

WORKING TOWARDS A GREATER USE OF TAX DATA IN STATISTICS CANADA

A: Introduction

As the tax data program has achieved its first round of objectives within Statistics Canada, the time is right to pursue an expanded mandate for the program. This mandate is focused on the greater use and integration of tax data in economic data programs, from which will derive the benefits of reduced respondent burden, improved quality of statistical data, additional data compared to what now exists, and possibly lower costs compared to the current costs of survey-based data.

In order to move forward, we need to map out where we want to go and how we propose to get there. This document contributes to this process. It is not a blueprint, but rather a beginning of a discussion process

It first outlines four broad goals in pursuit of greater use and integration of tax data into data programs— a very general statement that represents the ideas AND the ideals the Bureau has held for a long time. The core of this document is what follows – the vision of how we can now bring critical teams together and a detailed map of the projects that will let us realize all of these goals. The final section outlines how we (collectively) are proceeding. In particular, it will address what we hope to accomplish during the next few months in terms of development of a strategy and implementation plan which will then be presented to Policy Committee in the fall as the strategic part of the Tax Data Division (TDD) quadrennial program review.

B: The Goals

There are four broad goals which can be set in support of this expanded view of the better use of tax data in the bureau.

1. Identify and implement new survey data replacement by tax data. This goal is directed towards response burden reduction and includes the partial or entire replacement of surveys, both annual and sub-annual, by T1, T2 and GST data as appropriate.
2. Improve our current use of tax data in statistical processes leading to the production of estimates, such as sampling, edit and imputation. This includes setting appropriate expectations for what tax data can do for BTS business surveys.

3. Using tax data, provide additional annual and sub-annual information, where gaps in the survey system currently exist. This goal represents output enhancement without additional survey activity.
4. Disseminate cross-economy, tax-based industry data and demonstrate its potential by using it for business demographic studies and other analytical purposes.

C: The General Approach

Consultations: The first step will be to plan and carry out extensive consultations with users of tax data in the Bureau. The first three goals above will be used to structure the consultations although there will be a focus on BTS and on goals 1 and 2, the current use of tax data in statistical processes within BTS and for survey replacement. All consultations will be held with both subject matter divisions and infrastructure divisions such as BSMD, BR and ESD.

Building Key Partnerships: This is the key to the success of this initiative. Divisions, including TDD, must work on the intellectual side first to understand what should and can be done. But that's only half the job. To meet the goals stated above requires this thinking as a 'part 1', but then requires several strong partnerships in order to carry out implementation. From the point of view of TDD, we cannot simply look to what we can accomplish by ourselves working within our division to achieve these goals. This project embodies much interdependence.

Thus, we must define win-win situations so partner divisions want to participate. Once we can demonstrate success, then we hope to have Bureau-mandated encouragement to continue.

One set of partnerships will be with the infrastructure areas named above (BSMD, BRD, ESD). Methodological work on the tax files is critically important as is using the Business Register to represent the structure of the tax data.

A second set of partnerships will be with subject matter divisions. The goal of one set of these partnerships would be survey replacement, starting with monthly survey replacement using GST data. It is important to demonstrate that this replacement can be done by working closely and co-operatively with a subject matter division AND by choosing a narrow - as opposed to broad - area to demonstrate and advertise success. Other partnerships will be built around the further development and use of the T1 and T2 data systems.

In general, we see the role of TDD as first that of a researcher, entering into discussions and thinking about what should be done and how, and then that of a champion of an inter-field project, bringing many players together to commit to the goals and set a timetable to move toward these goals.

D: The Scope of this Initiative

We can identify 5 projects (and several sub-projects) which, when completed, will achieve the four major goals set forth above. The 5 projects are listed below, followed by a more in-depth look at each under the following headings: Objectives; The Approach; Issues; and Progress to date.

Project I: Replacing sub-annual surveys with GST data

Project II: Producing sub-annual industry data using GST for industries which may or may not have annual surveys

Project III: Replacing annual revenue and expense data from industry surveys with annual income statements produced using tax data

Project IV: Improving the effectiveness and efficiency of the use of T1 and T2 data in the BTS industry programs

Project V: Disseminating tax-based industry data and using it for business demographics and other analytical purposes

E: A Closer Look at the Projects:

Project I: Replacing sub-annual surveys with GST data

Projects 1 and 2 together are what we are referring to as “the GST project”. They are the beginning of the application phase of the GST work in the Bureau, and take as a starting point the methodological development work BSMD has done in the recent past and continues to build upon. The ongoing BSMD work and these two projects, all contribute to a better-developed, more usable GST file. Thus, as a side benefit to these specific activities, the Bureau will have available an ever-improving quality core GST file, enabling many different users to experiment with it for their own purposes.

Objective: The goal in this project is to obtain all, or much of, the data currently obtained through three large monthly industry surveys from GST data instead. The benefits include reduced response burden and reduced survey costs. The three surveys proposed to be completely or partially replaced are MRCTS, MWRTS and MSM.

The Approach: MRCTS has been chosen as the first, or test, case to be examined. A team consisting of key members from TDD, SID and BSMD is required (and as will be reported below, it has already been formed.)

The team must first understand the current state of the file from the perspective of the development work that BSMD has been doing. It must then determine its approach, and indeed, may test two or more approaches in order to get a good understanding of the strengths and weaknesses of each. It may be the case that one approach is the obvious choice for MRCTS, but a more difficult survey may require a different method. It may be beneficial if this other method has been developed in the simpler environment of MRCTS. Three types of approaches are outlined below simply to give some idea of the

range of possibilities that may be tried as we are learning how to best use the GST file for our Bureau goals.

One possible approach is modeling, wherein the file is edited and imputed to a certain stage by BSMD, but then a statistical model (regression, ratio or other) is used to benchmark and model the *current* GST data against the existing monthly survey data. This was the approach tested in Services a few years ago, with excellent results.

Another approach could be a regression model which relies heavily or entirely on *lagged* GST data for the current period estimate. This approach could prove very beneficial if timeliness issues arise in the production of GST-based estimates

A third approach is the “micro” approach in which the BSMD development work continues to the point at which monthly micro data is filled in for each firm in the universe. Then, the file could be summed and adjusted for dead and out-of-scope units to give industry results. Alternatively, a weighted sample could be drawn from this new ‘universe’ file and out-of-scope and dead units etc. could be handled in the normal manner in order to obtain industry estimates.

These approaches are mentioned here simply to give some idea of the range of possibilities that exist, all of which must be investigated and debated by the GST team.

Issues: The timeliness of the GST-based estimates must be monitored. There is a certain range over which the timeliness-quality-cost tradeoff is valid for this exercise, but there is some point at which the substitution by GST data is not useful if it is not timely. At present the data arrive in Statistics Canada between the 10th and 16th of the second month following the reference month (i.e. mid-March for January remittances). If we could process and disseminate the data in 7 calendar days, monthly data would be released between the 17th and 23rd days of the second month following the reference month (i.e. 7 to 8 weeks after the end of the reference month). Although in principle we could ask CCRA for earlier delivery, these current dates align well with the receipt at CCRA of a critical mass of GST files and earlier delivery would require us to work with considerably less data, likely at the expense of quality.

One important point to note is the flexibility of our options in order to reach response burden and cost goals. This is not an ‘all or nothing’ approach. We may find that we have early success in replacing the small, simple units which will allow us to reap considerable response and cost benefits before we have dealt with some of the more complicated issues associated with the replacement of data from complex enterprises.

The GST data project can also integrate with surveys in a more complex manner if there appear to be impediments to dropping large monthly samples. The GST data can replace survey questions on revenue from economic activity. If a monthly survey includes other questions (e.g. commodity or inventory information) the GST revenue estimates could be one part of a survey system. For example, a much smaller monthly sample could be taken in order to get some limited information on the other variables. An alternative

could be to ask the other variables of the large firms on a monthly basis and the smaller firms could be modeled, perhaps with a small sample of these firms surveyed very occasionally to validate and update the modeled results. Thus, we have many possible options before us, all of which allow us to meet our goals to a considerable extent.

The last issue relates to the reporting frequency and timeliness of the smaller GST filers. These filers typically are later in their filing and tend to file on a less frequent basis, both of which will make our use of GST data at the smaller end of the spectrum more difficult to achieve.

Progress to date: This project has gotten off to a strong start as an excellent team has been formed and is already working well together. A project manager is in place - namely Marie Brodeur, Assistant Director of TDD. Service Industries Division has one of its most experienced and knowledgeable chiefs dedicated half-time to this project as that division has made commitments to replace sample within a fairly short time period. Other members of both divisions are also on the team. While BSMD is focused on meeting the specific goals of the second phase of activity to *develop* the basic file, they also sit as members on the team dedicated to the *application* part of the project. The directors of TDD, SID and BSMD and IEAD constitute the steering committee for the GST application project.

It may be the case that when TDD presents the strategic part of its QPR to Policy Committee in the fall, we will be able to give a preliminary report on this project, as early indications are that it is progressing well.

Further, although we are mindful of the importance of getting results for Service Industries Division quite quickly, we may also start preliminary work on a second application, even as the MRCTS exercise is in its final stages. The sooner we can get preliminary work underway on another industry, the sooner we may have a second application developed which will further reduce response burden and cost.

This initiative has become very popular very quickly! We have many users lined up and knocking on our door as they wish to apply the GST data to their needs. As BSMD continues the basic development of the file on behalf of the Bureau, and as we further progress within the application phase of the work, we will create a good quality general-use file which will be available to all divisions to work with for their own purposes. Further, as we get more and more users working with the GST database, we will accelerate the pace of quality improvement of this file to the benefit of the Bureau.

Project II: Producing sub annual industry data using GST for industries which may or may not have annual surveys

Objective: The goal of this project is (i) to produce sub-annual (likely quarterly initially and eventually monthly) revenue data for industries which are now subject to annual surveys and (ii) provide sub-annual estimates of revenue for industries that are now outside of the survey system. In both cases, important information is lacking for the

SNA, and the gap is especially problematic in industries with a strong seasonal component (e.g. traveler accommodation). Indicator data are available, but not direct measurements of the level and change in revenue from economic activity on a sub-annual basis. Such information would improve the quality of input into the quarterly accounts. It would also enhance the output of subject matter areas which do not presently cover all of their area of responsibility with surveys. While there are many programs in this situation, perhaps the best example is Service Industries Division, as they have strategically decided to not survey a large portion of their universe in favour of a tax-use strategy. This project would enable them to produce sub-annual data where neither annual nor sub-annual data now exist.

The Approach: For the first case (in which we have annual data) the exercise would be one of modeling the GST data to the annual estimates and then of interpolation using the reported GST data to obtain quarterly or monthly movements. This has been tried on a test basis in the BSMD-SID activity several years ago with promising results. For those industries without annual data that can be used to benchmark the revenue estimates, a variety of other sub-annual indicators may have to be used together with the GST data to generate period to period changes, rather than levels. Discussions with SNA and subject matter divisions would have to be carried out to determine the demand for this kind of work and to set priorities for which industries should be addressed first. Following that, a team consisting of BSMD, the TDD GST team and the relevant subject matter division would be formed.

Progress to date: As indicated above, some earlier work suggests that this is 'doable'. Further, TDD realizes the importance of obtaining these 'missing' data for the SNA and industry programs. As much of the work in this broad tax data initiative is oriented towards BTS, we feel it is important to bring forward an SNA-oriented goal in order to bring that field fully into the initiative. With these thoughts in mind, TDD has taken the decision to redirect a small amount of divisional resources in order to add another person to the GST application team. This will enable us to establish a second arm of the GST application project, oriented towards producing sub-annual data from the GST files. This individual has been identified and the work will soon start in earnest, with consultation amongst divisions and identification of the first industries to be addressed. The relevant SNA and subject matter divisions, together with BSMD and TDD, would then form the team to move forward on this activity. It should be noted that the Time Series Research and Analysis Group within BSMD continues to think about this application, so the methodology work will not be starting from the very beginning.

Issues: We again have several levels of 'success' that can possibly be attained, each of which will give us benefits to a varying degree. If we can generate reasonable quality quarterly industry estimates, they will be valuable in the quarterly accounts. If they are of high quality, we can not only use them for these internal purposes but divisions can also publish these quarterly estimates within their industry dissemination programs. If good quality monthly estimates can be generated, that multiplies the benefits. Thus, we believe that we can realize some benefits early on as we continue to improve and extend the work.

While this work is in development mode, it will always be essential to have the relevant subject matter divisions on the team as the data results must be evaluated by subject matter eyes. It is the subject matter experts who understand the industry and its evolution, even if that knowledge was previously applied to annual data. If the sub-annual data are to be externally released, the GST-generated data should be treated like survey estimates that are reviewed by subject matter officers and released by the relevant divisions. As this work moves into production mode, roles and responsibilities will have to be defined – i.e. to what point does TDD develop the data and when does the subject matter division take over to produce or validate estimates and then release the estimate to the public?

It should be noted that this work is not yet formally funded. TDD is using its core GST team that was partly funded for the MRCTS application and adding some marginal funds in order to start this additional stream of work. We expect to make progress this fiscal year. This activity and a second phase of survey-replacement work will then likely form the basis of a request for a second year of funding for the ‘application phase’ of the GST project. In this second phase, we would develop a second and possibly third survey-replacement application and we would formally undertake to develop a range of sub-annual data.

Projects III and IV and the Consultation Process: Projects III and IV which follow are focused on the use of T2 and T1 files for the dual purpose of survey replacement and improved usage within the survey process. They will be the subject of in-depth study in the next 5 months, focusing on how this work can be carried out to maximize the benefits to the Bureau.

Project III: Replacing annual revenue and expense data from industry surveys with annual income statements produced using tax data¹

This large and complex project, which is composed of three sub-projects, has been spoken about as a Bureau goal for many years. The goal is to produce consistent, high quality income statements (and eventually balance sheet information) on an industry and provincial basis across the industry spectrum. This will permit entire or partial replacement of individual surveys which should gain us considerable response burden reduction and likely cost savings. It will also permit BTS to fill current data gaps in the industry survey program.

¹ The focus of this work initially will be on the income statement of corporations (the revenue and expense accounts) because of its high payoff. The income statement is a part of the very large number of annual industry surveys currently carried out in BTS. If we can replace this revenue and expense information with tax-based estimates, that will have huge benefits in response burden reduction. Similarly, revenue and expenses are the core information required for industries that are currently gaps in our statistical system, and thus this work can fill in these data gaps. The other major part of the set of financial statements provided to CCRA is the balance sheet, used primarily by IOFD and BOP. We will continue to work directly with these divisions to develop the balance sheet accounts.

Project III takes as its starting point the current state of the T1 and T2 files in Statistics Canada. These files are now subject to editing, imputation, estimation, outlier detection, allocation, etc and their quality continues to improve. But having a 'clean' T2 or T1 file does not attain the goal of producing tax-based estimates for each industry. It is only the first step – not the last. The three component projects which move us from the basic 'clean' file to tax-based revenue and expense estimates of corporations are presented below.

Project III.1: Finalizing the Chart of Accounts (COA)

What is the Chart of Accounts? The COA is a standard classification of business financial accounts with a hierarchical structure. It was developed by integrating existing financial, production and national accounting models currently in use in the Bureau. Once completed and populated with data, it will facilitate the production of consistent data across industries, thus servicing the user of economic data within or outside of the Bureau.

Objective: The goal of this project is to finalize the conceptual mapping of BTS surveys to the standardized Chart of Accounts (COA), thereby finalizing the development of the COA.

The Approach: The work to develop the COA was led by Standards Division. The top levels of the hierarchy have been accepted by the Methods and Standards Committee. What is left to do is the fine tuning – that is, a detailed comparison of the COA with existing survey programs, to ensure that it will represent a standardized set of economic accounts which can serve all major survey programs. The manufacturing surveys (ASM and MSM) have been compared in detail and we are now in the process of comparing Transportation surveys. The wide variety of surveys found in the Services program will follow. Each comparison will likely marginally improve the existing COA to make it more useful over a wide range of surveys

Issues: One issue that may arise in this project is the appropriate level or degree of detail we target in the COA. At too high a level of detail, there is too little tax data information to be viewed as a substitute for the revenue and expense statements that are collected as survey data. At too fine a level of detail, the COA becomes impractical as it often cannot be filled with tax data.

Progress to date: There is excellent progress to report because of the high quality TDD team that has been put in place, with a mandate focused on completing this work in a defined period of time. It consists of three accountants (including Jack Wilson a former AD of IOFD, a former private sector accountant and one from CCRA) and one supporting member, and is known as the COA Implementation Group. A steering committee for all of the work related to the COA will be formed consisting of the Directors of Standards, TDD and perhaps one other division.

The implementation team is meeting with divisions to review their surveys against the standardized COA. As divisional staff are very busy and might have difficulty in diverting time to a new project such as this, we have set it up such that the TDD team meets the division once to obtain the information it needs to carry out the necessary comparative work and then later to present their results to the division. The subject matter division thus does not have to devote a large amount of time to this work, but can bring their knowledge to bear where it is required. In this manner, refinements can be made to the COA which render it more applicable across the entire spectrum of industries.

Although the goal is to improve the Chart of Accounts and prepare to provide tax data to the subject matter divisions, a side benefit is emerging. As the team of accountants is working with a division on one of their surveys, they can, if requested, provide advice about the formulation of the questions and the terminology used on divisional questionnaires. We find that divisions are interested in changing their questionnaires to make them more compatible with the COA in the interests of standardized information on industries. Thus, divisions are often using this work as a catalyst to more closely align their questionnaires with the standardized chart of accounts.

Project III.2: Populating the Chart of Accounts (COA)

Objective: Once the COA mapping is finished, the goal of this sub-project is to populate the COA with data (both T1 and T2) either directly, or through data allocation and/or imputation. This will result in a standardized micro record for each firm, with a known relationship between this record and the corresponding tax record and survey questionnaire.

The Approach: The COA implementation team has made good progress in mapping the existing T2 data (in the GIFI format) to the income statement of the chart of accounts as it now stands. This work will continue as the COA is refined. At this point, we can start to fill in the COA with data. Unique matches or many-to-one matches from the T1 or the T2 GIFI to the COA give us the data directly. (Recall that we are assuming a fairly clean T2 file at this point because of all the editing, imputation, estimation etc. that it has been subject to.) There are then two situations which require further data work.

T2 GIFI has a characteristic form in which the first line in a block of related information is the 'generic' category which is variously interpreted as a 'total' or 'other' category. It has been decided that an effective way to deal with this is the 'generic to detail' allocation: where appropriate the top line is allocated across all the lines of the block, based on distributions from 'donor' firms that have fully completed the block. This fills in the block in an efficient manner, thus making synthetic data available to be mapped onto the COA categories. The COA implementation team will continue to improve and, together with subject matter divisions, to test this allocation in order to make it as accurate as possible. Again, TDD will try and do most of the preparatory work so subject matter divisions have to spend as little time as possible evaluating the results for their industries.

After the blocks of data have been filled in this manner, there will still be cases where data is mapped from GIFI to the COA only at high level categories. It will still be conceptually important to fill in data for sub-categories under these totals in the COA. In these cases, it is likely that imputation would be carried out on the detailed accounts. It would be necessary to fund BSMD to carry out the necessary work to complete this important step.

These steps should result in the COA income statement being filled with good quality data.

Issues: The main issue is whether or not everything we have outlined here is affordable and do-able within current resource availability. If not, some additional resources may be needed to complete the imputation and the divisional review stages.

A technical issue is whether or not to alter the COA in cases where the data cannot be reasonably manipulated to fit the COA model. In other words, do we compromise on theoretical appropriateness or data quality and availability?

Progress to date: The team of accountants is in place and will direct this work. The COA work is well established and progressing very well. Evaluation of the changes being suggested for the COA is will be carried out on a continual basis.

Project III.3: Producing revenues and expenses on an industry basis

The work on finalizing the chart of accounts and populating it with data is technical work and thus is moving ahead even before the consultations regarding the use of the T1 and T2 files commence. This work may (and should) become a part of the consultations to the extent a debate over different approaches may emerge or if funding is needed to complete the work. However, it is the final sub-project that will be at the core of the consultation on the use of T1 and T2. This is the work that produces data that has the potential to be used for survey replacement in a variety of ways.

Objective: The final step of project III is to use tax data (1) to produce core financial data (revenue and expenses) for industries at a national and provincial level, by size and other major characteristics and (2) to produce revenue and expenses for subsets of an industry such as smaller, simple-structure firms, or unincorporated businesses. The dual goals are oriented to the different needs that can arise in different parts of the BTS program, as it is unlikely that the same approach will be optimal for all industries. They are meant to be very broad at this point as the consultation exercise will define specifically how divisions wish to best use tax data. The data produced as a result of this project will be consistent and comparable across industries, and will be on a similar conceptual basis. Thus, the Bureau will be better able to serve client needs, to replace in whole or in part some current surveys and to provide industry data where it is now lacking.

The Approach: The consultation process will define in detail the approach to be followed. One key element of whatever approach we choose is to realize that we cannot simply add up the results of the two sub-projects above on the basis of NAICS codes and expect to have good quality estimates for industries or parts of industries. We know from the survey process that out-of-scope units, dead units etc. must be adjusted for in a survey estimate. An estimate based upon taxation data must similarly have such an adjustment incorporated into it. We must therefore use the existing BR information to adjust the industry estimates. These adjustment factors will likely be at the aggregate industry level rather than at the micro level.

This implies that we must fully integrate the BR with the tax files that have been cast into COA form and filled with data. The BR will give the tax data structure and characteristics and the tax data will add economic information to the units on the BR. TDD and BR must continue their strong partnership to achieve this.

More fundamentally, however, we must define where we are headed in terms of how we wish to use these tax data, as this will define the approach we take. For example, we might initially define two very different models. One model, the mass-production model (Ford Model T) might be to have these data largely machine-produced in TDD and then briefly reviewed by subject matter. This would achieve a relatively low work burden and optimize the timeliness of these annual data. Indeed, this approach might be used for producing across-the-board industry data designated as ‘preliminary’ as they could likely be produced relatively early in the survey cycle by historical standards, and would have the benefit of consistency across the industry universe.

The alternate or ‘Cadillac’ model could see a data file produced by TDD to a certain point and then provided to each division. For example, the file could be the micro-data file produced at the end of sub-project two, perhaps with the match to the BR also completed. The subject matter division could then produce its own industry estimates and review them as if they were data just arrived from OID, as in the survey process.

For the efficiency of resource usage in the Bureau and the associated cost reduction goals, it may be a worthy goal to aim for model one. For data quality reasons we might argue for model two. It is likely that, at least initially, we will reside somewhere between the two models. As outlined above, we might use both models to deliver a preliminary-final cycle of data. In either case, the roles and responsibilities of the various players at the various stages of the process would have to be outlined.

Issues: It will be very important to define exactly the way in which the tax data are to be used in order to determine the best approaches. Further, subject matter divisions and infrastructure divisions must work closely in this project. Although it may be the case that TDD carries out much of the hands-on development or prototype work, the industry estimates and their evaluation, validation and dissemination are the domain of the subject matter divisions.

There will also likely be the need for a collective will (or managerial direction) to standardize around as few variations in approaches as possible across all industries and to

adopt a simpler approach (that can be realized in a finite time horizon) rather than something more complex.

Progress to date: Considering specifically this sub-project, we have not yet started on it as it requires the completion (or at least good progress) of sub-projects 1 and 2, and, more importantly, it requires the consultation process to define more specifically the parameters of how we will proceed together with BTS, BSMD, BR and all other partner divisions. However, from the perspective of TDD, important progress has been made in finding someone to head up this critically important project. In September, a senior professional from INSEE will be coming on exchange to Statistics Canada for at least two years. He will be working in TDD and will be responsible for this critical project. He is extremely well respected for a wide range of knowledge, abilities and technical skills. As he was previously on exchange in Statistics Canada, he knows our environment.

Project IV: Improving the Effectiveness and Efficiency of the use of T1 and T2 data in the BTS Industry programs

Project III above is oriented towards a range of types of survey data replacement. But even the most administrative data-oriented industry data program will continue to have some surveys. Thus, project IV is intended to promote the use of tax data in the most effective and efficient manner possible within these survey processes.

Objective: After 5 years of development of the UES surveys, in which one of the important goals was to increase the use of tax data, we now have a much better idea of what works and what doesn't and the constraints and practicalities underlying the use of tax data in the survey process. The goal of this project is to thoroughly review how we presently use tax data (particularly T1 and T2) in the BTS survey processes, including the UES. This review is intended to lead to a revised plan in which we would use these administrative data in a more effective and efficient fashion.

The Approach: There are many topics that should be revisited, including the use of T1 and T2 in the edit, imputation and sample selection processes, the timing of the availability of files and the most effective method of obtaining estimates for the small T1 units. These are just a few examples of many aspects of our tax-use practices that should be reviewed.

Progress to date: This project will move forward under the umbrella of the extensive consultation process that will be outlined in section F of this document.

Issues: It will be important for divisions to take the time necessary to think through their internal issues regarding their current and desired use of tax data, and to come to a consensus of how they would like things to improve for them. Once all divisions have done this, the task force carrying out the consultation must mesh the needs and preferences of many divisions into one comprehensive and cohesive strategy that may require some compromises from divisions.

Above all, we must manage expectations and avoid the temptation to use this as an occasion to create a wish list of things divisions would like to have done for their program. The consultations must focus on the goals 1 and 2 (to identify additional survey data replacement and to improve our current use of tax data in statistical processes) together with an action plan to implement them. We all have very fertile minds and are very good at thinking of all the things that might possibly be done. Setting the limits on what is reasonable to do is part of the goal set by BTS senior management. Narrower boundaries are acceptable if we can realistically achieve them.

Project V: Disseminating tax-based industry data and using it for business demographics and other analytical purposes

Objective: The goal of this project is to publicize and disseminate the wealth of industry data produced in project III, and to show their potential through selected analytical activities.

The Approach: As indicated above, revenue and expense statements could be produced in a consistent fashion across the industry universe. In an analogous fashion to the small business data (the LEAP database) that was disseminated in the past by SSBS, we could envisage industry tabulations that span the economy on an industry and geographic basis. There are probably several ways that the subject matter divisions could come together in such as initiative – this is only one possibility. Business demographic tabulations could be produced across the industry spectrum and business demographic studies carried out to demonstrate the huge analytical potential of this file. Further, analytical tables could be derived from these basic tables. These might include analytical financial ratios of the type now produced in IOFD's 'FPICB' system. Such a system could possibly be maintained and disseminated by IOFD working together with the other subject matter divisions. It is hoped that recognized Statistics Canada analysts would participate in some demonstration analysis using these databases in order to show their potential.

Last but not least, these data could be disseminated externally to enable clients to benefit from their availability. This has long been a goal in Statistics Canada but we lacked high quality, conceptually consistent, universe data, strongly integrated with the Business Register.

Issues: The key elements in the success of project V are the universality of the industry data, its conceptual cohesiveness and the potential for relatively timely annual data. As in project III above and the GST projects, roles and responsibilities will have to be outlined to reach the optimal balance between developmental, prototype activity being carried out by infrastructure areas (BSMD, TDD, BR) on the one hand, and the subject matter role of taking final responsibility for the estimates of their industry, on the other. It is only by many divisions working closely together that we can reach the common goals represented by this project.

F: The Way Ahead

This document has attempted to outline a comprehensive set of goals oriented towards a greater and better use of tax data in the Bureau, and a set of projects that can achieve these goals. Information has been provided on the approach that can be, or already is being, taken to achieve them. In many cases, issues are identified to indicate the need for, and indeed the wide scope for, discussion on what should be done and how.

But most of this is “HOW” to proceed at the technical level. What has been alluded to and must now be addressed is “HOW” to proceed in terms of moving these ideas forward organizationally. This involves wide discussion, formulation of strategies and plans, and finally presentation, discussion and acceptance at the senior levels of the Bureau.

To address this, let's consider the above projects as three major segments: Segment 1 is composed of projects III and IV which together will be the subject of the T1/T2-BTS consultation. Segment 2 includes all the GST work (outlined in projects I and II) and segment 3 is project V, the widespread dissemination of tax-based industry data on a comprehensive and timely basis and its use for business demographics and other analysis.

Segments 1 and 2 will come first and will likely move forward at the same time, as they are now being developed through two different teams and somewhat different approaches. Clearly, they are not independent from each other and the two teams will have to interact as needed for the common good of the entire initiative. However, each has its own imperative: the GST project is the one in which Dr. Fellegi has frequently expressed a high degree of interest and he wishes to be kept up-to-date on its progress; the T1/T2 –related work is a focus of BTS as it has huge potential benefits for that field, both through survey reduction and improvement in how tax data are used in the large annual survey processes.

Let's start with the BTS Consultation task force which will have the use of T1 and T2 as its focus.

The approach is to form a small, dedicated ‘task force’ to spend about 5 months examining how programs now make use of tax data and how it can be improved. The task force will be led by a BTS assistant-director, namely Bob Pagnutti. The membership will consist of key individuals from several BTS divisions (key in the sense of their knowledge and experience with respect to the use of tax data in business surveys), an SNA member, someone from BSMD and one or more members from TDD. These individuals must thoroughly understand both the operational aspects of BTS survey processes and the conceptual basis of these business surveys. The task force will have in-depth discussions with each BTS division and likely several infrastructure and SNA divisions. Out of these discussions would come recommendations and a BTS action plan (by and for BTS) for improvements in the way tax data will be used in their survey programs.

The steering committee, consisting of the appropriate ACS and DG's, would first set the parameters of the consultations and then would be the recipient of the recommendations and the action plan. They would take responsibility for accepting and, ultimately, for committing to them.

The results of this consultation, the recommendations and the action plan would form an important component of the strategic part of the TDD quadrennial program review which will be presented in the fall. As the comprehensive proposals for addressing the Bureau's tax-related goals involve not only TDD but also other infrastructure divisions and, especially, BTS divisions, budgetary implications for these divisions will be brought forward.

Notwithstanding the consultation process that has just started, it should be noted that we (TDD) are already making progress on some of the work: we are putting a very high priority on finalizing the chart of accounts and then filling it with data from GIFL. This is technical work that can proceed now for the most part, but it is likely that it will come into scope for the consultation process at some point. The work is progressing very well.

Next, let's look at the GST segment. It is composed of a project to develop monthly survey replacement and a project to create sub-annual data where none previously existed.

The former project is now well under way, with MRCTS as the prototype case. Early progress is very good. The funding for this work came from two LTP decisions (giving funding to TDD, BSMD and SID) and some reprofiling of funds in TDD. At the time of the presentation of the TDD strategic QPR in the fall, we will likely have some early results which can be presented.

At some point, we must take the next step. Is progress on MRCTS solid enough that BTS will choose a second monthly survey for possible replacement by GST? If so, which one? This will require some discussion in BTS, which may have some links to the ongoing and broader T1/T2 discussions. Certainly the two areas of discussion will share the same overall vision. Once this decision is taken, project plans and funding needs can be outlined, either for submission to this year's LTP process or to be brought forward with the fall presentation of the TDD QPR

We have recently also decided to start the second GST project, that of creating sub-annual data where only annual industry exists or where no data exist. This work is largely oriented towards SNA needs for sub-annual data. Consultation will be carried out, starting with the SNA, to determine the industrial priorities for this work. Funding was found by reprofiling budget from within TDD and an ES recruit has been hired to get the work underway under the leadership of the project leader. As we develop this work, a request for ongoing funding for participating divisions may be put forward in the fall.

A steering committee has been formed to oversee the above GST work. It consists of the Directors of TDD, BSMD, SID and IEAD. As these projects continue, the subject matter

director will likely rotate, depending on which monthly survey application is being developed.

The final segment – the dissemination and promotion of the tax-based universe industry data – is a bit down the road. It should ideally be discussed after the action plan for our improved usage of T1/T2 information is showing results and we are ready to look farther ahead. However, it has been included in this document in order to establish it as part of the vision of what we can do with tax data in Statistics Canada.

Janice McMechan
Tax Data Division
April, 2002