

STATISTICS SWEDEN

28TH Voorburg Group Meeting

Tokyo, Japan
7-11 October 2013

Cross Cutting Topics

Extracting data from enterprise accounting systems in Sweden

Barbro von Hofsten (barbro.vonhofsten@scb.se)



Contents

1 Introduction	3
2 Actions taken by the project.....	3
2.1 Mapping of variables	3
2.2 Adjustment of the data collection tool and SBS questionnaire.....	4
2.3 Development of a communication plan with respondents.....	4
2.4 The pilot study	4
2.4.1 Scope and design of the pilot study.....	4
2.4.2 The specifications in the questionnaire	5
2.5 The evaluation study.....	6
2.5.1 About the experiment	6
2.5.2 Response rates and response burden	6
2.5.3 The quality of the technical solution	6
2.5.4 The quality in the data provided via SIE-files.....	7
2.5.5 Feedback from the respondents.....	8
2.5.6 Production costs	9
2.6 Contacts with the accounting sector.....	9
2.6.1 Contacts with operators within the accounting sector.....	9
2.6.2 Software providers	10
3 Future plans.....	10

1 Introduction

At Statistics Sweden (SCB) there is a concerted effort to find ways to simplify the whole process of data collection and thereby reducing the response burden. In the Structural Business Survey (SBS) one potential action is to collect the data via the use of standardised format. Some work has already commenced in this area of XBRL and the mapping of statistical variables to the Swedish XBRL taxonomy. At this point in time this method of data collection is not optimal due to the relatively low uptake of the XBRL protocol by the Swedish business community in their accounting systems. In view of this we wanted to investigate the possibilities to collect data in another common format already used in the accounting systems. The format in question is Standard Import Export (SIE). It is an open standard for transferring accounting data between different software products by different software suppliers, for example between the enterprise and the accountant. Due to the fact that the SIE standard is so widespread in the software business it has become a de facto standard in Sweden for transferring accounting data. Another standard, in principal, is the Swedish chart of accounts (BAS). About 95 per cent of the businesses in Sweden use BAS. The suggestion to use SIE in statistical reporting has also come up in discussions with the stakeholders association of the Swedish accounting plan (BAS-gruppen).

During 2012/2013 Statistics Sweden conducted a pilot to study the use of file transfer. The target of the project were to find out how SIE format combined with the option for respondents to complete the web questionnaire with the missing data in the SIE-file would work. The web application for electronic data collection (SIV) developed by SCB already contained functionality to receive and process data files. The project was carried out within one of the three specially designed sample surveys which together with administrative data from the Swedish Tax Agency form the Swedish Structural Business Statistics (SBS). The pilot study was to be conducted via a subsample within the SBS survey ordinary autumn collection.

2 Actions taken by the project

2.1 Mapping of variables

Prior to this project, in the autumn of 2010, minor tests were carried out where a limited number of enterprises were asked to supply data files in SIE format. The first step was to create a data file where the SBS survey variables were mapped with the BAS Accounting Plan for 2009. The files were sent to Statistics Sweden on a voluntary basis via e-mail. The mapping that was done in 2010 was far from complete due to time constraints.

Our ambition this time around was much higher than in 2010 and as we expected we encountered a few problems. First of all, the Swedish SBS survey is quite detailed. Therefore some of the accounts in the BAS Accounting Plan for 2011 would ideally have to be divided so they could be mapped with SBS variables. Other problems that we have dealt with includes that the data files sometimes contains minus signs for revenues and that they sometimes doesn't state which version of the accounting plan they are based on. There are also some industry-specific accounting plans. It would be very time consuming to map the SBS variables to the different industry specific Accounting plans, which would be needed to create a SIE-supported questionnaire for a particular industry.

We concentrated on the mapping of variables (SBS-variables to BAS2011 accounts) and design of the special web questionnaire. We made changes right up to the dissemination of the questionnaires to the respondents, however neither the mapping nor the questionnaire was 100 per cent functional. Our experience of the mapping exercise is that the work is very time consuming and difficult with high risks for errors. The mapping between the SBS variables and the BAS accounts have to be exactly right for the uploading of SIE-files to function. This is more difficult compared to the work

done with a traditional questionnaire, where we just give information about BAS accounts as a guidance to the respondents.

2.2 Adjustment of the data collection tool and SBS questionnaire

The data collection tool (SIV), developed by SCB, already contains functionality to receive and process data files. However in this project we had to make some adjustments to SIV since it was not sufficient to only receive data files. The idea was that enterprises would be able to upload a data file in the electronic SBS questionnaire using the SIV tool. In the questionnaire, we extracted information from the SIE files and the appropriate boxes of the questionnaire were filled in. The respondents were asked to complete the areas that were not found in the SIE file (for example breakdown of net turnover). The adjustments that were made in the SIV tool were not extensive.

However, the adjustments that were needed for the specific SBS questionnaire were more extensive. The designer of the questionnaire worked with the staff responsible for mapping the variables. There were several rounds of designing, testing and editing before we conducted the survey.

2.3 Development of a communication plan with respondents

The most important lesson from the small test that was carried out in 2010 was the importance of communication with the enterprises. An important part of this project has therefore been to improve the communication with the respondents. A plan was developed that contained experiences from the previous test, ways to communicate, information from software suppliers about SIE files and the contents on the web site. A small pre-test was conducted before sending out the questionnaires and cognitive interviews were made with a small number of companies afterwards.

2.4 The pilot study

2.4.1 Scope and design of the pilot study

The core of the Swedish SBS survey is the administrative data obtained from the Swedish Tax Agency. The data consists of Income statements and Balance sheets from virtually all active enterprises (approximately one million enterprises). The register does not contain data for all required variables in the SBS survey, therefore we complete the data with three specially designed sample surveys as well as with already collected data from other surveys.

Automated data collection with SIE files is primarily a tool for the completion of the so called SpecRR-questionnaire that is sent to a sample of approximately 16 500 enterprises each autumn (one of the three specially designed sample surveys). It is a specification of the Income Statement where enterprises are asked to specify income and costs. Totals from the administrative register are pre-printed. The SpecRR-sample in the SBS survey 2011 constituted the base for the experiment with SIE-files, i.e. the enterprises selected for the experiment were a subset of the enterprises that were included in the regular SpecRR-sample. For the enterprises included in the experiment the data from the uploaded SIE-file replaced the data from the Tax Agency.

The sample is stratified by industry in over 200 different strata. Within each strata is then a Pareto π ps sample drawn with a measure of size based on the enterprises income and costs.

There were several reasons why the whole SpecRR-sample did not have the opportunity to test the SIE-supported questionnaire. Firstly, it is very time consuming to construct the industry-specific account mapping that are needed to create a SIE-supported questionnaire for a particular industry. We did not have the resources to carry out this work in the pilot study. Another reason for limiting the scope of the experiment was that it was an integral part of the regular SBS survey and there was

a risk that the experiment would reduce the quality of the statistics produced on the basis of SpecRR-sample. An advantage with choosing only a part of the survey to participate is that we can compare the results with the enterprises not included in the experiment. It should be noted that the idea from the beginning was that the experiment was to include more enterprises (approximately 1000), but because of delays in the project (mainly due to the time consuming work with mapping the variables), it was decided to cut down the number of enterprises that would be included in the experiment.

We also limited the sub sample by the type and size of the enterprise and whether it was included in any of the other two specially designed surveys or not. Sole proprietors were excluded. Furthermore, only medium sized enterprises were included and defined as enterprises with 10-249 employees. These restrictions were made on the assumption that small businesses do not always manage their accounts themselves, and that it was likely that they would refrain from providing SIE files if support from bookkeepers or accountants became necessary. The largest enterprises were excluded because they do not always use BAS accounting plans. They often use group-specific or international account plans. In the SpecRR-sample there are 268 enterprises with a net turnover of more than 120 million Euro, which is less than 2 per cent of the total SpecRR-sample. Enterprises selected in one (or both) of the other specially designed samples in the SBS survey were excluded. This was done because we wanted the SIE questionnaire to be a pure SpecRR questionnaire, without any other parts. Finally we excluded industries with specific accounting plans, for example the construction and the transport industries.

Almost half of the strata in the survey was part of the SIE experiments. It was only a small part of the SpecRR-sample who met criteria for inclusion. The focus was in trade industries. Of the 2609 enterprises that qualified for inclusion in the experiment, 658 were selected randomly.

The original plan was to start the pilot survey two weeks after the ordinary SBS-survey. Unfortunately the project was delayed, mainly due to the time consuming work with mapping variables and constructing the special SpecRR questionnaire for the SIE experiment. The pilot survey started on the 24th of October 2012, which was a month after the ordinary survey was launched.

2.4.2 The specifications in the questionnaire

The SpecRR questionnaire contains the following specifications:

Specification		
Net turnover	To be specified by activity and, for some industries, by category of customers/users	Must be manually completed
Other operating income	To be specified by kind	Almost fully completed by SIE file
Change in inventory of products in process	To be specified by kind	Almost fully completed by SIE file
Raw materials, commodity costs	To be specified by cost of raw material, kind of commodity cost and/or sub contracting	Must be manually completed
Other external costs	To be specified by kind	Almost fully completed by SIE file
Total personnel costs	To be specified by kind	Fully completed by SIE file
Other operating costs	To be specified by kind	Almost fully completed by SIE file
Total of depreciation and write-downs	To be specified by kind	Fully completed by SIE file

We can't get all information requested in the SBS survey from SIE files. On the income side the BAS accounts for the net turnover are free, i.e. enterprises can choose which type of income they want to assign to a certain BAS account. This information can therefore not be used. On the cost side there are some BAS accounts covering more than one variable, i.e. the BAS account needs to be split. Therefore a further distinction had to be made. The idea was to adapt SIV for uploading data into the web questionnaire via SIE files. The respondent is then directly complementing the questionnaire with the information that can't be extracted from the SIE file. The question is whether enterprises will see this as a relief of the burden, as they still have to complete the questionnaire manually.

2.5 The evaluation study

2.5.1 About the experiment

Given that only a small proportion (16 percent) of the SpecRR-sample fulfilled the criteria for inclusion in the experiment and that we only examined a sample of these, one must be careful not to generalize the findings to the whole SBS survey. This applies primarily to conclusions on enterprises attitudes towards, knowledge of and willingness to use SIE files as a help in completing SpecRR questionnaires. However, it should be possible to make such statements on the part that really met the criteria for participation in the experiment whether they were actually selected or not. The experiment, however, should be quite comprehensive in terms of generalized conclusions about how the technical solution worked.

2.5.2 Response rates and response burden

By giving enterprises the opportunity to provide data using SIE files, we hope that the respondents feel that burden decreases. Results from the study carried out after the test shows that most of the enterprises that used the SIE alternative found that file transfer made the provision of data easier (3.5.5). This could in the future also lead to increased response rates and that the unit non-response rate may decrease. In the case of SBS survey, previous experience shows that an important factor for responsiveness is the time when the questionnaire is sent out. The earlier the questionnaires are sent to the respondents, the greater the chance that they respond. From that perspective, it was unfortunate that the SIE pilot experiment started almost a month later than the ordinary SBS survey.

The response rate for the enterprises selected to be included in the SIE experiment is lower than the response rate for the non-selected enterprises. Approximately 30 per cent of the enterprises that responded uploaded a SIE-file. It was fewer than we expected. The assessment is that this is not due to the possibility of using SIE as such, but rather that it is related to the delayed start of the SIE experiment.

2.5.3 The quality of the technical solution

On the whole, the technical solution in SIV worked well. However, there were some flaws that need to be addressed to get the data collected this way to work even better:

- There was an opportunity for enterprises to choose the SIE entrance without uploading a SIE file, which complicated the data collection. This possibility needs to be removed for future surveys.
- When the SIE file had been loaded a residual remained in most of the questionnaires specifications, which was left for the respondent to distribute. When the respondent began to fill in the specification, the residual was constant, which was likely to be confusing. In the new version of the web application for electronic data collection (SIV) this problem is solved.
- A general problem in SIV has been that the error checking ends up on the last page of the questionnaire. A newer version of SIV is now available where this problem is solved. Error checks are displayed on the page in the questionnaire where they occur.
- In the project's initial stage we tested a technical solution where a message came up with information about the accounts that had been distributed in the questionnaire after loading the SIE

file. The project would like to continue working on this solution, but the layout needs to be revised to be more user friendly. If this can be done it would probably ease the burden for the respondents.

2.5.4 The quality in the data provided via SIE-files

A cornerstone in the design of the Swedish SBS survey is that we start from known totals at micro level for all items in the population (available for most enterprises as administrative data from the Tax Agency). These known totals are further specified through three samples. SpecRR is by far the largest. For all enterprises selected in SpecRR, we pre-print the totals they reported to the Tax Authority.

An interesting aspect of the quality is how well the pre-printed administrative data mapped the data contained in SIE-files. The evaluation study shows that for most items that are pre-printed, it is a significantly higher percentage of changed pre-printed values among the enterprises that uploaded a SIE file compared to other enterprises. Especially changes made on both Net turnover and Other operating income were very high (42 and 50 percent respectively). For Net turnover, the proportion decreases significantly when only changes of more than 5 percent of the value are taken into account. Another common change that the respondents did to pre-printed values were on the Changes in inventories and work in progress (34 percent of the values are changed). The sum for Costs of goods for resale and Raw materials is the only variable that showed a lower proportion of changed values among the SIE enterprises than the other enterprises. This information is unfortunately not right as changes only showed for the separate variables Costs of goods for resale and Costs for raw materials and not on the total, which was measured in the evaluation study.

One explanation to why the information is different when a SIE file is loaded by the respondent compared to data from the Tax Agency, may be that data from different dates may have been used. Changes are made in the closing procedure and need not be included in the basic accounts. If this is the case, it can be assumed that the tax data contains a more updated version than a SIE file provided to Statistics Sweden for the SBS survey. The project has not investigated these issues in more detail.

The test results indicate that the distribution of costs are improved with information from SIE files. This means higher quality in the data. Another positive effect is that we have learnt more about how enterprises use the BAS accounts. From these experiences we can improve our questionnaires and instructions to the enterprises, which of course also is very useful.

The evaluation study indicates that extra costs for checking the received data will not be that high. We can't see that we have to contact enterprises more often than we do today.

2.5.4.1 Frequent causes of deviation

To try to understand more about why these "changes" have been made, a review was made of all the enterprises that have changed pre-printed values (from the Tax Agency) exceeding 5 percent of the original value. This examination included a total of 98 enterprises. The most common causes of deviation was:

- On the variable *Changes in inventories and work in progress*. The use of certain BAS accounts differs from how it was originally thought of and SCB has identified a specific BAS account that is problematic.
- The respondent has changed the *Allocation of costs* between different types of costs. This may include moving costs from raw materials into commodities or vice versa, or a redistribution of raw materials/commodities and other external costs or personal costs.
- The variables concerning *Consumable Inventories* are the only two variables in the specification of external costs that are part of several BAS accounts, and therefore cannot be mapped automatically. Enterprises are expected to make a distribution based on the residual when all the

other items have been allocated. Some enterprises have distributed too much, while others distributed very small amounts. This observation makes it clear that there probably is a general problem in the SBS survey with these variables.

- The respondent has *improved the distribution of external costs*.
- The respondent has made a *redistribution between net sales and other operating income*. The sum of total turnover was in most cases correct.
- The respondent has made a *redistribution within other operating income*

2.5.5 Feedback from the respondents

As part of the evaluation of the SIE experiment an interview study was conducted with 30 respondents. There were two main purposes of this study. The first was to get an understanding of how respondents felt about the use of file transfer via SIE file. The second aim was to gain knowledge about the reasons why some respondents didn't chose to use the SIE alternative. Of the enterprises that were included in the interview study 15 had used the SIE alternative and 15 had chosen not to use SIE file.

From the study, we wanted to answer the following questions:

- Where the information and instructions for file transfer clear and sufficient?
- Have the respondent encountered any problems when he or she tried to upload files in the questionnaire?
- Did the respondent find it easier to provide data?
- What were the reasons why some respondents chose not to use SIE file?

The general impression was that the information and instructions on SIE files worked fine. This was mentioned by both interview groups. However, most of the enterprises that had not used the SIE alternative, admitted that they had not read the instructions especially carefully. Only one of the 15 enterprises that chose the SIE alternative had problems with uploading the file, for the other enterprises the uploading worked fine.

Most of the enterprises that used the SIE alternative found that file transfer made the provision of data easier. This applies especially to the specifications of costs, where they did not have to do so many manual additions. None of the respondents expressed that the data provision was more difficult or time consuming. All 15 respondents would use the SIE alternative again. There were two main reasons why enterprises were not using the SIE alternative. The first was that there was no technical possibilities for creating or using a SIE file. This was due to deficiencies in the enterprises accounting system and/or that the BAS Accounting Plan was too old. The other main reason was that the respondent found it easier to provide the data manually, primarily because he or she did not have much practical experience of SIE files. A third reason why the file transfer was not used was that two respondents thought that the file transfer could be inaccurate because they were uncertain whether or not the accounts were linked to the correct SRU codes (the codes used for report to the Swedish Tax Agency). However, this is a misunderstanding because the file transfer is based on the BAS Accounts, not on SRU codes.

The enterprises in the study were also able to make suggestions for improvements. Several of the enterprises requested that the residual amount would count down as they completed the specifications. Some wanted information about what kind of SIE file that was best to use and others desired information about the variables in the questionnaire to be filled in manually. A common misunderstanding was that enterprises expected that the SIE file would take care of the whole questionnaire and that no manual complements were needed.

2.5.6 Production costs

To be able to offer a large part of the enterprises in the SBS SpecRR survey the opportunity to provide data via a SIE-file, some preparations need to be done. The activity that demands large resources are mapping SBS variables to BAS accounts. The lesson learnt in the project is that this work is very time consuming with high risks for errors. If we were to carry out this activity on a larger scale, we recommend that a special application is developed, with the purpose to support this kind of work. In the project we used Excel for this activity. We must also take into account the cost of maintaining the mapped data files. On the other hand, we are already doing some of this work, since we provide information about the BAS accounts included in the different SBS variables in the ordinary SpecRR questionnaire.

There will be costs for maintaining two kind of SpecRR questionnaires, the ordinary one and the one for uploading SIE-files. These costs are probably not that high since most of the initial work has already been done in this project. If we in the future allow all enterprises in the SpecRR-sample to submit SIE-files, we may not have to maintain the non-SIE questionnaire.

There are also some development costs associated with improving cover letters and instructions to the enterprises, but they are probably quite small and doesn't need to be taken into account as an additional cost.

The evaluation study indicates that extra costs for checking the received data will not be that high. We can't see that we have to contact enterprises more often than we do today.

The conclusion is that, apart from costs for mapping SBS variables to BAS accounts, the costs for Statistics Sweden are relatively small.

2.6 Contacts with the accounting sector

2.6.1 Contacts with operators within the accounting sector

The project also investigated future cooperation with software providers in accounting if adjustments of output for statistical purposes may be a way forward. In this work, the project has had a little more broad scope than the SBS survey and also included the LCS survey and surveys on costs in the mapping of variables. SBS and some of the other surveys have information in the questionnaires about which BAS accounts are included in the different variables. The goal was to identify variables that could be linked to the new BAS Accounts. There are a few such variables that were discussed with BAS-stakeholder association (BAS) initially. Cooperation with BAS has been expanded during the project. We have been able to take the opportunity to ask questions to BAS that are linked to the project but also in the wider SBS perspective. The experience is that we need to have more regular contact with organizations in this area. BAS has recommended that in addition to the SIE group we also cooperate more with the Association of Swedish Accounting consultants (SRF).

The SIE group has received information about the project as well. They have been beneficial to our work and are in favour of increasing the cooperation with Statistics Sweden in order to decrease the burden on enterprises. The idea of developing a specific SIE-type for reporting data to Statistics Sweden has been raised.

2.6.2 Software providers

The project has made a list of suggestions to be used in communication with software providers in a potential next step to create statistical reports directly in the accounting systems, to further increase the number of variables that can be reported more automatically. These are as follows:

- Division of BAS-accounts.
- New types of SIE-files, for example SCB-SIE-type.
- Background information of BAS-account version in SIE-files.
- Clarify how charts of accounts in accounting programs are updated.
- Contact trade associations about accounting plans for different industries.
- Clarify any default connections with the software companies. Are there standard mappings between BAS accounts and SRU field codes directly in some accounting software?
- Can SIE-files be used in the data collection of short-term surveys? For example, clarify the presence of accruals to determine whether short-term surveys are suitable for SIE-loading.
- Discuss the use of BAS sub-accounts. Do enterprises normally use the finest level in their accounts? What is provided as "default" in different systems?
- Information for software producers on SCB's website.

3 Future plans

The project has brought both positive and negative experiences of collecting SIE files. Many problems can be considered as "teething problems" that can be overcome. The most negative aspects are the time-consuming work on mapping SBS variables to BAS accounts and the high risks of errors done in this kind of work. Another negative aspect is that enterprises tend to use some BAS accounts wrongly. The most positive aspect is that the enterprises that submitted SIE-files want to continue using this option. Furthermore, examination of SIE files and contacts with, among others, BAS-stakeholder association (BAS) has helped us gain a better understanding of how enterprises are using BAS accounts in their bookkeeping. Thus, we can become better at adapting the surveys of Statistics Sweden to businesses' ability to communicate information and get better quality output. Meanwhile, since the start of the project, new ideas and issues have emerged that we need to continue working with.

Our plans for this autumn is to continue to offer enterprises the choice to upload SIE-files in the SBS survey. We will not offer all enterprises that possibility yet. We don't have resources today to carry out the time-consuming mapping. It will probably be especially difficult to map the SBS variables to all the different industry-specific accounting plans. We will therefore include 4000 enterprises in the group that gets to upload SIE-files this year. This will give us time to further analyze the data collected and bit by bit expand the number of enterprises that will have the choice to upload SIE-files.

We will also use this kind of data collection for the survey Intermediate consumption for the service sectors (TFF). We are already offering enterprises taking part in the Labour Cost Survey (LCS) to upload a SIE-file.