

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Seminar on New Frontiers for Statistical Data Collection
(Geneva, Switzerland, 31 October-2 November 2012)

Topic (iv): Data collection using mixed modes and multiple sources

USE OF ADMINISTRATIVE DATA WITHIN DATA COLLECTION

Contributed Paper

Prepared by Olivier Goddeeris, National Statistical Institute, Belgium

I. Introduction

1. Each National Statistical Office is confronted with the demand to simplify questionnaires and, in doing so, to lower their administrative burden. Besides the administrative simplification techniques goldplating, questioning fewer enterprises and changing the frequency of surveys, avoiding asking for the same information twice by using administrative sources is one of the most widely used techniques.
2. Statistics Belgium has also felt the pinch already for a long time. Companies are becoming more assertive and do not accept that the government does not use the information that is already in its possession. This is the reason why the Belgian National Statistical Institute has a long tradition in using administrative sources in data collection. Nevertheless, only a few years ago, NSI Belgium decided to use them more actively and ‘visibly’.
3. In the first part of this document we will explain how we used administrative sources in the past. In the second part we will present the new use of those sources as well as the technology making this new use possible. In the last part of the document we will summarise some advantages and lessons learned.

II. The use of administrative sources until 2008

4. Until 2008, Statistics Belgium used administrative sources mainly in order to abolish questionnaires or reduce the number of questions asked in a survey. When at first sight both uses may seem effective and efficient, only the first use has been proven as really satisfying for enterprises.
5. Reducing the number of questions in a survey seems, from our experience, only to result in short-lived satisfaction for both statisticians and enterprises. Statisticians can present a shorter questionnaire, which results in fewer complaints about the administrative burden of statistics. Enterprises can spend more time on their core business instead of completing administrative tasks. Nevertheless, the positive effect of this form of simplification fades quickly. Statisticians are confronted with consistency problems when they compare, ex post, the survey data with the administrative data used to replace the questions left out of the questionnaire. Major divergences

are not exceptional and many efforts need to be made correcting the data or to find the reason of the divergences. This is, for example, the case when the logical structure of a questionnaire is lost after deleting headings. Suppose that $A + B + C + D = E$ and that B, D and E can be deleted because this information is available in administrative sources. When the survey data and the data from administrative sources are compared ex post, the sum will very likely be incorrect. It is then often necessary to contact the enterprise again, which is not always appreciated.

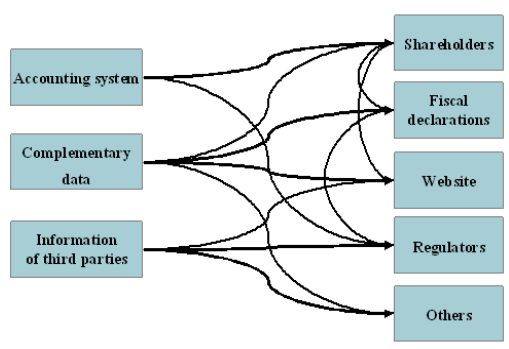
6. It also appears that, after having deleted variables from a survey form, enterprises often quickly start complaining again about the administrative burden of the survey and wonder if Statistics Belgium, when it simplified the survey, took into account the information the enterprise already transmitted to other administrations.

III. The use of administrative data in the data collection process from 2008 onward

7. Over the years Statistics Belgium discovered that it was best to delete variables only when entire sections of the survey form could be deleted. Randomly deleting variables from parts of a questionnaire should be avoided as much as possible.
8. With this information in mind, a project was launched in 2008 to find a solution for the following problems:
 - (a) How can we still delete variables from parts of a questionnaire for which the information is already available in an administrative source?
 - (b) How can we further reduce the administrative burden for enterprises?
 - (c) How can we ensure greater consistency between data from administrative sources and survey data?
9. We soon opted for the use of pre-filled questionnaires in combination with the XBRL standard.

A. What is XBRL?

10. XBRL is an electronic language that has been specially developed for the exchange of business economic information through the internet. Throughout the years many ways of exchanging information have surfaced, aimed at single or specific needs, which led to many different formats. These require the data to be entered again, which increases the error risk and causes double use.
11. Meanwhile economic actors are asking for information updates more frequently and require more transparency as part of “corporate governance”. Because of a lack of harmonisation the spaghetti model (see figure opposite) is the unfortunate result; the information exchange is slow, difficult and expensive.
12. A solution had to be found that meets the demands by enterprises on the one hand to rationalise internal and external business economic reporting, and by authorities and other users on the other hand to protect the quality of received data, whilst also limiting the total incurred costs. XBRL is an interesting option for these demands.

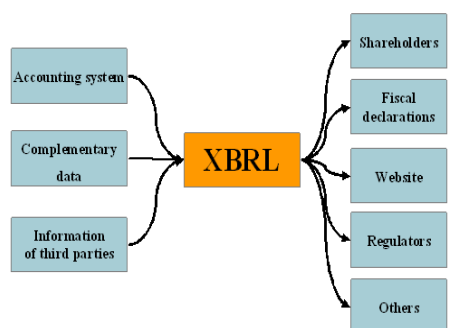


internal and external business economic reporting, and by authorities and other users on the other hand to protect the quality of received data, whilst also limiting the total incurred costs. XBRL is an interesting option for these demands.

13. XBRL (eXtensible Business Reporting Language) is an open standard based on XML (Extensible Markup Language) for the electronic collection and transfer of business economic data through the internet. The idea of the XBRL language is to identify each concept (e.g. ‘turnover’) and add it to a ‘taxonomy’, which is similar to a dictionary. These concepts, brought together in a structured way, can be

recognised, processed and represented in different ways, depending on the intended use (e.g. ‘annual accounts’ or ‘SBS’).

14. The use of XBRL has many advantages:
- (a) XBRL increases the quality of reported data because it renders manipulations such as manually re-entering data unnecessary;
 - (b) the XBRL format not only facilitates exchange in file format, but also processing and analysis of the data;
 - (c) XBRL also keeps down the costs for the provision of information, since its format enables the selection, re-use and presentation of data according to specific needs.



B. XBRL in Belgium

15. By choosing XBRL, Statistics Belgium joins the National Bank of Belgium, who also chose this standard for the filing of all annual accounts. Recently the Ministry of Finance has also adopted the standard for the declaration of corporate tax.

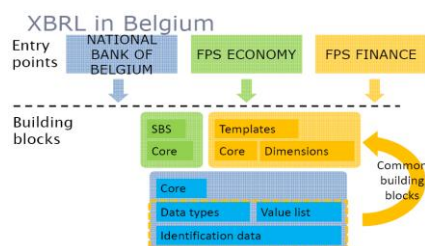
16. The 3 institutions have resolved to define concepts only once, as described above, and re-use them in different reports. For example, the element 'Street' was defined in the National Bank of

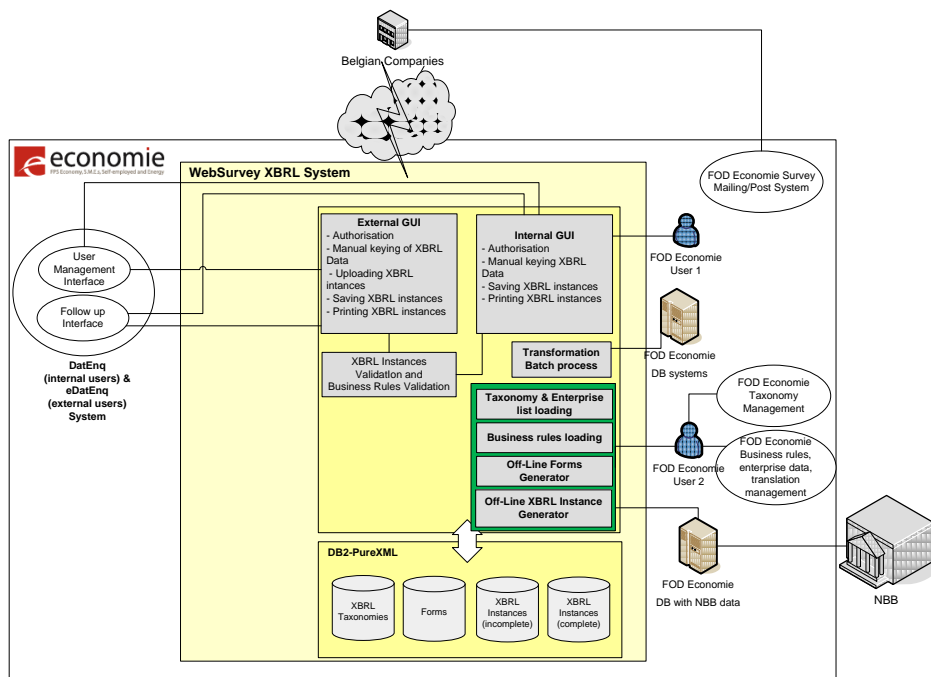
Belgium's taxonomy and re-used in the taxonomies of Statistics Belgium and of the Ministry of Finance's Corporate Tax. This allows enterprises to fill in their 'street' only once and, thanks to XBRL technology, the element will be filled in automatically in the different reports.

17. The use of XBRL by different public authorities has spurred software enterprises to master the standard quickly and integrate the different reports in their software. Certain accountancy software, for example, integrates XBRL reports for annual accounts, corporate tax and the structural business statistics survey. This detailed accountancy creates a 'push on the button' solution for the 3 reports, hereby reducing the burden to a minimum. The reports can afterwards be uploaded in the specific web portals.

C. Web architecture

18. XBRL already reduced the administrative burden on enterprises that use an integrated software solution to a minimum, as described above. However, not all enterprises are willing to invest in this solution. In 2008 Statistics Belgium therefore decided to provide a web application that is entirely based on the XBRL standard.





(a) Upload

19. The web application offers enterprises whose integrated software solution automatically generates an XBRL report the possibility to upload and validate the reporting file. Validation feedback is given immediately and, if required, the enterprise can re-upload a corrected version of the report.
20. The integration of surveys in existing software requires clear communication with private software enterprises. To this end, Statistics Belgium organises different meetings and provides a copy of the web application for software enterprises to run the necessary tests.

(b) Manual input

21. If an enterprise does not purchase an integrated solution, the survey can be filled in online. Based on the presentation Linkbase of a taxonomy (part of a taxonomy that represents the tree structure of a form), web pages are automatically generated, in which the necessary data can be entered.
22. To make the web application user-friendly and to provide a solution to the two aforementioned problems (how can consistency be improved between administrative data and survey data + how can we delete variables from parts of the questionnaire when this information is already available in administrative sources), Statistics Belgium also provides a module for the generation of web pages which pre-fills individual survey forms.
23. Through the "pre-fill module", it is possible to avoid duplication of data collection by pre-filling information already available in the web survey. When an enterprise opens its survey form, only some parts will need to be completed.
24. This method offers many advantages:
 - (a) Consistency between administrative data and survey data is considerably improved, partly because consistency checks can be carried out when data are entered.
 - (b) For enterprises it is unquestionable that Statistics Belgium uses data from administrative sources as they are displayed in the web form.

- (c) Variables from parts of the questionnaire are not deleted but are pre-filled, which avoids duplication of data collection and preserves the logical structure of the questionnaire.
 - (d) As both administrative and survey data are completed in the survey form, *ex-post* corrections have been considerably reduced. The necessary consistency checks are already carried out at the start.
25. The use of this module requires some preparation:
- (a) User ID: every respondent needs to be identified on the basis of a unique number. In Belgium every enterprise and every establishment of an enterprise has a unique number. This number is saved until the enterprise or the establishment ceases to exist. Enterprise numbers are assigned by the Ministry of Economy and are used by all public administrations.
 - (b) Consistency checks: when different administrative sources are used to pre-fill the questionnaire, a number of consistency checks should be carried out in advance between the different sources used. It is likely that data from different sources do not correspond. This should therefore be taken into account from the beginning.
 - (c) Format and data type: different administrative sources usually have different file formats and data types. From the beginning, the purpose was not to adjust the web application for each new file format or data type. To that end, Statistics Belgium has developed an SAS code to convert different file formats and data types into a standard txt file. This txt file serves as input for the pre-filled module. New file formats or data types are systematically added to the SAS code.
- (c) Upload, complete or correct**
26. The web application also allows the combination of the two aforementioned methods. Enterprises need only to upload an automatically generated XBRL file in the web application. If the file contains errors or if some fields need to be further completed, enterprises can open a form online to display errors and the fields that still need to be completed. In the first case, enterprises will be able to correct and send the survey on the basis of error messages. In the second case, the web application uploads an XBRL file containing certain information. In this way, the web application can be (further) completed. Enterprises need only to complete the fields that still need to be filled in.
27. It is not always necessary to upload an XBRL file to complete the questionnaire. A new copy/paste function has been recently added: it is now possible to copy tables and paste them directly into the web application. Through a java batch script, tables are converted into an XBRL format. In this way, it is possible to easily retrieve data from Excel or other applications and to paste them into the web application. Of course, the order of the columns and the data type need to correspond to what has been set in the XBRL taxonomy. This method avoids overwriting already available data.

IV. Advantages of the web application

28. The XBRL-based application offers different advantages:
- (a) Enterprises that decide to manually enter all data in the web application have less research to carry out and less information to enter thanks to the system of pre-filled surveys. Furthermore pre-filled data also serve as a basis to find the missing information.
 - (b) Through pre-filling administrative data, it is also possible to avoid asking variables from parts of survey forms twice and to maintain the logical structure of the questionnaire.
 - (c) This system also ensures consistency between survey data and administrative data.

- (d) Enterprises that integrate surveys in their software have access to a “push-on-the-button” solution. The software automatically generates an XBRL file which can be directly uploaded in the web application. Thanks to a close collaboration between software enterprises and Statistics Belgium, these enterprises assure their customers that the XBRL file developed through their software will pass all checks and that respondents will not need to manually enter data.
- (e) The web application leads to a win-win situation. Through the web application, enterprises have less work to do and they also save time. Moreover, authorities can access data more rapidly and this information is automatically checked and corrected.
- (f) Through the different taxonomies, data experts and consultants from Statistics Belgium can now manage metadata from the survey, which requires less interaction between enterprises and ICT. After modifications to taxonomies, “business users” from Statistics Belgium can generate, through this architecture, new web pages without any ICT intervention.

V. Lessons learned

29. For three years Statistics Belgium has intensively used the aforementioned web application. In this section, some factors that have contributed to the success of this project are presented:
- (a) When uploading files, particular attention should be paid to the consistency between pre-filled data and uploaded data. Statistics Belgium allowed until recently overwriting of pre-filled data when an XBRL file was uploaded. Analyses showed that it was not effective. Pre-filled data were often overwritten with divergent information, which ex post led to other consistency problems between administrative data and survey data. To avoid such problems, it is recommended that pre-filled data should not be overwritten. Attention should also be paid to the pre-fill for manual input.
 - (b) The “push-on-the-button” feature is the best solution for enterprises as the administrative burden is the lowest possible. However, some factors need to be taken into account when using this solution:
 - *The price of the integrated solution*: the solution cannot be more expensive than the price that enterprises would pay to have this survey completed manually. If the price is higher, it will not be more profitable for enterprises to purchase this integrated package.
 - *Stability through an integrated solution*: integrated surveys should not be adjusted, if possible. It will make the integrated solution more cost-effective.
 - *Frequency*: the more enterprises need to fill in a survey, the more they will try to find solutions to make the data transmission to Statistics Belgium easier. As they are likely to often use the integrated solution, it will be more profitable to invest in this solution.
 - *Simplification*: the purchase of an integrated solution should represent a real simplification for enterprises. Halfway solutions will not work. For example, despite the purchase cost, staff members of an enterprise still need to make many manual modifications.
 - (c) If more and more public administrations re-use the same standard, it will be more profitable for enterprises to adopt this standard. In Belgium software enterprises can offer integrated XBRL packages through which enterprises can automatically complete and send annual accounts, corporate taxes and statistical obligations. All these factors act as a powerful incentive for enterprises to purchase an XBRL package.
 - (d) It is imperative to carry out consistency checks in advance for the data that will be pre-filled in the web application. As mentioned before, inconsistencies (major or not) may often be found when data are derived from two different administrative sources.
 - (e) There is need for sufficient support to promote the use of the web application. Therefore Statistics Belgium will systematically contact sector federations when a new survey is added to

the application. Together with these federations, the best solution is sought. As they are closely involved in this project, sector federations will probably promote it actively among their members.

V. Results

30. From the first year, the web application was an immediate success. While about 10% of enterprises used to complete surveys electronically, this number immediately grew to 90% after the launch of the new application. Three years later this number stabilised at about 95%. Through an ongoing collaboration with software enterprises and sector federations, the web application was implemented smoothly. Enterprises appreciate the efforts made by Statistics Belgium to minimise the administrative burden.
31. Statistics Belgium also benefits from this success. As in other Member States, Belgium needs to cut back on expenditure. One of the cost-saving measures is to limit the number of new employees as much as possible in order to reduce public-sector bureaucracy. Through an extensive use of the web application, a large part of manual inputs and checks is not needed anymore. A reduction of the workload led to a new distribution of work allocations among employees, compensating the departure of employees internally.
32. The web application represents cost savings not only in terms of the number of employees, but also in time spent for processing surveys. Through automated checks in the web application, the pre-fill of administrative data and the significant reduction of manual inputs, the processing time for some surveys has been reduced by half.