Distr. GENERAL

8 October 2012

WP. 13
ENGLISH ONLY

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 (ii): New methods and technologies

RATIONALISING DATA COLLECTION: AUTOMATED DATA COLLECTION FROM ENTERPRISES

Contributed Paper

Prepared by Mr Juha-Pekka Konttinen, Statistics Finland

I. Introduction

1. The European Statistical System (ESS) is increasingly being called upon to reduce the administrative burden of European enterprises. Therefore, initiatives need to be taken to find a better balance between the user needs for statistics and the burden put on the producers. The latter group consists in the first place of the reporting enterprises, but also includes the statistical authorities in charge of data collecting and compiling of statistics. One approach to alleviate the burden is to reduce the volume of collected information; another approach is to introduce more efficient data collection methods and/or sources.

2. In the recent context of re-engineering official statistics and making optimal use of technological developments, Eurostat has supported the development of more efficient ways of collecting data. An example of the new methods and technologies is a system of automated data collection. This paper examines the current situation, experiences, challenges and benefits concerning automated data collection in the EU Member States and especially at Statistics Finland and in its accommodation statistics.

II. Automated Data Collection

A. Introduction

3. Automated Data Collection (ADC) is used for instant and automatic capturing of various data and information from sources such as databases, electronic files, coupons, forms and handwritten documents. It grants the user of the data capture software the power to process and capture entire text (or other electronic) documents. Moreover, all these processes should be done by pushing a single button. In the past, this could only be done manually, so by using automated data collection software tremendous amounts of time will be saved. In addition, the rate of errors will also be lowered substantially by using automated data collection.

4. Automated data collection can be utilised in many areas of society and it consists of many technologies. This paper examines the use of automated data collection in the context of statistical production. Automated data collection can be utilised in many domains of statistics. In

short, statistical information is generated automatically, e.g. from the respondent's management system into a specified file, and sent direct as an encrypted electronic transmission to the National Statistical Institute's database. As described, the procedure should be more or less automatic. In the optimal case, such as in Statistics Finland's accommodation statistics, the respondents simply push a single button in their hotel management system and the data file is sent immediately and automatically to Statistics Finland's database.

5. The National Statistical Institutes of the EU Member States and the Statistical Office of the European Communities, Eurostat, collectively make up the European Statistical System (ESS). Its task is to produce statistics for EU organs, develop and harmonise the statistics compiled in the Member States, and prepare statistical information. Reliable, detailed and comparable statistics of high quality provided by the Member States are vitally important for decision-making, monitoring and evaluation within the European Union. On the other hand, the European Statistical System is increasingly being called upon to reduce the administrative burden of European enterprises. This concerns all enterprises, but particularly small and medium-sized businesses (SMEs).

6. In addition, there is a certain need to decrease the burden on the producers of statistics. Therefore, initiatives need to be taken in order to find a better balance between the user needs (international and national legislation, other essential needs) and the burden on the producers (reporting, collecting data, and compiling and publishing statistics). The latter group consists in the first place of the reporting enterprises, but it also includes the statistical authorities in charge of data collecting and compiling of statistics.

7. The benefits from automated data collection to statistical production are obvious. The introduction of a widespread system of automated data collection can lead to a significant reduction in the reporting burden of enterprises and in the processing and compilation burden of statistical authorities. Moreover, the automation can improve the timeliness, quality and comparability of statistics.

B. Accommodation Statistics at Statistics Finland

8. Tourism statistics in the European Union are currently covered by Regulation (EU) No 692/2011 of the European Parliament and of the Council of 6 July 2011 concerning European statistics on tourism. The regulation establishes a common framework for systematic development, production and dissemination of European statistics on tourism supply and demand. This paper focuses on statistics relating to the supply side, that is, statistics on the capacity and occupancy (arrivals, nights spent) of tourist accommodation establishments.

9. Statistics Finland has been producing accommodation statistics since 1971. Nowadays the data are collected monthly and the sample comprises (depending on the season) 900 to 1,400 establishments. In the past, there were more or less two types of respondents: i) those who answered by faxing the reports generated from their hotel management system to Statistics Finland, and ii) those who filled in paper questionnaires. Either way, the data had to be recorded manually at Statistics Finland. In addition, the response burden on accommodation establishments was high because of the manual work involved.

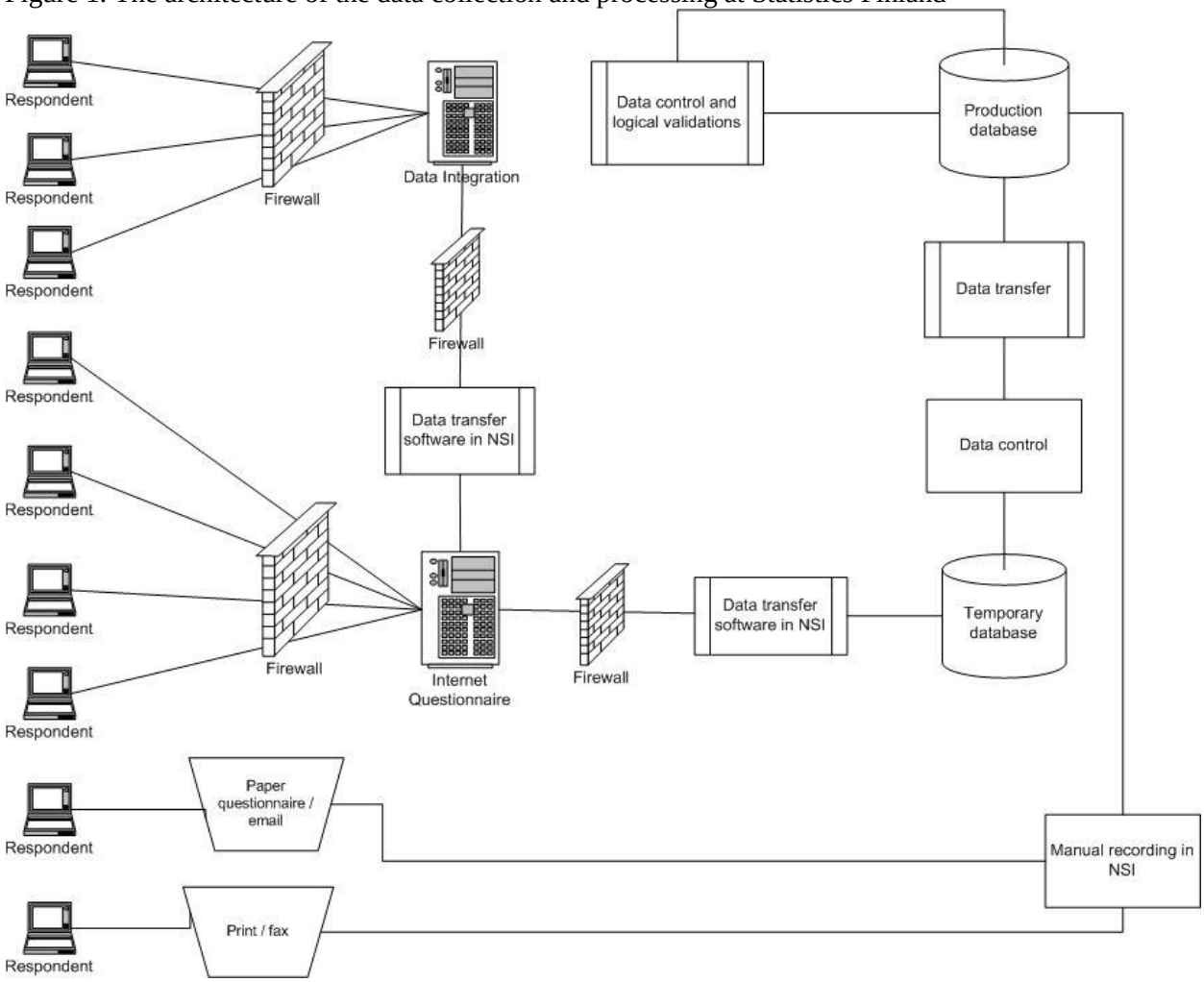
10. Statistics Finland therefore decided to develop new modes of responding that would be less burdensome both to the data suppliers and to the NSI. In 2005, a pilot project was carried out and the result of the study was a collective mode of an XML-based (Extensible Markup Language is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable) Internet questionnaire which could be filled in manually. Accommodation statistics were part of the pilot study and implemented the system in early 2005. In addition, this collection mode already included an application for mass dispatching of emails as well as an application for transferring data from the collection database to the production database. At the same time, the logicality of the data could be verified.

11. The Internet questionnaire was introduced alongside the previously used answering modes. Nevertheless, all of these modes of response involved manual phases, either for the data suppliers or for Statistics Finland, or for both. Thus, the need for automated data collection was recognised. This led to the launching of a pilot project on automated data collection in accommodation statistics. Statistics Finland already had existing architecture resulting from the pilot study. The major challenge was to find a way in which XML files could be formed from the data suppliers' information systems.

12. Statistics Finland and its accommodation statistics, and representatives of software suppliers concluded that the shared objective was to make data collection easier and quicker. It was agreed that Statistics Finland would draw up the required specifications and documents, such as a description of the XML file according to which the automated data reporting could be implemented. Accordingly, the participating software suppliers agreed to add a new reporting facility to their own software. A system of automated data collection was introduced in autumn 2005, and with the help of the system, the data suppliers could now transmit the XML file to Statistics Finland direct from the hotel management/booking system of the accommodation establishment simply by pushing a single button.

13. As a result, electronic data collection at Statistics Finland comprises two alternative modes. One is the Internet-based questionnaire and the other is automated data collection. Figure 1 describes the process of the data collection and processing at Statistics Finland, including both electronic and manual data collection.

Figure 1. The architecture of the data collection and processing at Statistics Finland

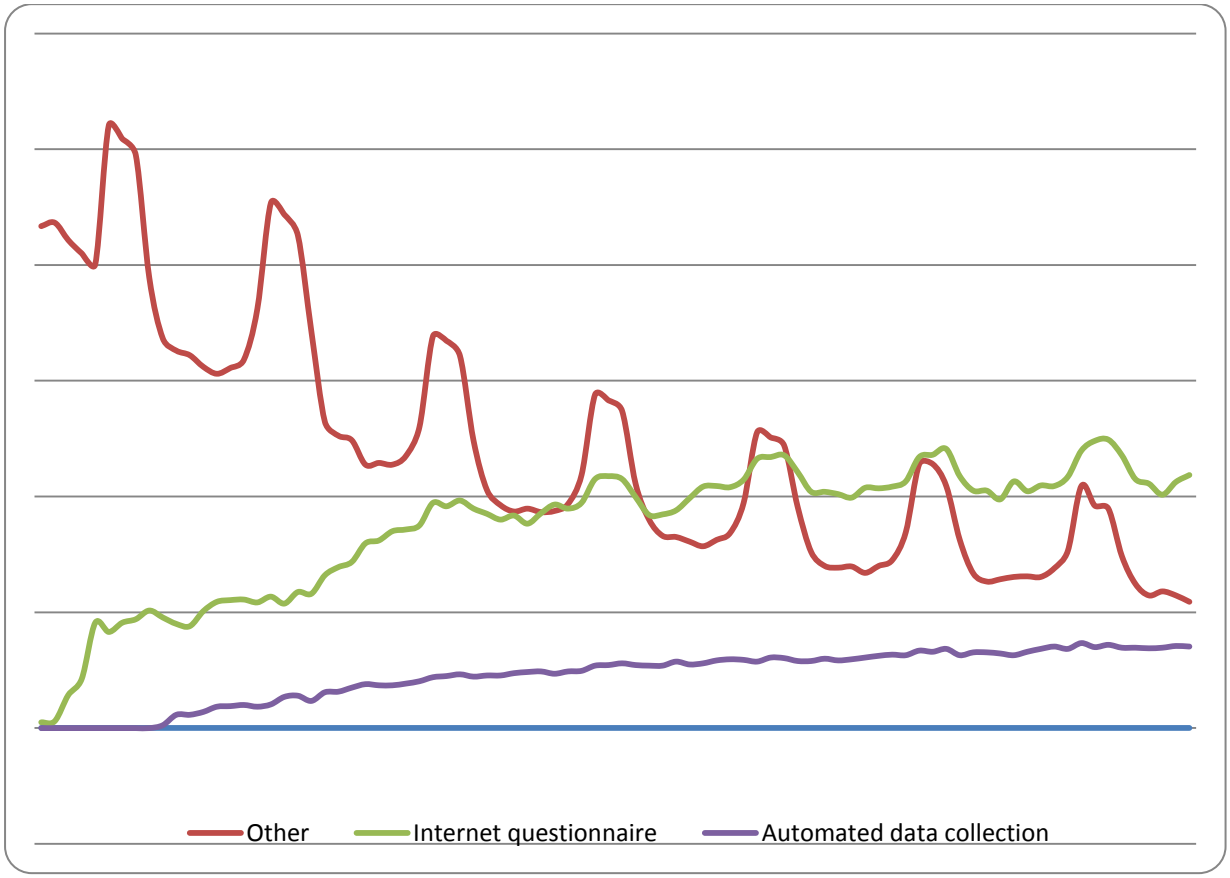


14. Figure 1 indicates that the process is quite similar in the two alternatives of electronic data collection. If the respondent delivers data by using automated data collection, the received data

file is transferred with only a short delay to the Internet questionnaire. Thus, the respondent can quite easily view and use the data sent. Otherwise, the respondent must manually key in the data to the Internet questionnaire. After this step, the process is similar for the two alternatives. The data are transferred from the Internet questionnaire to the temporary database and then further to the production database after data control and logical verifications. The respondents also receive an automated feedback report quarterly.

15. In 2011, Statistics Finland received approximately 66 per cent of the data direct electronically (automated data collection and the Internet questionnaire) and 34 per cent by other modes of reporting (paper questionnaire, email, fax, etc.). Overall, about 16 per cent (approximately 140 accommodation establishments) of the data were received automatically. Figure 2 shows that the development has been encouraging, but unfortunately quite slow, at least in automated data collection.

Figure 2. Number of respondents by reporting method 2005-2012



16. The introduction of both electronic and automated data collection has led to a notable reduction in the processing and compilation burden of Statistics Finland. During the years 2004-2010 the working hours spent on data collection, editing, reminders and feedback decreased by 35 per cent. In other words, the compilation burden of Statistics Finland and its accommodation statistics has been reduced significantly. Moreover, the timeliness has improved greatly since Statistics Finland receives data much more quickly than by other modes of data collection. This means that the NSI has more time to analyse and go through the data, which has led to improved quality of the statistics.

17. Moreover, the response burden of the respondents having implemented the system of automated data collection has practically vanished. Earlier, and in other reporting modes, the response burden per month was, on the average, between 30 minutes and 2 hours. The response burden of the establishments using automated data collection is 1 minute or practically zero. It is also

encouraging that since the new system was created, there have been no major technical problems during the implementation and the delivery process.

18. Overall, the experiences from electronic and automated data collections have been quite encouraging. The benefits are apparent, as discussed. Nevertheless, it must be noted that the implementation of automated data collection seems to be surprisingly slow. There are various reasons for this. In an overall picture, it seems that one of the most important ones is that the accommodation establishments either have many different management systems and software applications or else no software at all. This indicates that it takes time and money to update all the software although the resources needed should be quite reasonable compared to the benefits. In Finland, there are three major software vendors and, so far, only one of them has implemented the system.

19. Another major issue is that it seems to be fairly challenging to get updated versions of the system to the customers (respondents) and to introduce new functions. The reporter of the data and the person who is responsible for the updates is not necessarily the same person and, as a result, the reporters (respondents) of the data might not be aware of the new functions.

20. It is also quite challenging to persuade major hotel chains to implement the system. Global software houses, which often supply the management systems to these hotel chains consider that one country is a small market and are not easily convinced of the necessity to update their systems. In addition, small (and often seasonal) establishments do not have the appropriate software, resources and/or interest to invest money and effort in a new system. In addition, there is still the possibility to send the data by not using electronic data collection, which is often easier than the implementation of the new system.

21. There are a couple of possibilities for fostering the implementation of automated data collection. First, NSIs must promote and encourage the implementation of the system by highlighting the obvious benefits. This work has to be done on every occasion possible and among different interest groups, such as accommodation establishments (respondents), federations, hotel chains and associations. Second, NSIs must work together and create common systems that can be utilised in other countries. Third, NSIs must perhaps consider the possibility of sending data only electronically/automatically. In some cases, for instance, it is easier just to send the booking system print by fax. This is obviously a difficult issue because it would probably lead to more significant non-response. Fourth, in other domains, such as business reporting in certain European countries, legislation, that is, making it mandatory has been the most effective way.

B. Accommodation Statistics in other EU Member States

22. Statistics Finland has been pioneering automated data collection since 2005, but other Member States have also been testing and implementing more or less similar systems in recent years. INE Spain (The National Institute of Statistical) launched preparatory work in 2004 and this lead to the introduction of a system of automated data collection in 2008. Since then hotels have had the opportunity to send data either automatically from their management systems (if implemented) or through INE Spain's web page. In 2012, holiday dwellings and rural accommodation establishments can also send data using an XML file.

23. INE Spain has been compiling accommodation statistics since 1960. As in Finland, a sample survey is also used in Spain, but given the differences in the tourism economies of Spain and Finland, the sample includes (depending on the season) 9,200 to 11,200 establishments and is therefore much more extensive. In Spain accommodation establishments can also choose the way of reporting the data to the NSI either by using post, email, Internet questionnaire, fax or automated data collection. There has been some interest in using automated data collection, but the implementation phase has been even slower than in Finland. Nowadays some 100 establishments use automated data collection every month.

24. Encouraging results have led to further development also at other National Statistical Institutes. Eurostat launched an 'ESSnet project' involving eight Member States of the EU in 2010. ESSnet projects are instruments for putting together expertise distributed throughout the ESS organisations in order to develop specific actions that would benefit the whole system. Spain took the role of the co-ordinator and seven other countries (Belgium, Bulgaria, Finland, Latvia, Lithuania, Poland and Slovakia) decided to participate in the project as co-partners. The project is due to end in October 2012 but most of the actual work has already been done. The aims of this project were to i) reduce response burden ii) improve timeliness and iii) enhance international comparability and quality of statistics on accommodation establishments by developing a common system that can generate statistical information automatically from management systems. In other words, the aim of this project was to create a common and more or less similar European system that was implemented at Statistics Finland in 2005.

25. The goal of the project was ambitious but the results have been quite successful. During the project, it was understood that it is fairly challenging to create a common system that functions perfectly in different countries. As a result, an applicable common system was created. In consequence, a common part of the XML file, which contains all the mandatory variables required to satisfy the needs of the EU regulation, was created. In addition, all participants created an additional part of the XML file to ensure that national demands are met.

26. Another issue was the technology of the system. Different NSIs have deviating software available and/or purchased. Therefore, it is practically impossible to create a common software solution for the receiving and handling of the data. This meant that all participants created their own architecture and processes for the receiving and handling of the data. On the other hand, this implicates that these solutions are applicable in other domains of statistics at each NSI.

27. Currently six of these countries (Bulgaria, Finland, Latvia, Lithuania, Poland and Spain) have implemented the system and the other two (Belgium and Slovakia) are preparing the implementation which should take place in 2013.

28. The experiences of the ESSnet project and of the participating countries have been quite promising although all the countries encountered similar challenges as Finland and Spain did earlier. The consensus of the project is that the planning and execution have been fairly easy from the technical point of view (creating an XML file and appropriate software, etc.) at least at the statistical offices. The main challenge is once again finding software houses and getting accommodation establishments to implement the system in their hotel management systems.

29. The project group concluded that the recommendations for future work are that i) each country should continue advertising automated data collection to different interest groups and define and use appropriate strategies to get more establishments involved ii) each country should find out whether they are able to expand automated data collection to other types of accommodation establishments than hotels as well as to other surveys iii) the possibility to use automated data collection should be promoted also in other countries if e.g. an international hotel chain has implemented the system in one country.

30. In conclusion, international co-operation is essential if NSIs want to introduce a widespread system of automated data collection which could lead to a significant reduction in the reporting burden of enterprises, and in the processing and compilation burden of the statistical authorities of the Member States. The results of the ESSnet project can be evaluated in a broad sense only in the future. In short term, the results have been satisfactory at least in the countries that implemented the system of automated data collection during the project.

C. Automated data collection in other domains

31. Tourism statistics is not the only field of statistics in which automated data collection could be utilised. Statistics Finland collects some of the data for agricultural statistics automatically and is examining the possibility to expand the mode to data collections from enterprises by using XBRL

reporting in business statistics. XBRL (eXtensible Business Reporting Language) is one of the 'XML'-based languages and is becoming a global standard for exchanging business information. XBRL is commonly used in business reporting and can be used to e.g. define and exchange financial information, such as a financial statement.

32. The current work of Statistics Finland on business statistics is closely related to the European Commission programme for the 'Modernisation of European Enterprise and Trade Statistics' (MEETS). The programme was adopted by the European Parliament and Council in December 2008. The programme, which is to run over a period of five years from 2009 to 2013, will help to adapt business statistics to new needs and also adjust the system for the production of statistics to reduce the burden on enterprises.

33. The main idea is to implement a more efficient way to produce enterprise statistics by better exploitation of the data that already exist in the economy. New information and communication technologies offer opportunities for simplified data collections that should reduce the reporting burden on businesses and improve the quality of statistical information. The workshop on Efficient Ways of Statistical Data Collection from Enterprises, organised by Eurostat in March 2012, showed many new developments in this area. The creation of electronic data collection and processing systems that cover different statistical domains, wide use of the data provided by companies in their annual financial statements and the creation of integrated XBRL taxonomies will all simplify data transfer from companies to NSIs and make statistical output more consistent.

34. In Finland, the first major step towards automated data collection in business statistics was made in 2009 when the 'Fully Integrated Accounting' (FIA) project was launched. The project was co-ordinated by the Aalto University School of Economics and its Real-Time Economy Programme. The programme was studying the requirements to integrate the whole value chain: order-invoicing-payment-accounting-reporting. Its main objectives were to create tools for minimising the administrative burden of enterprises and move towards more efficient financial administration processes within enterprises and between interest groups.

35. In addition, a National Standard for Mandatory Business Reporting was created. That is, an accounting and statistical taxonomy, which was an outcome from a joint project of data users and other associations. The key members included the Finnish Tax Administration, the Trade Register, the Company Analysis Advisory Board, Statistics Finland, the Association of Finnish Accounting Firms and the Aalto University School of Economics. The emphasis of the work was on external reporting between enterprises and public sector organisations. The main idea was to offer faster, cheaper and more consistent official reporting for businesses. The taxonomy could be easily linked when automating financial administration processes. In addition, a similar system can be used for several purposes such as financial statements, tax returns and statistical surveys.

36. The idea of the National Standard for Mandatory Business Reporting is to control the core of public reporting for businesses. The implementation does not need large investments from businesses but their accounting needs to be linked to the taxonomy. The benefits are especially useful for small and medium size enterprises since it enables automatic calculation of certain financial indices and can be used as a chart of accounts for some small enterprises. The national standard was piloted with real data and operator in December 2009, and presented to the public in September 2010.

37. After the previous work the consensus was that the most prominent vehicle for handling the taxonomy would be XBRL/XML reporting. The Aalto University's Real-Time Economy programme held a seminar in Helsinki in September 2011 on XBRL jurisdiction in Finland and the work was launched in March 2012 when the kick-off meeting was held. The XBRL jurisdiction is a joint organisation of 15 members (private and public organisations). The objective of the jurisdiction is to promote the use of XBRL reporting in Finland and to promote

the use of the National Standard for Mandatory Business Reporting. Statistics Finland has decided to accede to the jurisdiction in 22 August 2012.

38. The challenges in the XBRL reporting of the National Standard for Mandatory Business Reporting are very similar to those in accommodation statistics. The XBRL schema can be developed and programmed in about two months but the real challenge is to introduce the taxonomy/XBRL schema to businesses. Co-operation between several interest parties is vital and it is essential to foster the use of the taxonomy/XBRL schema by strongly promoting knowledge about it. In certain European countries, legislation, that is, making it mandatory has been the most effective way. One major issue is also to define the reporting via either a centralised or a dispersed system. There are also other remaining issues such as different needs of data users in the area of e.g. timeliness and a certain need for a centralised data pool. Statistics Finland has an interface for receiving XML/XBRL data (see accommodation statistics) so technically it is possible to receive data at this moment.

39. Current work at Statistics Finland is related to the MEETS programme and one of its main objectives is supporting the implementation of a more efficient way of producing enterprise and trade statistics. EU grants will be provided to national statistical authorities for implementing more efficient ways of producing enterprise statistics and for improving the quality of statistical data. Statistics Finland has applied for a grant in the framework of the Programme for the Modernisation of the European Enterprise and Trade Statistics (MEETS). If accepted, the work will begin in the near future. The description of the work would be similar to those done earlier. The main objectives would include e.g. i) creation and development of electronic systems for efficient statistical data collection from enterprises ii) creation of statistical taxonomies that are integrated into the financial/administrative reporting of companies iii) creation and development of statistical extensions to XBRL financial taxonomies iv) creation of XBRL taxonomies for different business statistics surveys. In short, this would ensure the possibility to report business data automatically from enterprises to NSIs.

III. Conclusions

40. As already discussed above, automated data collection can be utilised in many areas of society including official statistics. The possibilities and, therefore, the benefits of automated data collection in statistical production are obvious and various. The introduction of a widespread system can lead to a significant reduction in both the reporting burden of enterprises and the processing and compilation burden of statistical authorities. It can also improve the timeliness, quality and comparability of statistics.

41. All in all, the results from automated data collection have been quite promising in various EU Member States, such as Finland. In the case of the accommodation statistics of Statistics Finland, the introduction of electronic and automated data collection has led to a significant reduction in the reporting burden of accommodation establishments and to a notable reduction in the processing and compilation burden of the agency. Moreover, it seems that the reduction of the burden has gone hand in hand with improvement of quality. The biggest challenge is to expand the use of the current systems to all or at least to majority of enterprises.