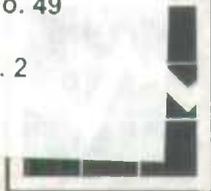


68N0003XPB

no. 49

c. 2



**Project to Improve Provincial
Economic Statistics**

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

**The UES Strategy to
Combine Enterprise,
Establishment and
Legal-Entity Data**

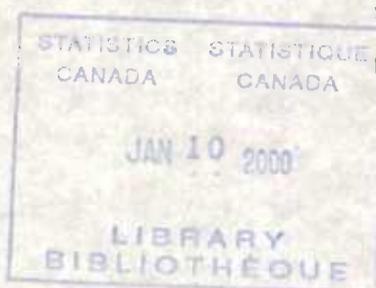
**Stratégie de l'EUE en vue
de combiner les données
au niveau de l'entreprise,
de l'établissement et de
l'entité légale**

Technical Series

Number 49

Série technique

Numéro 49



Internet: www.statcan.ca

Intranet: <http://pipes>



Statistics
Canada

Statistique
Canada

Canada

The UES Strategy to Combine Enterprise, Establishment and Legal Entity data

The Unified Enterprise Survey integrates survey and tax data, for simple and complex enterprises and their associated establishments and legal entities. The way in which this integration is accomplished has changed over the past 2½ years as the UES has developed, the Quarterly Financial Survey has been hooked to the business register and moved to the North American Industrial Classification System, and the old sample of corporate income tax (T2) records has been replaced by the General Index of Financial Information census. The purpose of this paper is to review the present character of this integration, focussing especially on the enterprise-level estimates and their relationship to tax data sources and the QFS.

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

The UES Strategy to Combine Enterprise, Establishment and Legal-Entity Data

Philip Smith
November 29, 1999

1. The Unified Enterprise Survey (UES) integrates survey and tax data, for simple and complex enterprises and their associated establishments and legal entities (LEs). The way in which this integration is accomplished has changed over the past 2½ years as the UES has developed, the Quarterly Financial Survey (QFS) has been hooked to the business register (BR) and moved to the North American Industrial Classification System (NAICS), and the old sample of corporate income tax (T2) records has been replaced by the General Index of Financial Information (GIFI) census.
2. The purpose of this paper is to review the present character of this integration, focussing especially on the enterprise-level estimates and their relationship to tax data sources and the QFS. Included in the paper are discussions of (i) the role of tax data in the overall strategy, (ii) shared responsibility for the consolidation and allocation steps, (iii) UES variables as defined in the Chart of Accounts, (iv) the issue of mapping between GIFI and the UES Chart of Accounts, (v) the GIFI edit and imputation strategy, (vi) enterprise data linkage issues, (vii) systems for effectively utilizing the GIFI database and (viii) the role of the UES "Part 1" survey and its evolutionary path since 1997. Special emphasis is given to the problem of how we will derive our estimates for key enterprise-level variables in the new post-GIFI environment. The paper concludes with a list of tasks emerging from the discussion that are now being undertaken.

UES enterprise- and establishment-level estimates

3. Our UES strategy is to collect information at both the *enterprise* level and the *establishment* level, and to check for possible inconsistencies between the two. Data reported for the individual establishments of an enterprise should cohere, or roll up properly, to the data we have obtained for the enterprise as a whole, after allowing for inter-establishment transactions and extraordinary financial statement items. Coherence of establishment and enterprise data can be readily verified by checking whether certain variables do, or do not, add up as they should. Variables such as value added, wages and salaries, or employment serve this purpose very well. Other variables, such as operating revenues or expenses, serve the purpose less well because of inter-establishment, intra-enterprise transactions. For variables of this latter type, the value reported for the enterprise as a whole can be smaller than the sum of the corresponding values for its establishments. Enterprise Statistics Division (ESD) does *coherence analysis* to check the consistency of the enterprise-level and establishment-level data.
4. *Ancillary units* are a special type of establishment for which most or all of the transactions are confined within the enterprise. In many cases the transactions are unrecorded, with no prices being explicitly charged. Head offices are perhaps the best-known example of ancillary units, typically providing a variety of "free" accounting, marketing, legal and management services to the other establishments of the enterprise. Warehouses and transportation units are other

common examples. There are about 3000 ancillary units on the BR. The services provided by ancillary units are not really free, of course, and the costs incurred by the units are fully reflected in the enterprise-level financial statements. In order to ensure that value added is properly allocated by industry and province, ESD imputes revenues to the ancillary units, and imputes equal and offsetting expenses to the establishments served.

5. Data are collected through our own surveys and from tax sources. The Industrial Organization and Finance Division (IOFD) and the Balance of Payments and Financial Flows Division (BPFFD) produce estimates from the *enterprise* data, while the industry-specializing subject matter divisions (Services, MCED, DTD, Transportation, Agriculture, SIEID and ICSD) produce estimates from the *establishment* data.

The role of tax data in our strategy

6. Income tax regulations do not allow corporations to file tax returns on a consolidated basis. Rather, filing must be done on a legal entity basis. Accordingly, tax data apply to individual *legal entities*.
7. Enterprises are directly associated with LEs, with each enterprise having one or more LEs. Similarly, enterprises are also directly associated with establishments, with each enterprise having one or more establishments. However, the relationship between LEs and establishments is many-to-many – in general, a single establishment is associated with one or more LEs, and a single LE is associated with one or more establishments. Certainly it is true that the vast majority of enterprises have only one establishment and only one LE. For these, the relationship is one-to-one and accordingly, for these enterprises the use of tax data is straightforward. However, the small minority of enterprises having more than one establishment and/or more than one LE account for about half of total business sector production in Canada. For these enterprises (there are approximately 10,000) the effective use of tax data is more challenging.
8. As just noted, an enterprise may have one, or more than one LE, but of the approximately 2 million enterprises currently listed on the business register (approximately 1 million of which are incorporated), only about 1500 have more than one LE. While few in number, many of these multi-LE enterprises ('multis') are very large and therefore very important. For these enterprises, we need to *consolidate* the tax data, which means assembling the tax records for all LEs in the enterprise and adding them together, while allowing for inter-LE transactions and extraordinary financial statement items. For the larger multi-LE enterprises, this consolidation step is accomplished by using enterprise-level *survey* data, collected via IOFD's Quarterly Financial Survey, in place of the tax data. For the smaller multi-LE enterprises, consolidation must be done in an approximate way by adding up the GIF1 financial statements, ignoring the possibility of inter-LE transactions and extraordinary financial statement items. For all the other enterprises (the remaining 2 million or so), the enterprise-level data are the same as the tax data. No consolidation is required since they each have only one LE.
9. Paragraph 3 talked about the process of rolling up establishment data to the enterprise level for purposes of coherence analysis checking. This rolling up process can be considered a form of consolidation, although the latter term is normally reserved for the similar but different problem of rolling up the financial statements of LEs to the enterprise level. Just as in rolling up establishments, so too in the consolidation of LEs some variables are more easily dealt with than others. Value added and wages and salaries can be consolidated by simply adding them up, whereas in the consolidation of revenues and expenses, one must deduct inter-LE transactions during the rolling-up process.
10. For purposes of our *enterprise* statistics programs in IOFD and BPFFD, the tax data meet most of our needs. The main problems we face are (1) how to consolidate the tax data

pertaining to the 1500 or so multi-LE enterprises, and (2) how to proceed when we need estimates for particular variables that are not directly available in the tax records. Generally speaking, the enterprise-level programs do *not* seek to produce estimates by industry or by province. Indeed, strictly speaking one *cannot* use enterprise-level data to produce estimates by industry and province, although in practice the two divisions do produce a few, high-level industry breakdowns by classifying enterprises, in a rather loose manner, to high-level industry categories. The point here is that since a single enterprise, by definition, can have establishments in more than one industry and/or province, data at the enterprise level cannot be directly aggregated by industry or province. A corollary is that since the relationship between LEs and establishments is many-to-many (see paragraph 7), one also cannot, in general, attach an industry code to an LE. Of course one *can* do so in most instances, since most enterprises have one LE and one establishment. However, one *cannot* do so, in principle, for the 10,000 or so enterprises that have more than one establishment.

11. In our *establishment* statistics programs (managed in the industry-specializing subject matter divisions: Services, MCED, DTD, Transportation, Agriculture, SIEID and ICSD) we also make use of the tax data. However, here the problems are somewhat different. First, since the primary goal is to produce estimates by province and by industry, we need a means of breaking down the enterprise-level data into separate province-industry elements. We do this by breaking each enterprise down into its component *establishments*. For *simple* enterprises, having all of their establishments in the same industry and province, we do not really need an establishment breakdown (unless we seek to produce sub-provincial tabulations) and we can use the tax data directly. Fortunately, over 99% of all enterprises are simple. However, for the relatively few (approximately 10,000, as noted) but frequently very large and important enterprises that are *complex*, having establishments in more than one industry and/or more than one province, we have a special problem. The tax data provide little information to assist in the industry/province breakdown. Occasionally it is possible to associate a particular LE with a particular establishment, or group of establishments, and in such instances the tax data may have direct use for a complex business even below the enterprise level. However, this is the exception rather than the rule and our statistical model does not normally exploit such possibilities. Thus, for complex enterprises we need survey information to break down the consolidated, enterprise-level data into establishment parts. According to the business register, there are approximately 70,000 establishments associated with the 10,000 complex enterprises.
12. The corporate tax supplementary form T2S-TC does provide limited provincial and territorial information for individual legal entities, although of course it provides no breakdowns by industry. Specifically, this particular tax form provides provincial data on (1) salaries and wages, (2) gross income attributable to each jurisdiction, (3) taxable income and (4) provincial or territorial taxes. These provincial distributions are valuable in some parts of our statistical program, such as the provincial allocation of corporation profits for purposes of the equalization formula and the Provincial Economic Accounts. They may also find use in deriving provincial allocations for industrial tabulations, although here they must be used more cautiously since they apply to legal entities rather than establishments. The T2S-TC form is mandatory for any corporation with a permanent place of business in more than one jurisdiction and must be filed regardless of whether there are income taxes payable. Statistics Canada will have access to a census of these forms via the CORTAX system starting in year 2000.
13. In the past, IOFD made use of a sample consisting of some 40,000 T2 tax records. These records came from three sources, which IOFD merged together in a single file. The first of these sources was IOFD's own sample, purchased from Canada Customs and Revenue Agency (CCRA) as a set of photocopied T2 forms, and transcribed and captured by IOFD's own employees. The second source was a sample purchased from CCRA by Business Register Division (BRD), again as a set of photocopied T2 forms, and transcribed and captured by BRD's Tax Record Access employees. This source was used by a number of

Statistics Canada programs, including those of the Small Business and Special Surveys Division and the Agriculture Division. The third and final source was another CCRA sample of T2 forms, in this case drawn for the Department of Finance, and transcribed and captured by CCRA staff. These three sources of T2 data no longer exist, as they have been replaced by GIFI. However, it is interesting to note that the problems we currently face with respect to GIFI are in some ways quite similar to those we used to face with respect to the third T2 source just referred to. The two files are or were transcribed and captured externally, using a set of variable definitions that do not conform perfectly to Statistics Canada requirements. In the past, IOFD coped with these problems through the use of concordance rules and a standardization process and we need to adopt a similar approach for the future.

Responsibility for consolidation and allocation

14. To be useful for purposes of our enterprise-level and establishment-level statistical programs, tax data must first be *consolidated* – for multi-LE enterprises, all of the LE financial statements must be rolled up, or aggregated, to a single enterprise financial statement – and then *allocated* – for complex enterprises, the enterprise financial statement must be broken down into separate establishment financial statements. (As noted in paragraph 11, we do not need to allocate the data for simple enterprises, even simple enterprises with more than one establishment, unless we are seeking to calculate estimates for sub-provincial areas. Nevertheless, we do allocate them, in order to form consistent micro-databases based on the establishment as the common statistical unit.)
15. It is important these two steps – consolidation and allocation – be done consistently across the business statistics program. ESD is responsible for the *allocation* step, although BRD, BSMD and the subject matter divisions are quite closely involved as well. ESD and IOFD are jointly responsible (see paragraph a) for coordinating the *consolidation* step, although this work must also closely involve TDD, BRD and BSMD to a very considerable extent (and other subject matter divisions to a lesser degree). IOFD has a very important role to play in the consolidation step since it directly monitors and interprets the relationships between LEs and enterprises for purposes of its Corporations Returns Act and QFS programs. These relationships change constantly as a results of births, deaths, mergers, acquisitions and divestitures.
16. The consolidation step entails creating and maintaining a list of the 1500 or so multi-LE enterprises, consistent with the UES Survey Universe File (SUF). Data from the QFS, for as many of these multi-LE enterprises as possible, must be merged with data from the GIFI database for the single-LE enterprises. In effect, we must create a synthetic GIFI statement for the enterprise as a whole using the QFS data, and we must substitute that synthetic statement inside a new “view” of the GIFI database, in place of the multiple GIFI statements that apply to the individual LEs of the enterprise. In addition, for the multi-LE enterprises not covered by the QFS, a cruder form of consolidation must be implemented by gathering together all the GIFI statements for the corresponding LEs and adding them up. Corporate mergers and divestitures and the creation of new legal entities are all very important in this context. This challenging task demands good collaboration among the five divisions primarily involved. Income tax schedules T2S(9) and T2013 include lists of related and associated corporations and these forms are of enormous potential help in consolidating legal entities.

Charts of accounts

17. IOFD has had a ‘Chart of Accounts’ (COA) for many years. This COA includes a list of approximately 1000 financial variables (primarily components of revenues, expenses, assets and liabilities), with associated definitions. The variables are arranged in a hierarchical list and each one is assigned a unique “SICA code”. Identities linking the variables together are also spelled out within the COA. IOFD has documented the COA in a thick loose-leaf binder.

18. GIFI is, in effect, an alternative COA. It includes a list of approximately 700 financial variables with associated definitions, arranged in a hierarchical list with uniquely assigned "GIFI codes." The GIFI COA is documented in TDD meta data documents and spreadsheets, and on the CCRA web site.
19. IOFD produces, and has produced for over thirty years, financial statistics covering the entire Canadian economy. It produces these statistics at the level of some 340 individual financial variables. These 340 variables are a subset within the IOFD COA and indeed, each of them can be expressed as an arithmetic function of other, lower-level variables within the COA. The 340 variables consist of (i) about 165 published variables that are derived from IOFD's quarterly survey and, in the past, from IOFD's annual sample of T2 financial statements, (ii) about 60 variables made available to the SNA Branch but not published, that are also derived from IOFD's quarterly survey and, in the past, from IOFD's annual sample of T2 financial statements, and (iii) about 120 additional variables that in the past came directly from the CCRA CORPAC file (via our so-called ADMIN file) or from the T2 sample record file, and starting with fiscal periods ending in year 2000 will come from the CCRA CORTAX file. IOFD's (and BSMD's) problem, in developing enterprise-level statistics, is therefore to find estimators for some $165+60=225$ variables, using information from the QFS and the GIFI.
20. This is not to say that 225 variables must be estimated for each industry. On the contrary, many of these variables apply to a single industry only. For example, "deposits" are an important form of liability for the banks, but they are not a liability for most other industries. Generally speaking, the statistical picture that IOFD develops for a particular industry consists of a small subset of 50-100 of the 225 variables. A similar statement applies to the GIFI Chart of Accounts: of the 700 or so variables in that COA, only 50-100 might apply to a particular business.
21. As noted, the UES seeks to collect consistent data at both the enterprise and the establishment levels. Coherence analysis is done by ESD in an effort to check and ensure consistency between the data for an enterprise and the data for its corresponding establishments (see paragraph 3). The 225 variables that have made up the core of IOFD's output in the past now also make up the core of the UES enterprise-level output. Coherence analysis is done in terms of these variables. Similarly, consolidation of the financial statements of individual LEs is done in terms of these same 225 variables. In effect, these variables, originating from the IOFD COA, have become the core of the new UES COA, which also includes further financial variables, such as expense item details, and a variety of other, primarily establishment-level variables such as sales by class of customer or by commodity class, as set out in the UES model questionnaire.
22. While the UES COA may be said to exist implicitly, it has not been explicitly defined and elaborated and there is a need for this to be done. To what extent should the UES COA mirror the old IOFD COA and to what extent should it be altered in light of GIFI? How should the establishment-level variables of the UES model questionnaire be brought into the UES COA? These questions are being examined and resolved by a working group including broad membership from amongst the various subject matter and infrastructure divisions involved in the UES (see paragraph b).

Mapping and imputing from GIFI to the COA

23. With the departure of IOFD's annual T2 sample starting in tax year 1999 and the arrival of GIFI in the 1998 reference year, we need a new means of estimating the 225 primary COA variables using GIFI. Previously, data were assembled directly by IOFD (and for 1997 and 1998, by TDD) transcription and capture staff, from the non-standard financial statements of the T2 sample to the standard variables of the COA. Now, the same T2 data are being transcribed by CCRA staff, not to the IOFD COA, but to the GIFI COA. The transcription is an

interim measure pending the start-up of mandatory GIFI filing by corporate tax-filers starting with fiscal periods ending in year 2000.

24. Our problem may be stated in the following terms. Given the set of 225 financial variables for which we desire estimates, and given the set of 700 (or so) GIFI variables for which, in the case of any given enterprise, TDD's GIFI databases may or may not have a value, what formulas should we use to calculate estimates for the 225 variables?
25. This problem can be thought of as a combination of two sub-problems, one a *mapping* problem and the other, an *imputation* problem. The first sub-problem might be stated this way: What is the one-to-many *conceptual reconciliation* between the UES COA definition of variable X (one of the 225) and the set of all 700 GIFI variables? The other sub-problem might be stated as: Given the conceptual reconciliation just referred to between UES COA variable X and the 700 variables in the GIFI COA, and given the practical problem that individual GIFI records do not, in general, provide details for each and every one of the 700 variables in the GIFI COA (and indeed, never will), how should the required, detailed GIFI variables be imputed, modelled or estimated so as to permit the calculation of variable X?
26. The *mapping* problem just referred to is a real one, but is not a particularly large or difficult one at the conceptual level. IOFD staff played an important role in the original design of GIFI and the two COAs are quite similar. Some progress has already been made in accomplishing this mapping, with leadership from TDD.
27. The *imputation* problem referred to here, on the other hand, is a more difficult one. In many instances the GIFI COA includes both several *detailed* categories, 'A₁', 'A₂' and 'A₃' say, and a more *generic*, or *block-total*, all-encompassing category, 'A' say. GIFI filers are encouraged to provide numbers for as many of A₁, A₂ and A₃ as possible, while lumping any remaining value into A. Whenever some value shows up in A, we statisticians have the problem of deciding how to distribute (impute) it over A₁, A₂ and A₃. Unless we make such a distribution, we cannot calculate the 225 variables in the UES COA. Fortunately, there are always some GIFI filers who *do* provide the required breakdown across A₁, A₂ and A₃. We therefore have the option of using their reported distribution to impute the distribution for others who lump some or all of their value into A. We can do this in one global operation applied to the entire GIFI database or, to do a better job, we can calculate the A₁, A₂ and A₃ shares and make the distributions separately by province-industry-size stratum. Unfortunately these imputation approaches, like any other imputation method, are imperfect and may, under some circumstances, generate obviously inappropriate results and/or balancing problems. They may also cause less obvious biases. We could conduct additional surveys and/or spend more effort analyzing the returns from previous surveys in order to assess and compensate for these biases, but such action is not currently in the plan. For the time being, we must proceed using the simple distribution method and taking steps to rebalance the records after imputation.

GIFI edit and imputation strategy

28. TDD obtains GIFI records from CCRA and compiles them in a raw view of the integrated GIFI database, a view that is "untouched by StatCan hands" (although TDD does pre-process the incoming data to remove any duplicates and merge multiple files). Let us call this view of the database GIFI-RAW.
29. The division then does a preliminary "cleaning" operation to the raw data, checking and correcting obviously wrong values. For example, it takes action if a reported value is negative but the variable definition does not allow values less than zero. It checks identity relationships and ensures they hold true. These various operations yield a second, clean view of the integrated database that is distinct from the first, raw view. Let us call this view GIFI-CLEAN.

30. TDD then produces a third view of the database containing edited and imputed records. This view includes synthesized records, generated by donor imputation or other means, to fill in for legal entities found on the UES Survey Universe File but not, as yet, otherwise represented in the GIFI database. Each record in this third view includes estimates for each of a number of designated *generic variables* or *block totals*, thereby permitting consistent aggregations for these variables across the database. Let us call this view of the database GIFI-E&I.
31. Currently, TDD has no plans to estimate details for A₁-, A₂- and A₃-type variables below the generic level (see paragraph 27) as part of its E&I strategy. When a GIFI filer reports such details, they will of course be included in all three views of the database. And when a donor record includes such details they will of course be propagated to any associated donor-imputed records. However, when a GIFI filer reports some or all of the value in a generic category A, the E&I strategy does not call for any distribution to be made (i.e. no imputed values of A₁, A₂, and A₃ will be calculated).
32. In addition, the current E&I strategy does not encompass *historical edits*. Of course, the 1998 GIFI record for a particular legal entity cannot, in general, be compared with a similar record for 1997 since the GIFI transcription process began in 1998. However, it is highly desirable that historical edits be introduced, where feasible, for the 1999 edition of GIFI. For example, we should check for each LE that the accumulated amount of depreciation reported for the end of the current period is equal to the amount reported for the end of the previous period plus the depreciation expense in the current period. However, since many financial details are reported only in the notes to the financial statements, and since only some and not all of the information in the notes is presently being transcribed to the GIFI format, it will not always be possible to reconcile two balance sheets for the same LE over consecutive accounting periods. Such reconciliations can be quite challenging, depending on the particular accounting concept involved. It might be possible to introduce a few, very limited, historical edits for the 1998 GIFI database, by making comparisons with records from the 1997 ADMIN database, but this has not as yet been established as a priority and would certainly imply additional costs and, presumably, delays in our schedule.
33. In cases where donor imputation is used to substitute for a missing GIFI record, historical continuity will be lost. This seems unlikely to cause significant problems as long as the donor imputation method is used only for relatively small LEs. For medium and large size LEs it is preferable to impute by applying cell-average growth rates to the previous year's record, thereby maintaining historical continuity.
34. Finally, while not currently in the plans it is desirable to produce a fourth view of the GIFI database. Starting from the GIFI-E&I view, this additional view would exclude all records that are associated with multi-LE enterprises and in their place would substitute one combined, enterprise-level record, developed in the manner discussed briefly in paragraph 16. Let us call this view GIFI-ENTERPRISE.

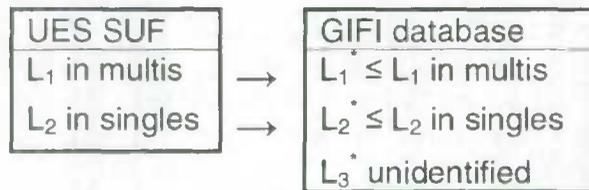
GIFI linkage issues

35. Each GIFI record should be linked to a corresponding enterprise on the BR, and conversely each legal entity associated with one of the enterprises on the UES SUF should be located within the GIFI database.
36. One particular issue that arises under this heading is that of LEs reporting multiple GIFI statements within the same year. This might happen if an LE was wound up or merged into another LE during the year. For example, an LE with its fiscal period ending June 30 might file a regular T2 form, accompanied by a GIFI statement, in August and might then be merged with another LE in October, leading to the filing of a second T2 and accompanying GIFI statement in December. If undetected, this could lead to double-counting of assets and liabilities. Occasionally a single LE will file two or more T2 forms within a given calendar year

and we need to take steps to avoid the double-counting. A simple interim approach could be to drop all records except the one with the longest reporting period. A more sophisticated method would be to merge the several records by adding the flow variables together from all records, taking the stock variables from the last record, and rebalancing the resulting merged record.

37. Another particular problem that can arise is the situation where one BN is linked to two or more enterprises on the BR. This might happen, for example, if the BN was associated with a joint venture. All such cases should be interpreted as erroneous, and should be corrected. For example, joint ventures should be recorded as separate enterprises on the BR.
38. The linkage issue might be pursued in the following way. Based on the UES SUF, we *expect* to find L (approximately 1 million) legal entities on the GIFI database. Of these, L_1 are legal entities within multi-LE enterprises and the remaining L_2 are LEs associated with single-LE enterprises ($L_1 + L_2 = L$). The GIFI records come rolling in steadily from CCRA as the year progresses and TDD compares them to the L records we are expecting to receive. At various points along the planning schedule, TDD will find only L_1^* of the expected L_1 legal entities, $L_1^* \leq L_1$, and only L_2^* of the expected L_2 legal entities, $L_2^* \leq L_2$. It will also find L_3^* unidentified and possibly new legal entities that were not present on the original UES SUF. The situation is depicted in Chart 1.

Chart 1
Legal Entities in the UES



TDD will forward the L_3^* unidentified LEs to the Business Register Division with minimal delay, so the register can be updated prior to the creation of the UES post-SUF. Meanwhile, at the appropriate time on the planning schedule TDD and BSMD will add one imputed record to the GIFI-E&I view of the database for each of the $(L_1 - L_1^*) + (L_2 - L_2^*)$ missing GIFI records. It must be presumed, unless other information has been received via survey feedback or other means, that these LEs continued to be active during the year despite the fact their GIFI record has not yet flowed through to Statistics Canada. It will be desirable to add imputed records in this manner at least twice per year – once fairly early on in the process (for example, after about 75% of records have been received) for purposes of calculating preliminary estimates, and again at a later stage after most records are in.

39. Of course, the BN linking problem has broader significance for a great many programs. The Income and Expenditure Accounts Division must link T4 records to the BR via the BN, to obtain industry codes for purposes of estimating labour income by industry, and it spends substantial resources each year coping with records for which there are no BR links. Similarly, the Labour Statistics Division (LSD) must link PD7 records to the BR, via the BN, in order to derive industry codes. Again, many records are unlinked and LSD invests substantial resources every year to cope with the problem. The GIFI linkage issue is but the latest manifestation of a common and increasingly worrisome problem. It stems in part from the fact that the business register tracks only the current BN and does not attempt to stock information about old, no-longer-valid BNs, and in part from the fact that there are about 200,000 enterprises on the register for which no NAICS codes have yet been assigned. A working group has been formed to address this issue (see paragraph c).

Estimating the UES COA variables

40. To meet our need to derive estimates for the 225 UES enterprise-level COA variables, two options are suggested.
41. The first option is to expand the E&I strategy (see paragraphs 30 and 31) in order to include, within the GIFI-E&I database, estimates for all detail variables of the type illustrated earlier by A_1 , A_2 and A_3 . This would allow us to derive our estimates of the 225 variables directly, using the conceptual mapping (see paragraphs 25 and 26), although we will, of course, have to finish our work to build that mapping. By providing richer and more consistent detail, this option would also greatly facilitate the use of the GIFI database in conjunction with other, establishment survey programs. However, this option might entail a considerable amount of additional work, thereby adding cost and delaying our effort to prepare the GIFI-E&I database within previously established deadlines.
42. A second option is to specify a set of 225 "magic formulas", in effect combining a solution to the mapping problem with a specially tailored solution to the imputation problem. For example, the magic formula for variable X (one of the 225) might be:

$$X = \text{if } (A_1 = 0) \text{ and } (A \neq 0) \text{ then } 0.25 * A, \text{ else } A_1$$

The magic formula option has the advantage that it requires effort only to specify the particular imputations needed to calculate the 225 variables. As such, it is presumably less costly than the first option. A major disadvantage is that it offers few if any spin-off benefits for other users of the database and risks inconsistency across our program, as others generate their own magic formulas for other purposes.

43. In the example of paragraph 42, the coefficient 0.25 is applied to A. This coefficient is intended to represent the estimated share of the aggregate A that is accounted for by the sub-aggregate A_1 in the case of an average enterprise. In the past, coefficients of this broad type were arrived at judgementally and applied on-the-fly by the expert IOFD (now TDD) T2 transcription and data capture staff, when they were transcribing from non-standard financial statements to standard forms based on the IOFD COA. We must make an effort to take advantage of the accumulated knowledge, wisdom and expertise of these staff as we derive our new imputation rules or magic formulas.
44. In sum, the need for an explicit plan to develop, for official release, enterprise- and establishment-level estimates from GIFI, with an associated assignment of responsibilities, is increasingly urgent. We need a short-term plan for UES-98 and a longer term plan for succeeding years. A working group has been established to tackle this problem, with IOFD and TDD co-leading the group (see paragraph d).

Systems for accessing, processing and tabulating GIFI

45. Now that the GIFI data have arrived, the challenge of making good use of them is at hand. The volume of GIFI data is enormous. There are over 800,000 records for the 1998 tax year already accumulated and several hundred thousand are still to come, each with many different variables. These records need and are receiving a considerable amount of editing. Missing records must be identified and imputed for, while duplicates must be removed. Some of these issues were alluded to in earlier sections (paragraphs 28-39 in particular).
46. TDD has invested considerable resources over the past several months to develop a first-class relational database to house the GIFI data. Subject matter analysts and methodologists are now drawing upon that database for a variety of purposes related to the estimation of UES outputs. As each analyst approaches the task, he or she must find or develop a set of software tools for purposes of querying the database, linking to survey data in the Central Data Store, processing the data in various ways and producing tabulations.

47. Both to improve efficiency and to foster greater consistency in our usage of the common data resources, it is desirable that users of the GIFI database be provided with a set of software application tools. These tools might well be adapted from ESD's existing U-Start system, which is already being used by subject matter analysts to access, process and tabulate data records in the Central Data Store. Shirley Dolan, chairperson of the UES Systems Architecture Working Group, has discussed the matter with the key players and is now seeking a resolution of the issue (as noted in paragraph e).

The role of the Part 1 survey – reference information

48. There has been much discussion and some misunderstanding of the UES "Part 1" survey over the past few years. The character and role of this survey have changed greatly since the original plan for the UES was drafted in December 1996. This concluding section of the paper recaps the history of the Part 1 survey and explains how and why the purpose of the survey has changed over time.
49. The first edition of the UES, for reference year 1997, included a Part 1 questionnaire sent to most of the large and complex enterprises, requesting data sufficient to generate the 225 variables referred to in paragraph 19. The Part 1 questionnaire was particularly important for purposes of estimating value added at the enterprise level – a key variable for coherence analysis and SNA purposes – since the variables required were not present on the old Quarterly Financial Survey questionnaire. An expanded T2 sample was also planned, although it did not materialize due to technical problems with the sampling job specification. This Part 1 survey provided the enterprise-level component of the UES for 1997, for complex enterprises in non-financial industries.
50. With the arrival of GIFI in 1998, the full Part 1 survey is no longer required. We are able to get all the annual enterprise-level data we need from the GIFI database, with one exception pertaining to multi-LE enterprises. A smaller edition of the Part 1 survey is still needed in order to acquire consolidated data for these few but important cases. Thus, for 1998 the enterprise-level estimates of UES will be derived by drawing upon GIFI and a slimmed down Part 1 survey.
51. For reference year 1999, no Part 1 survey whatsoever is needed. This is because the Quarterly Financial Survey has now been hooked to the business register, converted to NAICS and had its sample expanded to include all multi-LE enterprises of significant size. Starting in this year, the existing annual enterprise programs in IOFD, including those related to the Corporations Returns Act and the Federal-Provincial Fiscal Arrangements formula, will come from the fully integrated UES-QFS-GIFI program.

Concluding remarks

52. This paper is an attempt to summarize in very broad terms the manner in which enterprise-level data are being collected and processed and enterprise-level aggregates are being estimated under the UES. It is hoped the paper will help in reducing misunderstandings and fostering solutions to the numerous challenging problems we face on this front.
53. Discussion of earlier versions of this paper led to the creation of five working groups to address particular issues.
- a. Divisions (chiefly BRD, BSMD, ESD, IOFD and TDD) must prepare a plan to ensure that the GIFI and other tax data for the LEs of the approximately 1500 multi-LE enterprises are properly consolidated. We need clear procedures and responsibilities if we are to generate an enterprise-level database view from the LE-level information that comes to

us via GIFI and other administrative data sources. ESD and IOFD are co-leading a working group with members from BRD, BSMD and TDD to resolve the issue.

- b. We need a more clearly defined UES Chart of Accounts that embraces in some manner the old IOFD Chart of Accounts, the UES model questionnaire and the GIFI Chart of Accounts, and that relates directly to the System of National Accounts. ESD is coordinating a working group with members from a selection of subject matter, infrastructure (Standards Division, in particular) and SNA divisions to define and implement the new UES COA.
 - c. We need to review the status of the BR linkage problem and to seek solutions that will economize on the use of scarce resources in dealing with this unavoidable issue. The incidence in administrative files of BNs that are missing from the BR, and of multiple BNs linked to the same enterprise in the BR, is far too great and must be reduced. André Monty and the author of this paper are co-leading a working group with members from a selection of subject matter and infrastructure divisions to develop a plan to deal "once and for all" with the linkage problem.
 - d. With the processing and estimation stages of UES-98 about to begin, we need a plan for using the GIFI data to produce enterprise- and establishment-level estimates for a variety of financial variables. This plan might use "magic formulas" based on generic variables (block totals), or it might work off a more detailed set of variables derived through a more intensive GIFI imputation process. TDD and IOFD are co-leading a working group with members from a selection of subject matter and infrastructure divisions to decide how and by whom the tax-based estimates will be produced.
 - e. Finally, we need common software, possibly adapted from ESD's existing U-Start system, to help subject matter analysts access, process and tabulate data from TDD's new relational GIFI database, and to link the tax information to survey records in the Central Data Store. Shirley Dolan, chairperson of the UES Systems Architecture Working Group, has discussed the matter with the key players and is now seeking a resolution of the issue at the SAWG.
54. Finally, the author wishes to thank the numerous individuals who corrected errors and suggested valuable improvements to the earlier drafts of this paper.

Technical Series - Index

December 23, 1999

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)*

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. 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 Crysdale. *(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 Crysdale. *(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 – 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.

Stratégie de l'EUE en vue de combiner les données au niveau de l'entreprise, de l'établissement et de l'entité légale

L'Enquête unifiée auprès des entreprises intègre des données d'enquête et des données fiscales, pour les entreprises simples et complexes et leurs établissements et entités légales. La façon dont cette intégration est assurée s'est modifiée au cours des deux dernières années et demie, au fur et à mesure de l'évolution de l'EUE, l'Enquête trimestrielle sur les états financiers ayant été intégrée au Registre des entreprises et étant passée au Système de classification des industries de l'Amérique du Nord, et l'ancien échantillon de dossiers de l'impôt sur le revenu des sociétés (T2) ayant été remplacé par le recensement de l'Index général des renseignements financiers. Le présent document vise à passer en revue le caractère actuel de cette intégration, en mettant l'accent particulièrement sur les estimations au niveau de l'entreprise et leur lien avec les sources de données fiscales et l'ETEF.

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

Stratégie de l'EUE en vue de combiner les données au niveau de l'entreprise, de l'établissement et de l'entité légale

Philip Smith
Le 29 novembre 1999

1. L'Enquête unifiée auprès des entreprises (EUE) intègre des données d'enquête et des données *fiscales*, pour les *entreprises simples et complexes* et leurs *établissements et entités légales* (EL). La façon dont cette intégration est assurée s'est modifiée au cours des deux dernières années et demie, au fur et à mesure de l'évolution de l'EUE, l'Enquête trimestrielle sur les états financiers (ETEF) ayant été intégrée au Registre des entreprises (RE) et étant passée au Système de classification des industries de l'Amérique du Nord (SCIAN), et l'ancien échantillon de dossiers de l'impôt sur le revenu des sociétés (T2) ayant été remplacé par le recensement de l'Index général des renseignements financiers (IGRF).
2. Le présent document vise à passer en revue le caractère actuel de cette intégration, en mettant l'accent particulièrement sur les estimations au niveau de l'entreprise et leur lien avec les sources de données fiscales et l'ETEF. Le présent document aborde : i) le rôle des données fiscales à l'intérieur de la stratégie globale; ii) la responsabilité partagée pour les étapes de la consolidation et de la répartition; iii) les variables de l'EUE définies dans le plan comptable; iv) la question de la mise en correspondance de l'IGRF et du plan comptable de l'EUE; v) la stratégie de vérification et d'imputation de l'IGRF; vi) les questions de couplage des données sur les entreprises; vii) les systèmes permettant l'utilisation efficace de la base de données de l'IGRF; viii) le rôle de l'enquête de la « Partie 1 » de l'EUE et son évolution depuis 1997. Un accent particulier est mis sur le problème de calcul de nos estimations pour les variables clés au niveau de l'entreprise, dans le nouveau contexte postérieur à l'IGRF. Le document conclut avec une liste des tâches qui ont été entreprises, par suite de discussions.

Estimations au niveau de l'entreprise et de l'établissement de l'EUE

3. Notre stratégie de l'EUE consiste à recueillir des données, tant au niveau de l'*entreprise* qu'au niveau de l'*établissement*, ainsi qu'à vérifier les incohérences possibles entre les deux. Les données déclarées pour les divers établissements d'une entreprise devraient coïncider, ou pouvoir être cumulées de façon appropriée, en vue de correspondre aux données que nous avons obtenues pour l'entreprise dans son ensemble, après avoir pris en compte les transactions entre les établissements et les postes extraordinaires des états financiers. La cohérence des données sur les établissements et sur l'entreprise peut facilement être vérifiée. À cette fin, on doit déterminer si le total de certaines variables correspond ou non à ce qu'il devrait être. Des variables comme la valeur ajoutée, les salaires et les traitements ou l'emploi servent très bien cet objectif. D'autres variables, comme les revenus ou les dépenses d'exploitation, répondent moins bien à ce besoin, en

raison des transactions entre les établissements et à l'intérieur de l'entreprise. Pour ces dernières variables, la valeur déclarée pour l'ensemble de l'entreprise peut être inférieure à la somme des valeurs correspondantes pour les divers établissements. La Division de la statistique des entreprises (DSE) procède à une *analyse de la cohérence*, afin de vérifier l'uniformité des données au niveau de l'entreprise et de celles au niveau de l'établissement.

4. Les *unités auxiliaires* sont un type spécial d'établissement, dont la majorité ou la totalité des transactions est effectuée à l'intérieur de l'entreprise. Dans nombre de cas, ces transactions ne sont pas enregistrées, aucun coût n'étant imputé explicitement. Les sièges sociaux sont peut-être l'exemple le mieux connu des unités auxiliaires, et ils fournissent généralement une gamme variée de services « gratuits » de comptabilité, de marketing et de gestion, ainsi que de services juridiques, aux autres établissements de l'entreprise. Les entrepôts et les services de transport sont d'autres exemples courants. Il existe environ 3 000 unités auxiliaires dans le RE. Il va sans dire que les services fournis par les unités auxiliaires ne sont pas réellement gratuits, et les coûts que ces unités prennent en charge sont comptabilisés entièrement dans les états financiers de l'entreprise. Afin de s'assurer que la valeur ajoutée est répartie de façon appropriée selon l'industrie et la province, la DSE impute des revenus aux unités auxiliaires et des dépenses équivalentes et compensatoires aux établissements servis.
5. Les données sont recueillies grâce à nos propres enquêtes et aux sources de données fiscales. La Division de l'organisation et des finances de l'industrie (DOFI) et la Division de la balance des paiements et des flux financiers (DBPFF) produisent des estimations à partir des données sur l'*entreprise*, tandis que les divisions spécialisées selon l'industrie (Services, DFCE, DSC, Transport, Agriculture, DSIIE et DISC) produisent des estimations à partir des données sur les *établissements*.

Rôle des données fiscales à l'égard de notre stratégie

6. Les règlements de l'impôt sur le revenu ne permettent pas aux sociétés de soumettre des déclarations sur une base consolidée. La déclaration doit plutôt se faire sur la base de l'entité légale. Les données fiscales se rapportent donc à des *entités légales* individuelles.
7. Les entreprises sont directement liées aux EL, chaque entreprise en comptant une ou plusieurs. De même, les entreprises sont aussi liées directement aux établissements, chacune en comptant un ou plusieurs. Toutefois, les EL et les établissements ont un rapport de plusieurs à plusieurs ou multivoque, un seul établissement étant de façon générale lié à une ou plusieurs EL, et une seule EL étant liée à un ou plusieurs établissements. Il est vrai que la grande majorité des entreprises ne compte qu'un établissement et qu'une EL. Dans ces cas, le rapport est de un à un ou biunivoque. Conséquemment, pour ces entreprises, le recours aux données fiscales est simple. Toutefois, une petite minorité d'entreprises qui comptent plus d'un établissement ou plus d'une EL représentent environ la moitié de la production totale du secteur des entreprises au Canada. Dans ces cas (environ 10 000 entreprises), le recours efficace aux données fiscales est plus problématique.
8. Comme nous venons de le souligner, une entreprise peut compter une EL ou plus d'une EL, mais parmi les quelque deux millions d'entreprises figurant actuellement dans le Registre des entreprises (dont environ 1 million sont constituées en société), seulement 1 500 environ ont plus d'une EL. Nombre de ces entreprises comportant plusieurs EL sont très importantes, même si leur nombre est restreint. Dans ces cas, nous devons *regrouper* les données fiscales, ce qui signifie réunir les dossiers fiscaux de toutes les EL de l'entreprise et les additionner, tout en tenant compte des transactions entre les EL et des postes extraordinaires des états financiers. Pour les grandes entreprises comportant plusieurs EL, cette étape de consolidation est effectuée au moyen des données d'*enquête*

au niveau de l'entreprise, lesquelles sont recueillies grâce à l'Enquête trimestrielle sur les états financiers de la DOFI, plutôt qu'au moyen des données fiscales. Pour les petites entreprises comptant plusieurs EL, la consolidation doit être effectuée de façon approximative, grâce à la totalisation des états financiers de l'IGRF, sans tenir compte de la possibilité de transactions entre les EL, ni des postes extraordinaires des états financiers. Pour toutes les autres entreprises (les 2 millions environ qui restent), les données au niveau de l'entreprise sont les mêmes que les données fiscales. Aucune consolidation n'est nécessaire, étant donné que chacune ne compte qu'une EL.

9. Le paragraphe 3 aborde le processus de cumul au niveau de l'entreprise des données sur les établissements, aux fins de l'analyse de la cohérence. Ce processus cumulatif peut être considéré comme une forme de consolidation, même si ce terme est généralement réservé au problème similaire, mais différent, du cumul des états financiers des EL au niveau de l'entreprise. Comme c'est le cas pour le cumul des établissements, lorsque l'on procède à la consolidation des EL, certaines variables sont plus faciles à traiter que d'autres. On peut regrouper simplement la valeur ajoutée et les salaires et traitements en en faisant la somme, mais lorsque l'on cumule les revenus et les dépenses, on doit déduire les transactions entre les EL.
10. Aux fins de nos programmes de statistiques sur les *entreprises*, à la DOFI et à la DBPFF, les données fiscales répondent à la majorité de nos besoins. Les principaux problèmes auxquels nous faisons face sont les suivants : 1) comment regrouper les données fiscales relatives aux quelque 1 500 entreprises comportant plusieurs EL; et 2) comment procéder lorsque nous avons besoin d'estimations pour des variables particulières, qui ne sont pas disponibles directement dans les dossiers fiscaux. De façon générale, les programmes au niveau de l'entreprise ne visent *pas* à produire des estimations selon l'industrie ou la province. En fait, à strictement parler, on *ne peut* utiliser les données au niveau de l'entreprise pour produire des estimations selon l'industrie et la province, même si en pratique les deux divisions compétentes produisent quelques ventilations de haut niveau selon l'industrie, grâce à la classification des entreprises, de façon plutôt vague, dans des catégories d'industries de haut niveau. Étant donné qu'une entreprise simple peut, par définition, comporter des établissements dans plus d'une industrie ou plus d'une province, les données au niveau de l'entreprise ne peuvent être agrégées directement selon l'industrie ou la province. Cela a pour corollaire qu'on ne peut pas non plus, de façon générale, attribuer un code d'industrie à une EL, étant donné que le rapport entre les EL et les établissements est multivoque (voir le paragraphe 7). Évidemment, on *peut*, dans la plupart des cas, le faire, étant donné que la plupart des entreprises ne comptent qu'une EL et qu'un établissement. Toutefois, on *ne peut* le faire, en principe, pour les quelque 10 000 entreprises qui comptent plus d'un établissement.
11. Dans le cadre de nos programmes de statistiques sur les *établissements* (qui est géré par les divisions spécialisées : Services, DSCE, DSC, Transport, Agriculture, DSIE et DISC), nous utilisons aussi les données fiscales. Toutefois, dans ce cas, le problème diffère quelque peu. Tout d'abord, étant donné que l'objectif premier consiste à produire des estimations selon la province et l'industrie, nous avons besoin d'un moyen de ventiler les données au niveau de l'entreprise en éléments distincts province-industrie. À cette fin, nous ventilons chaque entreprise selon les *établissements* qui la composent. Dans le cas des entreprises *simples*, dont tous les établissements se trouvent dans la même industrie et dans la même province, nous n'avons pas réellement besoin d'une ventilation selon l'établissement (à moins de vouloir produire des totalisations infraprovinciales), et nous pouvons utiliser les données fiscales directement. Heureusement, plus de 99 % de toutes les entreprises sont simples. Toutefois, pour les entreprises *complexes*, qui sont souvent très grandes et importantes, mais relativement peu nombreuses (environ 10 000, comme il est indiqué), et qui comptent des établissements dans plus d'une industrie ou dans plus d'une province, le problème qui se pose est particulier. Les données fiscales comprennent peu de renseignements permettant d'effectuer la ventilation selon l'industrie/la province. À

l'occasion, il est possible de lier une EL particulière à un établissement particulier, ou à un groupe d'établissements, et dans ces cas, les données fiscales peuvent être utilisées directement pour une entreprise complexe, même à un niveau inférieur à celui de l'entreprise. Toutefois, il s'agit d'une exception plutôt que d'une règle, et notre modèle statistique ne tient pas compte généralement de telles possibilités. Ainsi, dans le cas des entreprises complexes, nous avons besoin de données d'enquête pour ventiler les données consolidées au niveau de l'entreprise selon les établissements. On compte environ 70 000 établissements liés à 10 000 entreprises complexes dans le Registre des entreprises.

12. Le formulaire supplémentaire de l'impôt des sociétés T2S-TC comprend des données limitées au niveau provincial et territorial pour des entités légales, mais ne comporte évidemment pas de ventilations selon l'industrie. Plus précisément, ce formulaire particulier fournit des données provinciales sur : 1) les salaires et les traitements, 2) le revenu brut de chaque secteur de compétence, 3) le revenu imposable, et 4) les impôts provinciaux ou territoriaux. Ces répartitions provinciales sont valables pour certains aspects de notre programme statistique, par exemple, la répartition provinciale des bénéficiaires des entreprises, aux fins de la formule de péréquation et des Comptes économiques provinciaux. Elles peuvent aussi être utilisées pour calculer les répartitions provinciales des totalisations selon l'industrie, avec davantage de précautions dans ce cas, toutefois, étant donné qu'elles s'appliquent à des entités légales plutôt qu'à des établissements. Le formulaire T2S-TC est obligatoire pour toute société qui a une place d'affaires permanente dans plus d'un secteur de compétence, et il doit être produit peu importe que des impôts doivent être versés ou non. Statistique Canada aura accès à un recensement de ces formulaires au moyen du système CORTAX à partir de l'an 2000.
13. Par le passé, la DOFI utilisait un échantillon comprenant quelque 40 000 formulaires T2. Ces enregistrements provenaient de trois sources, que la DOFI a fusionnées en un seul fichier. La première de ces trois sources était le propre échantillon de la DOFI, acquis auprès de l'Agence des douanes et du revenu du Canada (ADRC), sous forme d'un ensemble de formulaires T2 photocopiés, transcrits ainsi que saisis par les employés de la DOFI. La deuxième source était un échantillon acquis auprès de l'ADRC par la Division du Registre des entreprises (DRE), encore une fois sous la forme d'un ensemble de formulaires T2 photocopiés et transcrits ainsi que saisis par les employés de l'Accès aux dossiers fiscaux de la DRE. Cette source a été utilisée pour un certain nombre de programmes de Statistique Canada, y compris ceux de la Division des petites entreprises et des enquêtes spéciales et de la Division de l'agriculture. La troisième et dernière source est un autre échantillon de l'ADRC de formulaires T2, dans ce cas obtenus du ministère des Finances, et transcrits ainsi que saisis par le personnel de l'Agence. Ces trois sources de données T2 n'existent plus, ayant été remplacées par l'IGRF. Toutefois, il est intéressant de noter que les problèmes auxquels nous faisons face actuellement à l'égard de l'IGRF sont, dans une certaine mesure, assez similaires à ceux auxquels nous étions confrontés relativement à la troisième source T2 dont nous venons de parler. Les deux fichiers sont ou ont été transcrits et saisis au niveau externe, à partir d'un ensemble de définitions de variables qui ne correspondent pas parfaitement aux exigences de Statistique Canada. Par le passé, la DOFI a résolu ces problèmes grâce à l'application de règles de concordance et d'un processus d'uniformisation, et nous devons adopter une approche similaire pour l'avenir.

Responsabilités relatives à la consolidation et à la répartition

14. Pour que les données fiscales soient utiles aux fins de nos programmes statistiques au niveau de l'entreprise et de l'établissement, elles doivent être tout d'abord *consolidées* — pour les entreprises comportant plusieurs EL, tous les états financiers de ces EL doivent être cumulés, ou agrégés, en un seul état financier pour l'entreprise — puis *répartis* — pour les entreprises complexes, l'état financier de l'entreprise doit être ventilé en états financiers

distincts pour chaque établissement. (Comme il est indiqué au paragraphe 11, nous n'avons pas besoin de répartir les données pour les entreprises simples, même celles qui comptent plus d'un établissement, à moins de vouloir calculer des estimations pour des régions infraprovinciales. Néanmoins, nous les répartissons, afin de constituer des bases de microdonnées cohérentes, l'établissement servant d'unité statistique commune.)

15. Il est important que ces deux étapes — consolidation et répartition — soient appliquées de façon uniforme dans l'ensemble du programme de la statistique des entreprises. La DSE est responsable de l'étape de la *répartition*, même si la DRE, la DMEE et les autres divisions spécialisées interviennent aussi de très près. La DSE et la DOFI sont responsables conjointement (voir le paragraphe 53.a.) de la coordination de l'étape de la *consolidation*, même si ces travaux doivent faire intervenir étroitement et dans une très large mesure la DDF, la DRE et la DMEE (et les autres divisions spécialisées, dans une moindre mesure). La DOFI a un rôle très important à jouer relativement à l'étape de la consolidation, étant donné qu'elle contrôle directement et interprète les liens entre les EL et les entreprises, aux fins de ses programmes CALURA et ETEF. Ces liens évoluent constamment, par suite de la création et de la disparition d'entreprises, ainsi que des fusions, acquisitions et dessaisissements.
16. L'étape de la consolidation englobe la création et le maintien d'une liste des quelque 1 500 entreprises comportant plusieurs EL, en conformité du Fichier de l'univers d'enquête (FUE) de l'EUE. Les données de l'ETEF, au sujet du plus grand nombre possible de ces entreprises comptant plusieurs EL, doivent être fusionnées avec celles de la base de données de l'IGRF pour les entreprises ne comportant qu'une EL. En effet, nous devons créer un état synthétisé IGRF pour l'entreprise dans son ensemble, à partir des données de l'ETEF, et nous devons substituer cet état synthétisé, à l'intérieur d'une nouvelle « version » de la base de données de l'IGRF, aux états multiples de l'IGRF qui s'appliquent aux diverses EL de l'entreprise. Par ailleurs, dans le cas des entreprises comptant plusieurs EL qui ne sont pas assujetties à l'ETEF, une forme plus brute de consolidation doit être appliquée, soit la consolidation de tous les états de l'IGRF pour les EL correspondantes, et leur somme. Les fusions et les dessaisissements d'entreprises ainsi que la création de nouvelles entités légales, sont tous très importants dans ce contexte. Cette tâche nécessite une bonne collaboration des cinq divisions principalement concernées. Les annexes T2S (9) et T2013 comprennent des listes des sociétés liées et associées, et elles constituent une source d'aide considérable pour la consolidation des entités légales.

Plans comptables

17. La DOFI a un « plan comptable » depuis de nombreuses années. Ce plan comprend une liste d'environ 1 000 variables financières (principalement des composantes du revenu, des dépenses, de l'actif et du passif), ainsi que les définitions s'y rapportant. Ces variables sont organisées à l'intérieur d'une liste hiérarchique, et chacune comporte un code « SICA » unique. Les similitudes entre les variables sont aussi énoncées dans le plan comptable. La DOFI a documenté ce plan dans un épais cahier à relier.
18. L'IGRF représente en fait un substitut du plan comptable. Il comprend une liste d'environ 700 variables financières, ainsi que les définitions s'y rapportant, organisées dans une liste hiérarchique comportant des « codes IGRF » uniques. Le plan comptable de l'IGRF est documenté dans les documents et les tableurs électroniques de métadonnées de la DDF, ainsi que dans le site Internet de l'ADRC.
19. La DOFI produit, et ce depuis plus de 30 ans, des statistiques financières couvrant l'ensemble de l'économie canadienne. Ces statistiques sont produites au niveau des quelque 340 variables financières qui existent. Ces 340 variables constituent un

sous-ensemble à l'intérieur du plan comptable de la DOFI et, en fait, chacune d'entre elles peut être exprimée comme une fonction arithmétique d'autres variables de niveau inférieur du plan comptable. Les 340 variables comprennent : i) environ 165 variables publiées qui sont calculées à partir de l'enquête trimestrielle de la DOFI et, par le passé, de l'échantillon annuel des états financiers T2 de la DOFI; ii) environ 60 variables qui sont mises à la disposition de la Direction du Système de comptabilité nationale, mais qui ne sont pas publiées, et qui sont aussi calculées à partir de l'enquête trimestrielle de la DOFI et, par le passé, à partir de l'échantillon annuel d'états financiers T2 de la DOFI; et iii) environ 120 variables additionnelles qui, par le passé, provenaient directement du fichier CORPAC de l'ADRC (par l'entremise de notre fichier ADMIN), ou du fichier des enregistrements de l'échantillon T2. Pour les exercices qui prendront fin en l'an 2000, les données proviendront du fichier CORTAX de l'ADRC. Le problème de la DOFI (et de la DMEE), du point de vue de l'élaboration de statistiques au niveau de l'entreprise, consiste par conséquent à trouver des estimateurs pour $165 + 60 = 225$ variables, à partir des renseignements de l'ETEF et de l'IGRF.

20. Cela ne veut pas dire que les 225 variables doivent être estimées pour chaque industrie. Au contraire, nombre de ces variables s'appliquent à une seule industrie. Par exemple, les « dépôts » constituent une forme importante de passif pour les banques, mais pas pour la plupart des autres industries. De façon générale, le portrait statistique que la DOFI dresse pour une industrie particulière englobe un petit sous-ensemble de 50 à 100 des 225 variables. Il en va de même pour le plan comptable de l'IGRF : des quelques 700 variables qu'il comprend, seulement 50 à 100 s'appliquent à une entreprise particulière.
21. Comme il a été mentionné, l'EUE vise à recueillir des données cohérentes, tant au niveau de l'entreprise qu'à celui de l'établissement. L'analyse de la cohérence est effectuée par la DSE, en vue de vérifier les données et d'en assurer la cohérence pour une entreprise et les établissements correspondants (voir le paragraphe 3). Les 225 variables qui ont servi de base aux produits de la DOFI par le passé constituent aussi maintenant la base des produits au niveau de l'entreprise de l'EUE. L'analyse de la cohérence est effectuée à l'égard de ces variables. De même, la consolidation des états financiers des EL est effectué à l'égard des mêmes 225 variables. En effet, ces variables, qui proviennent du plan comptable de la DOFI, constituent la base du nouveau plan comptable de l'EUE, qui comprend aussi d'autres variables financières, par exemple, des détails sur les postes de dépenses, et diverses autres variables au niveau de l'établissement, par exemple, les ventes selon la catégorie de clientèle ou la catégorie de produits, conformément au questionnaire modèle de l'EUE.
22. Même si le plan comptable de l'EUE existe implicitement, il n'a pas été défini, ni élaboré de façon explicite, et cela doit être fait. Dans quelle mesure le plan comptable de l'EUE reflète-t-il l'ancien plan comptable de la DOFI, et dans quelle mesure devrait-il être modifié en tenant compte de l'IGRF? Comment les variables au niveau de l'établissement du questionnaire modèle de l'EUE peuvent-elles être intégrées dans le plan comptable de l'EUE? Ces questions font l'objet d'un examen et seront résolues par un groupe de travail comprenant des membres provenant des diverses divisions spécialisées et divisions d'infrastructure participant à l'EUE (voir le paragraphe 53.b.).

Mise en correspondance et imputation à partir de l'IGRF en fonction du plan comptable

23. Avec la suppression de l'échantillon T2 annuel de la DOFI, à partir de l'exercice 1999, et compte tenu de l'avènement de l'IGRF pour l'année de référence 1998, nous avons besoin d'une nouvelle façon pour estimer les 225 variables primaires du plan comptable, à partir de l'IGRF. Auparavant, les données étaient regroupées directement par le personnel

chargé de la transcription et de la saisie à la DOFI (et à la DDF, pour 1997 et 1998), à partir des états financiers non uniformes de l'échantillon T2, en fonction des variables types du plan comptable. Maintenant, les mêmes données T2 sont transcrites par le personnel de l'ADRC, non pas en fonction du plan comptable de la DOFI, mais de celui de l'IGRF. La transcription constitue une mesure provisoire en attendant que soit imposée la production selon l'IGRF aux sociétés déclarantes, pour les exercices se terminant en l'an 2000.

24. Notre problème peut être énoncé dans les termes suivants. Compte tenu de l'ensemble de 225 variables financières pour lesquelles nous avons besoin d'estimations, ainsi que de l'ensemble de 700 (à peu près) variables de l'IGRF pour lesquelles, dans le cas de toute entreprise donnée, les bases de la DDF et de l'IGRF peuvent présenter une valeur, quelles formules devrait-on utiliser pour calculer les estimations relatives aux 225 variables?
25. Ce problème peut être perçu comme une combinaison de deux sous-problèmes, un problème de *mise en correspondance* et un problème d'*imputation*. Le premier sous-problème peut être énoncé de la façon suivante : en quoi consiste le *rapprochement conceptuel* co-univoque entre la définition de la variable X comprise dans le plan comptable de l'EUE (une des 225), et l'ensemble des 700 variables de l'IGRF? L'autre sous-problème peut être énoncé de la façon suivante : compte tenu du rapprochement conceptuel dont on vient de faire mention entre la variable X du plan comptable de l'EUE et les 700 variables du plan comptable de l'IGRF, ainsi que du problème particulier des enregistrements individuels de l'IGRF, qui ne comprennent pas de façon générale de détails pour chacune des 700 variables du plan comptable de l'IGRF (et en fait n'en comprennent jamais), comment les variables nécessaires détaillées de l'IGRF peuvent-elles être imputées, modélisées ou estimées, afin de permettre le calcul de la variable X?
26. Le problème de *mise en correspondance* dont on vient de parler est bien réel, mais il n'est pas particulièrement important ou difficile à résoudre au niveau conceptuel. Le personnel de la DOFI a joué un rôle important quant à la conception originale de l'IGRF, et les deux plans comptables sont assez similaires. Des progrès ont déjà été accomplis relativement à cette mise en correspondance, sous la direction de la DDF.
27. Par ailleurs, le problème d'*imputation* dont il est question ici est plus complexe. Dans nombre de cas, le plan comptable de l'IGRF inclut à la fois plusieurs catégories *détaillées*, « A₁ », « A₂ » et « A₃ », par exemple, et une catégorie englobante plus *générique*, ou bloc-total, « A », par exemple. Les déclarants de l'IGRF sont encouragés à fournir des chiffres pour le plus grand nombre possible de A₁, A₂ et A₃ et à inclure la valeur restante dans A. Lorsque certaines valeurs sont indiquées en A, nous les statisticiens devons décider comment répartir (imputer) cette valeur entre A₁, A₂ et A₃. Si nous ne procédons pas à une telle répartition, nous ne pouvons calculer les 225 variables du plan comptable de l'EUE. Heureusement, il y a toujours certains déclarants de l'IGRF qui *fournissent* la ventilation nécessaire entre A₁, A₂ et A₃. Nous avons par conséquent la possibilité d'utiliser la répartition fournie pour imputer la répartition dans le cas des autres qui indiquent une partie ou la totalité de leurs chiffres en A. Nous pouvons le faire en une opération globale appliquée à l'ensemble de la base de données de l'IGRF ou, mieux encore, nous pouvons calculer les proportions de A₁, A₂ et A₃, et procéder aux répartitions séparément, selon la strate de province, d'industrie et de taille. Malheureusement, ces approches en matière d'imputation, comme toute autre méthode d'imputation, sont imparfaites et peuvent, dans certaines circonstances, produire des résultats inappropriés ou des problèmes d'équilibrage. Elles peuvent aussi comporter des biais moins évidents. Nous pourrions procéder à des enquêtes additionnelles ou consacrer plus d'efforts à l'analyse des déclarations des enquêtes précédentes, en vue d'évaluer ces biais et de les compenser, mais de telles mesures ne sont pas prévues actuellement dans le plan. Pour le moment, nous devons utiliser la méthode de répartition simple et prendre des mesures pour rééquilibrer les enregistrements après imputation.

Stratégie de vérification et d'imputation de l'IGRF

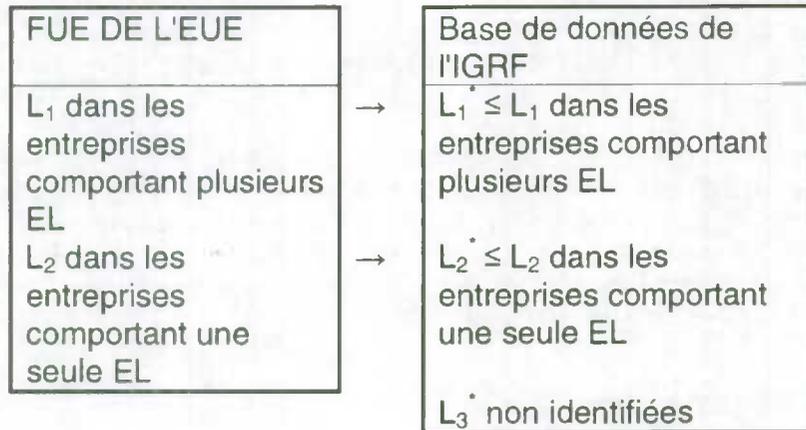
28. La DDF obtient des enregistrements de l'IGRF auprès de l'ADRC, et les compile en une version brute de la base de données intégrée de l'IGRF, une version qui n'a pas été « retouchée par Statistique Canada » (même si la DDF procède au traitement préalable des données qui entrent, afin de supprimer les données en double et de fusionner les fichiers multiples). Appelons IGRF-BRUT cette version de la base de données.
29. La division procède alors à une opération préliminaire « d'épuration » des données brutes, et vérifie ainsi que corrige les valeurs erronées évidentes. Par exemple, elle prend des mesures si une valeur déclarée est négative, mais que la définition de la variable ne permet pas de valeur inférieure à zéro. Elle vérifie les rapports d'identité et les confirme. Ces diverses opérations donnent lieu à la production d'une deuxième version épurée de la base de données intégrée, distincte de la première version brute. Appelons cette version IGRF-ÉPURÉ.
30. La DDF produit alors une troisième version de la base de données comprenant les enregistrements vérifiés et imputés. Cette version comprend les enregistrements synthétisés, produits au moyen de l'imputation par donneur ou d'autres moyens, en vue de rendre compte des entités légales qui figurent dans le Fichier de l'univers d'enquête de l'EUE, mais qui ne sont pas encore représentées autrement dans la base de données de l'IGRF. Chaque enregistrement de cette troisième version comprend des estimations pour chacune d'un certain nombre de variables *génériques* désignées ou blocs-totaux, ce qui permet par conséquent des agrégations cohérentes pour ces variables dans l'ensemble de la base de données. Appelons cette version de la base de données IGRF-Vet I.
31. À l'heure actuelle, la DDF n'a pas de plan pour estimer les détails des variables de type A_1 , A_2 et A_3 , sous le niveau générique (voir le paragraphe 27), dans le cadre de sa stratégie de vérification et d'imputation. Lorsqu'un déclarant de l'IGRF rend compte de tels détails, on les inclue évidemment dans les trois versions de la base de données. Et lorsqu'un enregistrement donneur comprend de tels détails, ceux-ci sont évidemment communiqués à tous les enregistrements connexes imputés par donneur. Toutefois, lorsqu'un déclarant de l'IGRF déclare une partie ou la totalité de ses chiffres dans une catégorie générique A, la stratégie de V et I ne prévoit pas de répartition (c'est-à-dire que l'on ne calculera aucune valeur imputée de A_1 , A_2 et A_3).
32. En outre, la stratégie actuelle de vérification et d'imputation ne comprend pas de *vérifications historiques*. Évidemment, l'enregistrement de l'IGRF de 1998 pour une entité légale particulière ne peut, de façon générale, être comparé à un enregistrement similaire pour 1997, étant donné que le processus de transcription de l'IGRF a commencé en 1998. Toutefois, il serait très souhaitable de procéder à des vérifications historiques, lorsque cela est faisable, pour l'édition 1999 de l'IGRF. Par exemple, nous devrions vérifier pour chaque EL que le montant cumulé d'amortissement déclaré pour la fin de la période en cours équivaut au montant déclaré pour la fin de la période précédente, plus les dépenses d'amortissement de la période en cours. Toutefois, étant donné que de nombreux détails financiers ne figurent que dans les notes des états financiers, et du fait que seulement certains des renseignements compris dans les notes sont transposés actuellement dans l'IGRF, il ne sera pas toujours possible de rapprocher les deux bilans concernant la même EL pour des périodes comptables consécutives. De tels rapprochements peuvent poser un défi important, selon le concept comptable particulier utilisé. Il serait possible d'appliquer quelques vérifications historiques très limitées pour la base de données de l'IGRF de 1998, grâce à des comparaisons avec les enregistrements de la base de données ADMIN de 1997, mais cela ne fait pas encore partie des priorités, et entraînerait certainement des coûts additionnels et, peut-être des retards dans notre calendrier.

33. Dans les cas où l'on utilise l'imputation par donneur pour remplacer un enregistrement manquant de l'IGRF, on perd la continuité historique. Cela ne devrait pas poser de problème significatif, à condition que la méthode d'imputation par donneur ne soit utilisée que pour des EL relativement petites. Dans le cas des EL de taille moyenne et de grande taille, il est préférable de procéder à l'imputation en appliquant des taux de croissance moyens pour la cellule à l'enregistrement de l'année précédente, ce qui permet d'assurer la continuité historique.
34. Enfin, même si cela n'est pas actuellement prévu dans les plans, il serait souhaitable de produire une quatrième version de la base de données de l'IGRF. Sur la base de la version IGRF-V et I, cet aperçu additionnel exclurait tous les enregistrements qui sont liés à des entreprises comportant plusieurs EL, auxquels on substituerait un enregistrement combiné au niveau de l'entreprise, élaboré de la façon abordée brièvement au paragraphe 16. Appelons cette version IGRF-ENTREPRISE.

Questions de couplage de l'IGRF

35. Chaque enregistrement de l'IGRF devrait être lié à une entreprise correspondante dans le RE, et, parallèlement, chaque entité légale liée à une des entreprises figurant dans le FUE de l'EUE devrait se trouver à l'intérieur de la base de données de l'IGRF.
36. Une question particulière se pose à ce chapitre, à savoir les EL qui produisent des déclarations multiples de l'IGRF pour la même année. Cela se produit si une EL est liquidée ou fusionnée à une autre EL au cours de l'année. Par exemple, une EL dont l'exercice se termine le 30 juin peut produire un formulaire régulier T2, accompagné d'une déclaration IGRF en août, puis être fusionnée par la suite avec une autre EL en octobre, ce qui mène à la production d'un deuxième T2 et de la déclaration IGRF qui l'accompagne en décembre. Si cette situation n'est pas décelée, cela pourrait mener au dénombrement en double d'actif et de passif. À l'occasion, une EL produira deux formulaires T2 ou plus à l'intérieur d'une année civile donnée, et nous devons prendre des mesures pour éviter le dénombrement en double. Une approche provisoire simple consisterait à supprimer tous les enregistrements, à l'exception de ceux comportant la période de déclaration la plus longue. Une méthode plus élaborée consisterait à fusionner les divers enregistrements, c'est-à-dire additionner les variables relatives aux flux de tous les enregistrements, prendre les variables relatives au stock du dernier enregistrement, et rééquilibrer l'enregistrement fusionné en découlant.
37. Il peut survenir un autre problème particulier, si un NE est lié à deux entreprises ou plus dans le RE. Cela peut se produire, par exemple, si le NE correspond à une coentreprise. Tous ces cas doivent être interprétés comme erronés et doivent être corrigés. Par exemple, les coentreprises doivent être inscrites comme des entreprises distinctes dans le RE.
38. La question du couplage peut être abordée de la façon suivante. Selon le FUE de l'EUE, on s'attend à trouver L entités légales (environ 1 million) dans la base de données de l'IGRF. Parmi celles-ci les L_1 sont des entités légales à l'intérieur d'entreprises comportant plusieurs EL, et les L_2 qui restent sont des EL liées à des entreprises ne comportant qu'une EL ($L_1 + L_2 = L$). Les enregistrements de l'IGRF nous parviennent constamment de l'ADRC, au fur et à mesure que l'année progresse, et la DDF les compare aux enregistrements L que nous prévoyons recevoir. À divers moments du calendrier de planification, la DDF ne trouvera que L_1 des entités légales L_1 prévues, $L_1 \leq L_1$, et seulement L_2 des entités légales L_2 prévues, $L_2 \leq L_2$. Elle trouvera en outre L_3 entités non identifiées, et peut-être de nouvelles entités légales qui ne figuraient pas dans le FUE original de l'EUE. Cette situation est démontrée dans le graphique 1.

Graphique 1
Entités légales de l'EUE



La DDF enverra les EL non identifiées L₃^{*} à la Division du Registre des entreprises, sans délai, afin que le Registre puisse être mis à jour avant la création du FUE postérieur à l'EUE. Dans l'intervalle, au moment approprié du calendrier de planification, la DDF et la DMEE ajouteront un enregistrement imputé à la version IGRF-V et I de la base de données pour chacun des enregistrements IGRF manquants (L₁ - L₁^{*}) + (L₂ - L₂^{*}). On doit partir du principe, à moins d'avoir reçu d'autres renseignements dans le cadre de la rétroaction d'enquête ou par d'autres moyens, que ces EL ont continué à être actives au cours de l'année, en dépit du fait que leur enregistrement IGRF n'est pas encore parvenu à Statistique Canada. Il serait souhaitable d'additionner les enregistrements imputés de cette façon au moins deux fois par année — une fois assez tôt dans le processus (p. ex., lorsqu'environ 75 % des enregistrements sont reçus), aux fins du calcul des estimations préliminaires, et encore une fois à une étape ultérieure, lorsque tous les enregistrements sont reçus.

39. Évidemment, le problème de couplage du NE comporte des répercussions plus larges pour bon nombre de programmes. La Division des comptes des revenus et des dépenses doit coupler les enregistrements T4 au RE au moyen du NE, afin d'obtenir des codes d'industrie pour l'estimation du revenu de la main-d'œuvre selon l'industrie, et elle consacre des ressources substantielles chaque année à traiter des enregistrements pour lesquels il n'existe pas de liens avec le RE. De même, la Division du travail (DT) doit coupler les enregistrements PD7 avec le RE, au moyen du NE, afin de calculer les codes d'industrie. Encore une fois, de nombreux enregistrements ne sont pas couplés, et la DT consacre des ressources substantielles chaque année pour régler ce problème. La question du couplage de l'IGRF constitue l'une des dernières manifestations d'un problème courant de plus en plus inquiétant. Elle découle en partie du fait que le Registre des entreprises ne tient compte que des NE courants et ne vise pas à stocker de l'information au sujet des NE anciens, qui ne sont plus valides, et en partie du fait qu'il existe environ 200 000 entreprises dans le Registre, auxquelles aucun code du SCIAN n'a encore été affecté. Un groupe de travail a été constitué pour résoudre cette question (voir le paragraphe 53.c.).

Estimation des variables du plan comptable de l'EUE

40. Afin de répondre à nos besoins quant au calcul des estimations pour les 225 variables du plan comptable au niveau de l'entreprise de l'EUE, deux options sont proposées.

41. La première option consiste à élargir la stratégie de vérification et d'imputation (voir les paragraphes 30 et 31), en vue d'inclure, à l'intérieur de la base de données IGRF-V et I, des estimations pour toutes les variables détaillées du type illustré précédemment par A₁, A₂ et A₃. Cela nous permettrait de calculer nos estimations des 225 variables directement, au moyen de la mise en correspondance conceptuelle (voir les paragraphes 25 et 26), même si nous devons, évidemment, finir nos travaux pour procéder à cette mise en correspondance. Du fait qu'elle fournit des détails plus riches et cohérents, cette option faciliterait grandement l'utilisation de la base de données de l'IGRF, de concert avec d'autres programmes d'enquête axés sur les établissements. Toutefois, elle pourrait entraîner une somme considérable de travaux additionnels, ce qui pourrait faire augmenter les coûts et retarder nos efforts en vue de l'élaboration de la base de données IGRF-V et I selon les échéances établies précédemment.
42. Une deuxième option consiste à préciser un ensemble de 225 « formules magiques », c'est-à-dire combiner une solution au problème de mise en correspondance à une solution sur mesure au problème d'imputation. Par exemple, la formule magique pour la variable X (une des 225) pourrait être la suivante :

$$X = \text{si } (A_1 = 0) \text{ et } (A \neq 0) \text{ alors } 0.25 \cdot A, \text{ autrement } A_1$$

L'option de la formule magique a pour avantage que des efforts ne sont nécessaires que pour préciser les imputations particulières requises pour calculer les 225 variables. À ce titre, elle est peut-être moins coûteuse que la première option. Elle a pour inconvénient majeur qu'elle ne comporte que quelques retombées positives, voire pas du tout, pour les autres utilisateurs de la base de données et qu'elle présente des risques d'incohérence entre nos programmes, d'autres intervenants pouvant produire leurs propres formules magiques à leurs propres fins.

43. Dans l'exemple du paragraphe 42, le coefficient 0,25 est appliqué à A. Ce coefficient vise à représenter la proportion estimée de A agrégé, dont la sous-agrégation A1 tient compte, dans le cas d'une entreprise moyenne. Par le passé, les coefficients de ce type étaient établis de façon discrétionnaire et appliqués au fur et à mesure des besoins par le personnel spécialisé dans la transcription et la saisie des données T2 de la DOFI (maintenant la DDF), au moment de la transposition des données des états financiers non uniformes dans les formulaires types, à partir du plan comptable de la DOFI. Nous devons faire un effort pour tirer parti des connaissances, de la sagesse et de l'expertise accumulées de ces employés, lorsque nous élaborerons nos nouvelles règles d'imputation ou formules magiques.
44. En résumé, la nécessité d'un plan explicite en vue de l'élaboration, pour diffusion officielle, d'estimations au niveau de l'entreprise et de l'établissement à partir de l'IGRF, et une répartition connexe des responsabilités, deviennent de plus en plus urgente. Nous avons besoin d'un plan à court terme pour l'EUE-98 et d'un plan à plus long terme pour les années subséquentes. Un groupe de travail a été établi pour résoudre ce problème, sous la direction conjointe de la DOFI et de la DDF (voir le paragraphe 53.d.).

Systemes relatifs à l'accès, au traitement et à la totalisation pour l'IGRF

45. Maintenant que les données de l'IGRF sont arrivées, le défi consiste à en faire bon usage. Le volume de données de l'IGRF est énorme. Plus de 800 000 enregistrements pour l'année d'imposition 1998 qui ont déjà été accumulés, et plusieurs centaines de milliers sont encore à venir, chacun comportant de nombreuses variables différentes. Ces enregistrements doivent faire l'objet d'une somme considérable de vérifications. Les enregistrements manquants doivent être identifiés et imputés, tandis que les

enregistrements en double doivent être supprimés. Certaines de ces questions ont été abordées dans les sections précédentes (dans les paragraphes 28 à 29 plus particulièrement).

46. La DDF a consacré des ressources considérables au cours des derniers mois à élaborer une base de données relationnelle de première catégorie, en vue de recueillir les données de l'IGRF. Les analystes spécialisés et les méthodologistes dépendent maintenant de cette base de données pour toute une gamme de fins relatives à l'estimation des produits de l'EUE. Au fur et à mesure que les analystes abordent cette tâche, ils doivent trouver ou élaborer un ensemble d'outils logiciels aux fins de la consultation de la base de données, du couplage des données d'enquête figurant dans l'entrepôt central de données, du traitement des données de diverses façons et de la production de totalisations.
47. Tant pour améliorer l'efficacité que pour susciter une plus grande cohérence quant à l'utilisation de nos ressources communes en matière de données, il est souhaitable que les utilisateurs de la base de données de l'IGRF obtiennent un ensemble d'outils logiciels d'application. Ces outils pourraient bien être adaptés du système OARS-U existant de la DSE, qui est déjà utilisé par les analystes spécialisés pour accéder aux enregistrements de données de l'entrepôt central de données. Shirley Dolan, présidente du Groupe de travail de l'architecture des systèmes de l'EUE, a discuté de la question avec les intervenants clés, et tente maintenant de trouver une solution à ce problème (comme il est indiqué au paragraphe 53.e.).

Rôle de l'enquête de la Partie 1 – Données de référence

48. Il y a eu beaucoup de discussions et quelques méprises concernant l'enquête de la « Partie 1 » de l'EUE au cours des dernières années. Le caractère et le rôle de cette enquête se sont modifiés considérablement depuis l'élaboration du plan original de l'EUE en décembre 1996. La conclusion du présent document vise à récapituler l'historique de l'enquête de la Partie 1 et à expliquer comment et pourquoi l'objectif de l'enquête s'est modifié au fil des ans.
49. La première édition de l'EUE, pour l'année de référence 1997, comprenait un questionnaire de la Partie 1, qui avait été envoyé à la majorité des grandes entreprises complexes, en vue d'obtenir des données suffisantes pour produire les 225 variables dont il est question au paragraphe 19. Le questionnaire de la Partie 1 était particulièrement important aux fins de l'estimation de la valeur ajoutée au niveau de l'entreprise – une variable clé pour l'analyse de la cohérence et le SCN – étant donné que les variables nécessaires ne figuraient pas dans le questionnaire de l'Enquête trimestrielle sur les états financiers. Un échantillon élargi T2 a aussi été prévu, même s'il ne s'est pas matérialisé en raison de problèmes techniques touchant la spécification de la tâche d'échantillonnage. Cette enquête de la Partie 1 a fourni la composante au niveau de l'entreprise de l'EUE pour 1997, dans le cas des entreprises complexes des industries non financières.
50. Avec l'avènement de l'IGRF en 1998, l'enquête complète de la Partie 1 n'est plus nécessaire. Nous pouvons obtenir toutes les données annuelles au niveau de l'entreprise dont nous avons besoin à partir de la base de données de l'IGRF, avec une exception en ce qui a trait aux entreprises comportant plusieurs EL. Une édition plus limitée de l'enquête de la Partie 1 est toujours nécessaire, en vue d'acquiescer des données consolidées pour ces cas importants, mais peu nombreux. Ainsi, pour 1998, les estimations au niveau de l'entreprise de l'EUE seront calculées à partir de l'IGRF et d'une enquête de la Partie 1 réduite.
51. Pour l'année de référence 1999, aucune enquête de la Partie 1 n'est nécessaire. Cela vient du fait que l'Enquête trimestrielle sur les états financiers a maintenant été intégrée au

Registre des entreprises, et convertie en fonction du SCIAN, et que son échantillon s'est élargi pour inclure toutes les entreprises comportant plusieurs EL de taille significative. À partir de cette année, les programmes annuels existants concernant les entreprises de la DOFI, y compris ceux ayant trait à la CALURA et à la formule des accords fiscaux entre le gouvernement fédéral et les provinces proviendront du programme pleinement intégré EUE-ETEF-IGRF.

Conclusion

52. Le présent document vise à résumer en termes généraux la façon dont les données au niveau de l'entreprise sont recueillies et traitées, et dont les agrégations au niveau de l'entreprise sont estimées en vertu de l'EUE. Il est à souhaiter qu'il contribuera à réduire certaines erreurs de perception et à favoriser la recherche de solutions aux nombreux problèmes auxquels nous faisons face à cet égard.
53. Les examens des versions précédentes du présent document ont entraîné la création de cinq groupes de travail chargés d'aspects particuliers.
 - a. Les divisions (principalement la DRE, la DMEE, la DSE, la DOFI et la DDF) doivent élaborer un plan, afin de s'assurer que l'IGRF et les autres données fiscales concernant les EL, dans le cas des quelque 1 500 entreprises comportant plusieurs EL, sont regroupées de façon appropriée. Nous avons besoin de procédures et de responsabilités claires si nous voulons produire une base de données au niveau de l'entreprise, à partir des données au niveau de l'EL qui nous parviennent par l'entremise de l'IGRF et d'autres sources de données administratives. La DSE et la DOFI dirigent conjointement un groupe de travail dont les membres proviennent de la DRE, de la DMEE et de la DDF, afin de résoudre cette question.
 - b. Nous avons besoin d'un plan comptable de l'EUE plus clairement défini, qui englobe dans une certaine mesure les anciens plans comptables de la DOFI, le questionnaire modèle de l'EUE et le plan comptable de l'IGRF, et qui se rapporte directement au Système de comptabilité nationale. La DSE coordonne un groupe de travail dont les membres proviennent de diverses divisions spécialisées et divisions d'infrastructure (la Division des normes, plus particulièrement), ainsi que des divisions du SCN, en vue de définir et de mettre en œuvre le nouveau plan comptable de l'EUE.
 - c. Nous devons revoir la situation du problème de couplage du RE et trouver des solutions qui nous permettront d'économiser des ressources rares, lorsque viendra le temps d'aborder cette question incontournable. Il y a beaucoup trop de fichiers administratifs de NE qui sont absents du RE, et de NE multiples couplés à la même entreprise du RE, et il faut en réduire le nombre. André Monty et l'auteur du présent document dirigent conjointement un groupe de travail dont les membres proviennent de diverses divisions spécialisées et divisions d'infrastructure, en vue d'élaborer un plan pour résoudre « une fois pour toute » le problème de couplage.
 - d. Étant donné que les étapes du traitement et de l'estimation de l'EUE-98 sont sur le point de commencer, nous avons besoin d'un plan pour utiliser les données de l'IGRF en vue de produire des estimations au niveau de l'entreprise et de l'établissement pour une gamme variée de variables financières. Ce plan pourrait avoir recours aux « formules magiques », à partir de variables génériques (blocs-totaux), ou il pourrait reposer sur un ensemble plus détaillé de variables calculées à partir d'un processus d'imputation de l'IGRF plus intensif. La DDF et la DOFI dirigent conjointement un groupe de travail dont les membres proviennent d'une gamme variée de divisions spécialisées et de divisions d'infrastructures, afin

de décider comment les estimations fondées sur les données fiscales seront produites et par qui.

- e. Enfin, nous avons besoin d'un logiciel commun, adapté peut-être du système OARS-U existant de la DSE, afin d'aider les analystes spécialisés à accéder à la nouvelle base de données relationnelle IGRF de la DDF, ainsi que de traiter et de totaliser les données qu'elle contient, et de coupler les données fiscales aux fichiers d'enquête de l'entrepôt central de données. Shirley Dolan, présidente du Groupe de travail de l'architecture des systèmes de l'EUE, a discuté de la question avec des intervenants clés et cherche maintenant une solution au problème.

- 54. Finalement, l'auteur souhaite remercier les nombreuses personnes qui ont corrigé les erreurs dans les ébauches précédentes du présent document et qui ont suggéré des améliorations valables.

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*)
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 générique – 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 – 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*)

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.



Projet d'amélioration des statistiques
économiques provinciales

**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

Numéro 49

Project to Improve Provincial
Economic Statistics

**The UES Strategy to
Combine Enterprise,
Establishment and
Legal-Entity Data**

Technical Series

Number 49



Internet : www.statcan.ca

Intranet : <http://pasep>



Statistique
Canada

Statistics
Canada

Canada

1010297585



STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA

CA 005

Handwritten scribble