paper and also in the section of this paper entitled 'What are the principal challenges facing Statistics Canada when using GIFI?'*

variables have both a long name and a short name. The advantage to using a long name is that it makes the variable easier for the user to identify. The advantage to using the shorter name is that they are eight characters or fewer, the maximum length of characters allowed in SAS programs. The tables and their fields are outlined in a diagram on the TDD web site,¹³ and there is also a detailed description of these fields and variables on the TDD shared drive.¹⁴

The user must decide which records he/she needs to access. In order to do this, the user needs to understand what information is in which tables and how the tables are linked to each other. Presently, there are 11 GIFI tables altogether, as shown within the table below. The first three are general identification tables, the next four are tables containing balance statement information and the last four are tables containing income statement information.

Table Description	Long Name in GIFI
Revenue Canada Business Information	Revenue_Canada_Business
T2 Tax and Financial Declaration	T2_Tax_and_Financial_Decl
T2 Declaration in the 1998 tax year	T2_Decl_1998
T2 Balance Sheet of Variables with long names – 1998	T2_Balance_Sheet_1998
T2 Balance Sheet of Variables with short names – 1998	T2_Balance_Sheet_1998_Shrtn_Vw
T2 Balance Sheet of Block Totals with long names – 1998	T2_Balance_Sheet_1998_Blkl_Tot
T2 Balance Sheet of Block Totals with short names – 1998	T2_Bs_1998_Blkl_Tot_Shrtn_Vw
T2 Income Statement of Variables with long names – 1998	T2_Income Stmt_1998
T2 Income Statement of Variables with short names – 1998	T2_Income Stmt_1998_Shrtn_Vw
T2 Income Statement of Block Totals with long names – 1998	T2_Income Stmt_1998_Blkl_Tot
T2 Income Statement of Block Totals with short names – 1998	T2_Is_1998_Blkl_Tot_Shrtn_Vw

Within the 'Revenue Canada Business Information' table, the user can get the NAICS code, Business Number, Province, Postal Code, Legal Name and many other pieces of information. The 'T2 Tax and Financial Declaration' table is mainly used to link the BN between the 'Revenue Canada Business Information' table and the next table, the 'T2 Declaration in the 1998 tax year' table. The 'T2 Declaration in the 1998 tax year' table has information about when the business' fiscal period started and ended, about the record's imputation donor if one was used, and about the stage of process of the record. The information in this table is current, not historical. The last eight tables contain either income statement or balance sheet variables or fields. Half of these tables contain data on the approximately 700 GIFI variables, while the other half contain data for the various block totals. Within each of these two groups, there are separate tables using short and long variable names.

For both the balance sheet and income statement there is a BNPID variable, which is the Business Number Program Identification Code. The BNACCT variable is a Business Number Account Sequence Number, the BN is the CCRA Business Registration Number, the FPEDT is the T2 Financial Declaration fiscal period end date, the FPSDT is the T2 Financial Declaration fiscal period start date, and T2STG is the Stage Identifier of a record. Any short name starting with the letter P is one containing information about donor imputation. Finally, the income statement contains information about the business' operating name and description. The balance sheet has a field called OBSI, which stands for opening balance sheet indicator and applies to businesses that began their operations during the fiscal period.

The user must select which variables to use. These might include assets, liabilities, retained earnings information, revenue, expenses and so on. There are some 685 items in the balance

¹³ Refer to footnote 8.

¹⁴ Refer to footnote 2.

sheet and income statement. Each of these items has both a long name and a short name. For example, Cash and Deposits can be accessed using either the long name Csh_and_Dpst_Amt or the short name L1000. In a short name, an 'L' always precedes the four-digit GIFI code. Similarly for block totals, the short name starts with a BT and then contains a four-digit code.

The easiest way to understand the metadata involved in accessing GIFI is to run a sample piece of SAS code and view the results. TDD is pleased to supply samples of code on request.

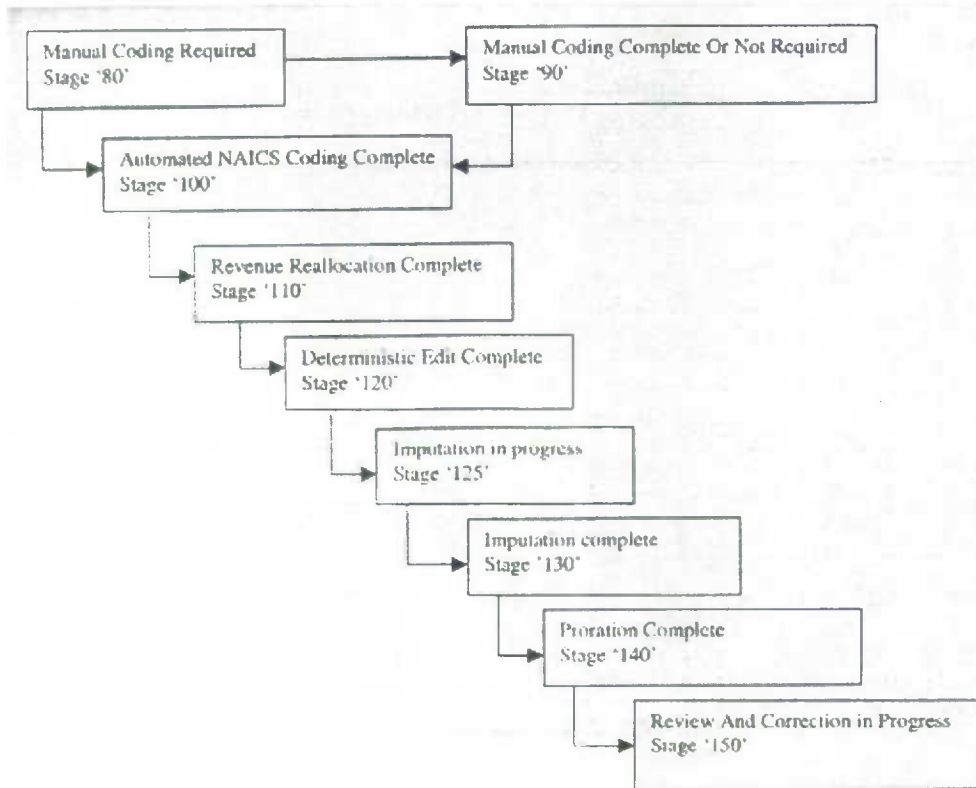
Processing Stages

It is very important to understand which process stage a record is in. The user needs to make an informed decision about which records to use in their analysis. The processing stage code for each record can be located in either the T2 declaration table or the balance sheet and income statement tables. The short name for the associated variable is T2STG and the long name is T2_Prcs_Stg_Id. Once data are loaded onto the Oracle database, they go through many different processes. These processes include edit and imputation, and they are done sequentially, possibly over a lengthy period time, so it is important for the user to know which records have gone through which processes.

The stage ID is different, depending on whether it is located on the balance sheet and income statement or on the T2 declaration table. The stage ID on the balance sheet and income statement contains information about the last change that affected the record. On the other hand, the stage ID on the T2 Declaration is the last process that the record underwent, even if the data in the record are not altered. The latter is used for processing purposes and the former is used for analytical purposes.

It is instructive to compare the stages of processing that apply to unincorporated T1 businesses (for purposes of the Tax Estimates Program sample) with those for T2 corporations (in the GIFI census). In the case of T1 businesses, all records are loaded at the '80' or '90' processing stage, depending on whether or not they have a SIC code. After a record receives a NAICS code, it is put into stage '100'. At stage '110' all records that are invalid are flagged as such and revenue is split into the five different types (farming, fishing, commissions, rental, professional). Stage '120' is reached when the deterministic edit process is complete. This means that all data balance, according to an accountant's point of view. When a record is in the process of imputation the stage identifier is set to '125'. This involves the preparation for GEIS and flagging records that are out-of-sample. A record can also be at the '125' stage if an imputation donor has not yet been found. Once an imputation donor has been found and the records have gone through the GEIS imputation, the stage is set to '130'. As soon as the record is ready for allocation, it is set to '140'. Finally, stage '150' is an ongoing process involving many different types of changes by many different users. Diagram 3 shows the T1 processing stages.

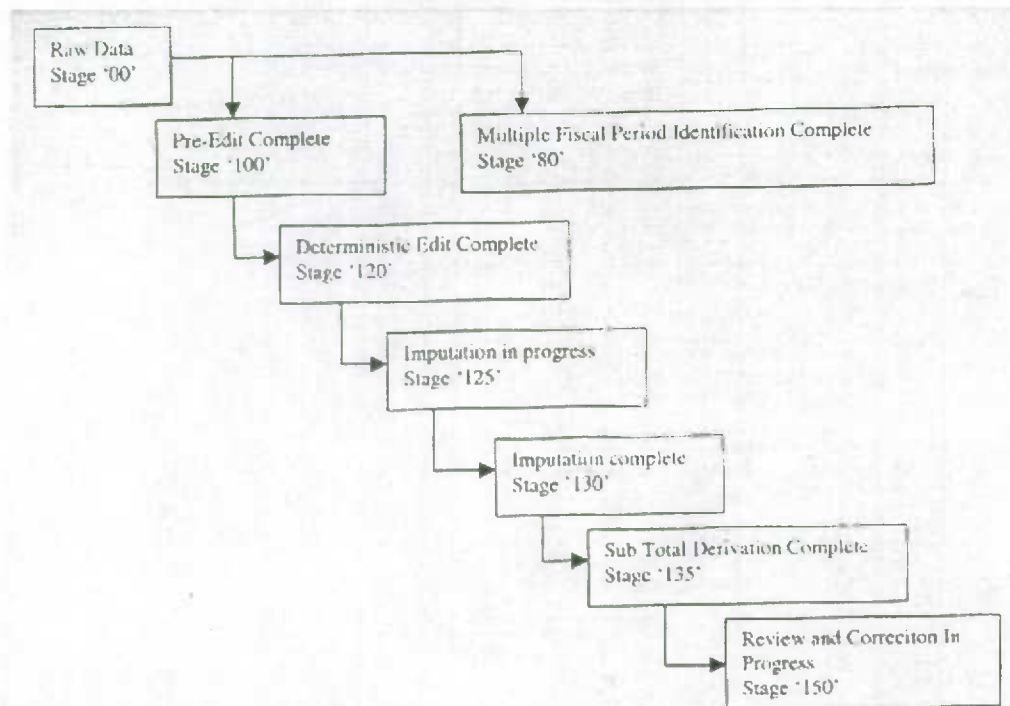
Diagram 3 – T1 Processing Stages



* TDD Web Site under Metadata and Processing Stages and Journal Tables

The processing steps for T2 corporations differ slightly from those in the T1 process. For example, all data are loaded into stage '00', which is the raw data stage. After a pre-edit is done to check for multiple fiscal periods, records are divided up into two groups. If a corporation has just one fiscal period, it goes into stage '100'. If a corporation has multiple fiscal periods, the record with the longest fiscal period goes into stage '100' and the ones with the shorter periods go into stage '80'. All records in stage '80' stay there and are not processed further, but are available if users would like to access them. The remainder of the process is similar to the T1 process. Diagram 4 shows the T2 processing stages.

Diagram 4 – T2 Processing Stages



* TDD Web Site under Metadata and Processing Stages and Journal Tables.

Journal tables

'Journal tables' are tables, much like those described in the section above, containing historical information about a record. For instance, a journal table indicates when the record was inserted into, deleted from or updated within a particular table. In effect, the journal table has information about each staging process the record has gone through. The journal table is stored as <table_name>_JN on the Oracle database. This journal table also contains other information, such as the user that altered the record, the date and time of the alteration, which application was executed, an ID number for that particular session and any other pertinent information. The T2 Declaration table always has information about the status of a record, whereas the T2 Declaration Journal has the history of all changes to that record.

Which industries are more accurately portrayed using GIFI?

The current GIFI database contains detailed financial information about Canadian incorporated businesses. One of the drawbacks to GIFI is that it only contains information about certain forms of organizations. On the Statistics Canada Business Register, businesses are divided up into 5 different categories. They are:

- 1) incorporated business,
- 2) joint ventures, where the parties each fill out T2 tax forms,
- 3) unincorporated businesses that fill out T1 tax forms,

- 4) government business enterprises that do not fill out a T2 or a T1 tax form, and may or may not complete another tax form related to its specific activities, and
- 5) government establishments, which fill out no tax form whatsoever.

The chief limitation of GIFI is that it only includes T2 financial information (although as noted, eventually it will cover T1 information as well). This means that GIFI only contains financial information for incorporated, joint venture and *eventually* unincorporated businesses. In the present situation, if an industry is made up mainly of incorporated or joint venture businesses, then it can be better portrayed by GIFI than an industry that is made up mostly of unincorporated businesses of government enterprises.

Businesses are also broken down into simple and complex enterprises.¹⁵ Both simple and complex enterprises are presently included on GIFI. If an industry is mainly comprised of simple enterprises, it will be a good candidate for portrayal using GIFI. Simple enterprises have only one LE and all of the enterprise's establishments are in the same province and industry, so assigning a NAICS code and producing accurate data by industry is straightforward. No consolidation or allocation process is required. Complex enterprises, on the other hand, are more problematic because getting establishment level data involves consolidation and allocation.¹⁶ An industry consisting of mainly complex enterprises may not be the best industry to portray using GIFI. Although the LEs of complex enterprises are included in the GIFI database, they must be consolidated to the enterprise level first and then allocated to the establishment level in a second step before they can be effectively utilized in producing industrial or provincial breakdowns.

Another point to note is the exclusion of insurance agencies from GIFI. Caution is required when dealing with all industries within NAICS codes 5241 (Insurance Carriers) and 5242 (Agencies, Brokerages, and Other Insurance Related Activities). There are also NAICS exclusions imposed by TDD. These include 526112 (Non-Trusteed Pension Funds), 8132 (Grant-Making and Giving Services), 8133 (Social Advocacy Organizations), 8134 (Civic and Social Organizations) and 91 (Public Administration). This is because GIFI is not designed to meet the needs of these industries.

Included, as an annex to this paper, is a set of rough indicators of how well GIFI portrays various industries. These coverage indicators will only be appropriate until the day arrives when unincorporated businesses are required to file with GIFI. Available on the BR is information about each industry on the number of establishments, number of employees and total revenue in that industry, and whether the enterprise is simple or complex. This information can be broken down separately for incorporated businesses, unincorporated businesses, joint venture businesses, government enterprises and other forms of organizations. With 921 industries and 3 variables (establishments, employees and revenue), these tables are too large to be included in this paper. Instead, for each of the three variables, the tables have been collapsed into a single, smaller table (Table 2) showing twenty-one selected industries for illustrative purposes. Eight of the twenty-one industries shown have been chosen from the top of the GIFI coverage index, eight from around the median point and five from the bottom. The full, detailed version of these tables is available from the author on request.

The first column of Table 2 shows a NAICS code and the second a short description of the industry. The remaining columns show two indexes of the coverage of GIFI, the first including complex enterprises and the second excluding complex enterprises, in terms of establishment counts, total employees and total revenue. The first index is calculated by adding together both the simple and complex incorporated and joint venture businesses, and dividing by the total. This is referred to in Table 2 as problematic data because of the complex enterprises involved. To

¹⁵ Simple and complex enterprises are defined in the section of this paper entitled 'The Business Register'.

¹⁶ The consolidation and allocation processes are outlined in the section entitled 'Consolidation and allocation of Legal Entity data to the establishment level'.

calculate the second index, simple incorporated and simple joint venture businesses are added together and divided by the total. This is referred to in Table 2 as non-problematic because data can be easily extracted. These calculations produce a percentage coverage of a particular industry by GIFI.

Table 2 is constructed using the number of establishments, total employees and total revenue in each industry. Ideally it would be based on value added information, but such data are not available. A problem with deriving the index using establishment counts is that all establishments are considered equal, regardless of size. A problem with using total employees as an indicator is that industries vary substantially in their labour-intensities. Finally, revenue is not a perfect indicator either, since the ratio of revenue to value added varies greatly by industry. Accordingly, the best approach is to consider all three indicators together, as a group.

Table 2 – Industries by Expected ‘Coverage’ on GIFI – By Establishment Counts, Total Employment and Total Revenue

NAICS	Industry Description	Establishment Counts		Number of Employees		Total Revenue	
		Problematic	Non-Problematic	Problematic	Non-Problematic	Problematic	Non-Problematic
212392	Diamond Mining	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
561330	Employee Leasing Services	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
212299	All Other Metal Ore Mining	100.0%	90.5%	100.0%	51.5%	100.0%	24.2%
311221	Wet Corn Milling	100.0%	46.7%	100.0%	8.9%	100.0%	0.9%
212393	Salt Mining	100.0%	10.0%	100.0%	0.0%	100.0%	0.1%
452110	Department Stores	100.0%	0.7%	100.0%	0.2%	100.0%	0.1%
441110	New Car Dealers	99.4%	96.4%	99.6%	94.6%	99.5%	93.9%
325120	Industrial Gas Mfg.	95.9%	17.8%	99.6%	2.7%	99.9%	5.3%
...
622112	Paediatric Hospitals	81.8%	36.4%	99.9%	47.2%	99.9%	57.0%
418990	All Other Wholesaler-Distributors	80.7%	78.6%	96.9%	70.1%	97.6%	36.5%
231410	Construction Management	80.0%	79.3%	97.1%	93.4%	95.5%	76.7%
336611	Ship Building and Repairing	79.9%	72.6%	99.7%	25.3%	99.6%	22.8%
713930	Marinas	79.4%	78.9%	99.7%	25.3%	88.2%	87.1%
332710	Machine Shops	79.0%	78.0%	98.2%	90.8%	97.7%	86.1%
112330	Turkey Production	78.2%	67.9%	87.8%	74.6%	94.6%	32.3%
312130	Wineries	77.4%	72.2%	94.3%	49.6%	97.1%	26.2%
...
722330	Mobile Caterers	29.0%	28.8%	74.1%	72.8%	61.4%	55.9%
111140	Wheat Farming	19.2%	19.2%	45.5%	45.2%	28.0%	27.9%
485310	Taxi Service	18.9%	18.8%	65.8%	61.0%	60.5%	58.7%
911410	Foreign Affairs	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
912110	Provincial Courts of Law	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
...
Total		66.8%	63.2%	85.1%	59.3%	82.1%	38.1%

Diamond Mining and Employee Leasing Services are both composed solely of simple incorporated businesses, making these industries excellent candidates for portrayal using GIFI. The 'All other Metal Ore Mining' industry is composed of simple incorporated businesses but also has a small number of complex incorporated establishments, producing a non-problematic coverage index of 90.5%. However, the influence of the complex establishments is grossly underestimated when deriving the index using total establishments, as illustrated by comparison with the revenue-based index. The latter is only 24.2%, which suggests the data is more problematic than the establishment counts index implies. In the case of Wet Corn Milling, 100% of the industry is made up of incorporated businesses. Of these, 46.7% are simple establishments but the proportion of the total revenue accounted for by these simple units is only 0.9%. Salt mining and department stores are both examples of industries that are made up entirely of incorporated businesses, but where only a very small percentage is attributable to simple enterprises, meaning that a large percentage of the data (around 99.9% according to total revenue) is problematic. New Car Dealers pose an interesting situation. Even though there are some unincorporated business establishments, about 0.6%, the industry is made up mostly of simple incorporated businesses, making this industry a relatively good one to portray using GIFI. According to the revenue-based non-problematic index, New Car Dealers are expected to have 93.9% coverage. In yet another illustrative case, the Industrial Gas and Manufacturing industry is dominated by complex incorporated businesses.

The middle group of industries presents a similar story, although they seem to have more simple unincorporated businesses. These industries tend to contain smaller, less capital-intensive businesses. An interesting point to note about these industries is that the revenue index is often higher than the establishment index. This suggests that incorporated businesses tend to be larger compared to unincorporated and government businesses. Although 99.9% of total revenue of all Paediatric Hospitals is available on GIFI, only 57% of this revenue is non-problematic. Interestingly, while simple incorporated businesses are only 36.4% of the total number of establishments, they account for 57% of total industry revenue. All Other Wholesaler-Distributors, Ship Building and Repairing, and Wineries all have non-problematic establishment count coverage indexes above 70%, but their total revenue indexes are much lower.

The last section of Table 2 is made up of industries mainly comprised of unincorporated businesses and governments. Mobile Caterers, Wheat farming and Taxi Service all have very low GIFI coverage indexes. This is because these types of businesses are run by few people, often out of personal residences and require little overhead to get started. Government industries are not covered at all by GIFI.

Overall, the index derived using establishment counts suggests that 66.8% of establishments are covered by GIFI, or 63.2% if only simple businesses are included in the index. Looking at the number of employees suggests that 85.1% of employees are covered on GIFI, or 59.3% based on simple businesses only. Finally, the total revenue index indicates that 82.1% of the economy is well covered by GIFI, although this measure drops to just 38.1% if only simple incorporated enterprises are included in the index.

Questions which help to show what is currently in the database

This section is intended to demonstrate to the reader what information is presently available on the GIFI database and what can be done with those data. There are two tables below, one displaying the depth and the other the breadth of GIFI.

One of the critical issues when using the database is the extent to which GIFI records are successfully assigned an industry code. Table 3 shows the total number of corporations that had filled out the mandatory GIFI variables at time of writing. This total is broken down based on whether the record had been assigned a NAICS code – by successfully linking the record to a

coded business on the BR – or not. Also included in the table is the percentage of missing information due to cases where no NAICS code is assigned. The percentage of missing information because no NAICS code is assigned is calculated by dividing the 'without NAICS' column by the 'total count' column. The percentage of records without NAICS codes was around 14% for each mandatory variable at time of writing.

Table 3 – Total Records Available for GIFI Mandatory Items

Code	Code Name	With NAICS	Without NAICS	Total Count	% Missing
L2599	Total Assets	741,653	126,663	868,316	14.6%
L3499	Total Liabilities	741,609	126,654	868,263	14.6%
L3620	Total Shareholder Equity	741,755	126,659	868,414	14.6%
L8299	Total Revenue	749,161	125,275	874,436	14.3%
L9368	Total Expenses	749,198	125,284	874,482	14.3%
L9659	Total Farm Revenue	743,594	124,097	867,691	14.3%
L9898	Total Farm Expenses	743,593	124,096	867,689	14.3%
L3849	Retained Earnings/Deficit-End	698,412	108,303	806,715	13.4%
L9999	Net Income/Loss after Taxes and Extraordinary Items	749,225	125,292	874,517	14.3%

Table 4 shows the total amount for each mandatory item on the GIFI database. The total amount is broken down by records with a NAICS code and records without a NAICS code. The percentage of missing information because no NAICS code is assigned is calculated by dividing the 'without NAICS' column by the 'total amount' column. Here, the percentage missing varies considerably by item. Item 3620, total shareholder equity, is missing the most information of all mandatory variables, around 20.9% of its total. Item 9659, total farm revenue, is missing the least amount of information, only 2.5% of its total. Although only 4.7% of total expenses are missing because no NAICS code has been assigned, it remains disconcerting because there is just over one trillion dollars of total expenses unaccounted for. Lastly, the total amount for expenses is \$24 trillion, which is unreasonably high, suggesting there are significant editing problems still to be resolved in this, the first edition of the GIFI database.¹⁷

Table 4 – Total Dollar Amount by GIFI item (billions of dollars)

NAICS	Code Name	With NAICS	Without NAICS	Total Amount	% Missing
L2599	Total Assets	\$1,932	\$291	\$2,223	13.1%
L3499	Total Liabilities	\$1,404	\$154	\$1,558	9.9%
L3620	Total Shareholder Equity	\$534	\$141	\$674	20.9%
L8299	Total Revenue	\$1,002	\$112	\$1,114	10.1%
L9368	Total Expenses	\$22,871	\$1,123	\$23,994	4.7%
L9659	Total Farm Revenue	\$17	\$0.4	\$17	2.5%
L9898	Total Farm Expenses	\$1,574	\$76	\$1,650	4.6%
L3849	Retained Earnings/Deficit-End	\$171	\$21	\$192	11.2%
L9999	Net Income/Loss after Taxes and Extraordinary Items	\$1,574	\$76	\$1,650	4.6%

¹⁷ Manual editing performed on the 1998 GIFI data subsequent to the calculation of these totals resulted in the correction of some very large data values found in a few specific records. Total expenses, as a result, have been reduced to a more reasonable figure.

The next example, displayed in Table 5, takes an industry and presents tabulations by province. This example is designed to show the reader the type of specific information that can be extracted from the GIFI database. The industry chosen is Home and Auto Supply Stores, NAICS code 452991, and was selected because it did well according to the GIFI coverage index.¹⁸ The user can specify other industries to examine and which variables they would like extracted. By linking the information in the Revenue Canada Business table to that in the other tables it is possible to create breakdowns by province.

The total number of legal entities recorded in the database for this industry is 482. All numbers in Table 5 have been rounded to the nearest thousand dollars. Total assets are \$1,192,000,000, total liabilities are \$1,015,000,000, total revenue is \$3,346,000,000, and total expenses are \$3,237,000,000.

Table 5 - NAICS 452991 - Home and Auto Supplies Stores (millions of dollars)

Province	Count	Total Assets	Total Liabilities	Total Revenue	Total Expenses
AB	36	\$82	\$69	\$222	\$211
BC	25	\$97	\$89	\$281	\$284
MB	14	\$25	\$22	\$68	\$58
NB	19	\$43	\$35	\$118	\$116
NF	14	\$36	\$34	\$86	\$82
NS	24	\$52	\$48	\$135	\$133
NT	X	X	X	X	X
ON	227	\$535	\$453	\$1,639	\$1,571
PE	X	X	X	X	X
QC	108	\$276	\$224	\$663	\$650
SK	11	\$30	\$27	\$85	\$84
YT	X	X	X	X	X
TOTAL	482	\$1,192	\$1,015	\$3,346	\$3,237

X = Suppressed to protect confidentiality

Table 5 could not be released in its original form because it would reveal confidential information about individual businesses. This is because certain cells contain one or two legal entities. In addition, if a single corporation's activity accounts for a large percentage of the total amount in a cell, then suppression is deemed necessary. There are also specific conditions, known as the Duffet Rules, for more complex situations. The user has to be careful when suppressing a cell. If a cell is suppressed but the total is still shown, then readers will be able to figure out the missing information. To avoid this, one or more additional cells must be suppressed to maintain the confidentiality of the original suppressed cell.

¹⁸ GIFI coverage index is outlined in more detail in the section entitled, 'Which industries are more accurately portrayed using GIFI?' of this paper.

Personal and Laundry Services Industry

An example of the statistical use of GIFI

To help the reader understand what GIFI has to offer the statistical system, an examination of the industries within the Personal and Laundry Services sub-sector, NAICS sub-sector 812, is now presented. The sub-sector was chosen quite arbitrarily for this purpose, although it is considered to lend itself fairly well to a GIFI-based analysis in light of its atomistic nature and the fact that it is not dominated by establishments of complex enterprises. The analysis is divided into four parts. First, the NAICS structure of the Personal and Laundry Services sub-sector is outlined and the question of why it might be a good industry to investigate using GIFI data is considered. A second section provides a breakdown of the sub-sector by number of legal entities (LEs) and total revenue, using the GIFI database. A third part then describes the sub-sector in terms of establishments and total revenue, drawing on information in the Business Register. Finally, a last section compares the information about the sub-sector that is conveyed by GIFI and the BR, and also points out possible problems and things to watch out for when analyzing a particular industry using GIFI data.

NAICS Sub-Sector 812 – Personal and Laundry Services

NAICS was conceived by statistical agencies in Canada, the US and Mexico. The goal was to produce a common framework of industry groups for each of the three countries. The classification is intended to cover all the economic activity of each of the three nations. It groups businesses by similarities in their production processes. This allows statistical agencies to maintain one questionnaire for many similar businesses and produce statistics that can be compared among these similar businesses.

The classification is based on a hierarchical structure. There are sectors, sub-sectors, industry groups, industries and national industries. The Canadian NAICS system is comprised of 20 sectors, 99 sub-sectors, 321 industry groups, 734 industries and 921 national industries. Sectors correspond to a 2-digit code, sub-sectors to a 3-digit code, industry groups to a 4-digit code, industries to a 5-digit code and national industries to a 6-digit code. The 6-digit code is commonly referred to within the confines of Statistics Canada as 'the industry code'. An industry code is assigned to each individual establishment within a business. An industry code can also be assigned to an enterprise. This is known as an enterprise NAICS code. In this case, the NAICS code is assigned to the enterprise based on the dominant activity of the establishments within the enterprise. The dominant activity is assigned by means of an algorithm, applied in the Business Register Division, that takes into account the industry code of each establishment within the enterprise and the relative size of that establishment, gauged by its share of overall enterprise revenue or employment.

The Personal and Laundry Services sub-sector belongs to the 'Other Services (Except Public Administration)' sector and is made up of 4 industry groups, 11 industries and 14 national industries. The NAICS breakdown of the Personal and Laundry Services sub-sector is presented below¹⁹.

¹⁹ For a more detailed definition of the Personal and Laundry Services Sub-sector, please refer to the North American Industry Classification System manual for Canada, published in 1997.

812 – Personal and Laundry Services

8121 Personal Care Services

- 81211 Hair Care and Esthetic Services
- 812114 Barber Shops*
- 812115 Beauty Salons*
- 812116 Unisex Hair Salons*
- 81219 Other Personal Care Services
- 812190 Other Personal Care Services*

8122 Funeral Services

- 81221 Funeral Homes
- 812210 Funeral Homes*
- 81222 Cemeteries and Crematoria
- 812220 Cemeteries and Crematoria*

8123 Dry Cleaning and Laundry Services

- 81231 Coin-Operated Laundries and Dry Cleaners
- 812310 Coin-Operated Laundries and Dry Cleaners*
- 81232 Dry Cleaning and Laundry Services (except Coin-Operated)
- 812320 Dry Cleaning and Laundry Services (except Coin-Operated)*
- 81233 Other Laundry Services
- 812330 Other Laundry Services*

8129 Other Personal Services

- 81291 Pet Care (except Veterinary) Services
- 812910 Pet Care (except Veterinary) Services*
- 81292 Photo Finishing Services
- 812921 Photo Finishing Laboratories (except One-Hour)*
- 812922 One-Hour Photo Finishing*
- 81293 Parking Lots and Garages
- 812930 Parking Lots and Garages*
- 81299 All Other Personal Services
- 812990 All Other Personal Services*

The Personal and Laundry Services industry is one on which GIFI might be expected to shed considerable illumination.²⁰ Indeed, when combined with data from other administrative sources the GIFI database might well enable Statistics Canada to terminate, or at the very least reduce substantially in scope, its existing survey of this sub-sector. That survey has a sample size of approximately 1000 units and a response rate of about 66%. Information requested on the survey questionnaire addresses the nature of business activity, form of organization, reporting period, revenues, operating expenses, inventories, employment, client base and the provincial/territorial distribution of revenue, expenses, wages and employment. Most of these data are available from GIFI and other administrative records such as the PD7/T4 system, GST records, T1 records and CORPAC/CORTAX. Moreover, the level of detail reported by the respondents to the survey is considered rather low. Being a census of incorporated legal entities, the GIFI data cover a very large part of the industry and provide considerable balance sheet and income statement detail. A survey might still be useful for purposes of obtaining other types of information related to the

²⁰ Refer to the section entitled 'Which industries are more accurately estimated using GIFI?'

client base, the provincial distribution of activity and the evolving establishment structure of complex businesses within the industry. Nevertheless, GIFI offers the possibility of a shorter survey questionnaire and a smaller sample, thereby reducing response burden while simultaneously increasing the quality of the resulting statistics.

To the extent that financial trends are fairly common across businesses within an industry, GIFI data from one group of businesses can be used to project financial data that are missing for other groups of businesses. This is a second important use of GIFI. Suppose, for example, the total revenue of simple incorporated businesses in the Personal and Laundry Services sub-sector is extracted from GIFI for a particular year. Suppose as well that total revenue for complex incorporated, simple and complex unincorporated, and government businesses for the Personal and Laundry Services sub-sector – businesses for which data are not available within GIFI – are somehow estimated for that same year, thereby allowing a full picture to be painted of total revenue of the industry. The following year, GIFI will again provide data for the total revenue of simple incorporated businesses. If data for the other components of the industry are highly correlated with the data for the simple incorporated business component, and especially if the other components are relatively small in comparison to the simple incorporated business component, then the observed change in GIFI from one year to the next can be used to project the financial information for the other components of the industry.

Below are Tables 6, 7 and 8, showing the GIFI coverage index based on establishment counts, employment and revenue respectively. Looking at the establishment counts table, it appears GIFI alone may be inadequate for purposes of estimating key financial variables in the Personal and Laundry Services sub-sector. GIFI is expected to cover just 50.6% of enterprises (complex and simple together) involved in the sub-sector and just 48.3% of the enterprises (simple only). There are a large number of unincorporated businesses in the sub-sector, not covered by GIFI as yet. However, GIFI coverage ratios for the sub-sector show a somewhat brighter picture when they are weighted by employment or revenue. In other words, the businesses that *are* covered by GIFI turn out to be relatively large compared to the ones not covered by GIFI. Tables 7 and 8 both indicate that about 65% of the activity of the sub-sector (measured by employment or revenue) is accounted for by simple incorporated businesses, for which GIFI data are available. The percentages are especially high for Dry Cleaning and Laundry Services, Coin-Operated Laundries, One-Hour Photo-Finishing and Funeral Homes.

The reasons why the four different industry groups have varying GIFI coverage are interesting. Industry group 8121 – Personal Care Services – is the one with the poorest GIFI coverage in the Personal and Laundry Services sub-sector. Having a small hair salon or barber shop in, or attached to, one's home need not involve much overhead and is relatively easy to do. Most likely, there will be many unincorporated businesses in this industry group. Funeral Services (industry group 8122) in contrast, tend to be larger operations involving greater overhead. As mentioned, funeral homes tend to be run on a larger scale compared to Cemeteries and Crematoria, so the latter may be characterized by relatively more unincorporated businesses. Overall, the industry group has a GIFI coverage rate above 75%. The Dry Cleaning and Laundry Services industry group (8123), in general, is covered quite well by GIFI. This industry group is similar to Funeral Services, both having at least 75% coverage. The last industry group, Other Personal Services (8129), is more tricky to analyze. This is because the group is made up of activities that are quite dissimilar. Parking lots/garages and photo developing shops are high on the list because property and equipment are expensive and, therefore, provide incentives for the owner to incorporate. Pet care services, on the other hand, are lower on the list and presumably are typically less capital intensive, smaller-scale operations, much like hair stylists. Each industry is unique in its own way and the GIFI data user must exercise discretion in deciding which are sufficiently well covered by the database to warrant general conclusions based solely on GIFI information.

Table 6 - Personal and Laundry Services Industries by GIFI 'Coverage' Based on Establishment Counts

NAICS	Industry description	GIFI Coverage Index with Complex Units	GIFI Coverage Index without Complex Units
812114	Barber Shops	26.3%	26.3%
812115	Beauty Salons	39.6%	37.1%
812116	Unisex Hair Salons	40.7%	40.4%
812190	Other Personal Care Services	56.2%	56.0%
812210	Funeral Homes	92.2%	75.1%
812220	Cemeteries and Crematoria	45.4%	40.1%
812310	Coin-Operated Laundries and Dry Cleaners	62.5%	61.9%
812320	Dry Cleaning and Laundry Services (except Coin-Operated)	68.3%	67.0%
812330	Other Laundry Services	80.8%	71.2%
812910	Pet Care (except Veterinary) Services	43.0%	42.8%
812921	Photo Finishing Laboratories (except One-Hour)	74.5%	71.4%
812922	One-Hour Photo Finishing	76.7%	73.0%
812930	Parking Lots and Garages	81.7%	74.1%
812990	All Other Personal Services	41.8%	41.1%
Total		50.6%	48.3%

Table 7 - Personal and Laundry Services Industries by GIFI 'Coverage' Based on Total Employees

NAICS	Industry description	GIFI Coverage Index with Complex Units	GIFI Coverage Index without Complex Units
812114	Barber Shops	48.7%	48.7%
812115	Beauty Salons	66.5%	61.7%
812116	Unisex Hair Salons	72.7%	63.8%
812190	Other Personal Care Services	86.2%	67.3%
812210	Funeral Homes	98.0%	76.3%
812220	Cemeteries and Crematoria	64.7%	31.3%
812310	Coin-Operated Laundries and Dry Cleaners	79.4%	78.2%
812320	Dry Cleaning and Laundry Services (except Coin-Operated)	89.4%	85.5%
812330	Other Laundry Services	99.2%	56.2%
812910	Pet Care (except Veterinary) Services	58.7%	55.6%
812921	Photo Finishing Laboratories (except One-Hour)	94.5%	59.8%
812922	One-Hour Photo Finishing	90.9%	75.2%
812930	Parking Lots and Garages	94.3%	37.1%
812990	All Other Personal Services	76.5%	53.9%
Total		79.9%	64.7%

**Table 8 - Personal and Laundry Services Industries by GIFI 'Coverage'
Based on Total Revenue**

NAICS	Industry description	GIFI Coverage Index with Complex Units	GIFI Coverage Index without Complex Units
812114	Barber Shops	45.0%	44.3%
812115	Beauty Salons	74.0%	70.5%
812116	Unisex Hair Salons	71.9%	65.4%
812190	Other Personal Care Services	77.9%	69.5%
812210	Funeral Homes	97.3%	70.6%
812220	Cemeteries and Crematoria	78.4%	34.4%
812310	Coin-Operated Laundries and Dry Cleaners	77.6%	77.0%
812320	Dry Cleaning and Laundry Services (except Coin-Operated)	87.2%	79.5%
812330	Other Laundry Services	98.8%	46.8%
812910	Pet Care (except Veterinary) Services	68.0%	66.6%
812921	Photo Finishing Laboratories (except One-Hour)	96.6%	60.0%
812922	One-Hour Photo Finishing	91.4%	78.3%
812930	Parking Lots and Garages	85.0%	49.3%
812990	All Other Personal Services	85.3%	71.6%
Total		82.1%	65.7%

Personal and Laundry Services – Statistical estimation using 1998 GIFI data

To explore the size of the Personal and Laundry Services sub-sector, the number of legal entities and the total revenue of each industry within the sub-sector were investigated via the GIFI database.²¹ Table 9, below shows the number of legal entities in each industry and Table 10 the total revenue in each industry.

As seen in Table 9, there are 15,898 legal entities on the GIFI database that are classified, according to the BR, to the Personal and Laundry Service sub-sector. Of these, 54% belong to the Personal Care Services industry group, 10% belong to the Funeral Services industry group, 24% to the Dry Cleaning and Laundry Services industry group and 12% to the All Other Personal Services industry group. The total counts in each industry group are also broken down by industry. Beauty Salons are by far the industry with the most legal entities, making up 25% of the total. The smallest industries are Cemeteries and Crematoria and Photo Finishing Laboratories (except One-Hour), each accounting for only about 1% of the total legal entities.

²¹ To find out how a NAICS code is assigned to a LE, refer to the section entitled 'What are the principal challenges facing Statistics Canada when using GIFI?'

Table 9 – Total Legal Entities in the 1998 GIFI Database

LE NAICS	Industry description	Total LEs	Percentage of LEs	Total LEs	Percentage of LEs
812114	Barber Shops	259	2%	8,594	54%
812115	Beauty Salons	4,045	25%		
812116	Unisex Hair Salons	2,033	13%		
812190	Other Personal Care Services	2,257	14%	1,506	10%
812210	Funeral Homes	1,342	9%		
812220	Cemeteries and Crematoria	164	1%	3,874	24%
812310	Coin-Operated Laundries and Dry Cleaners	846	5%		
812320	Dry Cleaning and Laundry Services (except Coin-Operated)	2,731	17%		
812330	Other Laundry Services	297	2%		
812910	Pet Care (except Veterinary) Services	400	3%	1,924	12%
812921	Photo Finishing Laboratories (except One-Hour)	223	1%		
812922	One-Hour Photo Finishing	510	3%		
812930	Parking Lots and Garages	463	3%		
812990	All Other Personal Services	328	2%		
Total		15,898	100%	15,898	100%

Table 10 shows the breakdown of total revenue by industry. The total revenue of the Personal and Laundry Services sub-sector is \$3,383 million. The Personal Care Service industry group contains the largest share of total revenue, 33% of the total. The Dry Cleaning and Laundry Services industry group is the second largest, comprised of 28% of total revenue. The Funeral Services industry group has 24% and the All Other Personal Care Services industry group makes up the remaining 15% of total revenue. The Funeral Homes industry has the largest proportion of revenue with 20% of the total, while the Barber Shop and Pet Care industries each account for only 1% of the total revenue.

Table 10 – Total Revenue in the 1998 GIFl Database (millions of dollars)

LE NAICS	Industry description	Total Revenue	Percentage of Revenue	Total Revenue	Percentage of Revenue
812114	Barber Shops	26	1%	1,127	33%
812115	Beauty Salons	510	15%		
812116	Unisex Hair Salons	371	11%		
812190	Other Personal Care Services	220	6%		
812210	Funeral Homes	667	20%	797	24%
812220	Cemeteries and Crematoria	130	4%	934	28%
812310	Coin-Operated Laundries and Dry Cleaners	87	3%		
812320	Dry Cleaning and Laundry Services (except Coin-Operated)	486	14%		
812330	Other Laundry Services	360	11%		
812910	Pet Care (except Veterinary) Services	28	1%	525	15%
812921	Photo Finishing Laboratories (except One-Hour)	149	4%		
812922	One-Hour Photo Finishing	92	2%		
812930	Parking Lots and Garages	194	6%		
812990	All Other Personal Services	61	2%		
Total		\$ 3,383	100%	\$ 3,383	100%

BR information about the Personal and Laundry Services sub-sector

Having examined the profile of the Personal and Laundry Services sub-sector provided by the GIFl database, it is interesting to have a look at the information available on the Business Register pertaining to these same industries. The purpose is to compare GIFl information – excluding unincorporated businesses and government entities, and covering incorporated legal entities rather than production establishments – with the more comprehensive and more statistically well-defined information existing on the BR. A question of particular interest: Is the Personal and Laundry Services sub-sector under-represented or over-represented on the GIFl database and by how much?

Total revenue and total establishments having one of the 14 Personal and Laundry Services NAICS codes were extracted from the BR. On the BR, a NAICS code is assigned first to the establishment level of a business. An enterprise-level NAICS code is then assigned, based on the dominant activity of the establishments of the business. This means that even if an enterprise is assigned the Personal and Laundry Services NAICS code, this does not necessarily mean all its activities are within the Personal and Laundry Services sub-sector. It could be that a particular enterprise has one or more establishments with one or more *other* NAICS codes, while having activities within the Personal and Laundry sub-sector as well. Such a case is represented in Table 11 as 'other' NAICS, in which the dominant activity is a NAICS code other than 812, but the business nevertheless has activities in one of the '812' industries as well.

It is important to keep this in mind when investigating an industry. In principle, to get accurate estimates for an industry, each enterprise must be broken down further according to the activity in question. As shown in column three of Table 11, there are 38,114 establishments (38,623 minus 509) on the BR that are part of enterprises classified to the Personal and Laundry Services sub-sector. As shown in the fourth column, most but not all of these establishments are themselves classified to the sub-sector. In addition, there are 509 establishments in other enterprises wherein the dominant enterprise activity is not that of the Personal and Laundry Services sub-sector, but at least one establishment is in this sub-sector. Of these 509 establishments, 35 are classified to the sub-sector. In total, as shown in the last row of column four in Table 11, there are 38,066 Personal and Laundry Service establishments on the BR.

Table 11 – Total Personal and Laundry Services Establishments on the BR

Enterprise NAICS	Industry description	Total Establishments	Total Personal and Laundry Services Establishments	Percentage of Total Personal and Laundry Services Establishments
812114	Barber Shops	1,557	1,557	100%
812115	Beauty Salons	12,681	12,677	100%
812116	Unisex Hair Salons	6,286	6,285	100%
812190	Other Personal Care Services	4,650	4,649	100%
812210	Funeral Homes	1,656	1,636	99%
812220	Cemeteries and Crematoria	375	371	99%
812310	Coin-Operated Laundries and Dry Cleaners	1,686	1,682	100%
812320	Dry Cleaning and Laundry Services (except Coin-Operated)	4,833	4,827	100%
812330	Other Laundry Services	351	347	99%
812910	Pet Care (except Veterinary) Services	1,287	1,287	100%
812921	Photo Finishing Laboratories (except One-Hour)	321	320	100%
812922	One-Hour Photo Finishing	726	725	100%
812930	Parking Lots and Garages	723	688	95%
812990	All Other Personal Services	982	980	100%
Other		509	35	7%
Total		38,623	38,066	99%

Table 12 shows the total revenue of the Personal and Laundry Services sub-sector as recorded on the BR. The procedure for constructing this table is similar to that for Table 11. Here total enterprise revenue of all enterprises having one or more establishments classified to the Personal and Laundry Services sub-sector is \$14,473 million. Of this amount, only \$6,875 million is accounted for by establishments within these enterprises that are classified to the Personal and Laundry Services sub-sector. The main reason why there is such a large difference is that enterprises with dominant activities in other industries that do participate in Personal and Laundry Services industries are numerous, yet tend to generate most of their revenues from activities other than those of the Personal and Laundry Services sub-sector.

**Table 12 – Total Personal and Laundry Service Revenue on the BR
(millions of dollars)**

Enterprise NAICS	Industry Description	Total Enterprise Revenue	Total Personal and Laundry Services Revenue	Percentage of Total Personal and Laundry Services Revenue
812114	Barber Shops	132	132	100%
812115	Beauty Salons	1,475	1,474	100%
812116	Unisex Hair Salons	979	979	100%
812190	Other Personal Care Services	429	428	100%
812210	Funeral Homes	806	793	98%
812220	Cemeteries and Crematoria	257	242	94%
812310	Coin-Operated Laundries and Dry Cleaners	158	158	100%
812320	Dry Cleaning and Laundry Services (except Coin-Operated)	913	900	99%
812330	Other Laundry Services	599	587	98%
812910	Pet Care (except Veterinary) Services	155	155	100%
812921	Photo Finishing Laboratories (except One-Hour)	142	137	97%
812922	One-Hour Photo Finishing	171	171	100%
812930	Parking Lots and Garages	545	520	95%
812990	All Other Personal Services	161	160	100%
Other		7,551	39	1%
Total		\$ 14,473	\$ 6,875	48%

Comparing the BR and GIFI data for Personal and Laundry Services

The next step is to compare the data on the BR to the data on the GIFI database. The third column of Table 13A shows the total revenue of establishments classified to the Personal and Laundry Services sub-sector, as recorded on the BR. The fourth column shows the *expected* total revenue on the GIFI database, calculated by adding up the total revenue shown on the BR for simple and complex incorporated enterprises in the Personal and Laundry Services sub-sector. Of course, the total revenue shown on the BR for the entire sub-sector, shown in the third column, is higher than this, because it includes unincorporated and government business enterprises as well. Displayed in the fifth column of Table 13A is the actual revenue for these industries as portrayed by GIFI. At the time of writing the GIFI database had just over 800,000 records, whereas the BR indicates that the economy as a whole has approximately 1.3 million incorporated legal entities. Therefore, the total revenue figure from GIFI is substantially lower than the expected total revenue, overall and in most individual industries.

**Table 13A – Comparing Total Revenue on the BR to GIFl
(millions of dollars)**

Enterprise NAICS	Industry Description	Total BR Revenue	Expected Revenue on GIFl	Actual Revenue In GIFl
812114	Barber Shops	132	59	26
812115	Beauty Salons	1,474	1,039	510
812116	Unisex Hair Salons	979	640	371
812190	Other Personal Care Services	428	298	220
812210	Funeral Homes	793	560	667
812220	Cemeteries and Crematoria	242	83	130
812310	Coin-Operated Laundries and Dry Cleaners	158	121	87
812320	Dry Cleaning and Laundry Services (except Coin-Operated)	900	716	486
812330	Other Laundry Services	587	275	360
812910	Pet Care (except Veterinary) Services	155	103	28
812921	Photo Finishing Laboratories (except One-Hour)	137	82	149
812922	One-Hour Photo Finishing	171	134	92
812930	Parking Lots and Garages	520	256	194
812990	All Other Personal Services	160	115	61
Other		39	0	0
TOTAL		\$6,875	\$4,481	\$3,383

The second half of the table (Part B) reports, in column three, the percentage of total revenue that *should* be on the GIFl database, according to the BR. This is calculated by dividing the expected revenue by the total revenue on the BR, and it shows by how much the GIFl data, alone, will under-cover the industry. The next column calculates the percentage of the total that is on the GIFl database at the time of writing. This is derived by dividing the total revenue now on the GIFl database by the total revenue shown on the BR, and it shows how much of the industry is presently covered by the GIFl database.

Table 13B – Comparing Total Revenue on the BR to GIFl (millions of dollars)

Enterprise NAICS	Industry Description	Percentage that Should be on GIFl	Percentage On GIFl
812114	Barber Shops	44%	19%
812115	Beauty Salons	70%	35%
812116	Unisex Hair Salons	65%	38%
812190	Other Personal Care Services	70%	51%
812210	Funeral Homes	71%	84%
812220	Cemeteries and Crematoria	34%	54%
812310	Coin-Operated Laundries and Dry Cleaners	77%	55%
812320	Dry Cleaning and Laundry Services (except Coin-Operated)	79%	54%
812330	Other Laundry Services	47%	61%
812910	Pet Care (except Veterinary) Services	67%	18%
812921	Photo Finishing Laboratories (except One-Hour)	60%	109%
812922	One-Hour Photo Finishing	78%	54%
812930	Parking Lots and Garages	49%	37%
812990	All Other Personal Services	72%	38%
Other		0%	0%
TOTAL		65%	49%

Of particular interest is the fact that for a few industries in Part B of Table 13, reported GIFl revenue accounts for a larger than expected percentage of total BR revenue. The Funeral Homes, Cemeteries and Crematoria, Other Laundry Services and Photo Finishing Laboratories (except One-Hour) are the cases in point. This likely is due to the fact that the BR data for total revenue are much less reliable and up-to-date than the reported GIFl figures.

As was mentioned earlier, GIFl and the BR cannot be compared using total employees because this variable is not an item on the typical financial statement and, therefore, is not available on GIFl. Likewise, it is not feasible to compare establishment counts on the BR with similar information on GIFl because the latter is based on legal entities rather than establishments, and the mapping between establishments and legal entities is many-to-many.

Possible solutions in comparing GIFl data with the BR

Another possible strategy would be to get the BN for each business in the Personal and Laundry Services sub-sector from both the BR and GIFl. By linking through the BN, the user could then compare revenue on the BR with revenue on the GIFl database, enterprise by enterprise. Due to time constraints, this type of analysis has not been done for this paper.

Problems encountered using GIFI data with respect to the Personal Laundry Services Industries

One particular problem encountered by the author while studying the GIFI data for the Personal and Laundry Service sub-sector related to computer handling. There are many variables and observations that the user may wish to look at. Depending on how large the industry or sub-sector is, a desktop computer may have inadequate capacity to manipulate large amounts of data. In some cases, it is better to run the GIFI database procedures on a more powerful computer. If the user is exploring a larger sub-sector than the Personal and Laundry Services, he/she may want to consider running the code off the server. Even with a sub-sector the size of Personal and Laundry Services, it can take a long time to run a piece of code.

Understanding the metadata is probably the toughest challenge the user faces when using the GIFI database. It is difficult to understand which variables are needed and how to retrieve the correct data. The user needs to make sure he/she comprehends what each variable is and how it is derived. Also, the user must be aware of the difference between a variable, a block total and a generic variable. The Oracle GIFI database is *relational*. There are many tables with many links, and this can be very complex and difficult. If not well understood, wrong data could inadvertently be used. Knowing the staging process of the data is also important, in order to ensure the records extracted are suitable ones. The user has to be careful when choosing a dummy variable name for each table. For example, 'is' and 'in' may seem like good short names for the income statement table, but they are in fact function names and if either is used the code will not run properly, either causing errors or producing wrong statistics. If the software used to access the data has never been used before, then there is an inevitable learning curve. SQL programming especially can be very difficult to understand and the results can be hard to interpret when errors occur.

More generally, the majority of problems, both small and large, stem from the fact that GIFI is very new to both Statistics Canada and CCRA. There will be problems when users initially dive in. In time, once this development period is complete, GIFI will be much easier to use and more documentation and technical support will be available.

Conclusion

GIFI is bringing many opportunities to Statistics Canada. The full census of financial statements for the entire incorporated and, eventually, unincorporated business population is greatly improving the range and quality of statistics the agency can produce. Not only is the accuracy and detail improving, both at the provincial and industry level, but there are big opportunities to reduce survey response burden. Further, the fact that the GIFI data is now beginning to arrive in a much more timely manner will enable many statistical programs to produce their outputs with a shorter time lag. Of course, to fully exploit the opportunities presented by GIFI, Statistics Canada must face many challenges. Dealing with the issues of complex enterprises, calendarization and the problem of multiple fiscal periods will not be easy. In addition, developing a good edit and imputation strategy will take time. Because of the inevitable learning curve, it will be a while before the advantages of GIFI can be fully realized. Statistics Canada is now investing the time and effort required to address these challenges and the future looks promising indeed.

References

- Bissett, Peter D. *Use of Tax Data in the Production of Provincial Economic Statistics*. PIPES Technical Series Number 42, October 1999.
- Farrall, Kim. *Changes to NAICS Canada Codes and Titles Between July 1997 and March 1998*. March 1st 2000. <http://stdsweb/naicstest/eng/changes.htm>.
- Hall, Karen. *The Harmonized Sales Tax Revenue Allocation Formula*. PIPES Technical Series Number 24, August 1998.
- Mach, Lenka. *Tax Estimates Program (TEP)*. September 1996.
- Mach, Lenka. *Tax Estimates Program (TEP): Sampling of T1 Businesses*. May 1997.
- Mach, Lenka. *Tax Estimates Program (TEP): Sampling of T2 Corporations*. June 1997.
- Canada Customs and Revenue Agency. *Corporate Income Tax*. May 18th 1998. <http://www.cca-adrc.gc.ca/>.
- Canada Customs and Revenue Agency. *Corporate Income Tax Redesign*. May 18th 1998. <http://www.cca-adrc.gc.ca/>.
- Canada Customs and Revenue Agency. *General Index of Financial Information*. BEST Course. July 1997.
- Smith, Philip. *Statistics Canada's Broad Strategy for Business Statistics*. PIPES Technical Series Number 57, May 25th 2000.
- Smith, Philip. *The UES Strategy to Combine Enterprise, Establishment and Legal-Entity Data*. PIPES Technical Series Number 49, November 29th 1999.
- Statistics Canada. *ADMIN FILE*. February 3rd 2000. http://tax/Tax_Component/Admin/Admin_index.htm
- Statistics Canada. *Annual Survey of Personal Services*. 1997.
- Statistics Canada. *Edit and Imputation of GIFI TY98 Proposed Strategy (Draft)*. May 25th 2000. F:/Laurie/Taxdata/Ty98/GIFIstrat2.4b.doc.
- Statistics Canada. *General Index of Financial Information (GIFI)*. January 10th 2000. http://tax/Tax_Component/FAQ/GIFIFAQ.htm.
- Statistics Canada. *GIFI Analysis Team Meeting Minutes*. January 11th 2000 to February 29th 2000. <http://tax/english/Minutes.htm>.
- Statistics Canada. *GIFI Table Layouts*. March 29, 2000. T:/External/Tax Dev External/GIFI Table Layouts.doc.
- Statistics Canada. *GIFI/CORTAX and Related Tax Services*. March 28, 2000. http://tax/Tax_Component/GIFI_T2_CORTAX/Bulletins/GIFI_Bulletins.htm.
- Statistics Canada. *List of Changes to the (Hard Copy) NAICS Canada 1997 Manual (From March 1998 to November 1999)*. March 1st 2000. <http://dissemination/english/Subjects/Standard/manualchanges.htm>.

Statistics Canada. *Mandate*. January 13th 2000. http://iofdweb/iofd_english_web/mandate.htm .

Statistics Canada. *NAICS Canada*. Occasional, Catalogue no. 12-501-XPE, March 1998.

Statistics Canada. *NAICS Canada*. February 29th 2000.
<http://dissemination/english/Subjects/Standard/introo.htm>.

Statistics Canada. *Procedures for gaining access to Tax Data*. March 15th 2000.
<http://tax/english/accesreng.html>.

Statistics Canada. *Processing Stages and Journal Tables*. March 10th 2000.
<http://tax/Tax%20Microdata%20System/English/Processing%20Stages%20and%20Journal%20Tables.htm>.

Statistics Canada. *Simple Entity Relationship Diagram for T2 Data*. March 29, 2000.
T:/External/Tax Dev External/Simple GIF Diagrams.doc.

Statistics Canada. *T2 Balance Sheet 1998*. March 29, 2000.
http://tax/Tax%20Microdata%20System/English/Current_Physical_diagram_frame.htm.

Statistics Canada. *T2 Income STMT 1998*. March 29, 2000.
http://tax/Tax%20Microdata%20System/English/Current_Physical_diagram_frame.htm.

Statistics Canada. *Tax Data Division Program Report*. May 31st 1999. http://tax/english/APR97-99_E.htm.

Statistics Canada. *Unified Enterprise Statistics Program – Project to Improve Provincial Economic Statistics*. PIPES Technical Series Number 1, May 1997.

Technical Series - Index

October 13, 2000

PIPES has a series of technical paper reprints primarily for internal purposes. A list of the reprints currently available is presented below. For copies, contact Bonnie Bercik at (613) 951-6790 or Diane Proulx at (613) 951-7192, fax number (613) 951-0411 or write to Statistics Canada, 13th Floor, Jean Talon Building, Tunney's Pasture, Ottawa, Ontario, K1A 0T6

1. Unified Enterprise Statistics Program – Project to Improve Provincial Economic Statistics – May 5, 1997 – PIPES Project Managers.
2. PIPES Evaluation Framework – September 15, 1997 – Philip Smith.
3. Report on the Unified Enterprise Survey & Reporting Arrangements Business Consultations – August 1997 – Guy Gellatly, Larry Murphy and Junior Smith.
4. Update on PIPES Progress: Notes for a Briefing for Federal and Provincial Finance Officials, Halifax, Nova Scotia, March 12, 1997 – Philip Smith.
5. An Overview of The Project to Improve Provincial Economic Statistics – November 1997 – George Beelen, Francine Hardy and Don Royce.
6. Using Databases to Design, Generate and Store Business Questionnaires at Statistics Canada – November 5, 1997 – Alana M. Boltwood.
7. The How and Why of Business Statistics – January 1999 – Elise Mennie. *(Not for external dissemination)*
8. An update on PIPES Fifteen Months into the Project – April 24, 1998 – Philip Smith.
9. Key Provider Manager (KPM) – 1997-98 Annual Report – May 1998 – Vicki Crompton.
10. A Framework for Planning Unified Enterprise Survey Data Collection – October 28, 1998 – Alana Boltwood.
11. Impact of the PIPES Funding on the Services Division Programme and Achievements in 1997-98 – April 1998 – Gordon Baldwin. *(Not for external dissemination)*
12. PIPES Organization and Decision-Making Structure – August 17, 1998 – Philip Smith. *(Not for external dissemination)*
13. The Central Goal of PIPES – November 17, 1997 – Philip Smith.
14. The Terminology and Framework of the Unified Enterprise Questionnaire – Revised March 1999 – Philip Smith.
15. Realizing and Measuring Quality Improvements in Provincial Economic Accounts – August 1998 – Philip Smith.
16. Annual Report 1997-98 – Ombudsman for Small Business Response Burden – July 1998, – Michael Issa. *(Not for external dissemination)*
17. Decision Making in PIPES – October 1, 1998 – Philip Smith.
18. Task Force on Electronic Data Reporting – April, 1998 – George Andrusiak, Monique Gaudreau, Laurie Hill, Anne Ladouceur, Denis Leblanc, Mario Ménard, Guy Parent, Joe Wilkinson, Doug Zinnicker.
19. PIPES Information Package – October 1998 – Philip Smith.
20. UES and the Non-Business Sectors – September 17, 1997 – Art Ridgeway.
21. CATS User Guide – April 1998 – Janet Howatson. *(Not for external dissemination)* Not available

22. Report on Collection and Data Capture Operation OID for UES 1997 – September 3, 1998 – Anne Ladouceur. *(Not for external dissemination)*
23. SDD Contribution to PIPES 1998-1999 – September 1998 – Shirley Dolan.
24. The Harmonized Sales Tax Revenue Allocation Formula – August 1998 – Karen Hall. *(Not for external dissemination)*
25. Task Group on Data Acquisition Strategy Report – July 22, 1998 – François Maranda and Don Royce.
26. Roles and Responsibilities in the Unified Enterprise Statistics Program – December 15, 1998 – George Andrusiak, Richard Barnabé, Albert Meguerditchian, Ray Ryan and Philip Smith. *(Not for external dissemination)*
27. Paper on the Project to Improve Provincial Economic Statistics from the Joint IASS/IAOS Conference – July 22, 1998 – Don Royce.
28. Respondent Relations Task Force – March 5, 1999 – Wayne Smith.
29. Response Analysis Follow-up Survey – March 1999 – Kristen Underwood.
30. Data Sharing Information Package – March 1999 – John Crysedale. *(Not for external dissemination)*
31. Coherence Analysis – Case Study from the Key Provider Manager Program – April 23, 1999 – Rachel Bernier and Julie Mandeville.
32. Evaluation of Collection Support Material used during the 1997 Unified Enterprise Survey – November 16, 1998 – Yvele Paquette.
33. Waiver Information Package – May 1999 – John Crysedale. *(Not for external dissemination)*
34. The PIPES Plan for 1999-00 – June 14, 1999 – Philip Smith. *(Not for external dissemination)*
35. BTS + Forum Post-conference Actions – April 1999 – Cornwall Conference Participants.
36. Report of the Task Force on Sources of Business Information – March 1999 – Vicki Crompton and Mark Marcogliese.
37. Field 5 Task Force Report on Improving Generic Boards – August 1999 – Mel Jones.
38. Study of Business Survey Questionnaires – June 1999 – Jason Gilmore.
39. Complexity Scale for Business Questionnaires – June 1999 – Jason Gilmore.
40. Update on PIPES – September 1999 – Philip Smith.
41. Exclusion Thresholds & Sampling Practices for Business Surveys – Implementation Strategy – September 1999 – Implementation Strategy Team.
42. Use of Tax Data in the Production of Provincial Economic Statistics – October 1999 – Peter Bissett.
43. Data Quality Survey 1996 – March 1999 – Ed Bunko. *(Not for external dissemination)*
44. Estimates of Information Cost to Business Respondents, 1998 – September 16, 1999 – Linda Grant and Michael Issa. *(Not for external dissemination)*
45. Data Security Task Force – January 1999 – Mark Steski. *(Not for external dissemination)*
46. Key Provider Manager Program – 1998-99 Program Report – June 1999 – Monique Gaudreau.
47. Treatment of Head offices and other support units in the UES – October 1999 – Charles Delorme.

48. UES Walkthrough – December 6, 1999 – Philip Smith.
49. The UES Strategy to Combine Enterprise, Establishment and Legal-Entity Data – November 29, 1999 – Philip Smith.
50. The Data Quality of the 1997 Unified Enterprise Survey (UES) Pilot – January 10, 2000 – Stuart Pursey.
51. Options for the Survey of Household Spending – January 2000 – Cynthia Baumgarten, David Binder, Louis-Marc Ducharme, Alison Hale, Albert Meguerditchian, Mike Sheridan, M.P. Singh, Philip Smith, Maryanne Webber, Brian Williams, Karen Wilson. *(Not for external dissemination)* AWAITING APPROVAL BY MIKE SHERIDAN
52. PIPES Budget Targets for 2001-02 and Beyond – March 28, 2000 – Project Management Team. *(Not for external dissemination)*
53. Unincorporated Business Statistics from the Tax Estimates Program – March 8, 2000 – Philip Smith.
54. PIPES and the Use of Statistics for Administrative Purposes – April 12, 2000 – Philip Smith.
55. Information Manual for Interviewers on the 1999 Survey of the Construction Industry – February 2000 – Claude Grenier. *(Not for external dissemination)*
56. The Conceptual and Practical Evolution of the Unified Enterprise Survey Integrated Questionnaire – May 2000 – Mario A. Vella and Annette Laurent.
57. Statistics Canada's Broad Strategy for Business Statistics – May 25, 2000 – Philip Smith.
58. Calendarizing Business Survey Data – June 2000 – Peter S. Tarassoff.
59. Report of the Task Force on Disclosure Avoidance Practices for Business Data – October 2000 – Daphne Bennett, John Crysdale, Ziad Ghanem, Pat Grainger, Tony Labilloy, Serge Lavallée, Jackie Leblanc, Robert Masse, Bruno Pépin, Randall Sheldrick, Leslie Shinder, Patricia Whitridge and Elaine Wilson. *(Not for external dissemination)*
60. Coherence Analysis of Large Enterprise Data – June 2000 – Monique Gaudreau and Jim Tebrake.
61. An Overview of the 1998 GIFI Database – July 26, 2000 – Jennifer Sarah Jones.

Other documents

PIPES Project Structure – Revised July 1999

Confidentiality of Statistical Data – April 1995 – D. Binder and L. Desramaux

The Business Register Hellerman Autocoder Project – November 1999 – Business Register Division
(English version only)

The Why of Business Surveys – March 2000

Autres documents

La structure du projet du PASEP – Révisé en juillet 1999

La confidentialité des données statistiques – avril 1995 – D. Binder et L. Desramaux

The Business Register Hellerman Autocoder Project – novembre 1999 – Division du registre des entreprises
(version anglaise seulement)

Le pourquoi des enquêtes-entreprises – mars 2000

44. Estimations des coûts d'information pour les entreprises répondantes, 1998 – le 16 septembre 1999 – Linda Grant et Michael Issa. (*Diffusion interne seulement*)
45. Groupe de travail de la sécurité des données – janvier 1999 – Mark Steski. (*Diffusion interne seulement*)
46. Programme des gestionnaires des répondants clés – Rapport de programme pour 1998-1999 – juin 1999 – Monique Gaudreau.
47. Traitement des sièges sociaux et autres unités de support dans l'EUE – le 28 octobre 1999 – Charles Delorme.
48. Tour d'horizon de l'EUE – le 6 décembre 1999 – Philip Smith.
49. Stratégie de l'EUE en vue de combiner les données au niveau de l'entreprise, de l'établissement et de l'entité légale – le 29 novembre 1999 – Philip Smith.
50. La qualité des données de l'Enquête-pilote unifiée de 1997 auprès des entreprises (EUE) – le 10 janvier 2000 - Stuart Pursey.
51. Options pour l'Enquête sur les dépenses des ménages – janvier 2000 – Cynthia Baumgarten, David Binder, Louis-Marc Ducharme, Alison Hale, Albert Meguerditchian, Mike Sheridan, M.P. Singh, Philip Smith, Maryanne Webber, Brian Williams, Karen Wilson. (*Diffusion interne seulement*) EN ATTENTE DE L'APPROBATION DE MIKE SHERIDAN
52. Budget projeté du PASEP pour les exercices 2001-2002 et ultérieurs – le 28 mars 2000 – Équipe de gestion de projet. (*Diffusion interne seulement*)
53. Statistiques sur les entreprises non constituées en société du Programme des estimations fiscales – le 8 mars 2000 – Philip Smith.
54. Le PASEP et l'utilisation des statistiques à des fins administratives – le 12 avril 2000 – Philip Smith.
55. Manuel d'information des intervieweurs – Enquête de 1999 sur l'industrie de la construction – février 2000 – Claude Grenier. (*Diffusion interne seulement*)
56. Évolution conceptuelle et pratique du questionnaire intégré de l'Enquête unifiée auprès des entreprises – mai 2000 – Mario A. Vella et Annette Laurent.
57. Stratégie générale de Statistique Canada relativement à la statistique des entreprises – le 25 mai 2000 – Philip Smith.
58. La calendrialisation des données des enquêtes-entreprises – juin 2000 – Peter S. Tarassoff.
59. Rapport du Groupe de travail sur les pratiques de contrôle de la divulgation des données sur les entreprises – octobre 2000 – Daphne Bennett, John Crysdale, Ziad Ghanem, Pat Grainger, Tony Labillois, Serge Lavallée, Jackie Leblanc, Robert Masse, Bruno Pépin, Randall Sheldrick, Leslie Shinder, Patricia Whitridge et Elaine Wilson. (*Diffusion interne seulement*)
60. Analyse de la cohérence des données sur les grandes entreprises – juin 2000 – Monique Gaudreau et Jim Tebrake.
61. Aperçu de la base de données de l'IGRF 1998 – le 26 juillet 2000 – Jennifer Sarah Jones.

Aperçu de la base de données de l'IGRF 1998

Le présent document vise à fournir un aperçu de l'IGRF, à un niveau très élémentaire, ainsi qu'à inciter les nouveaux utilisateurs à profiter de l'étendue et de l'exhaustivité de cet ensemble de données. Le document comporte quatre sections. La première constitue une introduction et décrit le contexte de l'IGRF et les avantages que comporte ce système pour Statistique Canada. La suivante sert à mettre l'IGRF en perspective, afin de donner à l'utilisateur une idée de la façon dont les données financières étaient recueillies par le passé, de même que de la situation actuelle et future de l'IGRF. La troisième section comprend des totalisations à partir de l'ensemble complet de données et décrit les métadonnées. La dernière section comprend des estimations fondées sur l'IGRF pour une branche d'activité particulière choisie comme exemple, à savoir les services personnels et services de blanchissage.

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Pour plus de renseignements sur ce document, veuillez communiquer
avec Bonnie Bercik (613) 951-6790 ou Diane Proulx (613) 951-7192
Télécopieur : (613) 951-0411

Aperçu de la base de données de l'IGRF 1998

Jennifer Sarah Jones
26 juillet 2000

Table des matières

Introduction	3
Objectif du document.....	3
Taxe de vente harmonisée (TVH).....	4
Rôle de Statistique Canada et du PASEP.....	4
Programme unifié des statistiques sur les entreprises (PUSE).....	5
En quoi consiste l'Index général des renseignements financiers (IGRF)?.....	6
Pourquoi mettre en œuvre l'IGRF?.....	7
Comment l'IGRF a-t-il été élaboré?.....	8
Qui profitera de l'IGRF?.....	9
Quelles sont les entreprises comprises et non comprises dans l'IGRF?.....	9
Quelle est la structure de l'IGRF?.....	10
Qu'est-ce que l'IGRF-abrégé?.....	13
Comment les données de l'IGRF sont-elles recueillies par l'ADRC?.....	13
Quels sont les avantages de l'utilisation des données de l'IGRF?.....	14
Quels sont les principaux défis auxquels Statistique Canada doit faire face pour l'utilisation de l'IGRF?.....	15
Comment peut-on accéder à l'IGRF?.....	16
Contexte	18
Registre des entreprises (RE).....	18
Structures légale et statistique des entreprises.....	19
Consolidation des données de l'entité légale et répartition au niveau de l'établissement.....	21
Comment les données financières étaient-elles recueillies par le passé?.....	22
Comment l'IGRF peut-il améliorer notre façon de faire?.....	23
Quelle est la situation actuelle de l'IGRF?.....	24
À quoi doit-on s'attendre pour l'avenir?.....	27
Totalisations	28
Métadonnées.....	28
Quelles sont les branches d'activité les mieux représentées par l'IGRF?.....	32
Démonstration de ce que comporte actuellement la base de données.....	37
Services personnels et services de blanchissage	40
Exemple de l'utilisation statistique de l'IGRF.....	40
Sous-secteur 812 du SCIAN — Services personnels et services de blanchissage.....	40

Services personnels et services de blanchissage — Estimation statistique au moyen des données de l'IGRF pour 1998	45
Information comprise dans le RE au sujet du sous-secteur des services personnels et services de blanchissage	46
Comparaison des données du RE et de celles de l'IGRF pour les services personnels et services de blanchissage.....	49
Solutions possibles pour la comparaison des données de l'IGRF et de celles du RE	51
Problèmes découlant de l'utilisation des données de l'IGRF relativement aux services personnels et services de blanchissage.....	52
Conclusion	52
Bibliographie.....	53

Introduction

Objectif du document

L'Agence des douanes et du revenu du Canada (ADRC), anciennement connue sous le nom de Revenu Canada, procède à de nombreux changements quant à sa méthode de collecte des données sur le revenu des entreprises. Grâce au remaniement des systèmes, l'ADRC sera en mesure de recevoir les états financiers des sociétés sous forme électronique, de façon sécuritaire et rapide. Afin d'appuyer ce nouveau système, l'ADRC a créé un index des postes financiers qui permettra d'uniformiser le processus de collecte tout en donnant aux comptables la souplesse nécessaire pour déclarer les données. Cet index est appelé Index général des renseignements financiers (IGRF), et les sociétés devront s'y conformer d'ici la fin de l'an 2000. Cela signifie que pour l'année d'imposition 2000, toutes les entreprises constituées en société soumettront leurs états financiers au moyen des codes de l'IGRF. L'IGRF deviendra aussi obligatoire pour les entreprises non constituées en société, et cela devrait se produire au cours des 3 à 5 prochaines années.

Statistique Canada est autorisé par la loi à accéder aux états financiers à des fins statistiques. Par le passé, il était coûteux et lent d'obtenir des données financières de l'ADRC. Par ailleurs, les données reçues ne comprenaient que peu d'information au sujet de nombreuses sociétés, ou encore comprenaient beaucoup d'information au sujet d'un petit échantillon de sociétés. Grâce à l'IGRF, on disposera non seulement d'un ensemble très détaillé de statistiques, mais on procédera aussi au recensement de 1,3 million de sociétés. Cet ensemble de données permettra d'améliorer considérablement la qualité des statistiques produites à Statistique Canada, de façon rapide et peu coûteuse.

Ce projet comporte de nombreux avantages pour les programmes de Statistique Canada. Le niveau de détail et la précision de ces statistiques sont élevés. Même si l'IGRF peut être utilisé pour améliorer la qualité statistique, il peut aussi contribuer à réduire le fardeau de réponse imposé aux entreprises. Les données devraient comporter des avantages, tant pour les anciens que pour les nouveaux programmes statistiques, et notamment ceux qui servent à produire les statistiques provinciales nécessaires pour la répartition de la taxe de vente harmonisée (TVH). Statistique Canada devra aussi surmonter de nombreux défis techniques et méthodologiques afin de bien utiliser ces données administratives. Il s'agit notamment de la consolidation des données des entités légales et de leur répartition au niveau de l'établissement. Il s'agit aussi de la confidentialité, de la mise en correspondance directe ou indirecte des données de l'IGRF avec celles des questionnaires d'enquête, de la modification des systèmes pour traiter des sommes importantes de données et des rapports avec un nombre accru d'utilisateurs.

Le présent document vise à fournir un aperçu de l'IGRF, à un niveau très élémentaire, ainsi qu'à inciter les nouveaux utilisateurs à profiter de l'étendue et de l'exhaustivité de cet ensemble de données. Le document comporte quatre sections. La première constitue une introduction et décrit le contexte de l'IGRF et les avantages que comporte ce système pour Statistique Canada. La suivante sert à mettre l'IGRF en perspective, afin de donner à l'utilisateur une idée de la façon dont les données financières étaient recueillies par le passé, de même que de la situation actuelle et future de l'IGRF. La troisième section comprend des totalisations à partir de l'ensemble complet de données et décrit les métadonnées. La dernière section comprend des estimations fondées sur l'IGRF pour une branche d'activité particulière choisie comme exemple, à savoir les services personnels et services de blanchissage. La consultation de l'ensemble du présent document fournira au lecteur un aperçu global de l'IGRF. Le lecteur peut aussi consulter des sections particulières correspondant à ses besoins propres.

Taxe de vente harmonisée (TVH)

La taxe de vente harmonisée (TVH) a été mise en œuvre le 1^{er} avril 1997, en Nouvelle-Écosse, à Terre-Neuve, au Nouveau-Brunswick et par le gouvernement du Canada. La TVH représente la combinaison de la taxe sur les produits et services (TPS) et de la taxe de vente provinciale (TVP) pour chaque province. Dans la plupart des provinces, on s'attend à ce que les sociétés produisent des formulaires de déclaration de revenu, non seulement pour chacune des provinces où elles ont des activités, mais aussi pour le gouvernement du Canada. Cela impose un fardeau considérable aux entreprises qui ont des activités dans plus d'une province. Par conséquent, le gouvernement fédéral et les trois provinces ont mis en œuvre la TVH. Celle-ci est perçue au même taux de 15 % dans chaque province et à l'égard des mêmes catégories de produits et de services. En outre, les sociétés n'ont qu'une déclaration de revenu à remplir, l'ensemble des opérations étant supervisé par un groupe d'administrateurs, et toutes les structures et règles étant les mêmes.

La TVH est une taxe sur la valeur ajoutée. Cela signifie que les producteurs reçoivent un crédit pour la TVH qu'ils versent à l'égard de leurs produits intermédiaires. Les produits intermédiaires sont ceux qui sont utilisés pour la production de produits ou de services finals. Si un autre producteur utilise ces derniers comme produits intermédiaires, en vue de la production d'un autre produit ou service, il reçoit aussi un crédit au titre de la taxe payable. Dans les faits, la taxe s'applique au produit final. Par conséquent, le consommateur final du produit débourse la totalité de la taxe sur la valeur ajoutée accumulée tout au long du processus de production.

L'ADRC est l'organisme chargé de percevoir la TVH. Des représentants du Trésor dans les divers gouvernements administrent la formule de répartition des recettes de la TVH entre les gouvernements. Le formulaire de TVH que les entreprises produisent comprend le total de la TVH qu'ils ont versé pour leurs achats et les crédits d'impôt auxquels ils ont droit à l'égard de leurs facteurs de production. Ce montant net est par la suite versé à l'ADRC, où l'ensemble de la TVH perçue va dans un compte commun de recettes. Le compte est actuellement constitué de la TVH, ainsi que de la TPS, dans le cas des provinces non harmonisées. L'ADRC reçoit ces recettes, mais elle ne dispose pas de données suffisantes pour savoir comment les diverses portions doivent être réparties entre les quatre gouvernements. Pour résoudre ce problème, on a créé une formule de répartition des recettes de la TVH. La formule nécessite des statistiques détaillées et fiables pour chaque province, afin de déterminer la répartition des recettes entre chaque gouvernement.

Rôle de Statistique Canada et du PASEP

Le compte de recettes de la TVH aurait pu être réparti entre les quatre gouvernements grâce à des données détaillées fournies par tous les inscrits aux fins de la TVH au sujet de leurs opérations taxables. Toutefois, cela aurait imposé un fardeau considérable aux entreprises. Par conséquent, les quatre gouvernements ont décidé d'envisager l'utilisation de statistiques macro-économiques en remplacement. Auparavant, Statistique Canada mettait l'accent sur les données nationales plutôt que provinciales, même s'il produisait déjà certaines données provinciales. Par ailleurs, on ne produisait pas de tableaux d'entrées-sorties pour chaque province, les détails et la fiabilité des données n'étaient pas très uniformes d'une province à l'autre, et les renseignements sur les biens et services et la branche d'activité étant insuffisants pour établir la formule de répartition. Statistique Canada avait beaucoup de chemin à faire pour produire des statistiques de grande qualité devant être utilisées pour la formule de répartition des recettes.

Le système des comptes provinciaux comporte beaucoup de détails et englobe une gamme importante de données économiques pertinentes au sujet de chaque province. Les données décrivent les biens ou les matières premières utilisés pour la production, les biens produits, les

acheteurs des produits, les sources de revenu dans la province, et les flux monétaires au sein des économies provinciales. Le système des comptes provinciaux comprend trois composantes : les comptes économiques provinciaux, les comptes de production provinciaux et les tableaux d'entrées-sorties provinciaux. Le ministère des Finances utilise ces statistiques pour la formule de répartition des recettes, en vue de calculer les parts pertinentes. L'utilisation de ces données pour la formule s'appliquant à la TVH sera plus efficace que la collecte des données directement auprès de toutes les entreprises.

Un projet a été établi à Statistique Canada afin que les statistiques provinciales atteignent le niveau de qualité approprié. Ce projet est appelé Projet d'amélioration des statistiques économiques provinciales (PASEP). Le projet vise à améliorer les comptes économiques provinciaux, à produire des tableaux annuels d'entrées-sorties pour chaque province, et à veiller à ce qu'il existe un niveau égal de statistiques détaillées et précises pour chaque province. Cela pose tout un défi, étant donné qu'aucun projet similaire n'a jamais été mené à Statistique Canada, ni par aucun autre organisme statistique dans le monde à cette fin. Le PASEP a été créé pour améliorer la qualité des données, et les divisions qui y participent doivent maintenant tenir à jour et améliorer encore davantage ces statistiques économiques provinciales pour l'avenir. Par suite du projet, on a assisté à une augmentation de la qualité des données d'enquête et des données administratives, ce qui devrait permettre de produire de meilleures données relativement à la TVH.

Le PASEP s'applique à la fois aux sources de données d'enquête et aux sources administratives. La source administrative la plus importante est le fichier des déclarations de revenu des sociétés. Par le passé, Statistique Canada recevait des états financiers de l'ADRC sous forme de fichier de l'univers « admin », avec un petit nombre de variables et un fichier échantillon comportant des renseignements plus détaillés. De fait, le fichier admin comprend peu de renseignements au sujet d'un grand nombre de sociétés, tandis que l'échantillon comprend une somme importante de données au sujet d'un petit nombre de sociétés. L'Index général des renseignements financiers (IGRF) remplacera à la fois le fichier administratif et l'échantillon. L'IGRF joue un rôle très important à l'égard du PASEP, du fait qu'il comprend beaucoup de renseignements au sujet d'un grand nombre de sociétés. En fait, il englobe toutes les données financières déclarées par les sociétés à l'ADRC, pour la population à peu près complète des sociétés. L'IGRF est une ressource très valable du fait qu'il s'agit d'un recensement de données détaillées et précises au niveau provincial, qui est essentiel à l'élaboration des comptes économiques provinciaux, ces derniers jouant un rôle crucial pour le processus de répartition de la TVH. En outre, l'IGRF comprend suffisamment de détails sur les coûts pour permettre le calcul du revenu brut, des dépenses intermédiaires et de la valeur ajoutée. Du fait de l'existence de cette ressource utile, il est possible, dans le cadre du PASEP, de réduire le fardeau pour les répondants et de limiter les coûts, tout en améliorant la qualité des statistiques produites.

Programme unifié des statistiques sur les entreprises (PUSE)

Pour mettre en œuvre la formule de répartition de la TVH, on a besoin de statistiques provinciales uniformes et cohérentes, de même que suffisamment étendues et exhaustives. Ces contraintes ont été prises en compte pour la conception du Programme unifié des statistiques sur les entreprises par le Secteur de la statistique du commerce et des entreprises. L'objectif de ce programme consiste à améliorer la qualité, l'uniformité et le niveau de détail des statistiques sur les entreprises, tout en réduisant le fardeau pour les répondants.

L'un des principaux objectifs du PUSE consiste à améliorer les relations avec les répondants grâce à l'unification des enquêtes. Le PUSE favorise l'intégration des questionnaires, de l'échantillonnage, de la collecte et du traitement postérieur à la collecte. À cette fin, tous les questionnaires peuvent être envoyés au siège social de l'entreprise, plutôt que séparément aux divers établissements qui la constituent, si tel est le souhait de l'entreprise. Cela contribue en

outre à améliorer la cohérence des données. La tenue des enquêtes au niveau de l'entreprise permet de vérifier le profil de chaque entreprise, pour une meilleure compréhension des entreprises complexes. Du fait de l'intégration des enquêtes, le Bureau améliore l'uniformité des données. En outre, l'adoption d'une terminologie et de concept uniformes pour les questionnaires contribue à l'atteinte de cet objectif. Ainsi, une entreprise reçoit de nombreux questionnaires dont la présentation est cohérente et raisonnable, sans chevauchement des questions. Parmi les autres éléments du programme figure la création d'une base unique de sondage. Il s'agit du Registre des entreprises (RE). Il existe aussi un système principal de gestion des contacts, qui sert à mesurer le fardeau de réponse imposé à chaque entreprise. Enfin, ce qui n'est pas à négliger, le programme est conçu pour favoriser une utilisation accrue des données administratives.

Comment les données administratives sont-elles utilisées?

En augmentant l'utilisation des données administratives, il est possible de produire davantage de données plus étendues et exhaustives, ainsi que d'assurer une couverture plus complète des statistiques sur la valeur ajoutée qui sont nécessaires pour les comptes provinciaux. L'exhaustivité des statistiques économiques est améliorée, du fait que les lacunes de données sont comblées et que la portée de celles-ci est améliorée, en raison du plus grand nombre d'enregistrements disponibles, ce qui permet la production d'estimations plus détaillées selon la branche d'activité et la province.

Les données administratives peuvent être utilisées directement pour les entreprises qui ne comportent qu'une entité légale (EL) et dont les activités se limitent à une branche d'activité et à une province, autrement dit, les entreprises « simples ». Les données peuvent être utilisées directement pour produire des statistiques sur la production. Lorsqu'il n'existe qu'une EL, mais que l'entreprise a des activités dans plus d'une branche d'activité ou d'une province, les données administratives peuvent être utilisées pour la production de statistiques sur les entreprises. Toutefois, elles ne comportent pas de renseignements sur la production selon la province ou la branche d'activité, à moins qu'il n'existe un profil de l'entreprise, ainsi que des données sur la répartition des variables du niveau de l'entreprise au niveau de l'établissement. Dans le cas des entreprises qui comptent plus d'une EL et qui ont des activités dans plus d'une branche d'activité ou d'une province, les données administratives ne peuvent être utilisées directement pour la production des statistiques au niveau de l'établissement ou de l'entreprise en l'absence d'un profil de l'entreprise et de renseignements devant servir au processus de consolidation et de répartition.

Les données de l'IGRF peuvent aussi être utilisées pour stratifier un échantillon, c'est-à-dire répartir l'échantillon entre les branches d'activité, les provinces et les groupes de taille. Cela vient du fait que le revenu brut d'entreprise (RBE), une variable de stratification clé, peut être obtenu directement à partir de l'IGRF. En outre, l'IGRF est utilisé parallèlement au numéro d'entreprise (NE) de l'ADRC, pour choisir des donneurs appropriés en vue de l'imputation dans les cas de non-réponse. Les données de l'IGRF combinées à d'autres données fiscales fournissent des renseignements précis pour les techniques de modélisation et l'estimation statistique. Cela contribue grandement à réduire le fardeau de réponse, à diminuer les coûts d'enquête et à accroître la fiabilité des estimations d'enquête détaillées. Cela a pour résultat que les enquêtes passent graduellement à ce cadre unifié.

En quoi consiste l'Index général des renseignements financiers (IGRF)?

La base de données de l'Index général des renseignements financiers, appelée aussi base de données de l'IGRF, représente un recensement des états financiers des sociétés disponible auprès de l'Agence des douanes et du revenu du Canada (ADRC). Il s'agit d'un index fondé sur

un système unique de codage qui permet aux sociétés de préparer leurs états financiers comme elles le veulent, et de les soumettre à l'ADRC de façon plus uniforme, efficace et rapide. L'index est constitué des postes les plus courants utilisés par les sociétés dans leurs états des résultats, leurs bilans et leurs états des bénéfices non répartis. Chaque poste comporte un code unique à quatre chiffres. Ces codes vont de 0000 à 9999. Le bilan comporte des codes qui vont de 1000 à 3849, et l'état des résultats, de 8000 à 9999. Globalement, 685 codes sont utilisés actuellement pour l'IGRF, ce qui signifie que parmi les 10 000 codes disponibles, de nombreux peuvent encore servir à une expansion future. Un graphique comportant quelques exemples figure ci-dessous et donne un aperçu des codes de variables qui sont disponibles dans l'IGRF :

BILAN		ÉTAT DES RÉSULTATS	
1000	Encaisse et dépôts	8000	Vente commerciale de biens et services
1120	Stocks	8160	Revenus provenant de la pêche
2360	Prêts à long terme	8910	Frais de location
2700	Dettes à court terme	9570	Remises
3849	Bénéfices non répartis/déficit - fin de l'exercice	9999	Revenu net/perte nette après impôts et éléments extraordinaires

Le système codé vise à permettre aux entreprises de structurer les données financières comme elles le souhaitent. Lorsqu'une entreprise prépare l'état financier qu'elle annexera à son formulaire T2 ou T1, chaque poste reçoit un code de l'IGRF; ce code, ainsi que le montant en dollars canadiens, sont alors communiqués sous forme électronique à l'ADRC, avec l'aide du logiciel informatique approprié. La première année de mise en œuvre a été fixée à l'année d'imposition 1998. En 1998 et 1999, toutes les sociétés doivent remplir leur déclaration de revenu comme d'habitude, et c'est l'ADRC qui se chargera de la transcription des postes selon les codes de l'IGRF. Pour l'année d'imposition 2000, les sociétés devront obligatoirement remplir leur déclaration de revenu sous forme électronique. Ce nouvel ensemble de données fournira à Statistique Canada l'ensemble de données financières sur les sociétés le plus détaillé, précis et actuel jamais produit.

Pourquoi mettre en œuvre l'IGRF?

L'ADRC modernise ses systèmes pour la perception de l'impôt des sociétés. L'Agence utilise les nouvelles technologies pour transmettre les données fiscales sur les sociétés sous forme électronique. Par conséquent, son système de transmission électronique des déclarations (TED) des sociétés sera remanié, afin de pouvoir transmettre des données fiscales, et son système de traitement le sera aussi, pour pouvoir traiter les dossiers fiscaux envoyés sous forme électronique. L'ADRC a élaboré l'IGRF à l'appui du nouveau système électronique. Comme il a déjà été mentionné, l'index fournit une façon de représenter les données financières sous forme électronique.

Le remaniement des systèmes de l'ADRC vient du fait que l'Agence souhaite disposer d'un moyen plus rapide, précis, sécuritaire et efficace de traiter les données fiscales des sociétés. Lorsque les sociétés envoient leurs formulaires de déclaration de revenu sous forme électronique, il n'est pas nécessaire de transposer à l'ordinateur les postes du formulaire sur papier. Cela augmente la précision et l'efficacité de l'ADRC. Par ailleurs, la vérification de l'information peut se faire de façon électronique plutôt que manuellement. L'ancienne méthode nécessitait beaucoup de main-d'œuvre et était lente. Aujourd'hui, les ordinateurs effectuent les vérifications presque instantanément, ce qui rend le processus plus précis et efficace. Dans le cas des entreprises, cela se révélera une façon beaucoup plus rapide de traiter les déclarations de revenu, le délai de traitement passant des 60 jours qu'il était auparavant à quelques jours

seulement. Les vérifications de l'ADRC seront aussi facilitées du fait de l'existence d'une base de données plus complète. On réalisera aussi des économies en raison du moins grand nombre de documents à traiter, et on réduira les coûts d'entreposage ainsi que les coûts d'exploitation. Enfin, un code chiffré permettra d'assurer la sécurité des données financières.

L'ensemble du processus sera plus efficace, non seulement pour l'ADRC, mais pour les entreprises, les fiscalistes, les comptables et les organismes gouvernementaux.

Comment l'IGRF a-t-il été élaboré?

L'IGRF a été élaboré par des représentants de l'ADRC, mais aussi de Statistique Canada, du ministère des Finances et des gouvernements provinciaux de l'Alberta et de l'Ontario, en consultation avec des associations comme l'Institut canadien des comptables agréés, l'Institut des dirigeants financiers du Canada, la Certified Accountants Association of Canada, le Tax Executives Institute, la Société des comptables en management du Canada et la Fédération canadienne de l'agriculture. On a aussi eu recours à des comptables et à des développeurs de logiciels, et on a examiné des centaines de déclarations de revenu des sociétés pour créer le système de l'IGRF.

Par le passé, l'ADRC recevait des états financiers correspondant au style propre aux comptables des entreprises concernées. Cette initiative permettra aux comptables d'adapter leurs produits aux besoins de leurs clients. Par ailleurs, l'ADRC souhaite tirer parti des possibilités d'efficacité accrue découlant de l'utilisation de formulaires envoyés sous forme électronique, et la seule façon d'y arriver consiste à avoir recours à des formulaires uniformisés. Si les sociétés continuaient toutes d'envoyer leurs états financiers selon leur propre style, cela irait à l'encontre de l'objectif visé par un système électronique. Au départ, l'ADRC avait retenu l'idée d'un formulaire entièrement uniformisé. Les comptables n'ont pas aimé cette idée parce que cela ne leur laissait pas suffisamment de souplesse. Les responsables de l'ADRC ont alors eu l'idée d'un index. Celui-ci donne une plus grande latitude aux comptables et permet d'assurer un certain niveau d'uniformité pour les formulaires de déclaration de revenu. Ainsi, l'ADRC peut traiter un certain nombre de variables tout en fournissant aux comptables la possibilité de choisir les variables qu'ils souhaitent utiliser. Évidemment, cet index comprend des variables génériques pour tous les postes qui ne comportent pas de catégorie type. Ces codes ont été établis afin que l'ajout de nouvelles variables ne nécessite pas la modification des codes précédents. Par exemple, le code 8299 correspondra toujours au total des revenus.

Les comptables peuvent percevoir les flux financiers d'une société de plusieurs façons. Il est important pour les personnes qui élaborent l'IGRF de comprendre ces flux, mais aussi d'en tenir compte. L'un de ces flux a trait à la nature des dépenses. Cela fait intervenir des éléments comme les salaires, les services publics, etc. Il existe une autre façon de faire, à savoir la division de l'organisation selon la gamme de produits. Cela nécessite de séparer chaque produit et de regrouper toutes les dépenses et les opérations s'y rapportant. Il existe une troisième classification, à savoir la gamme de processus. Elle consiste à diviser les comptes selon des processus de production distincts, comme c'est le cas pour les entreprises intégrées verticalement. On peut aussi examiner les différentes fonctions de l'organisation, par exemple, la formation, l'administration, le commerce de détail, etc., et structurer les comptes sur cette base. Les comptables choisissent l'une ou l'autre de ces approches lorsqu'ils établissent les états financiers. L'IGRF combine toutes ces approches pour permettre plus de souplesse.

Un autre aspect du processus de développement a trait au logiciel de production des déclarations de revenu. Pour que les sociétés puissent soumettre leurs déclarations de revenu sous forme électronique, elles doivent préparer leurs états financiers au moyen d'un logiciel fiscal. L'ADRC a incité des compagnies qui élaborent des logiciels comptables, comme ACCPAC et BEDFORD, à les remanier à l'appui de l'IGRF. Des liens doivent être établis entre les options qui figurent dans

les logiciels comptables et les postes de l'IGRF. Le comptable remplit la déclaration de revenu au moyen du logiciel, puis clique sur un bouton. Le logiciel attribue alors automatiquement un code de l'IGRF à chaque poste, et l'envoie à l'ADRC. Ce type de logiciel comptable est actuellement disponible, mais le système n'est pas encore en place à l'ADRC pour recevoir l'information électronique, ce qui fait que les sociétés doivent envoyer leurs codes de l'IGRF et les montants correspondants par la poste pour le moment, jusqu'à ce que le système de l'ADRC soit prêt à recevoir les formulaires en ligne.

Qui profitera de l'IGRF?

De nombreuses personnes et organisations sont susceptibles de profiter de l'IGRF. Les comptables recevront de l'ADRC des réponses rapides et détaillées à leurs questions parce que l'information sera obtenue et traitée rapidement et facilement par l'Agence. Le délai de traitement de leurs déclarations sera grandement réduit. Les vérificateurs pourront aussi utiliser l'IGRF pour les vérifications de conformité.

La Division des statistiques de l'ADRC utilise l'IGRF pour produire des statistiques et soumettre des données au ministère des Finances en vue de l'établissement de modèles de simulation fiscale. Ces modèles aident le ministère à élaborer des politiques et des dispositions législatives dans le domaine fiscal, et avec l'aide de l'IGRF, cela pourra se faire beaucoup plus rapidement. Statistique Canada profitera aussi de l'avènement de l'IGRF. Cette base de données contribuera à améliorer les statistiques utilisées pour le système de comptabilité nationale, les statistiques sur les branches d'activité, les statistiques financières et d'autres programmes statistiques.

Quelles sont les entreprises comprises et non comprises dans l'IGRF?

L'IGRF est constitué de la presque totalité de la population des entreprises qui produisent des déclarations T2 et T1. Les entreprises constituées en société produisent des déclarations T2. Une entreprise constituée en société est une entité légale distincte de son propriétaire. Les sociétés qui produisent des déclarations T1 (ou entreprises non constituées en société) ne sont pas des entités légales distinctes de leurs propriétaires. Il s'agit généralement de petites entreprises, et parfois simplement d'un expert-conseil qui travaille seul à domicile, ou d'une personne qui est propriétaire d'une ferme d'agrément. À l'heure actuelle, toutes les sociétés qui produisent des déclarations T2 figurent dans l'IGRF et, à l'avenir, les entreprises qui produisent des déclarations T1 seront aussi incluses. Les coentreprises ou les sociétés de personnes sont aussi couvertes par l'IGRF. Elles sont définies comme deux sociétés qui se sont unis à un moment donné, en vue d'exécuter une activité de concert, sans pour autant être considérées comme constituant une seule société. C'est pourquoi chaque société produit sa propre déclaration T2 pour ses propres activités, ainsi qu'une déclaration conjointe pour les fonctions combinées.

Tous les états financiers compris dans la base de données de l'IGRF se situent au niveau de l'entité légale. Une entité légale (EL) est une entreprise ou une compagnie constituée ou non en société, qui est définie techniquement comme une entreprise dont le droit de posséder des éléments d'actif, d'emprunter des ressources et d'exercer une activité économique est reconnu par la loi (document n° 1 de la série des documents techniques du PASEP, 1997). Les données de base de l'IGRF se situent au niveau de l'entité légale et n'ont pas été consolidées au niveau de l'entreprise, ni réparties au niveau de l'établissement¹.

¹ Les processus de répartition et de consolidation sont expliqués de façon plus détaillée dans la section intitulée « Consolidation des données de l'entité légale et répartition au niveau de l'établissement ».

Les déclarants T2 et T1 comprennent à la fois des entreprises simples et complexes. Une entreprise simple est une entreprise qui a des activités dans une province seulement et dans une seule branche d'activité. Une entreprise complexe, par ailleurs, est définie comme une entreprise qui a des fonctions dans plus d'une province ou dans plus d'une branche d'activité. La branche d'activité est définie par le Système de classification des industries de l'Amérique du Nord (SCIAN), un système de classification que Statistique Canada utilise pour regrouper les entreprises en différentes « classes ».

Tous les déclarants doivent produire une déclaration de l'IGRF, sauf les compagnies d'assurance et les sociétés non résidentes. L'IGRF n'est pas structuré pour répondre aux besoins des compagnies d'assurance qui ont des activités de souscription, par exemple, les assureurs sur la vie, les organismes d'assurance-dépôt, les assureurs de biens et les assureurs contre les risques divers. À l'heure actuelle, la Division de l'organisation et des finances de l'industrie (DOFI) procèdent à une enquête en vue de l'inclusion de ces compagnies d'assurance. Ces dernières seront ajoutées à l'IGRF une fois que l'on aura convenu de postes financiers pertinents pour elles.

Les sociétés inactives utilisent aussi l'IGRF. En outre, l'année suivant la constitution d'une entreprise en société, celle-ci produit deux bilans, à savoir un bilan d'ouverture et un bilan de fermeture. Toutefois, les sociétés n'utilisent pas l'IGRF lorsqu'elles produisent un état des charges reportées, du fait de leur participation à des activités de développement, ce qui a pour résultat qu'elles n'ont pas de revenu de ventes significatif et qu'elles ont des dépenses reportées.

Quelle est la structure de l'IGRF?

La base de données de l'IGRF comporte actuellement 685 postes financiers. La plupart des entreprises ne déclarent que de 40 à 50 de ces postes. Quelques-unes, soit environ 5 % de toutes les entreprises, ne déclarent que les variables obligatoires, tandis que les autres déclarent un peu plus que la moyenne des 40 à 50 postes. L'index comprend les postes qui figurent dans le bilan et l'état des résultats. Le bilan représente 48 % de tous les postes de l'IGRF et se divise en trois sections, comme il est indiqué ci-dessous. Chaque pourcentage correspond à la proportion du total des postes.

Actif	34 %
Passif	13 %
Bénéfices non répartis	2 %

L'état des résultats représente les 52 % qui restent des catégories de l'IGRF et se divise en deux sections :

Revenus et dépenses	34 %
Revenus et dépenses agricoles	17 %

À l'intérieur de ces sections figure des sous-sections. Le total de l'actif est constitué de l'actif à court terme, des immobilisations et de l'actif à long terme. Le total du passif est constitué du passif à court terme et du passif à long terme ainsi que des capitaux propres. L'information d'état des bénéfices non répartis ne comporte qu'un sous-secteur appelé bénéfices non répartis. La section des revenus et des dépenses de l'état des résultats comprend trois sous-sections: revenu non agricole, coût des ventes non agricoles et frais d'exploitation non agricoles. La section des revenus et des dépenses agricoles se répartit entre les sous-sections suivantes: revenus agricoles, dépenses agricoles ainsi qu'éléments extraordinaires et impôts sur le revenu. Une ventilation plus détaillée figure dans le tableau 1 ci-dessous.

Tableau 1 – Structure des postes du bilan et de l'état des résultats dans l'IGRF

Structure des postes du bilan dans l'IGRF			
Section	Sous-section	Blocs-totaux	Total des postes
Actif	Actif à court terme	12	61
	Immobilisations	14	124
	Actif à long terme	10	47
	Total	36	232
Passif	Passif à court terme	11	39
	Passif à long terme	12	38
	Capitaux propres	5	10
	Total	28	87
Bénéfices non répartis	Bénéfices non répartis	6	12
	Total	6	12
Total des postes du bilan :	Total	70	331

Structure des postes de l'état des résultats dans l'IGRF			
Section	Sous-section	Blocs-totaux	Total des postes
Revenus et dépenses	Revenu non agricole	10	67
	Coût des ventes non agricoles	10	47
	Frais d'exploitation non agricoles	25	121
	Total	45	235
Revenus et dépenses agricoles	Revenus agricoles	7	52
	Dépenses agricoles	5	59
	Éléments extraordinaires et impôts sur le revenu	7	8
	Total	19	119
Total des postes de l'état des résultats :	Total	64	354

Total des postes de l'IGRF :	Total	134	685
-------------------------------------	--------------	------------	------------

À l'intérieur de ces trois sous-sections, les blocs-totaux ont été définis et calculés par la Division des données fiscales (DDF). Le bilan comporte 70 blocs-totaux et l'état des résultats, 64, pour un total de 134. Ces blocs-totaux ne font pas partie des 685 variables totales de l'IGRF. Un bloc-total correspond à la somme des postes en dollars. Il peut s'agir de la somme de plusieurs postes ou d'un seulement. Il n'est pas toujours facile de déterminer quels postes sont inclus dans un bloc-total². Un bloc-total est facile à reconnaître parce qu'il est soit surligné dans une liste des variables de l'IGRF³ ou qu'il est représenté par un code unique à quatre chiffres précédé par les lettres BT dans les tableaux de l'IGRF⁴. On ne doit pas confondre les blocs-totaux et les postes génériques, même si les deux comportent des codes à quatre chiffres. Le poste générique

² Voir le « *Modèle conceptuel et le Modèle physique* » sous la rubrique « *Meta-données* » du site web de la DDF.

³ Une liste complète des variables de l'IGRF se trouve sous la section de l'Impôts des sociétés dans le site Internet de l'ADRC.

⁴ Les tableaux sont expliqués de façon plus détaillée dans la section Métadonnées du présent document.

représente une variable de l'IGRF. Il ne s'agit pas d'une somme comme les blocs-totaux. Il s'agit plutôt d'un poste que le comptable choisira lorsqu'il n'est pas sûr du meilleur poste à utiliser. Cela lui donne de la souplesse. Un poste générique comporte un code à quatre chiffres comme le bloc-total, mais il est précédé par un L. En résumé, le bloc-total représente la somme de nombreux postes, y compris les variables génériques, mais il s'agit d'une variable créée par Statistique Canada. Le graphique ci-dessous comprend des exemples de blocs-totaux et de variables génériques et de la façon dont les montants sont calculés.

Code	Nom de la variable	Bloc-total (calculé par SC)	Poste générique (disponible dans l'IGRF)
1000	Encaisse et dépôts	Cette zone correspond à la somme des postes 1000 à 1007.	Comprend tous les éléments d'encaisse et de dépôts qui ne peuvent être ventilés dans les zones 1001 à 1007.
2700	Dettes à court terme et prêts	Cette zone correspond à la somme des postes 2700 à 2706.	S'il n'est pas possible de déterminer la ventilation des dettes à court terme, on entre le montant total dans cette zone.*
8520	Publicité et promotion	Cette zone correspond à la somme des postes 8520 à 8524	Comprend les montants déclarés comme dépenses de publicité ou de promotion qui ne sont pas inclus dans les zones 8521 à 8524.
9180	Taxes foncières	Cette somme correspond uniquement au poste 9180.	**

* Cette zone comprend les exclusions qui n'ont pas été mentionnées dans ce graphique.

** Il n'existe pas de poste correspondant à cette zone.

Le déclarant de l'IGRF doit s'assurer que certaines variables de l'IGRF sont en équilibre. Il existe quatre règles pour l'équilibrage, à savoir :

Total de l'actif = total du passif + total des capitaux propres
 Total des revenus – total des dépenses = revenu non agricole net
 Total des revenus agricoles – total des dépenses agricoles = total des revenus agricoles nets
 Revenu net/perte nette = revenus net/perte nette après impôts et éléments extraordinaires

Chaque société doit aussi inclure toutes les variables obligatoires demandées par l'ADRC. Il existe neuf variables obligatoires. Parmi elles, six ou sept variables obligatoires doivent être déclarées par chaque société selon ses activités et sa situation. Ces variables sont les suivantes :

Postes obligatoires :

Poste 2599 - Total de l'actif
 Poste 3499 - Total du passif
 Poste 3620 – Total des capitaux propres
 Poste 9999 – Revenu net/perte nette après impôts et éléments extraordinaires

Postes obligatoires si la société a des activités non agricoles :

Poste 8299 – Total des revenus
 Poste 9368 – Total des dépenses

Postes obligatoires si la société a des activités agricoles :

Poste 9659 – Total des revenus agricoles
 Poste 9898 – Total des dépenses agricoles

Postes obligatoires si des montants sont inscrits pour les postes 3660 à 3744 (bénéfices non répartis)

Poste 3849 - Bénéfices non répartis/déficit – fin de l'exercice

D'autres postes verront le jour. En fait, certains codes sont déjà réservés à cette fin. Ils doivent être utilisés exclusivement par les entreprises qui produisent des T1 ou les entreprises non constituées en société, et les autres serviront au programme du Compte de stabilisation du revenu net (CSRN). Il s'agit d'un programme de filet de sécurité pour les agriculteurs qui est offert par le gouvernement fédéral et les provinces pour assurer la stabilité du revenu.

Qu'est-ce que l'IGRF-abrégé?

L'IGRF-abrégé a été conçu pour les sociétés qui n'utilisent pas de logiciel d'impôt pour produire leurs déclarations de revenu. Elles doivent avoir un revenu brut et un actif brut de moins de trois millions de dollars. L'IGRF-abrégé est une version papier de l'IGRF et comprend 100 des postes les plus couramment utilisés du bilan et de l'état des résultats. Ce formulaire comprend une section distincte qui englobe des postes essentiels pour les sociétés agricoles. Il s'agit d'une solution à court terme uniquement. À l'avenir, toutes les sociétés devront produire leurs états financiers au moyen du logiciel de l'IGRF, mais aucune date précise n'a été fixée à cette fin.

Comment les données de l'IGRF sont-elles recueillies par l'ADRC?

Chaque société doit envoyer ses états financiers chaque année à l'ADRC. Ces états financiers comportent quatre parties. La première est le bilan. Elle est constituée des détails de l'actif, du passif et de l'état des bénéfices non répartis. La deuxième partie est l'état des résultats, qui comprend des détails des revenus et des dépenses pour les sociétés agricoles et non agricoles. La « Liste de contrôle des notes » est le troisième formulaire et comprend quelques questions visant à aider l'ADRC à déterminer qui a préparé la déclaration de revenu et quel a été le niveau de leur participation. Elle sert en outre à déterminer le type d'information figurant dans les Notes aux états financiers. Ces notes, le dernier élément compris dans l'IGRF, sont aussi obligatoires. Elles expliquent tous les éléments des états financiers que la société juge bon d'expliquer. Par exemple, elles peuvent comprendre une explication de ce qui constitue un poste ou de la ventilation des postes. Lorsque la version électronique de l'IGRF deviendra obligatoire, en octobre 2000, ces notes devront être envoyées sous forme textuelle.

L'ADRC fait actuellement la transition de son ancien système manuel au nouveau système électronique. Les sociétés envoient leurs états financiers et leurs déclarations T2 aux directions générales locales de l'ADRC. Celles-ci évaluent chaque déclaration de revenu de la façon habituelle. À l'heure actuelle, les postes de la déclaration de revenu sont transcrits selon les postes de l'IGRF par l'ADRC. La personne qui saisit les données entre le code, suivi par le montant en dollar correspondant et un astérisque, par exemple, 1000 10000* 1060 25000*, etc. Ces données subissent alors une série de vérifications très limitées. Si le fichier réussit l'étape de la vérification, il va directement au Centre des données fiscales (CDF). Il existe six centres, et c'est là que les données sont entreposées. Si les données échouent l'étape de la vérification, elles subissent une nouvelle saisie. Qu'elles réussissent ou non cette étape, elles sont envoyées à un des CDF. Chaque fichier qui est rejeté à la vérification reçoit une étiquette invalide. À partir du centre de production, les fichiers sont envoyés à la Division du traitement des déclarations des entreprises du siège social de l'ADRC. Par la suite, elles sont envoyées à la Division des statistiques de l'ADRC, la dernière étape avant leur envoi à Statistique Canada. La Division des données fiscales (DDF) de Statistique Canada reçoit huit fichiers de données fiscales SAS. L'un d'eux est le fichier des données invalides. Les autres sont les fichiers qui ont réussi l'étape de la

vérification. Chaque fichier comprend un numéro d'entreprise (NE), un code de l'IGRF et le montant en dollar connexe pour chaque code. La DDF transfère toutes les données de ces fichiers dans une base de données Oracle qui est utilisée par les analystes spécialisés de Statistique Canada.

À l'avenir, il ne faudra plus transcrire les postes des déclarations de revenu selon les codes de l'IGRF, et les responsables du traitement des données n'auront plus à entrer les données sur papier à l'ordinateur. Toutes ces étapes se feront de façon électronique au moyen de logiciels de comptabilité. Au cours de la première année de l'IGRF, à savoir l'année d'imposition 1998, le logiciel de l'IGRF a été mis à la disposition des sociétés pour qu'elles l'utilisent. Même s'il n'a pas été disponible avant juin 1999, on a incité les entreprises qui n'avaient pas encore produit leur déclaration de revenu pour 1998 à l'utiliser. Pour 1998, environ 3 % des sociétés ont soumis leur déclaration de revenu en format IGRF, et environ 97 % des déclarations de revenu ont dû être transcrites par l'ADRC. Pour l'année d'imposition 1999, on s'attend à ce que le nombre de déclarants qui produiront leur déclaration de revenu selon le format de l'IGRF augmente, tandis que le nombre de fichiers devant être transcrits devrait diminuer. Tous les déclarants qui produisent des déclarations concernant des années prenant fin après le 31 décembre 1999 doivent utiliser l'IGRF. Toutefois, les états financiers ne pourront être classés de façon électronique avant octobre 2000, au moment où le nouveau système CORTAX entrera en vigueur à l'ADRC. Par la suite, les sociétés devront produire leurs états sous forme électronique.

Quels sont les avantages de l'utilisation des données de l'IGRF?

Statistique Canada peut tirer de nombreux avantages de l'avènement de l'IGRF. Globalement, la qualité des statistiques produites par le Bureau s'améliorera considérablement. Cela vient du fait que l'IGRF représente un recensement détaillé et précis des données financières. Le processus sera plus efficace parce que la presque totalité de la population des états financiers des entreprises qui produisent des déclarations T2 et des déclarations T1 figura dans l'IGRF. La base de données fournira aussi des statistiques très détaillées, en raison de l'augmentation du niveau de détail qui sera disponible selon la branche d'activité, la région et la taille de l'entreprise. La précision sera grandement améliorée, du fait que l'on ne devra pas procéder à une nouvelle entrée manuelle, ce qui devrait réduire les possibilités d'erreur humaine. Par ailleurs, toute la vérification des données se fera de façon électronique. Grâce à ces vérifications électroniques, les données seront reçues plus rapidement par Statistique Canada, ce qui augmentera l'actualité des statistiques que le Bureau produit. Il sera aussi possible de contrôler plus facilement les fermetures, les créations et les restructurations d'entreprises. On présume qu'une entreprise est inactive ou a cessé ses activités si elle ne produit pas de déclaration de revenu. Les entreprises de cette catégorie reçoivent une étiquette dans la base de données de l'IGRF de la DDF. Les créations sont notées dans une fonction d'indicateur de bilan d'ouverture (IBO), à laquelle on peut accéder dans le tableau de bilan de l'IGRF.

L'IGRF comporte d'autres avantages, dont le principal est la réduction du fardeau imposé aux répondants. Cela peut prendre trois formes. Le couplage des postes de l'IGRF et des questions des enquêtes peuvent contribuer à diminuer la longueur de ces dernières. On peut aussi remplacer les enquêtes, sur une base biennale, triennale ou complètement, grâce à l'utilisation des données de l'IGRF pour obtenir les données nécessaires les années où il n'y a pas d'enquête. En outre, l'IGRF peut être utilisé pour couvrir les très petites entreprises d'une branche d'activité. Ces entreprises, qui sont définies comme celles ayant des revenus annuels inférieurs à 30 000 \$, ne font jamais l'objet d'enquête. Enfin, cela augmente les possibilités des échantillons d'enquête pour les petites et moyennes entreprises. Il est aussi possible de réduire les coûts, grâce à l'élimination de la transcription et à la réduction des coûts de manutention des formulaires sur papier. La réduction de la paperasserie signifie en outre des résultats plus rapides, parce que l'on n'a plus à faire de recherches dans les états financiers, de photocopies ou de transmission de données à Statistique Canada. Parmi les autres aspects importants figure la

confidentialité. Les états financiers sont envoyés par les sociétés à l'ADRC, et par l'ADRC à Statistique Canada, au moyen d'un code chiffré, afin d'en assurer la sécurité.

Quels sont les principaux défis auxquels Statistique Canada doit faire face pour l'utilisation de l'IGRF?

Statistique Canada devra surmonter de nombreux défis pour exploiter pleinement la base de données de l'IGRF. Un des principaux obstacles a trait au traitement d'un volume important de données brutes de l'ADRC. Un nouveau système doit être conçu et élaboré pour résoudre ce problème. Jusqu'à maintenant, le système comporte une capacité d'entreposage suffisante, mais d'autres enregistrements doivent encore être fournis par l'ADRC. Le nombre de divisions qui tentent d'accéder à l'IGRF augmentera, ce qui aura un effet sur le nombre de demandes de soutien technique. Il existe aussi un problème de précision. Pour les années de référence 1998 et 1999, les données sont transcrites manuellement par l'ADRC. Cela signifie que des erreurs de codage humaines sont possibles, et on se demande si l'ADRC a disposé des ressources suffisantes pour lire toutes les notes accompagnant les états financiers. Même pour l'année de référence 2000, pendant laquelle on assurera la collecte et la vérification électroniques des données, on se demande encore si l'attribution automatique des codes de l'IGRF par le logiciel comptable sera précise pour toutes les entreprises, étant donné que des problèmes de transition sont susceptibles de se poser.

Un autre défi a trait au choix de la période de production des déclarations de revenu par les entreprises. Tout porte à croire que les entreprises choisiront leurs propres exercices comme période de déclaration. Ceux-ci varient d'une entreprise à l'autre. Aux termes de la loi, cette période doit porter sur 12 mois consécutifs. Évidemment, cela ne s'applique pas aux nouvelles entreprises, ni aux entreprises qui cessent leurs activités ou qui fusionnent, lesquelles peuvent produire une déclaration pour une période inférieure à 12 mois. Idéalement, Statistique Canada préférerait que chaque entreprise présente sa déclaration à la fin de l'année civile. Pour résoudre ce problème, le Bureau doit calendrialiser les données. Si un formulaire est reçu pour un exercice qui ne porte pas de janvier à décembre, les données doivent être transposées pour correspondre à l'année civile de référence. Par le passé, il est arrivé souvent que l'on n'a pas tenu compte de ce problème, et on a additionné des exercices comme s'il s'agissait d'années civiles. Cette approche constituait parfois une approximation acceptable, lorsque les statistiques en découlant ne s'appliquaient qu'à des niveaux élevés d'agrégation, mais cette solution n'est plus appropriée pour la production d'estimations pour un grand nombre de cellules de province ou de branche d'activité. Un autre problème se produit lorsque les EL comportent plus d'une période de déclaration. Cela peut se produire lorsqu'une EL se fractionne en deux EL, ou que deux EL fusionnent. À l'heure actuelle, le Bureau abandonne tous les enregistrements « supplémentaires » de l'exercice plus court afin d'éviter ce dénombrement en double. À l'avenir, nous aimerions fusionner tous les enregistrements et additionner toutes les variables de flux, tirer les variables de stocks du dernier enregistrement et rééquilibrer l'enregistrement final.

Il est difficile d'assurer la confidentialité des données, mais il s'agit là d'une exigence essentielle. Aux termes de la *Loi de l'impôt sur le revenu* et de la *Loi sur la statistique*, les sociétés sont protégées contre la divulgation de leurs états financiers. L'ADRC importe et exporte les états financiers au moyen d'un système informatique encodé très élaboré. Une fois que les données parviennent à Statistique Canada, deux stratégies principales sont utilisées. Dans un cas, les employés qui utilisent les données doivent être expressément autorisés à le faire par la direction. L'utilisateur des données doit signer un formulaire pour indiquer qu'il comprend que les données sont confidentielles et qu'elles doivent être traitées en conséquence. En outre, les règles de

Duffet sont utilisées pour éviter toute divulgation par recoupement non voulue de renseignements confidentiels⁵.

Les données de l'IGRF sont reçues de l'ADRC au niveau de l'entité légale. En général, les entités légales peuvent comporter plus d'un établissement et avoir des activités dans plus d'une branche d'activité. Cela rend difficile l'attribution d'un code unique du SCIAN dans certains cas. Si une entreprise n'a qu'un établissement pour chaque EL, c'est-à-dire un rapport biunivoque, l'attribution des codes du SCIAN ne pose pas de problème. La présence de plusieurs EL, ou d'une EL comportant plusieurs établissements dans différentes branches d'activité, peut poser des problèmes. On compte environ 1 500 entreprises au Canada qui comportent plus d'une EL, et nombre d'entre elles sont parmi les plus importantes au pays. Aujourd'hui, pour attribuer un code du SCIAN à une de ces EL, la DDF utilise le Registre des entreprises (RE). Dans le RE, chaque EL correspond à une entreprise, et chacune de ces entreprises se voit attribuer un code du SCIAN en fonction de sa principale activité. L'EL reçoit le même code du SCIAN que l'entreprise. Toutefois, cela pose un problème. Par exemple, une EL peut avoir des activités de commerce de détail, mais parce que l'activité prédominante de l'entreprise est la fabrication, lorsque l'on cumulera les enregistrements de l'IGRF, les données relatives au commerce de détail se retrouveront dans la branche de la fabrication. Cela signifie que les données sur la branche d'activité pourraient faire l'objet d'une surestimation ou d'une sous-estimation. Un groupe de travail examine actuellement les 1 500 entreprises comportant plusieurs EL et tente de déterminer la meilleure approche pour résoudre ce problème. Heureusement, pour la plupart des enregistrements de l'IGRF (qui portent sur les petites entreprises), ce problème ne se pose pas.

Le couplage des enregistrements de l'IGRF et du Registre des entreprises au moyen du numéro d'entreprise de l'ADRC, et l'attribution du code du SCIAN figurant dans le RE à l'enregistrement de l'IGRF ne fonctionnent de toute évidence pas en l'absence de liens. Il est important que de tels liens existent pour maximiser le nombre d'enregistrements disponibles comportant un code de branche d'activité. S'il manque des liens, cela aura essentiellement pour effet d'entraîner la perte d'enregistrements de l'IGRF. Il peut exister plusieurs raisons pour justifier l'absence de tels liens. Il peut arriver que des NE figurant dans l'IGRF ne se trouvent pas dans le RE. Et même lorsqu'il n'existe pas de problème de couplage des NE, il peut arriver qu'un code du SCIAN n'ait pas été attribué dans le RE. Un autre groupe de travail s'occupe actuellement du problème de couplage du RE.

En général, les données administratives n'englobent pas tous les renseignements dont Statistique Canada a besoin. Il arrive parfois que des unités statistiques que nous aimerions examiner ne soient tout simplement pas disponibles. Cela vient de ce que l'IGRF ne représente qu'un ensemble de données financières, ce qui fait que les données non financières, comme les entrées et les sorties de biens, ne sont pas disponibles. Pour combler ces lacunes, Statistique Canada aura toujours besoin des enquêtes. Il existe un autre problème, à savoir le couplage de variables administratives particulières avec des questions d'enquête. Il arrive parfois que les variables de l'IGRF ne correspondent pas exactement à nos variables cibles et doivent être extraites indirectement, par exemple, grâce à la somme de quelques postes. Enfin, même si un poste de l'IGRF correspond directement à une variable cible, il n'existe pas de garantie que le niveau de détail nécessaire sera disponible dans l'IGRF. L'attribution d'un code à un poste ne signifie pas nécessairement que toutes les entreprises y inscriront une valeur.

Comment peut-on accéder à l'IGRF?

Les données financières d'une entreprise sont des données très confidentielles. Le gouvernement fédéral a mis en place des politiques pour protéger la confidentialité des états

⁵ Les règles de Duffet sont mentionnées à nouveau dans la section Totalisations du présent document, dans l'analyse du Tableau 5 - Code 452991 du SCIAN, Magasins de fournitures pour la maison et l'auto

financiers des sociétés. Pour y avoir accès, il faut obtenir une autorisation de sécurité. L'utilisateur doit aussi être autorisé par son surveillant, et il doit signer le formulaire d'accès aux données indiquant qu'il comprend les problèmes de confidentialité liés à l'utilisation des ensembles de données de l'IGRF. Tous les utilisateurs de l'IGRF doivent se conformer à un certain nombre de procédures de sécurité importantes :

1. L'utilisateur ne peut se servir d'un ordinateur muni d'un commutateur A/B. Le commutateur B est externe et pourrait permettre à des personnes non autorisées de l'extérieur d'entrer dans le système et d'accéder à des données de nature délicate.
2. Les tableaux sommaires des ordinateurs personnels doivent comporter l'avertissement suivant : « Le présent document comprend des données semi-agrégées dont la confidentialité n'a pas été vérifiée. CES DONNÉES DOIVENT ÊTRE TRAITÉES COMME "CONFIDENTIELLES" ».
3. Il existe des procédures spéciales pour l'impression à partir du processeur central⁶.
4. Les utilisateurs doivent s'assurer que leur poste de travail est toujours verrouillé lorsqu'il est laissé sans surveillance. Cela dissuadera des personnes non autorisées de l'extérieur d'utiliser les données à mauvais escient.

Une fois que l'utilisateur comprend l'importance de la sécurité, il peut consulter le site Internet de la Division des données fiscales (DDF), sous Accès aux données, en vue d'obtenir plus de renseignements au sujet de l'accès aux données. Il existe deux documents. Le premier explique les procédures d'accès, et le deuxième est une copie électronique du formulaire d'autorisation d'accès aux données. Les formulaires sont aussi disponibles auprès de la DDF. L'utilisateur devra indiquer pourquoi il souhaite avoir accès aux données, et il devra fournir une description des données qu'il souhaite obtenir (dans ce cas, il s'agit des données de l'IGRF). La Division des données fiscales doit aussi savoir qui utilisera les données et quand l'accès aux données prendra fin. Le directeur dont relève l'utilisateur et le directeur de la Division des données fiscales doivent donner leur autorisation. Le délai d'attente pour l'accès aux données, une fois que le formulaire de demande a été rempli, dépend de la disponibilité des employés. Généralement, le temps d'attente va de quelques jours à un peu plus d'une semaine.

Une fois que l'accès a été accordé, la Division des données fiscales en informe l'utilisateur. Celui-ci devra choisir un mot de passe qui lui permettra d'accéder aux données. Le mot de passe doit être entré chaque fois qu'un programme est exécuté. Il existe deux façons d'entrer le mot de passe. Dans un cas, on accède à l'Interface universelle de connexion aux bases de données et on crée un message guide-opérateur. Celui-ci est plus sécuritaire parce qu'il fait en sorte que les utilisateurs doivent entrer leur mot de passe deux fois pendant l'exécution du programme. Par ailleurs, cette façon de faire prend du temps. La deuxième méthode consiste à accéder directement à Oracle. Dans ce cas, l'entrée en communication de l'utilisateur, le mot de passe et le chemin d'accès sont compris dans le code. Cette méthode est la plus facile, parce qu'elle ne nécessite qu'une seule entrée du mot de passe. Il faut toutefois bien s'assurer de garder ce mot de passe secret. Les deux instructions en SAS du message guide-opérateur PROC SQL se présentent ainsi :

```
Connect to ODBC as mydb (PROMPT);  
Connect to ORACLE as mydb (user=lastfir password=XXX path="...");
```

Un autre document sera remis à l'utilisateur lorsque l'accès sera autorisé. Il explique comment accéder à la base de données Oracle à partir d'un ordinateur personnel. Évidemment, l'administrateur de réseau de l'utilisateur peut aider ce dernier s'il a des questions ou des difficultés.

⁶ Ces procédures spéciales et d'autres renseignements figurent dans un document intitulé « Procédures à suivre pour obtenir l'accès aux données fiscales ». Celui-ci se trouve dans le site Internet de la DDF sous Accès aux données.

Pour commencer, l'utilisateur doit acquérir certaines connaissances. Les données se trouvent dans une base de données Oracle, mais elles ne sont pas manipulées directement au moyen d'outils Oracle. Pour manipuler les données, l'utilisateur devra apprendre SAS ou MS-Access. Pour accéder aux données, l'utilisateur devra aussi se familiariser avec SQL. MS-Access est un outil logiciel relativement facile à utiliser. Il comporte une interface conviviale et produit des tableaux plus agréables à consulter, mais SAS est souvent le logiciel privilégié, parce qu'il est plus souple et efficace pour traiter des ensembles importants de données. Tant pour SAS que pour MS-Access, les petits programmes qui ne comportent pas un nombre important de variables ou de manipulations de ces variables fonctionnent bien. Si de nombreuses variables sont nécessaires, les programmes devront être exécutés la nuit, ou mieux encore, indépendamment du serveur.

L'utilisateur devra aussi apprendre les métadonnées de l'IGRF⁷. Cela n'est pas simple. Il devra comprendre comment les codes de l'IGRF sont établis (tant pour les blocs-totaux que pour les variables), ainsi que les instructions relatives aux enregistrements. Il peut s'agir de choisir un code particulier du SCIAN, ou encore d'éliminer tous les enregistrements qui comportent plusieurs fins d'exercice. Enfin, il est très important de comprendre les éléments de données qui sont entreposés dans les différents tableaux et la façon dont ces tableaux sont couplés en Oracle⁸. Ce sont généralement les débuts qui sont les plus difficiles. Des modèles de programmes sont disponibles auprès de la DDF. Ils peuvent être copiés en SAS ou en MS-Access pour s'assurer que l'IGRF est connecté et fonctionne de façon appropriée.

Contexte

Registre des entreprises (RE)

L'univers complet des entreprises, c'est-à-dire toutes les entreprises au Canada, est constitué de trois à quatre millions d'entreprises. La grande majorité de celles-ci sont très petites, et Statistique Canada n'a pas les ressources pour conserver tous leurs enregistrements sur fichier. Par définition, une très petite entreprise est une entreprise dont le revenu total est inférieur à 30 000 \$ par année. Si l'entreprise a un revenu total inférieur à 30 000 \$ et qu'elle est *constituée en société*, elle figure dans le RE. Le RE comprend des renseignements sur toutes les entreprises constituées en société et certaines sociétés non constituées en société au Canada. Il a été créé initialement dans les années 80, à partir du Système de retenues sur la paye. Seules les entreprises comptant des employés figuraient dans le fichier. Le RE comportait des renseignements au sujet des créations d'entreprises, des fermetures d'entreprises et de la taille des entreprises. Au milieu des années 90, l'ADRC a créé le système de numéro d'entreprise (NE) unique. Un NE est un numéro qui est attribué à chaque entité légale au Canada et qui est utilisé pour identifier une entreprise aux fins de l'impôt. Le système des NE, combiné à l'utilisation des fichiers de la TPS qui englobent les entreprises sans salariés, a permis à la Division du RE de faire passer la taille du RE à deux millions d'entreprises. Environ un million d'entre elles sont des entreprises constituées en société, et un million sont des entreprises non constituées en société. À l'heure actuelle, les données de l'IGRF sont disponibles pour toutes les entreprises constituées en société, mais aucune donnée n'est disponible pour celles qui ne le sont pas.

Les entreprises sont aussi réparties entre les entreprises complexes et les entreprises simples. Une entreprise complexe est une entreprise qui comprend des établissements dans plus d'une province et dans plus d'une branche d'activité. Une entreprise simple se situe exactement à l'opposé. Elle a des activités dans une seule province et dans une seule branche d'activité. On

⁷ Voir la section du présent document intitulé *Métadonnées*. Pour plus de renseignements, voir la note numéro 2 en bas de page.

⁸ Voir la note numéro 2 en bas de page.

compte environ 10 000 entreprises complexes, qui comprennent près de 70 000 établissements. Parmi ces 10 000 entreprises, seulement 1 500 sont des entreprises comportant plusieurs entités légales. Même si les entreprises complexes sont relativement peu nombreuses, elles représentent près de la moitié de toute la production des entreprises, et la plupart d'entre elles sont constituées en société. Les entreprises simples, par ailleurs, représentent les quelque 1 990 000 entreprises qui restent dans le RE. Environ la moitié de toutes les entreprises simples figurant dans le Registre est constituée en société, et l'autre moitié ne l'est pas.

Structures légale et statistique des entreprises

Statistique Canada reçoit des données financières sur les entreprises à partir des données fiscales de l'ADRC. Les formulaires de déclaration de revenu remplis par les entreprises sont du ressort des EL de cette entreprise. La structure légale correspond aux formulaires de déclaration T1 ou T2. (Voir le diagramme 1 - Structure légale.) Le Secteur de la statistique du commerce et des entreprises s'intéresse à la collecte de statistiques sur la base de l'activité. Toutefois, la structure légale ne correspond pas toujours parfaitement à la branche d'activité. L'entreprise complexe qui a des établissements dans plus d'une branche d'activité peut avoir une EL ou plus, et si elle compte plus d'une EL, il se peut qu'il ne soit pas possible de mettre en correspondance les EL et les établissements. Par conséquent, étant donné que les entreprises complexes jouent un rôle important au sein d'une branche d'activité, il est difficile de produire des estimations statistiques uniquement à partir des données fiscales sur la base de l'EL.

Il existe aussi une structure d'exploitation dans le RE, qui correspond à la façon dont l'entreprise se perçoit et perçoit ses activités. À partir de la structure légale et de la structure d'exploitation d'une entreprise, Statistique Canada crée une ventilation statistique d'une entreprise que l'on désigne sous le nom de structure statistique. (Voir le diagramme 2 - Structure statistique.)

Diagramme 1 – Structure légale d'une entreprise

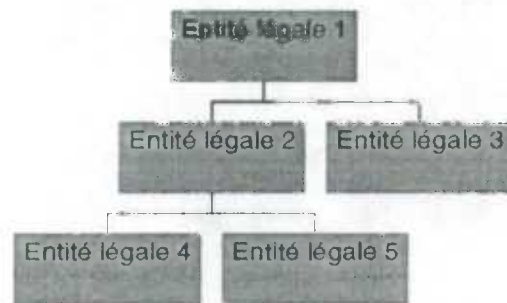
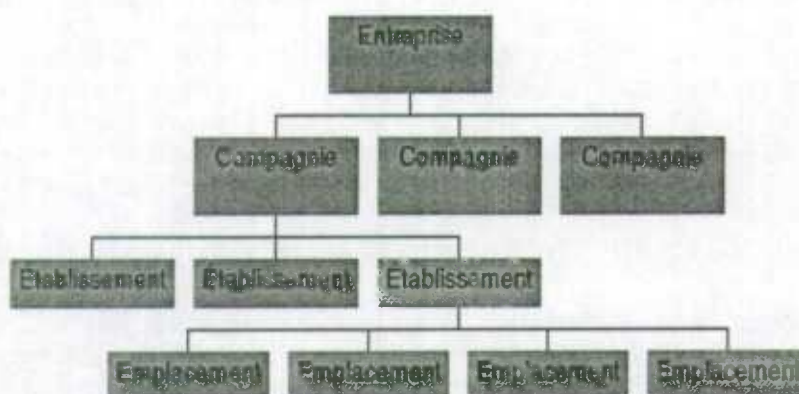


Diagramme 2 – Structure statistique d'une entreprise



Cette structure statistique comporte quatre parties. La meilleure façon de la consulter est de bas en haut. L'emplacement correspond au lieu d'affaires de l'entreprise. Généralement, un gestionnaire est autorisé à prendre des décisions pour cet emplacement exclusivement, mais il ne peut intervenir dans les décisions concernant la comptabilité ou le recrutement. L'établissement est constitué d'un ensemble d'emplacements de la même branche d'activité. Le gestionnaire est autorisé à prendre des décisions au sujet des activités au jour le jour, comme le recrutement, les ventes et les achats, ainsi que la comptabilité. Ce gestionnaire a lui aussi un patron et ne peut prendre de décisions à long terme. La compagnie est constituée de plusieurs établissements. À ce niveau, on prend des décisions sur les investissements, le financement, et toute autre décision à long terme ou décision qui a des répercussions sur la croissance de la compagnie. L'entreprise est constituée de toutes les compagnies ou essentiellement de l'ensemble de l'entreprise. L'entreprise gère tout et possède tout. Dans le cas d'une entreprise simple, tous ces niveaux coïncident au niveau de l'entreprise. Statistique Canada s'intéresse le plus à l'établissement, parce que c'est celui qui reçoit le code de branche d'activité. Les établissements d'une entreprise complexe importante sont comparables aux petites unités de production des entreprises simples. La branche d'activité où évolue un établissement sert à déterminer les enquêtes qui s'appliquent à chaque établissement. Nous avons besoin de données au niveau de l'établissement pour assurer l'uniformité. Si nous voulons additionner des variables pour l'ensemble d'une branche d'activité, d'une province ou du pays, ou si nous voulons comparer des variables, les données doivent être uniformes. Il est aussi important de souligner que seules les données de l'état des résultats sont réparties au niveau de l'établissement, cette répartition ne touchant pas les données du bilan. Cela vient du fait que les établissements, par définition, ne sont pas chargés d'éléments comme l'actif et le passif. Les entreprises et les compagnies sont responsables des données du bilan. Par conséquent, ces données sont regroupées au niveau de la compagnie ou de l'entreprise, tandis que celles de l'état des résultats sont réparties au niveau de l'établissement.

Auparavant, on utilisait la Classification type des industries (CTI) de 1980 pour attribuer les codes de branche d'activité. La CTI a vu le jour à la fin des années 70, mais les codes ne sont plus à jour parce que les branches d'activité sont différentes de ce qu'elles étaient il y a 20 à 25 ans. Statistique Canada utilise maintenant le Système de classification des industries de l'Amérique du Nord (SCIAN). Statistique Canada, l'Office of Management and Budget des États-Unis et l'Instituto Nacional de Estadística, Geografía e Informática du Mexique ont conçu ce système de codage qui a vu le jour en 1997. Les codes sont non seulement les mêmes pour l'ensemble de l'Amérique du Nord, mais les branches d'activité proprement dites sont plus à jour et sont par conséquent plus pertinentes. Pour attribuer les codes, nous devons passer des données au niveau de l'EL à celles au niveau de l'établissement. Toutefois, cela ne se fait pas toujours en une étape. Les données doivent être regroupées au niveau de l'entreprise, puis réparties de

nouveau au niveau de l'établissement, à partir des données d'enquête et du RE. Un code de branche d'activité peut alors être attribué.

Consolidation des données de l'entité légale et répartition au niveau de l'établissement

Un des défis que pose l'IGRF vient du fait que les données se situent au niveau de l'EL. Pour obtenir les données au niveau de l'établissement, il faut procéder à une consolidation au niveau de l'entreprise (si l'entreprise fait partie des quelque 1 500 entreprises comportant plusieurs EL), puis procéder à une nouvelle répartition au niveau de l'établissement (si l'entreprise est « complexe »). Pour les entreprises simples qui comportent une EL et un établissement, le processus de consolidation ou de répartition n'est pas nécessaire, ce qui fait qu'aucun problème ne se pose. Ces entreprises représentent 99 % de toutes les entreprises. Pour les 1 % qui restent, qui représentent environ la moitié de toute l'activité des entreprises, cette tâche n'est pas toujours facile. Le processus est décrit ci-dessous.

La consolidation des données de l'EL au niveau de l'entreprise s'effectue de deux façons différentes. Tout d'abord, on utilise les données de l'Enquête trimestrielle sur les états financiers de la DOFI. Cette enquête comporte des questions particulières au sujet de l'entreprise dans son ensemble, pour les entreprises comportant plusieurs entités légales. Lorsque ces données d'enquêtes sont disponibles, elles peuvent être utilisées comme données consolidées pour l'entreprise, en remplacement des données des EL de cette entreprise. La deuxième façon consiste à procéder à une approximation qui nécessite uniquement de faire la somme des données pour toutes les EL de l'entreprise. On peut déterminer quelles EL appartiennent à quelle entreprise à partir des données recueillies dans le cadre du programme de la *Loi sur les déclarations des personnes morales et des syndicats* de la DOFI.

La répartition des données de l'entreprise au niveau de l'établissement est un peu plus compliquée parce que la perception qu'a l'entreprise d'elle-même ne correspond pas à la structure statistique. Il existe trois façons d'obtenir des données au niveau de l'établissement. La première repose sur la Partie 2 du questionnaire de l'Enquête unifiée auprès des entreprises. Cette partie de l'enquête comprend des questions destinées à l'établissement d'une entreprise. Il s'agit de la méthode privilégiée parce que les données sont précises et détaillées. La deuxième façon de répartir les données fait intervenir ce que l'on appelle la feuille de répartition. Ce questionnaire est similaire à la Partie 2, mais a une portée plus générale. Il s'adresse aux entreprises qui comptent de nombreux établissements et qui ont besoin de plus de souplesse en raison de contraintes de temps et d'autres contraintes. Les données sont recueillies auprès d'une entité de collecte (EC) qui peut, par exemple, comprendre tous les établissements d'une province donnée. Une EC correspond à de nombreux établissements et est définie aux fins de la collecte. Évidemment, la répartition doit encore être effectuée du niveau de l'EC au niveau de l'établissement, et cela se fait au moyen des données de la feuille de répartition. Enfin, il existe une troisième méthode de répartition, à savoir le recours au RE où l'on trouve des données de stratification quantitative au sujet de chaque établissement, par exemple, le nombre d'employés. À partir de ces éléments, on peut déterminer la taille relative de l'entreprise, et d'autres variables peuvent être réparties entre chaque établissement à partir de l'indicateur de taille. Évidemment, ces variables ne sont que des estimations brutes et ne sont pas toujours à jour, ce qui fait que ce type de ventilation n'est pas toujours précis.

Un problème majeur se pose lorsqu'une EL produit une déclaration de revenu deux fois la même année. Cela peut venir du fait que l'EL tente de modifier son exercice, ou qu'une fusion s'est produite. La meilleure façon de traiter ces déclarants qui produisent plusieurs déclarations la même année consisterait à fusionner tous les enregistrements, c'est-à-dire à additionner toutes les variables de flux, à tirer les variables de stocks du dernier enregistrement et à rééquilibrer l'enregistrement final.

Il existe aussi des possibilités dont Statistique Canada devrait tirer parti. Par exemple, lorsqu'une EL correspond exactement à un établissement, il serait avantageux que le RE en tienne compte. Cela réduirait la somme de travail nécessaire pour la consolidation et la répartition, et permettrait d'obtenir la même ventilation. Toutefois, il est très long d'examiner ces éléments pour chaque entreprise complexe.

Comment les données financières étaient-elles recueillies par le passé?

Fichier de l'univers d'enquête (FUE) T2

Le FUE T2 de la DOFI a été créé par l'ADRC pour SC, plutôt qu'à partir du RE, comme cela a été le cas pour le Groupe de l'accès aux dossiers fiscaux. L'ADRC a produit un fichier appelé CORPAC, qui comprend certains éléments d'information au sujet d'une société, ainsi que cinq variables financières. Ces variables sont le total de l'actif, le total des revenus, les bénéfices totaux avant impôts et le revenu imposable. Ce fichier était envoyé à la DOFI sur bande et comprenait environ 1,5 million d'entreprises. Une fois le fichier épuré par la DOFI et les données manquantes imputées, on renommait le fichier ADMIN. Ce fichier ADMIN est essentiellement la version de la DOFI du RE. Ce processus se poursuivra jusqu'en octobre 2000, date à laquelle le fichier CORPAC sera remplacé par le fichier CORTAX.

Enquête trimestrielle sur les états financiers (ETEF)

L'Enquête trimestrielle sur les états financiers (ETEF) comprend un questionnaire que la DOFI envoie à un échantillon d'environ 6 000 entreprises constituées en société, en vue de recueillir des données financières trimestrielles. Jusqu'à la fin de 1998, l'échantillon était tiré du FUE T2, tandis que pour 1999, il sera tiré du RE, à partir du FUE de l'EUE. On fait enquête auprès des entreprises les plus importantes, ainsi que d'un tiers des entreprises de taille moyenne, mais aucune petite entreprise n'est visée. Le questionnaire est envoyé à toutes les entreprises comportant plusieurs entités légales, et celles-ci doivent fournir des données consolidées au niveau de l'entreprise. Des estimations pour les petites entreprises sont élaborées grâce à des projections à partir des données fiscales des années précédentes, les données d'enquête pour les entreprises de taille moyenne étant utilisées comme base de projection. Les estimations sont révisées par la suite, lorsque les données fiscales annuelles des petites entreprises deviennent disponibles. Une fois ces données recueillies, la DOFI utilise des poids pour procéder à des estimations pour l'ensemble de la population.

Le questionnaire comporte des questions au sujet de l'état des résultats et du bilan des entreprises. Cela vient du fait que dans le cadre du SCN, on a besoin de statistiques financières trimestrielles, en vue de la production de comptes, comme les comptes trimestriels des revenus et des dépenses, les comptes des flux financiers, les comptes nationaux de bilan et la balance des paiements internationaux. Ces comptes comprennent des données au sujet des activités d'investissement, des modèles de financement, des taux de rendement et d'autres ratios du rendement financier. Il existe d'autres utilisateurs pour ces données, par exemple, d'autres ministères gouvernementaux, le secteur privé et le secteur universitaire. Ces données d'enquête seront maintenues en permanence parce que le SCN a besoin de statistiques financières trimestrielles. En outre, les enquêtes en question ont été modifiées pour comprendre les questions au sujet de la valeur ajoutée, information essentielle pour les tableaux des entrées-sorties.

EFTA (États financiers trimestriels et annuels)

Les estimations annuelles pour les EFTA étaient produites de la façon suivante. Des données trimestrielles recueillies au niveau de l'entreprise étaient ramenées sur une base annuelle. En outre, les données fiscales au niveau de l'entité légale étaient « cumulées » au niveau de l'entreprise. La variable des bénéfices, qui était tirée de ces deux sources, était comparée au niveau de l'entreprise pour déceler tout problème de base de sondage ou de données. Les variables d'imposition qui n'étaient pas recueillies dans l'enquête trimestrielle étaient fusionnées avec les données trimestrielles annualisées. L'échantillon final consistait par conséquent en de « bonnes »⁹ données financières d'enquête trimestrielle annualisées, complétées par des données fiscales pour les entreprises plus grandes. Pour le reste des entreprises constituées en société, les données étaient tirées de trois échantillons déterminés par jadis le Groupe de l'accès aux dossiers fiscaux de la Division du Registre des entreprises. Les poids étaient recalculés et appliqués pour produire des estimations.

Comment l'IGRF peut-il améliorer notre façon de faire?

Le processus de collecte des données financières a beaucoup changé. Avec l'avènement du système de codage unique et du système de données électroniques de l'IGRF, les choses devraient s'améliorer considérablement. Les données des états financiers seront maintenant recueillies de façon plus opportune, globale, et efficace, ainsi que moins coûteuse. Les données seront à la fois détaillées et précises et permettront la production de statistiques à valeur ajoutée au niveau provincial. En outre, la fiabilité des données s'améliorera parce que l'IGRF englobera la presque totalité de la population des sociétés. Enfin, les données des états financiers adopteront une présentation relativement uniforme, ce qui facilitera la tâche des utilisateurs au moment de regrouper les statistiques pour leurs propres fins.

Les données au sujet des entreprises simples seront disponibles directement à partir de l'IGRF. On disposera aussi d'information au sujet des entreprises complexes, mais cela sera plus problématique. Il sera possible de réduire le fardeau d'enquête pour les répondants. Les enquêtes pourront maintenant être effectuées sur une base biennale ou triennale, plutôt qu'annuelle, du fait de l'utilisation des estimations fiscales les années où il n'y a pas d'enquête. Le Bureau pourra utiliser les données fiscales plutôt que les données d'enquête pour les branches d'activités plus petites, comme les services personnels et services de blanchissage. En outre, l'IGRF peut être utilisée pour couvrir les entreprises très petites d'une branche d'activité. Ces entreprises ont moins de 30 000 \$ de revenu et ne font par conséquent jamais l'objet d'enquêtes. Enfin, cela augmentera les possibilités de l'échantillon d'enquête pour les entreprises de petite et de moyenne tailles.

Un autre progrès imminent a trait à la modification des dispositions législatives touchant l'impôt sur le revenu pour les sociétés et les particuliers, ce qui a pour résultat que les représentants statistiques provinciaux ont accès aux données fiscales des entreprises. En vertu des dispositions législatives courantes, Statistique Canada a accès aux états financiers, mais ne peut pas partager les statistiques produites au moyen de ces données avec les autres représentants provinciaux. Auparavant, les statistiques que le Bureau produisait étaient créées principalement à partir de données d'enquête. Avec l'avènement de l'IGRF, les statistiques produites représenteront une combinaison de données d'enquête et de données administratives. Par conséquent, cette modification des dispositions législatives fédérales permet une plus grande exploitation de l'IGRF, du fait qu'elle augmente la qualité des statistiques tout en réduisant le fardeau de réponse et en préservant la confidentialité.

⁹ De « bonnes » données financières d'enquête trimestrielle annualisées signifient que l'on dispose d'au moins deux trimestres de chiffres réels, dont l'un correspond au dernier trimestre de l'année.

Quelle est la situation actuelle de l'IGRF?

Les employés de Statistique Canada ont eu accès pour la première fois à l'IGRF (en format Oracle) en octobre 1999. La Division des données fiscales conserve maintenant cette base de données dans le Système de microdonnées fiscales (SMF). À l'heure actuelle, seules les entreprises constituées en société et les coentreprises sont disponibles dans l'IGRF au niveau de l'entité légale. On compte 134 blocs-totaux et 685 éléments de variable, tandis que le nombre d'enregistrements dépend du tableau Oracle auquel les utilisateurs accèdent. En date du 1^{er} mars 2000, on comptait 881 298 lignes dans le tableau de déclaration des données fiscales et des données financières T2, chacune représentant une entreprise pour l'année d'imposition 1998. On compte 889 106 lignes dans le tableau des déclarations T2 de 1998, soit une pour chaque déclaration financière. Le tableau des déclarations T2 comporte davantage de lignes que celui de la déclaration des données financières et fiscales T2, parce que certaines entreprises produisent plus d'une déclaration par exercice. On compte 924 947 lignes dans le bilan de 1998, une pour chaque bilan d'ouverture et de fermeture. En 1998, les bilans d'ouverture étaient au nombre de 36 880, et les bilans de fermeture totalisaient 888 067. Pour l'état de résultats de 1998, on compte 883 930 lignes, chacune représentant un état des résultats.

Pour l'année d'imposition 1998, de 97 % à 99 % des données ont été transposées pour l'IGRF par l'ADRC. De même, moins de 3 % des déclarants ont fourni leurs données sur la base de l'IGRF. La DDF vise à établir un équilibre entre les données, afin de veiller à ce qu'elles soient uniformes. En 1999, le plan est le même, mais le nombre de déclarations qui devront être transcrites commencera à diminuer, et celui des formulaires comportant déjà des codes de l'IGRF augmentera. En outre, pour 1999, la Division entend se concentrer sur des vérifications plus approfondies, des vérifications historiques et des imputations partielles, ainsi que sur les entreprises qui ont cessé leurs activités et sur la révision des critères d'appariement et les exercices multiples. Les plus pessimistes se pencheront sur les données pour 1998, et dans une moindre mesure, sur celles pour 1999, et mettront l'accent sur les problèmes qu'elles comportent. Les plus optimistes s'attendent à toute une vague d'améliorations importantes, année après année.

Des aperçus de l'IGRF figurent dans la base de données de l'IGRF. Ces aperçus comprennent les enregistrements mentionnés ci-dessus, mais ils sont disponibles pour chaque phase du traitement. Il existe actuellement trois aperçus : IGRF-BRUT, IGRF-ÉPURÉ et IGRF-V et I (vérification et imputation). Ainsi, lorsqu'un utilisateur de la base de données de l'IGRF repère des résultats inhabituels dans la base de données V et I de l'IGRF, il peut faire des recherches pour déterminer d'où proviennent ces données.

L'IGRF-BRUT comprend des enregistrements de l'IGRF envoyés directement par l'ADRC. Certaines données brutes n'ont pas encore été traitées, sauf pour en exclure les enregistrements en double et ceux portant sur des exercices multiples. Cet aperçu a pour objectif de conserver les fichiers originaux. L'IGRF-ÉPURÉ comprend des enregistrements qui ont subi une épuration initiale et dont les erreurs évidentes ont été vérifiées. L'aperçu IGRF-V et I est constitué des enregistrements qui ont été vérifiés et imputés. Cet aperçu vise à combler les lacunes de données pour les EL qui sont comprises dans le Fichier de l'univers d'enquête (FUE) de l'EUE, mais pas dans la base de données de l'IGRF. À l'heure actuelle, les plans relatifs aux données pour 1998 visent à faire en sorte que l'on dispose d'estimations pour les blocs-totaux, mais il n'existe aucun plan pour l'imputation en dessous du niveau du bloc-total. Enfin, aucune vérification historique n'est effectuée parce que l'on ne dispose pas de données historiques pour le moment.

Le Système généralisé de vérification et d'imputation (SGVI) est utilisé par la Division des données fiscales pour effectuer des vérifications et des imputations relativement aux données de l'IGRF. Il permet de sélectionner des donneurs à partir d'autres enregistrements fiscaux, en vue d'imputer les enregistrements manquants au moyen du même code à six chiffres du SCIAN. En

1998, la transcription de l'IGRF s'est faite manuellement, ce qui fait que ce processus subira des réductions considérables une fois que l'ensemble de l'IGRF sera en place. Le système de vérification et d'imputation postérieur au SGVI est conçu pour calculer les variables au prorata, afin de s'assurer qu'elles sont parfaitement en équilibre. Il vise en outre à calculer des totaux partiels précis.

La stratégie de vérification et d'imputation pour l'année d'imposition 1998 consiste principalement à équilibrer les données. L'équilibrage permettra le maintien des équations suivantes.

Total de l'actif = total du passif + total des capitaux propres
Total des revenus – total des dépenses = revenu non agricole net
Total des revenus agricoles – total des dépenses agricoles = total des revenus agricoles nets
Revenu net/perte nette = revenu net /perte nette après impôts et éléments extraordinaires

Tout d'abord, on ne peut pas s'attendre à ce que ces équations soient toujours exactes, ce qui fait qu'un seuil doit être établi. Tant que les équations respectent un seuil de 100 \$, on convient qu'elles sont en équilibre. La stratégie d'équilibrage du bilan, de l'état des résultats et de l'état des bénéfices non répartis figure ci-dessous.

Équilibrage du bilan et l'état des résultats :

- Étape 1 : Certains postes doivent être positifs et certains autres négatifs. Ces postes sont automatiquement modifiés et étiquetés.
- Étape 2 : Calculer les variables pour le total de l'actif, le total du passif et le total des capitaux propres, le total des revenus, le total des dépenses, le total des revenus agricoles, le total des dépenses agricoles, le total des redressements, ainsi que le revenu net/la perte nette après impôts et éléments extraordinaires pour les postes fournis. Effectuer cette étape même si des totaux sont fournis.
- Étape 3 : S'assurer que les équations du bilan et de l'état des résultats sont en équilibre, selon le seuil de tolérance, puis passer à l'étape 4 ou à l'étape 5.
- Étape 4 : Si l'on peut utiliser des valeurs au prorata pour que les équations soient exactement en équilibre, le faire. Aucune étiquette n'est nécessaire. Si l'une ou l'autre des variables des équations est dérivée, l'indiquer avec une étiquette, puis laisser tomber les variables dérivées de l'étape 2. Établir au prorata les variables des équations « d'équilibrage », afin que leur somme soit exacte; ne pas les étiqueter. Puis, établir au prorata les postes qui constituent les variables de l'équation, s'ils dépassent le seuil. Indiquer ces changements avec une étiquette. Si aucun poste n'est déclaré, mais que les variables de l'équation ne sont pas égales à zéro, le poste doit être étiqueté pour imputation par donneur (SGVI), dans le cas des postes manquants de cette section.
- Étape 5 : S'il n'existe pas de façon d'équilibrer l'équation, étiqueter l'enregistrement en vue de l'imputation par donneur de l'ensemble du bilan ou de l'état des résultats. En ce qui a trait à l'état des résultats, les changements peuvent parfois se limiter à la section des redressements, et les données sur les revenus et les dépenses peuvent être conservées. Puis l'enregistrement peut être étiqueté pour imputation par donneur de la section des redressements.

En ce qui a trait à la section sur les bénéfices non répartis, le poste 3600 doit être égal au poste 3849, et le poste 3680, au poste 9999. S'ils ne sont pas égaux, une étiquette est établie.

Une imputation partielle est une imputation effectuée à un enregistrement de l'IGRF lorsqu'il manque tous les postes de détail d'une section. Ces imputations partielles sont effectuées au moyen du SGVI. A l'heure actuelle, la DDF en est à décider si elle doit investir des ressources pour des imputations partielles des données pour 1998. De toute façon, des imputations

partielles seront effectuées tôt ou tard. À cette fin, tous les enregistrements sont examinés du point de vue des valeurs aberrantes, afin de les exclure comme enregistrement receveur ou donneur dans le SGVI. Si un de ces enregistrements exclus nécessite un donneur, il fera l'objet d'une intervention manuelle. Si l'ensemble du bilan ou de l'état des résultats doit être imputé, il doit recevoir une étiquette d'imputation de masse. Selon la variable manquante, les variables obligatoires appropriées seront choisies par le SGVI pour l'imputation. Enfin, si un total est manquant, il sera imputé par donneur.

La vérification postérieure au SGVI sert à effectuer une imputation lorsqu'un poste manquant est imputé, mais que le total original est maintenu. Le poste imputé doit alors faire l'objet d'un calcul au prorata. En outre, des totaux partiels sont calculés pour les variables précisées et les blocs-totaux.

Une imputation de masse est effectuée pour les enregistrements totalement absents. Cela peut venir du fait que l'enregistrement figure dans le FUE de l'EUE, mais ne comporte pas de lien avec les données fiscales. Dans ces cas, les données pertinentes doivent être tirées du RE, en vue de sélectionner un donneur approprié. Si aucun enregistrement de l'IGRF n'est disponible, ou si l'enregistrement reçu ne comporte pas de données utilisables, il reçoit aussi une étiquette d'imputation de masse, et les renseignements le concernant sont obtenus à partir du FDR T2 de l'AI 98, afin de trouver un donneur. Enfin, les postes qui ne figurent ni dans l'échantillon de l'EUE, ni dans l'IGRF, nécessitent une imputation de masse, du fait qu'il faut procéder à une estimation pour les branches d'activité qui ne sont pas visées par l'EUE.

Une fois qu'un donneur approprié est trouvé, un calcul au prorata est effectué, afin de s'assurer que les données nouvellement imputées sont en équilibre avec le reste de l'enregistrement, à l'intérieur du seuil. Le calcul au prorata consiste à diviser la variable de l'enregistrement du FDR par la même variable de l'IGRF qui sert de donneur. Ce ratio est multiplié par la suite par la variable qui sert de donneur pour obtenir la valeur imputée. Par exemple, si la variable manquante correspond au total des dépenses, le total des revenus du FDR est divisé par le total des revenus de l'unité qui sert de donneur. Ce ratio est par la suite multiplié par la variable qui sert de donneur, c'est-à-dire le total des dépenses, pour obtenir la valeur imputée.

À l'heure actuelle, on n'effectue pas de calendrialisation des données de l'IGRF. On fait aussi face au problème des entreprises qui ont plusieurs exercices. Il s'agit d'entités légales qui ont plus d'une période de déclaration, soit parce qu'une autre EL a été créée à partir de l'EL existante, ou que deux EL ont fusionné. Cela pose un problème parce que des postes peuvent être comptés en double. La solution consiste actuellement à conserver l'exercice le plus long et à interrompre le traitement de l'enregistrement le plus court.

On tient un tableau de journal pour enregistrer l'historique de toutes les mises à jour de chaque enregistrement¹⁰.

Du fait des changements qui se produisent, des modifications doivent être apportées au plan d'enquête de l'EUE. La Partie 1 de l'EUE a été conçue pour recueillir des données financières annuelles pour les 225 variables nécessaires aux fins du plan comptable de l'EUE. Elle a été envoyée pour l'année de référence 1997, mais elle n'était pas nécessaire pour l'AR 1998, en raison de l'avènement de l'IGRF. Elle ne s'est toutefois pas appliquée aux entreprises comportant plusieurs entités légales. Une Partie 1 plus restreinte a été envoyée pour recueillir des données consolidées. En 1999, l'enquête de la Partie 1 n'a pas été effectuée parce que l'Enquête trimestrielle sur les états financiers est maintenant reliée au RE, et son échantillon comprend toutes les entreprises comportant plusieurs EL de taille importante.

¹⁰ Voir le paragraphe intitulé *Tableaux de journal* dans la section *Meta-données*.

À quoi doit-on s'attendre pour l'avenir?

Il sera obligatoire pour toutes les entreprises qui produisent des déclarations T2 et T1 de les produire sous forme électronique pour l'ADRC. Cela signifie que Statistique Canada recevra ce recensement de données financières sous forme électronique. Comme il a été mentionné dans la section précédente, Statistique Canada a reçu près de 900 000 enregistrements pour l'année d'imposition 1998. Ces enregistrements ont été transcrits manuellement à l'ADRC, à partir des états financiers qui ont été fournis sur papier par les entreprises, avec leurs déclarations de revenu pour 1998. On a procédé ainsi parce que le logiciel de l'IGRF n'est pas beaucoup utilisé, et que le nouveau système informatique « CORTAX » de l'ADRC n'est pas encore en place. La situation sera très similaire pour l'année d'imposition 1999. Toutefois, on s'attend à ce qu'un plus grand nombre de sociétés utilisent le logiciel de l'IGRF, ce qui fait qu'un moins grand nombre d'enregistrements devront être transcrits. Le fichier de l'IGRF pour 1999 devrait comprendre un peu plus d'un million d'enregistrements. À la fin de l'an 2000, toutes les sociétés qui produisent des déclarations T2 devront le faire au moyen du logiciel de l'IGRF. Cela signifie que plus d'un million d'enregistrements seront disponibles pour l'année d'imposition 2000, et qu'ils auront été codés au moyen du logiciel de l'IGRF et non pas transcrits manuellement. Cela éliminera les coûts de la transcription, tout en augmentant l'actualité et la précision des données. En octobre 2000, on s'attend à ce que CORTAX soit opérationnel, ce qui fait que les entreprises pourront soumettre leurs états financiers sous forme électronique. Ainsi, pour l'année d'imposition 2000, à peu près toutes les déclarations T2 seront reçues sous forme électronique. En outre, on s'attend à ce qu'il s'écoule encore de 3 à 5 ans jusqu'à ce que l'IGRF soit modifié pour répondre aux besoins des déclarants T1. Par la suite, Statistique Canada recevra les données financières sous forme électronique pour à peu près toutes les entreprises au Canada. Ainsi, le Bureau disposera de statistiques détaillées au sujet de variables comme les ventes, les dépenses, la valeur ajoutée et les bénéfices, tant selon la région que la branche d'activité.

Au fur et à mesure de l'élaboration et de l'amélioration des systèmes de l'IGRF à l'ADRC, Statistique Canada ne devra plus consacrer autant d'efforts à l'équilibrage et à la vérification des données administratives. L'accent sera davantage mis sur l'imputation historique et les étapes de traitement de pointe comme la calendrialisation. Lorsque l'on a recours à l'imputation par donneur, la continuité historique des micro-enregistrements se perd. L'objectif visé est d'utiliser l'imputation par donneur pour les petites EL, lorsque cela ne devrait pas poser de problèmes importants, mais pour les moyennes et les grandes EL, il sera plus approprié de procéder à l'imputation au moyen d'un taux de croissance moyen de cellule appliquée au micro-enregistrement de l'année précédente. On assurera ainsi la continuité historique.

La stratégie d'imputation pour l'avenir est la suivante. À partir du FUE de l'EUE, nous nous attendons à environ un million d'EL. Toutefois, certains des enregistrements attendus de l'IGRF pourraient ne pas se concrétiser, en raison de regroupements, de fermetures d'entreprises ou d'autres raisons. Il se peut aussi que d'autres enregistrements de l'IGRF que l'on attendait pas soient disponibles, du fait de la constitution de nouvelles entités légales qui ne figure pas dans le RE. Les responsables du RE seront informés de toutes les EL non identifiées, afin de permettre des mises à jour rapides aux fins de la stratification a posteriori du FUE de l'EUE. Parallèlement, la DDF et la DMEÉ ajouteront les enregistrements imputés à l'aperçu de V et I pour chacun des enregistrements attendus, mais manquants. Cette étape sera effectuée deux fois par année, c'est-à-dire une fois au début du processus, lorsque 75 % des enregistrements auront été reçus, et une deuxième fois plus tard dans le processus, lorsque presque tous les enregistrements auront été reçus.

À l'heure actuelle, trois aperçus de l'IGRF sont mis à la disposition de l'utilisateur. On les appelle parfois IGRF-BRUT, IGRF-ÉPURÉ et IGRF V et I¹¹. Il se peut qu'un autre aperçu s'ajoute, soit

¹¹ Ces aperçus de l'IGRF sont expliqués de façon plus approfondie dans la section intitulée « Quelle est la situation actuelle de l'IGRF? »

l'IGRF-ENTREPRISE. Il englobera tous les enregistrements d'entité légale unique et tous les enregistrements consolidés d'entités légales multiples. Par conséquent, cet aperçu comprendra tous les enregistrements de l'IGRF consolidés au niveau de l'entreprise. On vise ainsi à permettre le couplage d'un enregistrement de l'IGRF à une entreprise du RE, au moyen du NE. Par ailleurs, chaque EL liée à une entreprise comprise dans le FUE de l'EUE devrait pouvoir, soit être couplée à un enregistrement unique de l'IGRF, ou consolidé.

La stratégie comprend aussi une étape de calendrialisation. En outre, si un enregistrement comporte plusieurs exercices, tous les enregistrements seront fusionnés grâce à l'addition des variables de flux, à l'utilisation des variables de stocks du dernier enregistrement et au rééquilibrage de l'enregistrement final.

Statistique Canada a créé cinq groupes de travail pour résoudre un certain nombre de questions particulières qui se sont posées relativement à l'utilisation de la nouvelle base de données de l'IGRF¹². Le groupe de travail A a été établi pour résoudre la question de la consolidation. Il créera des procédures de consolidation et recommandera une répartition appropriée des responsabilités entre les divisions. Le groupe de travail B a pour mandat de définir un nouveau plan comptable pour l'EUE. À cette fin, il examinera le plan comptable de la DOFI, le questionnaire modèle de l'EUE, le plan comptable implicite de l'IGRF et la structure de la comptabilité du SCN. Le groupe de travail C s'occupe du problème de couplage des NE. S'il est impossible d'obtenir des couplages précis et exhaustifs à partir du RE, les données de l'IGRF ne peuvent être utilisées efficacement à des fins statistiques. Le groupe de travail D élabore une stratégie pour l'estimation du plus grand nombre possible des variables nécessaires pour le plan comptable de l'EUE, grâce aux renseignements figurant dans la base de données de l'IGRF. Cela contribuera à réduire le fardeau de réponse imposé aux entreprises. Enfin, le groupe de travail E trouvera des approches communes en matière de logiciel, afin d'aider les utilisateurs à accéder aux variables de l'IGRF et à les analyser.

Totalisations

Métadonnées

Lorsque l'on crée un aperçu des données de l'IGRF en SAS ou en MS-Access, il est très important de comprendre comment récupérer les données appropriées. Les données de l'IGRF ne figurent pas dans des tableurs simples, mais plutôt dans une base de données relationnelle beaucoup plus souple. La base de données Oracle de l'IGRF comprend de nombreux tableaux qui englobent de nombreuses zones et variables. On peut coupler tous les tableaux. Par exemple, le NE peut être utilisé pour coupler des enregistrements du tableau des déclarations T2 aux enregistrements correspondants des tableaux du bilan ou de l'état des résultats. Les variables comprennent des éléments comme l'actif, le passif, les bénéfices non répartis, les revenus et les dépenses. Ces zones ou variables comportent à la fois un nom complet et un nom abrégé. L'avantage de l'utilisation du nom complet est qu'elle facilite la tâche de l'utilisateur au moment de l'identification. Le nom abrégé, quant à lui, a pour avantage qu'il comporte huit caractères ou moins, à savoir le nombre maximum de caractères acceptés par les programmes SAS. Les tableaux et leurs zones sont décrits dans un diagramme du site Internet de la DDF¹³, et il existe aussi une description détaillée de ces zones et de ces variables dans l'unité partagée de la DDF¹⁴.

¹² Les objectifs de ces groupes de travail sont énoncés dans le document n° 49 de la série des documents techniques du PASEP, ainsi que dans la section du présent document intitulée : « Quels sont les principaux défis auxquels Statistique Canada doit faire face pour l'utilisation de l'IGRF? »

¹³ Voir la note numéro 8 en bas de page.

¹⁴ Voir la note numéro 2 en bas de page.

L'utilisateur doit décider quel enregistrement il souhaite obtenir. À cette fin, il doit comprendre ce qui figure dans les tableaux ainsi que la façon dont les tableaux sont liés les uns aux autres. À l'heure actuelle, il existe dans l'IGRF 11 tableaux, qui sont énumérés à l'intérieur du tableau ci-dessous. Les trois premiers sont des tableaux d'identification générale, les quatre suivants, des tableaux comprenant des données de bilan, et les quatre derniers, des tableaux comprenant des données d'état des résultats.

Description du tableau	Nom complet dans l'IGRF
Renseignements sur les entreprises de Revenu Canada	Revenue_Canada_Business
Déclaration des données financières et fiscales T2	T2_Tax_and_Financial_Decl
Déclarations T2 pour l'année d'imposition 1998	T2_Decl_1998
Bilan des variables T2 avec noms complets - 1998	T2_Balance_Sheet_1998
Bilan des variables T2 avec noms abrégés - 1998	T2_Balance_Sheet_1998_Shrtn_Vw
Bilan des blocs-totaux T2 avec noms complets - 1998	T2_Balance_Sheet_1998_Blks_Tot
Bilan des blocs-totaux T2 avec noms abrégés - 1998	T2_Bs_1998_Blks_Tot_Shrtn_Vw
État des résultats des variables T2 avec noms complets - 1998	T2_Income Stmt_1998
État des résultats des variables T2 avec noms abrégés - 1998	T2_Income Stmt_1998_Shrtn_Vw
État des résultats des blocs-totaux T2 avec noms complets - 1998	T2_Income Stmt_1998_Blks_Tot
État des résultats des blocs-totaux T2 avec noms abrégés - 1998	T2_Is_1998_Blks_Tot_Shrtn_Vw

À l'intérieur du tableau « renseignements sur les entreprises de Revenu Canada », l'utilisateur peut obtenir le code du SCIAN, le numéro d'entreprise, la province, le code postal, la raison sociale et de nombreux autres renseignements. Le tableau « déclaration des données financières et fiscales T2 » est utilisé principalement pour coupler le NE du tableau « renseignements sur les entreprises de Revenu Canada » et du tableau suivant, « déclarations T2 pour l'année d'imposition 1998 ». Le tableau « déclarations T2 pour l'année d'imposition 1998 » comporte des renseignements au sujet du début et de la fin de l'exercice de l'entreprise, de l'imputation par donneur de l'enregistrement, le cas échéant, ainsi que de la phase du traitement de l'enregistrement. Les renseignements compris dans ce tableau sont actuels, et non historiques. Les huit derniers tableaux comprennent des variables ou des zones de l'état des résultats ou du bilan. La moitié de ces tableaux comprennent des données sur les quelque 700 variables existantes de l'IGRF, tandis que l'autre moitié comprend des données pour les divers blocs-totaux. À l'intérieur de chacun de ces deux groupes, il existe des tableaux distincts qui utilisent les noms complets et les noms abrégés de variables.

Tant pour le bilan que pour l'état des résultats, il existe une variable BNPID, qui correspond au code d'identification du programme de numéro d'entreprise. La variable BNACCT correspond au numéro de séquence du compte de numéro d'entreprise, le NE correspond au numéro d'enregistrement d'entreprise de l'ADRC, la variable FPEDT correspond à la date de fin d'exercice de la déclaration financière T2, la variable FPSDT, à la date de début d'exercice de la déclaration financière T2, et la variable T2STG sert à identifier la phase du traitement de l'enregistrement. Tout nom abrégé commençant par la lettre P comprend des renseignements au sujet de l'imputation par donneur. Enfin, l'état des résultats comprend des renseignements au sujet de la raison sociale et de la description de l'entreprise. Le bilan comporte une zone appelée OBSI, qui correspond à l'indicateur de feuille de bilan et qui s'applique aux entreprises qui ont commencé leurs activités au cours de l'exercice.

L'utilisateur doit sélectionner les variables à utiliser. Il peut s'agir de l'actif, du passif, des bénéficiaires non répartis, des revenus, des dépenses, etc. Le bilan et l'état des résultats comportent quelque 685 postes. Chacun de ces postes est désigné à la fois par un nom complet et par un nom abrégé. Par exemple, on peut accéder au poste Encaisse et dépôts au moyen du

nom complet Csh_and_Dpst+Amt ou du nom abrégé L1000. Dans le nom abrégé, le code à quatre chiffres de l'IGRF est toujours précédé d'un L. De même, pour les blocs-totaux, le nom abrégé commence par les lettres BT et comprend un code à quatre chiffres.

Il existe une façon facile de comprendre les métadonnées auxquelles l'IGRF permet d'accéder, à savoir l'exécution d'un échantillon de code SAS et la consultation des résultats. La DDF fournit gracieusement des échantillons de codes sur demande.

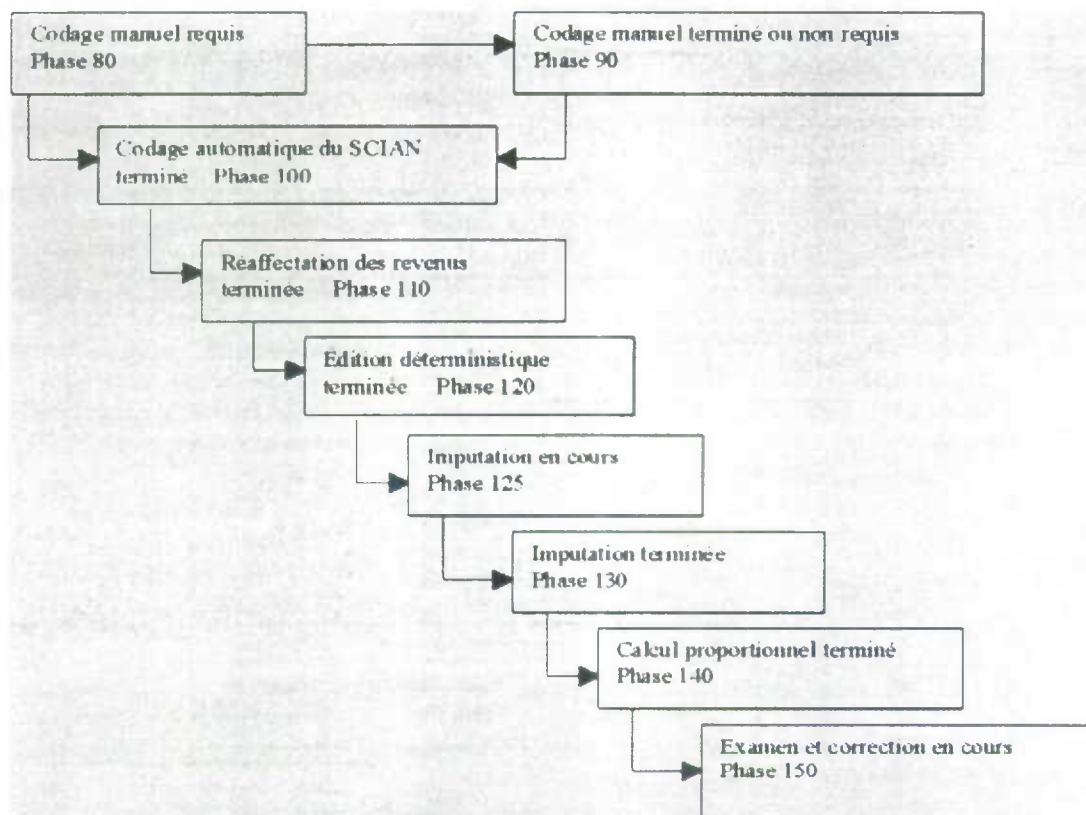
Phases du traitement

Il est très important de comprendre à quelle phase du traitement un enregistrement se trouve. L'utilisateur doit prendre une décision éclairée au sujet des enregistrements à utiliser pour l'analyse. Le code de la phase de traitement de chaque enregistrement peut être repéré dans le tableau des déclarations T2 ou dans les tableaux de bilan et d'état des résultats. Le nom abrégé de la variable connexe est T2STG et le nom complet, T2_Prcs_Stg_Id. Une fois que les données sont chargées dans la base de données Oracle, elles passent par différentes phases. Ces phases comprennent la vérification et l'imputation, et elles se déroulent de façon séquentielle, parfois sur une longue période, ce qui fait qu'il est important que l'utilisateur sache quelles sont les phases auxquelles les enregistrements ont été soumis.

Le numéro d'identification de la phase diffère selon qu'il figure dans le bilan et l'état des résultats ou dans le tableau des déclarations T2. L'identification de la phase du bilan et de l'état des résultats comprend des renseignements au sujet des derniers changements qui ont touché l'enregistrement. Par ailleurs, l'identification de la phase des déclarations T2 correspond à la dernière phase subie par l'enregistrement, même si les données de l'enregistrement n'ont pas été modifiées. Cette dernière phase sert au traitement, et la première sert à l'analyse.

Il est instructif de comparer les phases du traitement qui s'appliquent aux entreprises T1 non constituées en société (aux fins de l'échantillon du programme des estimations fiscales), et celles qui s'appliquent aux sociétés T2 (dans le cadre du recensement de l'IGRF). Dans le cas des entreprises T1, tous les enregistrements sont chargés à la phase de traitement « 80 » ou « 90 », selon qu'ils comportent ou non un code de la CTI. Lorsqu'un enregistrement reçoit un code du SCIAN, il se situe à la phase « 100 ». À la phase « 110 », tous les enregistrements invalides sont étiquetés, et le revenu est réparti entre cinq catégories différentes (revenus agricoles, revenus provenant de la pêche, commissions, revenus de location, services professionnels). La phase « 120 » correspond à la fin du processus de vérification déterministe. Cela signifie que toutes les données sont en équilibre d'un point de vue comptable. Lorsqu'un enregistrement est en voie d'imputation, l'identificateur de la phase est établi à « 125 ». Il s'agit de la préparation en vue du SGVI et de l'étiquetage des enregistrements qui ne font pas partie du champ de l'enquête. Un enregistrement peut aussi se situer à la phase « 125 », si un donneur n'a pas été trouvé pour l'imputation. Une fois que l'on trouve un donneur pour l'imputation, et que les enregistrements ont été imputés au moyen du SGVI, on arrive à la phase « 130 ». Dès que l'enregistrement est prêt pour la répartition, il se situe à la phase « 140 ». Enfin, la phase « 150 » est un processus permanent faisant intervenir de nombreuses catégories diverses de changements effectués par de nombreux utilisateurs différents. Le diagramme 3 montre les phases du traitement des T1.

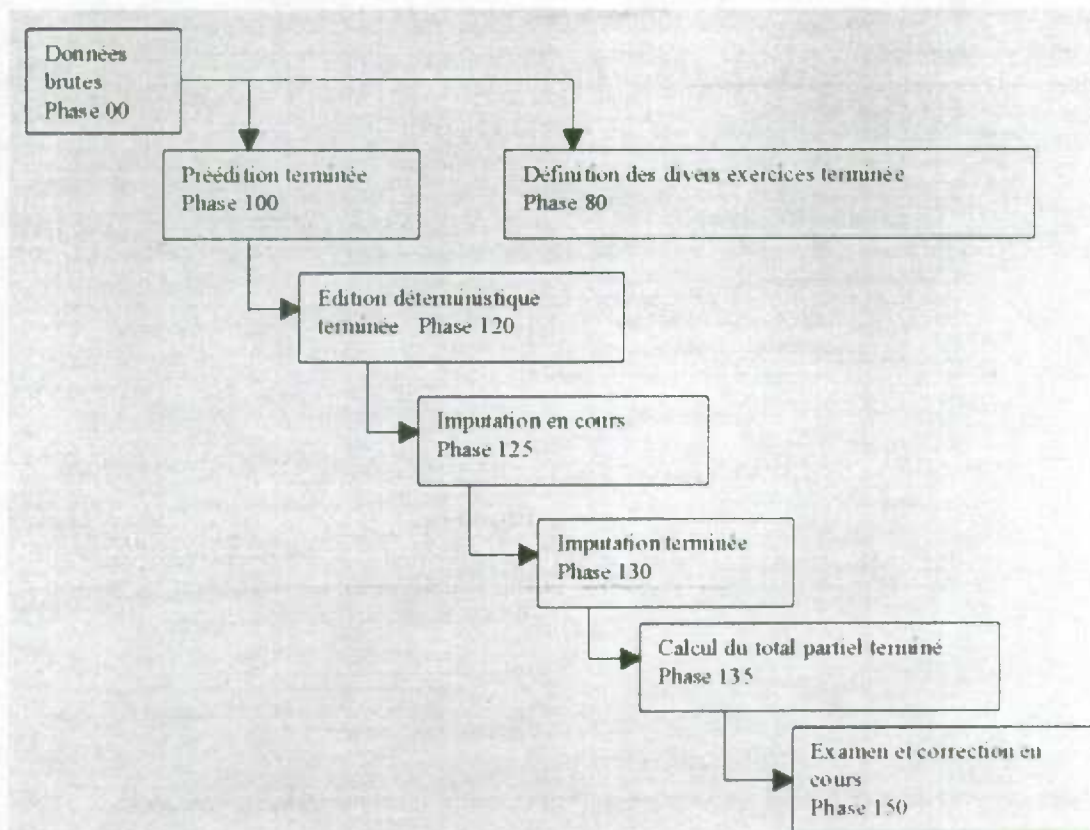
Diagramme 3 – Phases du traitement des T1



* Voir le site Internet de la DDF sous Méta-données, Phases de traitement et tableaux journaux.

Les phases de traitement pour les sociétés T2 diffèrent légèrement de celles des T1. Par exemple, toutes les données sont chargées à la phase « 00 », la phase des données brutes. Une fois une vérification préliminaire effectuée pour déterminer s'il existe des exercices multiples, les enregistrements sont répartis en deux groupes. Si une société n'a qu'un exercice, elle passe à la phase « 100 ». Si elle comporte plusieurs exercices, l'enregistrement correspondant à l'exercice le plus long passe à la phase « 100 », et celui correspondant aux exercices plus courts, à la phase « 80 ». Tous les enregistrements de la phase « 80 » y demeurent et ne sont pas traités davantage, et ils sont mis à la disposition des utilisateurs qui aimeraient les consulter. Le reste du processus est similaire au processus T1. Le diagramme 4 montre les phases du traitement des T2.

Diagramme 4 – Phases du traitement des T2



* Voir le site Internet de la DDF sous Méta-données, Phases de traitement et tableaux journaux.

Tableaux de journal

Les tableaux de journal sont des tableaux qui, comme ceux décrits dans la section précédente, comprennent des renseignements historiques au sujet d'un enregistrement. Par exemple, un tableau de journal indique quand un enregistrement a été intégré à un tableau particulier, supprimé et mis à jour. Le tableau de journal comporte des renseignements au sujet de chaque phase subie par l'enregistrement. Le tableau de journal est entreposé dans la base de données Oracle sous <table_name>_JN. Il comprend aussi d'autres renseignements, par exemple, l'identité de l'utilisateur qui a modifié l'enregistrement, la date et l'heure de la modification, l'application utilisée, le numéro d'identification de la séance particulière et tout autre renseignement pertinent. Le tableau des déclarations T2 comporte toujours des renseignements au sujet de l'état d'un enregistrement, tandis que le journal des déclarations T2 fait l'historique de toutes les modifications apportées à cet enregistrement.

Quelles sont les branches d'activité les mieux représentées par l'IGRF?

La base de données courante de l'IGRF contient des renseignements financiers détaillés au sujet des entreprises canadiennes constituées en société. Un des inconvénients de l'IGRF est qu'il ne contient que des renseignements au sujet de certaines formes d'organisations. Dans le Registre

formes d'organisations. Dans le Registre des entreprises de Statistique Canada, les entreprises sont divisées en cinq catégories différentes. Il s'agit :

- 1) des entreprises constituées en société,
- 2) des coentreprises, dont les parties constituantes produisent toutes des déclarations T2,
- 3) des entreprises non constituées en société qui produisent des déclarations T1,
- 4) des entreprises publiques qui ne produisent pas de formulaire T2 ni T1, et qui produisent ou non une autre déclaration en fonction de leurs activités particulières, et
- 5) les établissements gouvernementaux, qui ne produisent aucune déclaration.

La principale limite de l'IGRF vient du fait qu'il ne comprend que les renseignements financiers T2 (même s'il englobera tôt ou tard des renseignements T1, comme il a été souligné précédemment). Cela signifie que l'IGRF ne comprend des renseignements financiers que pour les entreprises constituées en société, les coentreprises et, *ultérieurement*, les entreprises non constituées en société. À l'heure actuelle, si une branche d'activité est composée principalement d'entreprises constituées en société ou de coentreprises, elle est mieux représentée par l'IGRF qu'une branche d'activité qui est composée principalement d'entreprises non constituées en société ou d'entreprises publiques.

Les entreprises sont aussi réparties entre les entreprises simples et complexes¹⁵. À l'heure actuelle, les entreprises simples et complexes sont comprises dans l'IGRF. Si une branche d'activité est constituée principalement d'entreprises simples, elle sera bien représentée par l'IGRF. Les entreprises simples n'ont qu'une EL, et tous les établissements de l'entreprise ont des activités dans la même province et branche d'activité, ce qui simplifie l'attribution du code du SCIAN et la production de données précises selon la branche d'activité. Aucun processus de consolidation ou de répartition n'est nécessaire. Les entreprises complexes, par ailleurs, posent davantage de problèmes, parce que l'obtention de données au niveau de l'établissement nécessite une consolidation et une répartition¹⁶. Une branche d'activité qui est composée principalement d'entreprises complexes ne sera peut-être pas bien représentée par l'IGRF. Même si les EL des entreprises complexes sont incluses dans la base de données de l'IGRF, elles doivent être consolidées au niveau de l'entreprise en premier lieu, puis réparties au niveau de l'établissement, en deuxième lieu, avant de pouvoir être utilisées efficacement pour la production de ventilations selon la branche d'activité ou la province.

Il convient de souligner un autre point, à savoir l'exclusion des compagnies d'assurance de l'IGRF. Il faut faire preuve de prudence lorsqu'on traite les branches d'activité codées 5241 (Sociétés d'assurance) et 5242 (Agences et courtiers d'assurances et autres activités liées à l'assurance) dans le SCIAN. La DDF exclut aussi certains codes du SCIAN. Il s'agit notamment du code 526112 (Caisses de retraite non fiduciaires), 8132 (Fondations et organismes de charité), 8133 (Organismes d'action sociale), 8134 (Organisations civiques et amicales), et 91 (Administrations publiques). Cela vient du fait que l'IGRF n'est pas conçu pour répondre aux besoins de ces branches d'activité.

En annexe au présent document figure un ensemble d'indicateurs bruts de la façon dont l'IGRF représente les diverses branches d'activité. Ces indicateurs de la couverture seront appropriés jusqu'au jour où les entreprises non constituées en société devront produire des déclarations de l'IGRF. Des renseignements au sujet de chaque branche d'activité, relativement au nombre d'établissements, au nombre d'employés et au revenu total de cette branche d'activité, ainsi qu'à la nature simple et complexe des entreprises, figurent dans le RE. Ces renseignements peuvent être ventilés séparément par les entreprises constituées en société, les entreprises non

¹⁵ Les entreprises simples et complexes sont définies dans la section du présent document intitulé « Registre des entreprises ».

¹⁶ Les processus de consolidation et de répartition sont énoncés dans la section intitulée « Consolidation des données de l'entité légale et répartition au niveau de l'établissement ».

constituées en société, les coentreprises, les entreprises publiques et les autres formes d'organisations. Étant donné que ces tableaux comportent 921 branches d'activité et 3 variables (établissements, employés et revenu), il sont trop importants pour être inclus dans le présent document. Pour chacune de ces trois variables, on a plutôt regroupé les tableaux en un seul tableau moins important (Tableau 2), qui utilise 21 branches d'activité pour illustrer le propos. Huit des 21 branches d'activité incluses ont été choisies à partir de la portion supérieure de l'index de couverture de l'IGRF, huit, à partir de la portion médiane, et cinq, à partir de la portion inférieure. La version complète et détaillée de ces tableaux peut être obtenue auprès de l'auteur sur demande.

La première colonne du tableau 2 montre le code du SCIAN, et la deuxième comporte une brève description de la branche d'activité. Les colonnes qui restent indiquent deux indices de couverture de l'IGRF, le premier incluant les entreprises complexes, et le deuxième excluant les entreprises complexes, du point de vue du nombre d'établissements, du total des employés et du revenu total. Le premier indice est calculé en additionnant les entreprises simples et complexes constituées en société et les coentreprises, et en divisant par le total. Ces données sont mentionnées comme problématiques dans le tableau 2 en raison des entreprises complexes concernées. Pour calculer le deuxième indice, on additionne les entreprises simples constituées en société et les coentreprises simples, et on divise par le total. Ces données sont jugées non problématiques dans le tableau 2 parce qu'elles peuvent être facilement extraites. Ces calculs visent à déterminer la couverture en pourcentage d'une branche d'activité particulière par l'IGRF.

Le tableau 2 est élaboré à partir du nombre d'établissements, du total des employés et du revenu total pour chaque branche d'activité. Idéalement, il devrait être fondé sur des données à valeur ajoutée, mais de telles données ne sont pas disponibles. Le calcul de l'indice au moyen du nombre d'établissements pose un problème parce que tous les établissements sont considérés comme égaux, peu importe leur taille. L'utilisation du nombre total d'employés comme indicateur pose aussi un problème parce que les branches d'activité varient substantiellement du point de vue de la main-d'œuvre qu'elles utilisent. Enfin, le revenu n'est pas non plus un indicateur parfait, étant donné que le ratio du revenu à la valeur ajoutée varie considérablement selon la branche d'activité. Par conséquent, la meilleure approche consiste à examiner les trois indicateurs ensemble, en tant que groupe.

Tableau 2 – Branches d'activité selon la « couverture » prévue dans l'IGRF — Nombre d'établissements, nombre total d'employés et revenu total.

SCIAN	Description de la branche d'activité	Nombre d'établissements		Nombre d'employés		Revenu total	
		Données problématiques	Données non problématiques	Données problématiques	Données non problématiques	Données problématiques	Données non problématiques
212392	Extraction de diamant	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %
561330	Location de personnel permanent	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %
212299	Extraction de tous les autres minerais métalliques	100,0 %	90,5 %	100,0 %	51,5 %	100,0 %	24,2 %
311221	Mouture humide du maïs	100,0 %	46,7 %	100,0 %	8,9 %	100,0 %	0,9 %
212393	Extraction de sel	100,0 %	10,0 %	100,0 %	0,0 %	100,0 %	0,1 %
452110	Grands magasins	100,0 %	0,7 %	100,0 %	0,2 %	100,0 %	0,1 %
441110	Marchands d'automobiles neuves	99,4 %	96,4 %	99,6 %	94,6 %	99,5 %	93,9 %
325120	Fabrication de gaz industriels	95,9 %	17,8 %	99,6 %	2,7 %	99,9 %	5,3 %
...
622112	Hôpitaux pédiatriques	81,8 %	36,4 %	99,9 %	47,2 %	99,9 %	57,0 %
418990	Tous les autres grossistes-distributeurs	80,7 %	78,6 %	96,9 %	70,1 %	97,6 %	36,5 %
231410	Gestion de construction	80,0 %	79,3 %	97,1 %	93,4 %	95,5 %	76,7 %
336611	Construction et réparation de navires	79,9 %	72,6 %	99,7 %	25,3 %	99,6 %	22,8 %
713930	Marinas	79,4 %	78,9 %	99,7 %	25,3 %	88,2 %	87,1 %
332710	Ateliers d'usinage	79,0 %	78,0 %	98,2 %	90,8 %	97,7 %	86,1 %
112330	Élevage de dindons	78,2 %	67,9 %	87,8 %	74,6 %	94,6 %	32,3 %
312130	Vineries	77,4 %	72,2 %	94,3 %	49,6 %	97,1 %	26,2 %
...
722330	Cantines et comptoirs mobiles	29,0 %	28,8 %	74,1 %	72,8 %	61,4 %	55,9 %
111140	Culture du blé	19,2 %	19,2 %	45,5 %	45,2 %	28,0 %	27,9 %
485310	Services de taxi	18,9 %	18,8 %	65,8 %	61,0 %	60,5 %	58,7 %
911410	Affaires étrangères	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %
912110	Tribunaux provinciaux	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %
...
Total		66,8 %	63,2 %	85,1 %	59,3 %	82,1 %	38,1 %

L'extraction de diamant et la location de personnel permanent sont deux branches d'activité composées exclusivement d'entreprises simples constituées en société, ce qui en fait d'excellentes candidates pour la représentation au moyen de l'IGRF. L'extraction de tous les autres minerais métalliques est composée d'entreprises simples constituées en société, mais elle comporte aussi un petit nombre d'établissements complexes constitués en société, ce qui permet la production d'un indice de couverture non problématique de 90,5 %. Toutefois, l'influence des établissements complexes est manifestement sous-estimée lorsque l'on calcule l'indice au moyen du total des établissements, comme le montre la comparaison avec l'indice fondé sur le revenu. Ce dernier n'est que de 24,2 %, ce qui laisse suggérer que les données posent davantage de problèmes que ne le laisse supposer l'indice du nombre d'établissements. Dans le cas de la mouture humide du maïs, la branche d'activité est composée à 100 % d'entreprises constituées en société. Parmi elles, 46,7 % sont des établissements simples, mais la proportion du revenu total que représentent ces unités simples n'est que de 0,9 %. L'extraction de sel et les grands magasins sont deux exemples de branches d'activité composées exclusivement d'entreprises constituées en société, où seulement un très petit pourcentage des données est le fait des entreprises simples, ce qui a pour résultat qu'un pourcentage important des données (environ 99,9 % selon le revenu total) posent un problème. Les marchands d'automobiles neuves présentent une situation intéressante. Même s'ils comptent quelques établissements non constitués en société, soit environ 0,6 %, la branche d'activité est composée principalement d'entreprises simples constituées en société, ce qui en fait une candidate relativement bonne pour la représentation au moyen de l'IGRF. Selon l'indice non problématique fondé sur le revenu, les marchands d'automobiles neuves devraient être couvertes à 93,9 %. Un autre cas représentatif est celui de la fabrication de gaz industriels, une branche d'activité qui est dominée par des entreprises complexes constituées en société.

Le groupe intermédiaire de branches d'activité présente un portrait similaire, même si ces branches d'activité semblent compter davantage d'entreprises simples non constituées en société. Elles ont tendance à englober des entreprises plus petites qui utilisent moins de capital. Il est intéressant de souligner que l'indice du revenu total est très souvent plus élevé que l'indice du nombre d'établissements. Cela laisse supposer que les entreprises constituées en société ont tendance à être plus importantes que les entreprises non constituées en société et les entreprises publiques. Même si le revenu total de tous les hôpitaux pédiatriques est représenté à 99,9 % dans l'IGRF, seule une proportion de 57 % de ce revenu ne pose pas de problème. Il est intéressant de constater que même si les entreprises simples constituées en société ne représentent que 36,4 % du nombre total d'établissements, elles constituent 57 % du revenu total de la branche d'activité. Par ailleurs, tous les autres grossistes-distributeurs, la construction et réparation de navires et les vineries ont tous des indices de couverture selon le nombre d'établissements non problématiques, c'est-à-dire supérieurs à 70 %, mais leurs indices de revenu total sont beaucoup plus faibles.

La dernière section du tableau 2 est constituée de branches d'activité comprenant principalement des entreprises non constituées en société et les gouvernements. Les cantines et comptoirs mobiles, la culture du blé et les services de taxi ont tous des indices très faibles de couverture par l'IGRF. Cela vient du fait qu'elles sont gérées par un petit nombre de personnes, souvent à partir du lieu de résidence personnelle, et que leur lancement n'entraîne que peu de frais généraux. Les entreprises publiques ne sont pas couvertes par l'IGRF.

Dans l'ensemble, l'indice calculé au moyen du nombre d'établissements laisse supposer que 66,8 % des établissements sont couverts par l'IGRF, ou 63,2 %, si l'on inclut uniquement les entreprises simples. L'examen du nombre d'employés laisse supposer que 85,1 % des employés sont couverts par l'IGRF, ou 59,3 % si l'on se fonde sur les entreprises simples seulement. Enfin, l'indice du revenu total indique que 82,1 % de l'économie est bien couverte par l'IGRF, même si cette mesure passe à seulement 38,1 % si l'on ne tient compte que des entreprises simples constituées en société.

Démonstration de ce que comporte actuellement la base de données

La présente section vise à démontrer au lecteur quels renseignements figurent actuellement dans la base de données de l'IGRF, et ce qu'on peut faire avec ces données. Deux tableaux figurent ci-dessous, un faisant état de l'exhaustivité, et l'autre de l'étendue de l'IGRF.

Parmi les enjeux essentiels liés à l'utilisation de la base de données figure l'attribution des bons codes d'activité aux enregistrements de l'IGRF. Le tableau 3 montre le nombre total de sociétés qui avaient déclaré les variables obligatoires de l'IGRF au moment de la rédaction du document. Ce total est ventilé selon que l'enregistrement a reçu ou non un code du SCIAN, grâce au couplage réussi de l'enregistrement à une entreprise codée dans le RE. Figure aussi dans le tableau le pourcentage de données manquantes découlant des cas où aucun code du SCIAN n'a été attribué. Le pourcentage de données manquantes en raison de la non-attribution d'un code du SCIAN est calculé en divisant la colonne « Sans code du SCIAN » par la colonne « Nombre total ». Le pourcentage d'enregistrements sans code du SCIAN se situait à 14 % environ pour chaque variable obligatoire au moment de la rédaction du présent document.

Tableau 3 – Total des enregistrements disponibles pour des postes obligatoires de l'IGRF

Code	Nom de code	Avec code du SCIAN	Sans code du SCIAN	Nombre total	% de données manquantes
L2599	Total de l'actif	741 653	126 663	868 316	14,6 %
L3499	Total du passif	741 609	126 654	868 263	14,6 %
L3620	Total des capitaux propres	741 755	126 659	868 414	14,6 %
L8299	Total des revenus	749 161	125 275	874 436	14,3 %
L9368	Total des dépenses	749 198	125 284	874 482	14,3 %
L9659	Total des revenus agricoles	743 594	124 097	867 691	14,3 %
L9898	Total des dépenses agricoles	743 593	124 096	867 689	14,3 %
L3849	Bénéfices non répartis/déficit - fin de l'exercice	698 412	108 303	806 715	13,4 %
L9999	Revenu/perte nette après impôts et éléments extraordinaires	749 225	125 292	874 517	14,3 %

Le tableau 4 montre le nombre total d'enregistrements pour chaque poste obligatoire dans la base de données de l'IGRF. Le nombre total est ventilé entre les enregistrements qui comportent un code du SCIAN et ceux qui n'ont pas de code du SCIAN. Le pourcentage de données manquantes en raison de la non-attribution d'un code du SCIAN est calculé en divisant la colonne « Sans code du SCIAN » par la colonne « Nombre total ». Dans ce cas, le pourcentage d'enregistrements manquants varie considérablement selon le poste. Le poste 3620, total des capitaux propres, est celui pour lequel il manque le plus d'information au sujet de toutes les variables obligatoires, soit environ 20,9 % du total. Le poste 9659, total des revenus agricoles, est celui où il manque le moins d'information, soit seulement 2,5 % du total. Même si seulement 4,7 % des données sur le total des dépenses manquent en raison de l'absence d'un code du SCIAN, cela reste surprenant parce que seulement un billion de dollars de dépenses totales n'est pas représenté. Enfin, le montant total des dépenses est de 24 billions de dollars, ce qui est trop élevé et laisse sous-entendre des problèmes importants de vérification, qui doivent être résolus dans le cadre de cette première édition de la base de données de l'IGRF.¹⁷

Tableau 4 – Montant total selon le poste de l'IGRF (en milliards de dollars)

SCIAN	Nom du code	Avec code du SCIAN	Sans code du SCIAN	Montant total	% de données manquantes
L2599	Total de l'actif	1 932 \$	291 \$	2 223 \$	13,1 %
L3499	Total du passif	1 404 \$	154 \$	1 558 \$	9,9 %
L3620	Total des capitaux propres	534 \$	141 \$	674 \$	20,9 %
L8299	Total des revenus	1 002 \$	112 \$	1 114 \$	10,1 %
L9368	Total des dépenses	22 871 \$	1 123 \$	23 994 \$	4,7 %
L9659	Total des revenus agricoles	17 \$	0,4 \$	17 \$	2,5 %
L9898	Total des dépenses agricoles	1 574 \$	76 \$	1 650 \$	4,6 %
L3849	Bénéfices non répartis/déficit - fin de l'exercice	171 \$	21 \$	192 \$	11,2 %
L9999	Revenu/perte nette après impôts et éléments extraordinaires	1 574 \$	76 \$	1 650 \$	4,6 %

¹⁷ Une révision manuelle du fichier de l'IGRF 1998, après les calculs des totaux, a permis de corriger des valeurs extrêmes qui affectaient seulement quelques enregistrements. Le résultat, suite à ces corrections, a rétabli le total des dépenses à un niveau raisonnable.

Le prochain exemple, qui figure au tableau 5, repose sur une branche d'activité et présente des totalisations selon la province. L'exemple est conçu pour montrer au lecteur le type de renseignements particuliers qui peut être tiré de la base de données de l'IGRF. La branche d'activité choisie est celle des magasins de fournitures pour la maison et l'auto, code 452991 du SCIAN, et elle a été sélectionnée parce qu'elle obtient un bon indice de couverture par l'IGRF¹⁸. L'utilisateur peut préciser d'autres branches d'activité à examiner ainsi que des variables à extraire. En couplant les données du tableau des entreprises de Revenu Canada et celles des autres tableaux, il est possible de créer des ventilations selon la province.

Le nombre total d'entités légales inscrites dans la base de données pour cette branche d'activité est de 482. Tous les chiffres du tableau 5 ont été arrondis au millier de dollars près. Le total de l'actif est de 1 192 000 000 \$, le total du passif, de 1 015 000 000 \$, le total des revenus, de 3 346 000 000 \$ et le total des dépenses, de 3 237 000 000 \$.

Tableau 5 — Code 452991 du SCIAN — Magasins de fournitures pour la maison et l'auto (en millions de dollars)

Province	Nombre	Total de l'actif	Total du passif	Total des revenus	Total des dépenses
ALB.	36	82 \$	69 \$	222 \$	211 \$
C.-B.	25	97 \$	89 \$	281 \$	284 \$
MAN.	14	25 \$	22 \$	68 \$	58 \$
N.-B.	19	43 \$	35 \$	118 \$	116 \$
T.-N.	14	36 \$	34 \$	86 \$	82 \$
N.-É.	24	52 \$	48 \$	135 \$	133 \$
T. N.-O.	X	X	X	X	X
ONT.	227	535 \$	453 \$	1 639 \$	1 571 \$
I.-P.-É.	X	X	X	X	X
QUÉ.	108	276 \$	224 \$	663 \$	650 \$
SASK.	11	30 \$	27 \$	85 \$	84 \$
YUKON	X	X	X	X	X
TOTAL	482	1 192 \$	1 015 \$	3 346 \$	3 237 \$

X = Supprimé pour protéger la confidentialité.

Le tableau 5 n'a pas pu être diffusé dans sa forme originale, parce que cela aurait entraîné la divulgation de renseignements confidentiels au sujet d'entreprises du fait que certaines cellules comprennent une ou deux entités légales seulement. En outre, si l'activité d'une seule entreprise représente un pourcentage important du montant total figurant dans une cellule, la suppression est jugée nécessaire. Il existe aussi des conditions spéciales, c'est-à-dire les règles de Duffet, pour les situations plus complexes. L'utilisateur doit être prudent lorsqu'il supprime une cellule. Si une cellule est supprimée mais que le total figure toujours dans le tableau, les lecteurs pourront déterminer les données qui manquent. Afin d'éviter cette situation, il faut supprimer une ou plusieurs autres cellules pour maintenir la confidentialité de la cellule supprimée au départ.

¹⁸ L'indice de couverture de l'IGRF est décrit de façon plus détaillée dans la section intitulée « Quelles sont les branches d'activité les mieux représentées par l'IGRF? » du présent document.

Services personnels et services de blanchissage

Exemple de l'utilisation statistique de l'IGRF

Afin d'aider le lecteur à comprendre ce que l'IGRF a à offrir au système statistique, un examen des branches d'activité comprises dans le sous-secteur des services personnels et services de blanchissage, code 812 du SCIAN, figure ci-après. Ce secteur a été choisi assez arbitrairement à cette fin, même si l'on considère qu'il se prête assez bien à une analyse fondée sur l'IGRF, du fait de sa nature atomistique et parce qu'il n'est pas dominé par des établissements d'entreprises complexes. L'analyse se divise en quatre parties. Tout d'abord, on fait état de la structure du sous-secteur des services personnels et services de blanchissage dans le SCIAN et on détermine ce qui fait qu'il peut s'agir d'une bonne branche d'activité à explorer au moyen des données de l'IGRF. Une deuxième section fournit une ventilation du sous-secteur selon le nombre d'entités légales (EL) et le total des revenus, à partir de la base de données de l'IGRF. Une troisième partie décrit le sous-secteur du point de vue des établissements et du total des revenus, à partir des renseignements figurant dans le Registre des entreprises. Enfin, une dernière section comprend une comparaison des renseignements au sujet du sous-secteur compris dans l'IGRF et dans le RE, souligne des problèmes possibles, ainsi que des éléments à vérifier au moment de l'analyse d'une branche d'activité particulière à partir des données de l'IGRF.

Sous-secteur 812 du SCIAN — Services personnels et services de blanchissage

Le SCIAN a été conçu par les bureaux statistiques du Canada, des États-Unis et du Mexique. L'objectif visé consistait à produire un cadre commun de groupes de branches d'activité pour chacun de ces trois pays. La classification vise à englober l'ensemble de l'activité économique de chacun de ces trois pays. Le système regroupe les entreprises sur la base des similitudes quant à leurs processus de production. Cela permet aux bureaux statistiques de prévoir un questionnaire pour de nombreuses entreprises similaires et de produire des statistiques qui peuvent être comparées entre ces entreprises similaires.

La classification est fondée sur une structure hiérarchique. Elle est composée de secteurs, de sous-secteurs, de groupes, de classes et d'industries nationales. Au Canada, le SCIAN est constitué de 20 secteurs, 99 sous-secteurs, 321 groupes, 734 classes et 921 industries nationales. Les secteurs correspondent au code à 2 chiffres, les sous-secteurs, au code à 3 chiffres, les groupes au code à 4 chiffres, les classes, au code à 5 chiffres et les industries nationales, au code à 6 chiffres. Le code à 6 chiffres est souvent désigné à l'intérieur de Statistique Canada comme le « code de branche d'activité ». Un code de branche d'activité est attribué à chaque établissement d'une entreprise. Un code de branche d'activité peut aussi être attribué à une entreprise. Il s'agit du code du SCIAN d'entreprise. Dans ce cas, le code est attribué à l'entreprise selon l'activité prédominante des établissements qu'elle comprend. L'activité prédominante est déterminée au moyen d'un algorithme qui est appliqué par la Division du Registre des entreprises, et qui tient compte du code d'activité de chaque établissement de l'entreprise ainsi que de la taille relative de cet établissement, mesuré selon la proportion qu'il représente du revenu ou de l'emploi total de l'entreprise.

Le sous-secteur des services personnels et services de blanchissage appartient au secteur des « autres services, sauf les administrations publiques » et est constitué de 4 groupes, 11 classes et 14 industries nationales. La ventilation du sous-secteur des services personnels et services de blanchissage à l'intérieur du SCIAN est présentée ci-dessous¹⁹.

¹⁹ Pour une définition plus détaillée du sous-secteur des services personnels et services de blanchissage, voir le manuel du Système de classification des industries de l'Amérique du Nord pour le Canada, publié en 1997.

812 – Services personnels et services de blanchissage

8121 Services de soins personnels

- 81211 Services de coiffure et d'esthétique
- 812114 *Salons de coiffure pour hommes*
- 812115 *Salons de beauté*
- 812116 *Salons de coiffure mixtes*
- 81219 Autres services de soins personnels
- 812190 *Autres services de soins personnels*

8122 Services funéraires

- 81221 Salons funéraires
- 812210 *Salons funéraires*
- 81222 Cimetières et crématoriums
- 812220 *Cimetières et crématoriums*

8123 Services de nettoyage à sec et de blanchissage

- 81231 Blanchisseries et nettoyeurs à sec libre-service
- 812310 *Blanchisseries et nettoyeurs à sec libre-service*
- 81232 Services de nettoyage à sec et de blanchissage, sauf le libre-service
- 812320 *Services de nettoyage à sec et de blanchissage, sauf le libre-service*
- 81233 Fourniture de linge et d'uniformes
- 812330 *Fourniture de linge et d'uniformes*

8129 Autres services personnels

- 81291 Soins pour animaux de maison, sauf les services vétérinaires
- 812910 *Soins pour animaux de maison, sauf les services vétérinaires*
- 81292 Services de développement et de tirage de photos
- 812921 *Laboratoires de développement et de tirage de photos, sauf le service en une heure*
- 812922 *Développement et tirage de photos en une heure*
- 81293 Stationnements et garages
- 812930 *Stationnements et garages*
- 81299 Tous les autres services personnels
- 812990 *Tous les autres services personnels*

Les services personnels et les services de blanchissage sont un sous-secteur sur lequel l'IGRF devrait fournir un éclairage considérable²⁰. En fait, les données de la base de données de l'IGRF combinées aux données d'autres sources administratives pourraient permettre à Statistique Canada de mettre un terme à son enquête existante sur ce sous-secteur, ou à tout le moins de la réduire de façon substantielle. Cette enquête comporte un échantillon d'environ 1 000 unités et obtient un taux de réponse d'environ 66 %. Les données recueillies au moyen du questionnaire d'enquête portent sur la nature de l'activité de l'entreprise, la forme d'organisation, la période de déclaration, les revenus, les dépenses d'exploitation, les stocks, l'emploi, la clientèle et la répartition provinciale/territoriales des revenus, des dépenses, des salaires et de l'emploi. La plupart de ces données sont disponibles à partir de l'IGRF et d'autres enregistrements

²⁰ Voir la section intitulée « Quelles sont les branches d'activité les mieux représentées par l'IGRF? »

administratifs comme le système PD7/T4, les enregistrements de la TPS, les enregistrements T1 et CORPAC/CORTAX. En outre, le niveau de détail fourni par les répondants de l'enquête est considéré comme assez faible. Du fait qu'il s'agit d'un recensement des entités légales constituées en société, les données de l'IGRF couvrent une proportion très importante de la branche d'activité et comportent des détails considérables sur le bilan et l'état des résultats. Une enquête pourrait continuer d'être utile pour obtenir d'autres genres de données relativement à la clientèle, à la répartition provinciale des activités et à la structure en évolution des établissements des entreprises complexes de cette branche d'activité. Néanmoins, l'IGRF devrait permettre de raccourcir le questionnaire d'enquête et de réduire l'échantillon, ce qui contribuera à diminuer le fardeau de réponse tout en améliorant parallèlement la qualité des statistiques produites.

Étant donné que les tendances financières sont assez répandues d'une entreprise à l'autre au sein d'une branche d'activité, les données de l'IGRF pour un groupe d'entreprises peuvent être utilisées pour projeter des données financières manquantes pour d'autres groupes d'entreprises. Il s'agit là de la deuxième utilisation en importance de l'IGRF. Supposons, par exemple, que le revenu total des entreprises simples constituées en société du sous-secteur des services personnels et services de blanchissage soit tiré de l'IGRF pour une année particulière. Supposons en outre que le revenu total des entreprises complexes constituées en société, des entreprises simples et complexes non constituées en société et des entreprises publiques du sous-secteur des services personnels et services de blanchissage, entreprises pour lesquelles des données ne sont pas disponibles dans l'IGRF, est estimé d'une façon ou d'une autre pour la même année, ce qui donne un aperçu global du revenu total de la branche d'activité. L'année suivante, l'IGRF fournira encore une fois les données sur le revenu total des entreprises simples constituées en société. Si les données pour les autres composantes de la branche d'activité comportent une corrélation étroite avec les données de la composante des entreprises simples constituées en société, et plus particulièrement si les autres composantes sont relativement petites en comparaison de la composante des entreprises simples constituées en société, le changement observé à l'intérieur de l'IGRF d'une année à l'autre peut être utilisé pour projeter les données financières des autres composantes de la branche d'activité.

Ci-après figurent les tableaux 6, 7 et 8, qui montrent l'indice de couverture de l'IGRF selon le nombre d'établissements, l'emploi et le revenu, respectivement. Lorsque l'on examine le tableau du nombre d'établissements, il ressort que l'IGRF ne peut à lui seul servir à estimer des variables financières clés du sous-secteur des services personnels et services de blanchissage. On s'attend à ce que l'IGRF ne couvre que 50,6 % des entreprises (complexes et simples regroupées) du sous-secteur, et seulement 48,3 % des entreprises (simples seulement). Il existe un nombre important d'entreprises non constituées en société dans ce sous-secteur, qui ne sont pas couvertes par l'IGRF non plus. Toutefois, les ratios de couverture de l'IGRF pour le sous-secteur semblent un peu meilleurs lorsqu'ils sont pondérés selon l'emploi ou le revenu. Autrement dit, les entreprises qui *sont* couvertes par l'IGRF ont tendance à être relativement importantes comparées à celles qui ne le sont pas. Les tableaux 7 et 8 indiquent tous les deux qu'environ 65 % de l'activité du sous-secteur (mesurée par l'emploi ou le revenu) est le fait des entreprises simples constituées en société, pour lesquelles des données de l'IGRF sont disponibles. Les pourcentages sont particulièrement élevés pour les services personnels et services de blanchissage, les blanchisseries libre-service, le développement et tirage de photos en une heure et les salons funéraires.

Les raisons qui font que ces quatre groupes différents sont couverts différemment dans l'IGRF sont intéressantes. Le groupe 8121 - Services de soins personnels - est celui qui a la couverture la plus faible dans l'IGRF pour le sous-secteur des services personnels et services de blanchisserie. L'exploitation d'un petit salon de coiffure pour hommes ou pour femmes dans son domicile ou à proximité n'entraîne pas beaucoup de frais généraux et est relativement facile. Il est très probable que ce groupe compte de nombreuses entreprises non constituées en société. Les services funéraires (groupe 8122), par contre, ont tendance à être composés d'entreprises importantes comportant des frais généraux élevés. Comme il a été mentionné précédemment, les salons funéraires ont tendance à être exploités à plus large échelle que les cimetières et les

crématoriums, ce qui fait que ces derniers peuvent être caractérisés par un nombre relativement plus élevé d'entreprises non constituées en société. Dans l'ensemble, le groupe a un taux de couverture supérieur à 75 % dans l'IGRF. Le groupe des services de nettoyage à sec et de blanchissage (8123) est couvert relativement bien par l'IGRF, en général. Ce groupe s'apparente aux services funéraire, les deux ayant une couverture d'au moins 75 %. Le dernier groupe, à savoir les autres services personnels (8129) est plus compliqué à analyser. Cela vient du fait qu'il est constitué d'activités qui sont assez semblables. Les stationnements et garages et les laboratoires de développement et de tirage de photos figurent en tête de liste parce que leur équipement et leurs immobilisations sont coûteux, ce qui incite par conséquent les propriétaires à se constituer en société. Les soins pour animaux de maison, par ailleurs, figurent plus bas dans la liste, et on présume qu'ils utilisent généralement moins de capital et qu'ils sont exploités à plus petite échelle, un peu comme les salons de coiffure. Chaque branche d'activité est unique et les utilisateurs des données de l'IGRF doivent faire preuve de discernement pour déterminer lesquelles sont suffisamment couvertes par la base de données pour permettre des conclusions générales fondées uniquement sur les données de l'IGRF.

Tableau 6 — Services personnels et services de blanchissage selon la « couverture » par l'IGRF à partir du nombre d'établissements

SCIAN	Description de la branche d'activité	Indice de couverture par l'IGRF en présence d'unités complexes	Indice de couverture par l'IGRF en l'absence d'unités complexes
812114	Salons de coiffure pour hommes	26,3 %	26,3 %
812115	Salons de beauté	39,6 %	37,1 %
812116	Salons de coiffure mixtes	40,7 %	40,4 %
812190	Autres services de soins personnels	56,2 %	56,0 %
812210	Salons funéraires	92,2 %	75,1 %
812220	Cimetières et crématoriums	45,4 %	40,1 %
812310	Blanchisseries et nettoyeurs à sec libre-service	62,5 %	61,9 %
812320	Services de nettoyage à sec et de blanchissage, sauf le libre service	68,3 %	67,0 %
812330	Fourniture de linge et d'uniformes	80,8 %	71,2 %
812910	Soins pour animaux de maison, sauf les services vétérinaires	43,0 %	42,8 %
812921	Laboratoires de développement et de tirage de photos, sauf le service en une heure	74,5 %	71,4 %
812922	Développement et tirage de photos en une heure	76,7 %	73,0 %
812930	Stationnements et garages	81,7 %	74,1 %
812990	Tous les autres services personnels	41,8 %	41,1 %
Total		50,6 %	48,3 %

Tableau 7 — Services personnels et services de blanchissage selon la « couverture » par l'IGRF à partir du nombre total d'employés

SCIAN	Description de la branche d'activité	Indice de couverture par l'IGRF en présence d'unités complexes	Indice de couverture par l'IGRF en l'absence d'unités complexes
812114	Salons de coiffure pour hommes	48,7 %	48,7 %
812115	Salons de beauté	66,5 %	61,7 %
812116	Salons de coiffure mixtes	72,7 %	63,8 %
812190	Autres services de soins personnels	86,2 %	67,3 %
812210	Salons funéraires	98,0 %	76,3 %
812220	Cimetières et crématoriums	64,7 %	31,3 %
812310	Blanchisseries et nettoyeurs à sec libre-service	79,4 %	78,2 %
812320	Services de nettoyage à sec et de blanchissage, sauf le libre service	89,4 %	85,5 %
812330	Fourniture de linge et d'uniformes	99,2 %	56,2 %
812910	Soins pour animaux de maison, sauf les services vétérinaires	58,7 %	55,6 %
812921	Laboratoires de développement et de tirage de photos, sauf le service en une heure	94,5 %	59,8 %
812922	Développement et tirage de photos en une heure	90,9 %	75,2 %
812930	Stationnements et garages	94,3 %	37,1 %
812990	Tous les autres services personnels	76,5 %	53,9 %
Total		79,9 %	64,7 %

Tableau 8 — Services personnels et services de blanchissage selon la « couverture » par l'IGRF à partir du revenu total

SCIAN	Description de la branche d'activité	Indice de couverture par l'IGRF en présence d'unités complexes	Indice de couverture par l'IGRF en l'absence d'unités complexes
812114	Salons de coiffure pour hommes	45,0 %	44,3 %
812115	Salons de beauté	74,0 %	70,5 %
812116	Salons de coiffure mixtes	71,9 %	65,4 %
812190	Autres services de soins personnels	77,9 %	69,5 %
812210	Salons funéraires	97,3 %	70,6 %
812220	Cimetières et crématoriums	78,4 %	34,4 %
812310	Blanchisseries et nettoyeurs à sec libre-service	77,6 %	77,0 %
812320	Services de nettoyage à sec et de blanchissage, sauf le libre service	87,2 %	79,5 %
812330	Fourniture de linge et d'uniformes	98,8 %	46,8 %
812910	Soins pour animaux de maison, sauf les services vétérinaires	68,0 %	66,6 %
812921	Laboratoires de développement et de tirage de photos, sauf le service en une heure	96,6 %	60,0 %
812922	Développement et tirage de photos en une heure	91,4 %	78,3 %
812930	Stationnements et garages	85,0 %	49,3 %
812990	Tous les autres services personnels	85,3 %	71,6 %
Total		82,1 %	65,7 %

Services personnels et services de blanchissage — Estimation statistique au moyen des données de l'IGRF pour 1998

Afin d'examiner la taille du sous-secteur des services personnels et services de blanchissage, on a évalué le nombre d'entités légales et le revenu total de chaque branche d'activité du sous-secteur, à partir de la base de données de l'IGRF²¹. Le tableau 9 ci-dessous montre le nombre d'entités légales dans chaque branche d'activité, et le tableau 10, le revenu total dans chaque branche d'activité.

Comme on peut le voir dans le tableau 9, la base de données de l'IGRF compte 15 898 entités légales classifiées selon le RE dans le sous-secteur des services personnels et services de blanchissage. Parmi elles, 54 % appartiennent au groupe des services de soins personnels, 10 %, au groupe des services funéraires, 24 %, au groupe des services de nettoyage à sec et de blanchissage, et 12 %, au groupe des autres services personnels. Les totaux pour chaque groupe sont aussi ventilés selon la branche d'activité. Les salons de beauté sont de loin la branche d'activité qui compte le plus d'entités légales, soit 25 % du total. Les branches d'activité plus petites sont les cimetières et crématoriums et les laboratoires de développement et de tirage de photos, sauf le service en une heure, chacune représentant 1 % environ seulement du total des entités légales.

Tableau 9 — Total des entités légales de la base de données de l'IGRF pour 1998

EL SCIAN	Description de la branche d'activité	Total des EL	Pourcen- tage d'EL	Total des EL	Pourcen- tage d'EL
812114	Salons de coiffure pour hommes	259	2 %	8 594	54 %
812115	Salons de beauté	4 045	25 %		
812116	Salons de coiffure mixtes	2 033	13 %		
812190	Autres services de soins personnels	2 257	14 %		
812210	Salons funéraires	1 342	9 %	1 506	10 %
812220	Cimetières et crématoriums	164	1 %	3 874	24 %
812310	Blanchisseries et nettoyeurs à sec libre-service	846	5 %		
812320	Services de nettoyage à sec et de blanchissage, sauf le libre service	2 731	17 %		
812330	Fourniture de linge et d'uniformes	297	2 %		
812910	Soins pour animaux de maison, sauf les services vétérinaires	400	3 %	1 924	12 %
812921	Laboratoires de développement et de tirage de photos, sauf le service en une heure	223	1 %		
812922	Développement et tirage de photos en une heure	510	3 %		
812930	Stationnements et garages	463	3 %		
812990	Tous les autres services personnels	328	2 %		
Total		15 898	100 %	15 898	100 %

²¹ Pour savoir comment on attribue un code du SCIAN à une EL, voir la section intitulée « Quels sont les principaux défis auxquels Statistique Canada doit faire face pour l'utilisation de l'IGRF? »

Le tableau 10 montre la ventilation du revenu total selon la branche d'activité. Le revenu total du sous-secteur des services personnels et services de blanchissage est de 3,383 millions de dollars. Le groupe des services de soins personnels représente la part la plus importante du revenu total, soit 33 %. Le groupe des services de nettoyage à sec et de blanchisserie est le deuxième en importance, représentant 28 % du revenu total. Le groupe des services funéraires compte pour 24 % du revenu, et les autres services personnels représentent les 15 % qui restent du revenu total. Les salons funéraires représentent la proportion la plus importante des revenus, avec 20 % du total, tandis que les salons de coiffure pour hommes et les soins pour animaux de maison ne représentent chacun que 1 % du revenu total.

Tableau 10 — Revenu total dans la base de données de l'IGRF de 1998 (en millions de dollars)

EL SCIAN	Description de la branche d'activité	Revenu	Pourcentage	Revenu	Pourcentage
		total	du revenu	total	du revenu
812114	Salons de coiffure pour hommes	26	1 %	1 127	33 %
812115	Salons de beauté	510	15 %		
812116	Salons de coiffure mixtes	371	11 %		
812190	Autres services de soins personnels	220	6 %		
812210	Salons funéraires	667	20 %	797	24 %
812220	Cimetières et crématoriums	130	4 %	934	28 %
812310	Blanchisseries et nettoyeurs à sec libre-service	87	3 %		
812320	Services de nettoyage à sec et de blanchissage, sauf le libre service	486	14 %		
812330	Fourniture de linge et d'uniformes	360	11 %		
812910	Soins pour animaux de maison, sauf les services vétérinaires	28	1 %	525	15 %
812921	Laboratoires de développement et de tirage de photos, sauf le service en une heure	149	4 %		
812922	Développement et tirage de photos en une heure	92	2 %		
812930	Stationnements et garages	194	6 %		
812990	Tous les autres services personnels	61	2 %		
Total		\$ 3 383	100 %	\$ 3 383	100 %

Information comprise dans le RE au sujet du sous-secteur des services personnels et services de blanchissage

Après avoir examiné le profil du sous-secteur des services personnels et services de blanchissage à partir de la base de données de l'IGRF, il est intéressant d'examiner les données disponibles dans le Registre des entreprises au sujet des mêmes branches d'activité. L'objectif visé est de comparer les données de l'IGRF, en excluant les entreprises non constituées en société et les entités publiques et en englobant les entités légales constituées en société plutôt que les établissements de production, et les renseignements plus exhaustifs et mieux définis au niveau statistique qui figurent dans le RE. Une question pose un intérêt particulier : le sous-secteur des services personnels et services de blanchissage est-il sous-représenté ou sur-représenté dans la base de données de l'IGRF et dans quelle mesure?

Le revenu total et le total des établissements comportant un des 14 codes du SCIAN relevant des services personnels et services de blanchissage ont été extraits du RE. Dans le RE, un code du SCIAN est attribué en premier lieu à l'établissement d'une entreprise. Un code du SCIAN au niveau de l'entreprise est par la suite attribué, en fonction de l'activité prédominante des établissements de l'entreprise. Cela veut dire que même si une entreprise reçoit un code du SCIAN correspondant aux services personnels et services de blanchissage, cela ne signifie pas nécessairement que toutes ses activités se trouvent dans ce sous-secteur. Cela peut venir du fait qu'une entreprise particulière compte un ou plusieurs établissements comportant un ou plusieurs autres codes du SCIAN, tout en ayant aussi des activités dans le sous-secteur des services personnels et services de blanchissage en même temps. Un tel cas figure au tableau 11 à la rubrique « autre », c'est-à-dire des branches d'activités dont l'activité prédominante correspond à un code du SCIAN autre que 812, mais dont les activités se retrouvent aussi dans une des branches d'activité « 812 ».

Il est important de se rappeler cela lorsque l'on examine une branche d'activité. En principe, pour obtenir des estimations précises pour une branche d'activité, chaque entreprise doit être ventilée selon l'activité en question. Comme le montre la colonne 3 du tableau 11, 38 114 établissements (38 623 - 509) du RE font partie d'entreprises classifiées dans le sous-secteur des services personnels et services de blanchissage. Comme le montre la quatrième colonne, la plupart de ces établissements sont classifiés dans ce sous-secteur, mais pas la totalité. En outre, 509 établissements d'autres entreprises dont l'activité prédominante ne fait pas partie du sous-secteur des services personnes et services de blanchissage, comptent au moins un établissement dans ce sous-secteur. Parmi ces 509 établissements, 35 sont classifiés dans le sous-secteur. Au total, comme le montre la dernière ligne de la colonne 4 du tableau 11, 38 066 établissements des services personnels et services de blanchissage figurent dans le RE.

Tableau 11 – Total des établissements des services personnels et services de blanchissage figurant dans le RE

Entreprise SCIAN	Description de la branche d'activité	Total des établissements	Total des établissements des services personnels et services de blanchissage	Pourcentage du total des établissements des services personnels et services de blanchissage
812114	Salons de coiffure pour hommes	1 557	1 557	100 %
812115	Salons de beauté	12 681	12 677	100 %
812116	Salons de coiffure mixtes	6 286	6 285	100 %
812190	Autres services de soins personnels	4 650	4 649	100 %
812210	Salons funéraires	1 656	1 636	99 %
812220	Cimetières et crématoriums	375	371	99 %
812310	Blanchisseries et nettoyeurs à sec libre-service	1 686	1 682	100 %
812320	Services de nettoyage à sec et de blanchissage, sauf le libre service	4 833	4 827	100 %
812330	Fourniture de linge et d'uniformes	351	347	99 %
812910	Soins pour animaux de maison, sauf les services vétérinaires	1 287	1 287	100 %
812921	Laboratoires de développement et de tirage de photos, sauf le service en une heure	321	320	100 %
812922	Développement et tirage de photos en une heure	726	725	100 %
812930	Stationnements et garages	723	688	95 %
812990	Tous les autres services personnels	982	980	100 %
Autre		509	35	7 %
Total		38 623	38 066	99 %

Le tableau 12 montre le revenu total du sous-secteur des services personnels et services de blanchissage figurant dans le RE. La méthode qui a servi à élaborer ce tableau est similaire à celle s'appliquant au tableau 11. Dans ce cas-ci, le revenu total de toutes les entreprises qui comptent un ou plusieurs établissements dans le sous-secteur des services personnels et services de blanchissage est de 14 473 millions de dollars. De ce montant, seulement 6 875 millions de dollars sont le fait d'établissements des entreprises classifiées dans le sous-secteur des services personnels et services de blanchissage. La principale raison qui justifie une différence si importante est que les entreprises qui ont des activités dominantes dans d'autres branches d'activité et qui ont aussi des activités dans les services personnels et services de blanchissage sont nombreuses, mais qu'elles ont tendance à tirer la majorité de leur revenu d'activités autres que celles comprises dans le sous-secteur des services personnels et services de blanchissage.

Tableau 12 — Revenu total pour les services personnels et services de blanchissage figurant dans le RE (en millions de dollars)

Entreprise SCIAN	Description de la branche d'activité	Revenu total des entreprises	Revenu total des services personnels et services de blanchissage	Pourcentage du revenu total des services personnels et services de blanchissage
812114	Salons de coiffure pour hommes	132	132	100 %
812115	Salons de beauté	1 475	1 474	100 %
812116	Salons de coiffure mixtes	979	979	100 %
812190	Autres services de soins personnels	429	428	100 %
812210	Salons funéraires	806	793	98 %
812220	Cimetières et crématoriums	257	242	94 %
812310	Blanchisseries et nettoyeurs à sec libre-service	158	158	100 %
812320	Services de nettoyage à sec et de blanchissage, sauf le libre service	913	900	99 %
812330	Fourniture de linge et d'uniformes	599	587	98 %
812910	Soins pour animaux de maison, sauf les services vétérinaires	155	155	100 %
812921	Laboratoires de développement et de tirage de photos, sauf le service en une heure	142	137	97 %
812922	Développement et tirage de photos en une heure	171	171	100 %
812930	Stationnements et garages	545	520	95 %
812990	Tous les autres services personnels	161	160	100 %
Autre		7 551	39	1 %
Total		14 473 \$	6 875 \$	48 %

Comparaison des données du RE et de celles de l'IGRF pour les services personnels et services de blanchissage

L'étape suivante consiste à comparer les données du RE à celles figurant dans la base de données de l'IGRF. La troisième colonne du tableau 13A montre le revenu total des établissements classifiés dans le sous-secteur des services personnels et services de blanchissage figurant dans le RE. La quatrième colonne montre le revenu total *qui devrait figurer* dans la base de données de l'IGRF, soit la somme du revenu total figurant dans le RE pour les entreprises simples et complexes constituées en société du sous-secteur des services

personnels et services de blanchissage. Évidemment, le revenu total figurant dans le RE pour l'ensemble du sous-secteur, qui est indiqué à la troisième colonne, est plus élevé, parce qu'il comprend aussi les entreprises non constituées en société et les entreprises publiques. À la cinquième colonne du tableau 13A figure le revenu réel de ces branches d'activité selon l'IGRF. Au moment de la rédaction du présent document, la base de données de l'IGRF comportait un peu plus de 800 000 enregistrements, tandis que selon le RE, l'économie dans son ensemble compte environ 1,3 million d'entités légales constituées en société. Par conséquent, le revenu total de l'IGRF est de beaucoup inférieur au revenu total qui devrait y figurer, globalement et pour la plupart des branches d'activité.

Tableau 13A — Comparaison du revenu total figurant dans le RE et dans l'IGRF (en millions de dollars)

Entreprise SCIAN	Description de la branche d'activité	Total du revenu du RE	Revenu qui devrait figurer dans l'IGRF	Revenu réel dans l'IGRF
812114	Salons de coiffure pour hommes	132	59	26
812115	Salons de beauté	1 474	1 039	510
812116	Salons de coiffure mixtes	979	640	371
812190	Autres services de soins personnels	428	298	220
812210	Salons funéraires	793	560	667
812220	Cimetières et crématoriums	242	83	130
812310	Blanchisseries et nettoyeurs à sec libre-service	158	121	87
812320	Services de nettoyage à sec et de blanchissage, sauf le libre service	900	716	486
812330	Fourniture de linge et d'uniformes	587	275	360
812910	Soins pour animaux de maison, sauf les services vétérinaires	155	103	28
812921	Laboratoires de développement et de tirage de photos, sauf le service en une heure	137	82	149
812922	Développement et tirage de photos en une heure	171	134	92
812930	Stationnements et garages	520	256	194
812990	Tous les autres services personnels	160	115	61
Autre		39	0	0
TOTAL		6 875 \$	4 481 \$	3 383 \$

La deuxième partie du tableau (partie B) rend compte, dans la colonne 3, du pourcentage du revenu total qui *devrait* figurer dans la base de données de l'IGRF, selon le RE. On obtient ce chiffre en divisant le revenu attendu par le revenu total figurant dans le RE, ce qui montre dans quelle mesure les données de l'IGRF prises isolément donneront lieu à un sous-dénombrement de la branche d'activité. La colonne suivante fait état du pourcentage du total qui figurait dans la base de données de l'IGRF au moment de la rédaction du présent document. Ce chiffre est calculé en divisant le revenu total figurant actuellement dans la base de données de l'IGRF et le revenu total figurant dans le RE, et montre la proportion de la branche d'activité actuellement couverte par la base de données de l'IGRF.

Tableau 13B — Comparaison du revenu total du RE et de l'IGRF (en millions de dollars)

Entreprise SCIAN	Description de la branche d'activité	Pourcentage devant figurer dans l'IGRF	Pourcentage figurant dans l'IGRF
812114	Salons de coiffure pour hommes	44 %	19 %
812115	Salons de beauté	70 %	35 %
812116	Salons de coiffure mixtes	65 %	38 %
812190	Autres services de soins personnels	70 %	51 %
812210	Salons funéraires	71 %	84 %
812220	Cimetières et crématoriums	34 %	54 %
812310	Blanchisseries et nettoyeurs à sec libre-service	77 %	55 %
812320	Services de nettoyage à sec et de blanchissage, sauf le libre service	79 %	54 %
812330	Fourniture de linge et d'uniformes	47 %	61 %
812910	Soins pour animaux de maison, sauf les services vétérinaires	67 %	18 %
812921	Laboratoires de développement et de tirage de photos, sauf le service en une heure	60 %	109 %
812922	Développement et tirage de photos en une heure	78 %	54 %
812930	Stationnements et garages	49 %	37 %
812990	Tous les autres services personnels	72 %	38 %
Autre		0 %	0 %
TOTAL		65 %	49 %

Il est particulièrement intéressant de constater que quelques branches d'activité de la partie B du tableau 13 ont déclaré des revenus dans l'IGRF représentant un pourcentage plus important que prévu du revenu total du RE. Les salons funéraires, les cimetières et crématoriums, la fourniture de linge et d'uniformes et les laboratoires de développement et de tirage de photos, sauf le service en une heure, en font partie. Cela est probablement dû au fait que les données du RE pour le revenu total sont beaucoup moins fiables et à jour que les chiffres déclarés dans l'IGRF.

Comme il a été mentionné précédemment, l'IGRF et le RE ne peuvent être comparés à partir du nombre total d'employés parce que cette variable ne figure habituellement pas dans les états des résultats et qu'elle n'est pas conséquent pas disponible dans l'IGRF. De même, il n'est pas possible de comparer le nombre d'établissements du RE avec les renseignements similaires de l'IGRF, parce que dans ce dernier cas, on se base sur les entités légales plutôt que sur les établissements, et que la mise en correspondance entre les établissements et les entités légales est multivoque.

Solutions possibles pour la comparaison des données de l'IGRF et de celles du RE

Il existe une autre stratégie possible, à savoir obtenir le NE pour chaque entreprise du sous-secteur des services personnels et services de blanchissage, à la fois à partir du RE et de l'IGRF. Grâce à un couplage au moyen du NE, l'utilisateur pourrait comparer le revenu du RE à celui figurant dans la base de données de l'IGRF, entreprise par entreprise. En raison de contraintes de temps, ce type d'analyse n'a pas été effectuée pour le présent document.

Problèmes découlant de l'utilisation des données de l'IGRF relativement aux services personnels et services de blanchissage

L'auteur a fait face à un problème particulier au moment de l'étude des données de l'IGRF pour le sous-secteur des services personnels et services de blanchissage, du point de vue du traitement informatique. Il existe de nombreuses variables et observations que l'utilisateur peut souhaiter consulter. Selon la taille de la branche d'activité ou du sous-secteur, il se peut qu'un ordinateur personnel n'ait pas la capacité suffisante pour manipuler des sommes importantes de données. Dans certains cas, il est préférable d'exécuter les procédures de la base de données de l'IGRF au moyen d'un ordinateur plus puissant. Si l'utilisateur examine un sous-secteur plus important que celui des services personnels et services de blanchissage, il pourrait envisager la possibilité de le faire indépendamment du serveur. Même avec un sous-secteur de la taille des services personnels et services de blanchissage, il peut falloir beaucoup de temps pour exécuter un code.

La compréhension des métadonnées est probablement le défi le plus important que les utilisateurs doivent relever lorsqu'ils utilisent la base de données de l'IGRF. Il est difficile de comprendre les variables qui sont nécessaires et la façon de récupérer les données correctes. L'utilisateur doit s'assurer qu'il comprend chacune des variables et la façon dont elle est calculée. En outre, il doit être conscient de la différence qui existe entre une variable, un bloc-total et une variable générique. La base de données Oracle de l'IGRF est *relationnelle*. Elle comporte de nombreux tableaux et de nombreux liens, ce qui peut la rendre plus complexe et difficile. En l'absence d'une bonne compréhension, on peut utiliser les mauvaises données par inadvertance. Il est aussi important de connaître les diverses phases du traitement des données, afin de s'assurer que les enregistrements extraits sont appropriés. L'utilisateur doit être prudent lorsqu'il choisit le nom d'une variable nominale pour chaque tableau. Par exemple, « is » et « in » pourraient sembler de bonnes désignations abrégées pour le tableau de l'état des résultats, mais il s'agit en fait de noms de fonctions qui, s'ils sont utilisés, entraîneront une mauvaise exécution du code, ce qui pourrait donner lieu à des erreurs ou à la production de statistiques erronées. Si le logiciel utilisé pour accéder aux données n'a jamais été utilisé auparavant, il existe de toute évidence une courbe d'apprentissage. La programmation SQL peut se révéler particulièrement difficile à comprendre, et les résultats peuvent être difficiles à interpréter lorsque des erreurs se produisent.

De façon plus générale, la majorité des problèmes, tant mineurs que majeurs, découlent du fait que l'IGRF est un système très nouveau, tant pour Statistique Canada que pour l'ADRC. Il va sans dire que les utilisateurs qui s'en servent pour la première fois feront face à des problèmes. Lorsque la période de développement sera terminée, l'IGRF sera beaucoup plus facile à utiliser, et la documentation et le soutien technique seront plus considérables.

Conclusion

L'IGRF offre de nombreuses possibilités à Statistique Canada. Le recensement complet des états des résultats pour l'ensemble des entreprises constituées en société et de celles non constituées en société devrait améliorer considérablement la gamme et la qualité des statistiques produites par le Bureau. Non seulement le niveau de précision et de détail augmenteront-ils, tant au niveau de la province que de la branche d'activité, mais il existe aussi des possibilités importantes de réduire le fardeau de réponse. En outre, du fait que les données de l'IGRF commencent maintenant à arriver beaucoup plus rapidement, de nombreux responsables de programmes statistiques pourront produire leurs sorties dans un délai plus court. Évidemment, pour exploiter pleinement les possibilités que présente l'IGRF, Statistique Canada doit relever de nombreux défis. Il faut résoudre la question des entreprises complexes et de la calendrialisation, ainsi que

le problème des exercices multiples, ce qui ne sera pas facile. En outre, il faudra du temps pour élaborer une bonne stratégie de vérification et d'imputation. En raison de l'inévitable courbe d'apprentissage, il faudra un certain temps avant que les avantages de l'IGRF puissent être pleinement réalisés. Statistique Canada investit maintenant le temps et les efforts nécessaires pour relever ces défis, et l'avenir semble tout à fait prometteur.

Bibliographie

- Bissett, Peter D. *Utilisation des données fiscales pour la production des statistiques économiques provinciales*. Série technique du PASEP, numéro 42, octobre 1999.
- Farrall, Kim. *Changes to NAICS Canada Codes and Titles Between July 1997 and March 1998*. 1^{er} mars 2000. <http://stdsweb/naicstest/eng/changes.htm>.
- Hall, Karen. *La formule de répartition des recettes de la taxe de vente harmonisée*. Série technique du PASEP, numéro 24, août 1998.
- Mach, Lenka. *Tax Estimates Program (TEP)*. Septembre 1996.
- Mach, Lenka. *Tax Estimates Program (TEP): Sampling of T1 Businesses*. Mai 1997.
- Mach, Lenka. *Tax Estimates Program (TEP): Sampling of T2 Corporations*. Juin 1997.
- Agence des douanes et du revenu du Canada. *Impôt des sociétés*. 18 mai 1998. <http://www.cca-adrc.gc.ca/>.
- Agence des douanes et du revenu du Canada. *Impôt des sociétés - Système de traitement révisé*. 18 mai 1998. <http://www.cca-adrc.gc.ca/>.
- Agence des douanes et du revenu du Canada. *Index général des renseignements financiers*. Cours PRISE. Juillet 1997.
- Smith, Philip. *Stratégie générale de Statistique Canada relativement à la statistique des entreprises*. Série technique du PASEP, numéro 57, 25 mai 2000.
- Smith, Philip. *Stratégie de l'EUE en vue de combiner les données au niveau de l'entreprise, de l'établissement et de l'entité légale*. Série technique du PASEP, numéro 49, 29 novembre 1999.
- Statistique Canada. *ADMIN FILE*. 3 février 2000. http://tax/Tax_Component/Admin/Admin_index.htm
- Statistique Canada. *Enquête annuelle sur le secteur des services personnels*, 1997.
- Statistique Canada. *Edit and Imputation of GIFI TY98 Proposed Strategy (Draft)*. 25 mai 2000. F:/Laurie/Taxdata/Ty98/IGRFstrat2.4b.doc.
- Statistique Canada. *IGRF - Index général des renseignements financiers*. 10 janvier 2000. http://tax/Tax_Component/FAQ/GIFIFAQF.htm.
- Statistique Canada. *GIFI Analysis Team Meeting Minutes*. 11 janvier 2000 au 29 février 2000. <http://tax/english/Minutes.htm>.

- Statistique Canada. *GIFI Table Layouts*. 29 mars 2000. T:/External/Tax Dev External/IGRF Table Layouts.doc.
- Statistique Canada. *IGRF/CORTAX et services associés de tax*. 28 mars 2000.
http://tax/Tax_Component/GIFI_T2_CORTAX/Bulletins/GIFI_Bulletins.htm.
- Statistique Canada. *Liste des changements au manuel SCIAN Canada 1997 (version publiée) (De mars 1998 à novembre 1999)*. 1^{er} mars 2000.
http://dissemination/francais/Subjects/Standard/manualchanges_f.htm.
- Statistique Canada. *Mandate* 13 janvier 2000. http://iofdweb//iofd_english_web/mandate.htm .
- Statistique Canada. *SCIAN Canada*. Occasionnellement, Catalogue no. 12-501-XPF, mars 1998.
- Statistique Canada. *SCIAN Canada*. 29 février 2000.
http://dissemination/francais/Subjects/Standard/introo_f.htm.
- Statistique Canada. *Procédures à suivre pour obtenir l'accès aux données fiscales*. 15 mars 2000. <http://tax/french/accesfre.html>.
- Statistique Canada. *Processing Stages and Journal Tables*. 10 mars 2000.
http://tax/Tax_%20Microdata_%20System/English/Processing_%20Stages_%20and_%20Journal_%20Tables.htm.
- Statistique Canada. *Simple Entity Relationship Diagram for T2 Data*. 29 mars 2000.
T:/External/Tax Dev External/Simple IGRF Diagrams.doc.
- Statistique Canada. *T2_Balance_Sheet_1998*. 29 mars 2000.
http://tax/Tax%20Microdata%20System/French/METADATA_titlepage_F111.htm
- Statistique Canada. *T2_Income_STMT_1998*. 29 mars 2000.
http://tax/Tax%20Microdata%20System/French/METADATA_titlepage_F111.htm
- Statistique Canada. *Tax Data Division Program Report*. 31 mai 1999. http://tax/english/APR97-99_E.htm.
- Statistique Canada. *Programme unifié des statistiques sur les entreprises - Projet d'amélioration des statistiques économiques provinciales*. Série technique du PASEP, numéro 1, 5 mai 1997.

Série technique - Index

13 octobre, 2000

Dans le cadre du PASEP, on a réimprimé une série de documents techniques, principalement pour usage interne. Voici la liste des réimpressions disponibles. Pour obtenir des copies communiquez avec Bonnie Bercik au (613) 951-6790 ou Diane Proulx au (613) 951-7192, numéro de télécopieur (613) 951-0411 ou écrire à Statistique Canada, 13^e étage, Immeuble Jean Talon, Parc Tunney, Ottawa, Ontario, K1A 0T6

1. Programme unifié des statistiques sur les entreprises – Projet d'amélioration des statistiques économiques provinciales – le 5 mai 1997 – Programme de Gestionnaire du PASEP.
2. Cadre d'évaluation du PASEP – le 15 septembre 1997 – Philip Smith.
3. Rapport de l'Enquête unifiée sur les entreprises et les modalités de déclaration – Consultations auprès des entreprises – août 1997 – Guy Gellatly, Larry Murphy et Junior Smith.
4. Bilan de l'évolution du PASEP : Notes d'une séance d'information à l'intention des représentants fédéral et provinciaux des finances, qui a eu lieu à Halifax (Nouvelle-Écosse), le 12 mars 1997 – Philip Smith.
5. Aperçu du Projet d'amélioration des statistiques économiques provinciales – novembre 1997 – George Beelen, Francine Hardy et Don Royce.
6. Des bases de données pour la conception, la génération et le stockage des questionnaires-entreprises à Statistique Canada – le 5 novembre 1997 – Alana M. Boltwood.
7. La statistique des entreprises : sa raison d'être – janvier 1999 – Elise Mennie. (*Diffusion interne seulement*)
8. Bilan du PASEP 15 mois après son lancement – le 24 avril 1998 – Philip Smith.
9. Programme des gestionnaires des répondants clés (GRC) – Rapport annuel pour 1997-1998 – mai 1998 – Vicki Crompton.
10. Un cadre de planification de la collecte des données de l'Enquête unifiée sur les entreprises – le 28 octobre 1998 – Alana Boltwood.
11. Répercussions du financement du PASEP sur le programme et les réalisations de la Division des services en 1997-1998 – avril 1998 – Gordon Baldwin. (*Diffusion interne seulement*)
12. L'organisation et la structure décisionnelle du PASEP – le 17 août 1998 – Philip Smith. (*Diffusion interne seulement*)
13. Les buts principaux du PASEP – le 17 novembre 1997 – Philip Smith.
14. Terminologie et cadre de référence du questionnaire de l'Enquête unifiée sur les entreprises – Révisé en mars 1999 – Philip Smith.
15. Amélioration de la qualité des statistiques économiques provinciales et mesure des changements apportés – août 1998 – Philip Smith.
16. Rapport annuel 1997-1998 – Médiateur – Fardeau de réponse de la petite entreprise – juillet 1998 – Michael Issa. (*Diffusion interne seulement*)
17. Le processus décisionnel du PASEP le 1^{er} octobre 1998 – Philip Smith.
18. Groupe de travail sur la déclaration électronique des données (DED) – avril 1998 – George Andrusiak, Monique Gaudreau, Laurie Hill, Anne Ladouceur, Denis Leblanc, Mario Ménard, Guy Parent, Joe Wilkinson, Doug Zinnicker.
19. Trousse d'information sur le PASEP – octobre 1998 – Philip Smith.
20. L'EUE et les secteurs non commerciaux – le 17 septembre 1997 – Art Ridgeway.

21. Le guide de l'utilisateur du SASC – avril 1998 – Janet Howatson. (*Diffusion interne seulement*) Non disponible
22. Compte rendu de la collecte et la saisie de données DOI pour l'EUE de 1997 – le 3 septembre 1998 – Anne Ladouceur. (*Diffusion interne seulement*)
23. Contribution prévue de la DDS au PASEP, 1998-1999 – septembre 1998 – Shirley Dolan.
24. La formule de répartition des recettes de la taxe de vente harmonisée – août 1998 – Karen Hall. (*Diffusion interne seulement*)
25. Groupe de travail sur l'acquisition des données auprès des entreprises – le 22 juillet 1998 – François Maranda et Don Royce.
26. Rôles et responsabilités dans le cadre du Programme unifié des statistiques sur les entreprises – le 15 décembre 1998 – George Andrusiak, Richard Barnabé, Albert Meguerditchian, Ray Ryan et Philip Smith. (*Diffusion interne seulement*)
27. Document sur le Projet d'amélioration des statistiques économiques Provinciales de la conférence mixte de l'AISE/AISO – le 22 juillet 1998 – Don Royce.
28. Groupe de travail sur les relations avec les répondants – le 5 mars 1999 – Wayne Smith.
29. Enquête de suivi et d'analyse des réponses – mars 1999 – Kristen Underwood.
30. Dossier d'information sur le partage des données – mars 1999 – John Crysdale. (*Diffusion interne seulement*)
31. Analyse de cohérence – Étude de cas du programme des Gestionnaires des répondants clés – le 23 avril 1999 – Rachel Bemier et Julie Mandeville.
32. Évaluation des documents de soutien de la collecte utilisés durant l'Enquête unifiée sur les entreprises de 1997 – le 16 novembre 1998 – Yvele Paquette.
33. Trousse d'information sur les renoncations – mai 1999 – John Crysdale. (*Diffusion interne seulement*)
34. Plan du PASEP pour 1999-2000 – le 14 juin 1999 – Philip Smith. (*Diffusion interne seulement*)
35. Forum SCE + Mesures de suivi de la conférence – avril 1999 – Participants de la conférence de Cornwall.
36. Rapport du groupe d'étude des sources d'information sur les entreprises – mars 1999 – Vicki Crompton et Mark Marcogliese.
37. Rapport du Groupe de travail du secteur 5 sur l'amélioration de la dotation générale – août 1999 – Mel Jones.
38. Étude des questionnaires des enquêtes auprès des entreprises – juin 1999 – Jason Gilmore.
39. Échelle de complexité des questionnaires sur les entreprises – juin 1999 – Jason Gilmore.
40. Rapport sur l'avancement du PASEP – septembre 1999 – Philip Smith.
41. Seuils d'exclusion et méthodes particulières d'échantillonnage pour les enquêtes-entreprises – Stratégie de mise en oeuvre – septembre 1999 – L'Équipe de la stratégie de la mise en oeuvre.
42. Utilisation des données fiscales pour la production des statistiques économiques provinciales – octobre 1999 – Peter Bissett.
43. Enquête sur la qualité des données de 1996 – mars 1999 – Ed Bunko. (*Diffusion interne seulement*)



Projet d'amélioration des statistiques
économiques provinciales

Project to Improve Provincial Economic
Statistics

Aperçu de la base de données de l'IGRF 1998

An Overview of the 1998 GIFI Database

Série technique

Technical Series

Numéro 61

Number 61



Internet : www.statcan.ca
Intranet : <http://pasep>



Statistique
Canada

Statistics
Canada

Canada

STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA



1010312591

e. 2

