

**Economic and Social Council**Distr.: General
31 March 2011

Original: English

Economic Commission for Europe

Conference of European Statisticians

Fifty-ninth plenary session

Geneva, 14-16 June 2011

Item 3 of the provisional agenda

Organization of data collection and sharing, and the management challenges for the implementation of Statistical Data and Metadata eXchange**Data collection and data sharing at Statistics Netherlands:
yesterday, today, tomorrow****Note by Statistics Netherlands***Summary*

This paper discusses the history of the organization of data collection activities at Statistics Netherlands, starting in the early 1990's. Our story ends with the establishment of a single Data Collection Division in 2011, to conduct social and business surveys. As to the organization of this Division a process-and-knowledge driven approach has been adopted, in which processes that require the same systems, tools and knowledge are combined. A number of shifts can be identified from the developments: a) from stove-pipes to a coordinated system of data collection; b) from single-survey statistics to multi-source/mixed-mode designs; c) from single-survey managers to managers of integrated sets of statistics; d) from local decision making to corporate decision making; e) disappearance of the traditional differences between social and business statistics.

I. Introduction

1. Until 2007, data collection activities at Statistics Netherlands (SN) have been organised separately for social and business statistics. In 2011, a new Division of Data Collection started, after being a pilot organisation for a number of years. In this Division data collection for all statistics is centralised.

2. The drivers for the centralisation and harmonisation of data collection activities have been twofold: internal and external reasons. An internal reason is a more cost-effective organisation of the data collection. The external reasons include the facts that by law (2004), Statistics Netherlands was commissioned to reduce response burden even more, and make use of available register data as much as possible, without affecting the quality of the statistics. On the basis of this law, a strategy on data collection was put in place. This strategy dictated the ways of data collection in general: use register data as much as possible; only if additional data are needed, surveys can be conducted, using Electronic Data Interchange (like Extensible Business Reporting Language (XBRL)), web surveys, paper, Computer-assisted telephone interviewing (CATI) and Computer-assisted personal interviewing (CAPI) (in this order). The result of this strategy is multi-source/mixed-mode data collection designs.

3. This strategy has far reaching consequences for Statistics Netherlands. We see a shift from single-survey managers to content matter experts, who manage integrated sets of statistics. Their statistics will be based on multiple sources and mixed-mode surveys, using advanced statistical modelling. Coordination of definitions of units and variables are important conditions for using administrative data and XBRL. Methodologists involved must be competent in all modes.

4. This strategy also brings about changes in the production processes and systems, e.g. with mixed-mode designs, survey complexity increases. Also (data warehousing) systems for managing all available information has to be present. Going over all developments, the picture emerges that it is no longer possible to design a single survey without taking all other data collections into account. This has consequences for the organization of data collection activities.

5. Currently, this strategy is taken two steps further: new sources for data, as well as new techniques for survey data collection are incorporated. New sources for data include among others, the use of data that are available on the internet (like consumer goods sold on the internet), mobile phone data, and Global Positioning System (GPS) data. New techniques for survey data collection include mobile phone interviewing, sms questionnaire, and skype interviews. The effects of these new technologies on statistics with regard to quality and costs, still have to be explored.

6. In this paper, we will discuss the establishment of a centralized data collection division at Statistics Netherlands and the arguments that have been put forward in favour or against a specific organizational model. We will put this discussion in a historical perspective, starting with the organization of data collection activities at SN in the past. Then we move to the present, and finally we will present some ideas about future developments. A discussion will end this paper.

II. The past

7. Until 1994 data collection activities at Statistics Netherlands (SN) have been organised according to the lines of the various statistics. SN consisted of many small

departments, and each of them was responsible for the production of a separate statistic, like agricultural statistics, statistics on the manufacturing industries, educational statistics, unemployment statistics, statistics on crime and victimization, time-use statistics, expenditure statistics, etc. To get the data, each department ran its own surveys. For each survey a separate process was active, which included determining stakeholders' demands, designing surveys, data collection preparations, survey conducting and monitoring, data validation, data processing, analysis and dissemination. Consequently, little activities were coordinated. Nowadays we would call this a stove-pipe approach.

8. A small number of components of the survey were coordinated, however. With regard to business surveys, a business register (to be used as a sampling frame) was coordinated, as well as unit definitions, but each department took care of the survey design, the fieldwork, and the data processing (like data editing). Each department e.g. had its own staff of field officers who visited businesses to assist in the completion of questionnaires. And each department designed its own questionnaires, as well as its own advance and reminder letters. E.g. the designs for the Short Term Statistics and the Structural Business Statistics differed from one department to another. In fact, some coordination here was carried out by the printing and dispatching department, since all questionnaires were to go through there.

9. In contrast to business surveys, for social surveys more activities were already centralised. The interviewing activities (for face-to-face and telephone interviewing) were centralised in a separate fieldwork department, with standardised interviewing procedures. In this department also some general questions that were used in all questionnaires were harmonised, e.g. the household box. Also, the sampling procedures were standardised, using one sample frame. This department was started in 1975, on the opening of the SN location in Heerlen (as well as the one in Voorburg, close to The Hague).

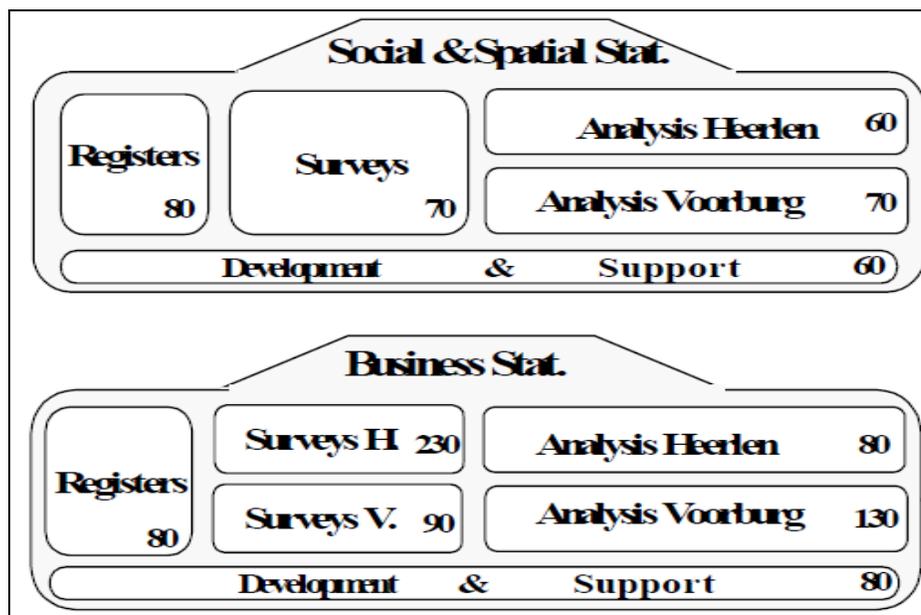
10. In 1994, SN was reorganised and a centralised Data Collection Division was established, next to five content matter divisions, and one supporting division. This Division was aimed at centralising all data collection activities that took place. It existed of five departments:

- a) A social survey data collection department, which was the fieldwork department as mentioned above.
- b) A business survey department, in which all business field officers were centralised.
- c) A business register department, that managed this register and coordinated business units.
- d) A department that took care of the development of electronic data collection, as well as managing registers.
- e) A methodology department that assisted in methodological issues with regard to data collection.

11. However, still many data collection activities took place within the content matter divisions. Within these divisions, it was the general feeling that by giving away data collection activities, they would lose control over the process and the data quality.

12. In 2000, SN was reorganised again. The data collection activities were split according to social and business statistics, and as for business statistics, it was split according to location. Also, separate departments aimed at managing registers were established. This is shown in figure 1 (taken from Keller & Willeboordse, 2000).

Figure 1.
The organization of data collection (surveys) at SN in 2000



13. In essence, the drivers for all these reorganisations have been cost-efficiency reasons (because of budget cuts). It had been argued that data collection activities could be organised in more efficient ways. Still, it was hard to find a way to make it work. Traditionally, managers of statistics are reluctant to change methods, because of changes to the time series. Also they fear that they might lose control over the data collection process, which –according to them– may have serious effects on the data quality.

14. A conclusion we can draw from these developments, is that centralising data collection activities only can succeed when the relationship between data collection and content matter departments is defined clearly and maintained. A second conclusion that comes to mind is that in the past the developments within social statistics were ahead of those in business statistics.

15. In a way, when looking at the organizational chart in figure 1, we could say that the reorganisation of 2000 is a move back to the stove-pipe approach. However, this is not what it seems. Both within the Social and the Business Statistics Division centralisation activities took place, aimed at reducing the overlap in activities, standardisation and harmonization of practices, processes, tools and systems, and thus increasing cost-efficiency. Apart from management reasons, these activities were also aimed at meeting new output demands, i.e. integrating related statistics. The first steps in this direction were made in the second half of the 1990's. Before the 2000 reorganisation, however, many of these standardization and harmonisation attempts failed because of strong obstinacy by local survey managers (as said above). With the 2000 reorganisation, data collection activities and data analysis were put in one hand, separately for social and business surveys. Thus, making integration of statistics at the output and the input level (separately for these statistics) possible.

16. In 2000, for the Short Term Statistics and the Structural Business Statistics e.g. a program was set up to integrate the various designs for these surveys: IMPEctation ECONomic Transformation process (IMPECT). This program was the follow-up of projects which started in the second half of the 1990's, aimed at redesigning economical statistics. Also for the Social Statistics, already before the 2000 reorganisation, programs were started

to integrate related surveys, like in POLS: Continuous Survey on Living Conditions. Also before 2000 (in 1998), the Social Statistics Division started with the development of a centralized Social Statistical Database (SSD), in which a lot of relevant information on persons, household, and quality of life was integrated (Al & Bakker, 2000; Bakker, 2002). Today, the SSD is a large database with standardized retrieval procedures in which social data, both from surveys and registers are stored and shared for many social statistics.

17. Looking back, the 2000 reorganization was a necessary step to achieve standardisation and centralisation of data collection activities, separately for social and business statistics. In the 1990's, the organisation (or should we say the people working at SN) simply was not ready for such a radical change. The more gradual developments in which related surveys were integrated, paved the road to a centralized data collection unit, both for social and business statistics.

III. The present

18. In 2007, the data collection program started. This pilot organisation was aimed at the establishment of a data collection division. At that time, all data collection activities were put together in a single program, with one program leader. This included the Social Surveys Department, and the Business Surveys Departments in Heerlen and Voorburg (see figure 1). At the beginning of 2011 the Division of Data Collection was established.

19. The drivers for the centralisation and harmonisation of data collection activities were twofold: internal and external reasons. Again, the internal reasons include a more cost-efficient organisation of the data collection activities. The external reasons include the 2004 law on Statistics, as well as professionalisation of the survey design, in order to reduce response burden. Also SN was faced with the challenge to produce more timely, coherent and integrated statistics.

20. The external reasons include the fact that by law (2004), Statistics Netherlands was commissioned to reduce response burden even more, and make use of available register data as much as possible, without affecting the quality of the statistics. On the basis of this law, a strategy on data collection was put in place (Göttgens, Snijkers et al., 2005). This strategy dictated the ways of data collection in general: use register data as much as possible; only if additional data are needed, surveys can be conducted, using Electronic Data Interchange (like XBRL), web surveys, paper, CATI, and CAPI (in this order). The result of this strategy is multi-source/mixed-mode data collection designs (Snijkers, 2009).

21. This strategy brings about changes in the production processes and systems: with mixed-mode designs, survey complexity increases. Also (data warehousing) systems for managing all available information has to be present. Going over all developments, the picture emerges that it is no longer possible to design a single survey without taking all other data collections into account. The extension of the use of registers (which started in the 1980's at SN), as well as the development of web surveys and mixed-mode designs, make a centralised approach necessary.

22. Another external reason for the establishment of the Data Collection Division was the professionalisation of survey designs in order to reduce response burden. This includes sample designs, questionnaire design and survey communication (Snijkers, 2008). As for sample design, politics in the Netherlands commissioned SN to implement a system for business survey holidays (even though the actual response burden in the Netherlands is less than .25% of the total caused by government regulations). This caused the necessity of coordinating samples for a number of business surveys (Structural Business Survey, Investment survey). Also questionnaires needed to be coordinated to make sure that information was asked only once.

23. As for survey communication, over the last years SN received more and more complaints about the tone-of-voice of its letters (both by politicians as well as businesses). Society clearly expects clear communication from governmental organisations. Different styles for communication from the same organization, as was the case in the early days with the stove-pipe approach, do not match that expectation. Instead of stating that a survey is mandatory, respondents expect some explanation about the survey and what the data are being used for. Furthermore, one expects communication by using modern techniques that match current social developments. All these reasons pointed in the direction of centralizing data collection activities.

24. But making this happen is not an easy job. We have already seen that local managers of statistics are reluctant to give away their control of surveys. Also, we have concluded that in the 1990's the same kinds of activities were taking place within many departments in the Bureau, using many systems and tools. As a consequence, a lot of SN employees were dealing with the same data collection issues independently, and a lot of effort was put in the development and maintenance of the same kind of systems and tools. This was considered to be inefficient and inflexible, not just for reasons of allocation of resources, and development and maintenance of systems and tools, but also with regard to centralization of knowledge, practices, processes, workflows, and training of staff, as well as establishing a modern attitude towards data collection and response burden.

25. Therefore, the establishment of a single data collection division was aimed at a number of goals:

a) Reorganization of data collection processes, workflows, and activities both for social and business in as few processes as possible. This is aimed at the abolishment of redundant processes, and the establishment of new, efficient processes. With the development of mixed-mode designs and the extension of the use of registers, many data flows have to be managed.

b) Reorganisation of systems and tools, which is aimed at the abolishment of redundant systems and tools, and the putting into practice of new, efficient systems and tools.

c) The clustering and combination of practices and knowledge within one unit, making these practices and knowledge transparent, and subject to improvement. Thus, achieving professionalisation of survey designs, both for social and business surveys, and making a more flexible allocation of staff possible.

d) As for data sharing, coordination of input variables should make sure that SN eventually does not ask for already known information.

26. An analysis revealed that these goals could be achieved for many processes, systems and practices (Program Data Collection, 2009):

a) Data collection management systems. Both for social and business surveys many systems were being used. This made it hard to get unified overviews of the field work status of the various surveys, and thus to adequately monitor and control the actual data collection.

b) Systems were inflexible, outdated and operated without interfaces. It showed that these systems often were developed for a single or a specific set of surveys, and for specific modes. Next, these systems were not fit for the application of new modes like web surveys. This made it almost impossible to apply these systems for other surveys, and new modes like the internet. Also it was hard to have these systems 'talk' to each other. This is a serious barrier for the development of mixed-mode survey designs.

c) Survey design practices. We have seen that many surveys were designed separately, using e.g. various style guidelines for questionnaire design, guidelines for designing survey communication.

27. This may not come as a surprise, considering the situation we came from (as discussed above). In general, we could say that all actions are aimed at establishing a coordinated system of data collection for all procedures, workflows, systems and tools to be used for surveys and registers. Our premises here is that both social and business surveys follow the same process and require the same kind of systems, tools, and procedures, as well as practices, skills and methodological knowledge. Of course, there are differences between social and business surveys (see e.g. Snijkers & Bavdaz, 2011; Rivière, 2002), but from a process and management perspective as well as a knowledge and skills perspective we feel that they are very much alike (except for the very large businesses). This view is backed-up by the Generic Statistical Business Process Model that can be applied to both kinds of surveys (UNECE, 2009).

28. These drivers, developments, strategies and principles gave guidance to the establishment of one Data Collection Division. This new division consists of two departments: Survey Design and Survey Deployment. The organizational setting is based on the way knowledge is being used. This means that knowledge of business and social surveys is combined into one unit focusing on synergy effects.

29. Let us illustrate this process-and-knowledge driven approach with two examples. Until 2007 it was obvious that we would combine the two contact centres at SN that had a role in the data collection process: the Inbound Centre (helpdesk for respondents) and Outbound Centre (CAPI unit) into a single unit. This view was primarily technology driven: both centres used the phone as their primary tool, and both units dealt with respondents. However, the required staff skills and communication process and knowledge differ enormously. Based on this analysis we concluded that the Outbound Contact Centre had to be combined with the fieldwork (CAPI) unit. The integration of these units was taken a few steps further in a very short time. Due to lack of office space we looked for alternatives and we came up with the idea to create a virtual Call Centre setting in which it is possible to work from various locations, including working from home. At the same time, we were looking for solutions to maintain a cost-effective fieldwork organisation that would cover the whole nation. The nationwide coverage came under pressure because of decreasing capacity needs, due to mixed mode surveys. The establishment of the virtual Call Centre also solved this problem by implementing this idea for the fieldwork organisation. This gave us the possibility to deploy field workers also for CATI interviewing. At this very moment we are in the middle of implementing this concepts and it will give us a very scalable, flexible and cost-effective field work organisation, which is ready for mixed-mode designs. To reduce costs even further we now use e-learning facilities to train interviewers in interviewing and refusal-conversion skills. Furthermore, because of these skills, the interviewing staff is not only used for personal interviewing but also for reminding businesses.

30. A second example of combining practices and knowledge relates to the way we finally decided to organize the Inbound Call Centre activities. The primary competence for these employees is giving adequate support to respondents, companies and institutions, with regard to the content of surveys as well as technical support. And, since this kind of support was also necessary for our interviewers to do their job, it was only logical that this Call Centre was merged with other supporting activities like technical assistance to the interviewers. The creation of a single support unit by itself creates an optimal feedback loop in further optimizing the data collection activities and an optimal use of knowledge, since all experiences on external feedback is concentrated in one single point.

31. An aspect that explicitly had to be dealt with was gaining support from local survey managers, i.e. establishing a relationship with the internal customers, the content matter departments. A separate unit, the Front desk (as part of the Design Department) consisting of account managers, was established to identify and manage the needs of our customers. Basically, this is a matter of trust!

32. With this newly established Data Collection Division SN is ready for future developments, even though there is still a lot of work to be done before we reach the stage we want to be. It may be clear that we are only at the beginning of this, and the question is: Will we ever be ready? In the future, the data collection processes will even be more diverse and complex to handle, than it is today. New technological developments and new demands for statistics bring this about.

IV. The future

33. The challenges SN, like any national statistical institute (NSI) in the western world, faces is the demand for more and integrated information. It's not separate social and business statistics anymore, but there is a demand for statistics that provide an overview of specific themes, like social coherence, globalisation, global warming, knowledge-based economy, etc. These demands require the integration of statistics that originally would be considered to be social or business statistics. The dividing line between social and business statistics becomes vague; it is not the measurement unit that is relevant but whether data are used for describing social or economic concepts. Also these statistics have to be published in a shorter period of time, with less money, and less response burden.

34. In the way we collect the data we need, NSI's are also faced with new techniques and technologies that can be implemented (see e.g. Conrad & Schober, 2008; Daas, 2011), and combined in multi-source/mixed-mode designs. This has consequences for the organization of data-collection processes and methodological knowledge: methodologists must be competent in all modes, and must have knowledge of the effects on data quality and response burden.

35. This requires a constant adaptation of the organization and the way we collect our data to the demands and requirements we face, both with regard to our outputs and our inputs. As a consequence, currently the data collection strategy of SN is taken two steps further: new sources for data, as well as new techniques and technologies for survey data collection are incorporated. New sources for data include, among others, the use of data that are available on the internet (like consumer goods sold on the internet), mobile phone data, and GPS data (Daas et al., 2011). New techniques for survey data collection include mobile phone interviewing, sms questionnaire, and skype interviews. The effects of these new technologies on statistics with regard to quality and costs, still have to be explored, as well as risk assessments for new sources with regard to e.g. stability of content, timely and reliable delivery/availability.

36. The new data collection strategy consists of three steps (Hermans, Snijkers & Roos, 2011):

a) In the first step, the data that are available within SN are investigated, maximizing the re-use of data. This requires the existence of a Data Service Centre for data warehousing and data sharing.

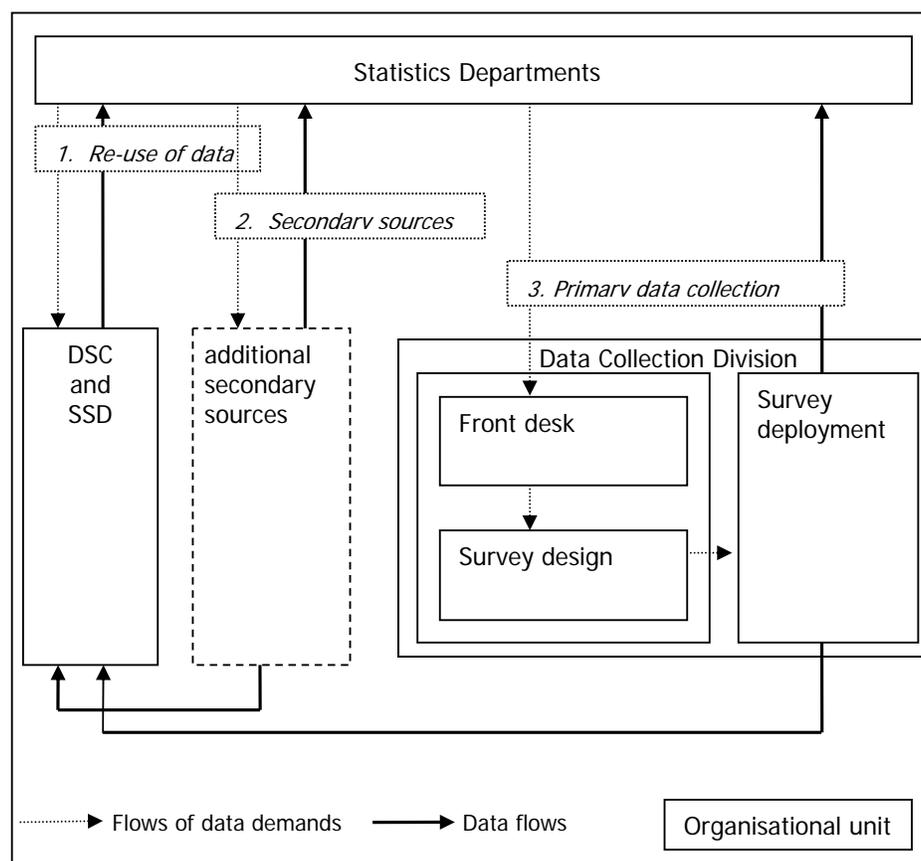
b) Secondly, data from registers and other sources are used. This includes the use of secondary sources in the broadest sense: registers, data from private businesses (like e.g. mobile phone providers), information that is available on the internet (using web-crawlers).

c) And finally, if still more data are needed, surveys are conducted. This includes in this order: Electronic Data Interchange using e.g. XBRL (Roos, 2010), web surveys, and finally traditional modes like paper, CATI and CAPI). New technologies open up new possibilities for these modes, like cell phones, questionnaires via text messages, skype interviews, etc.

37. In this strategy data sharing becomes an important step in the production of statistics. Currently, SN has a Data Service Centre (DSC) that operates as a kind of library for registers, next the Social Statistical Database (as mentioned in section 2). In the light of these developments, the DSC needs to be extended as a library of all available statistical data. From an organisational perspective we get a process of statistics production as is shown in figure 2.

Figure 2.

Producing statistics at Statistics Netherlands in the future



V. Discussion

38. Looking backward on where we came from, two important developments can be identified: the shift from single-survey statistics to multi-source/mixed-mode designs, and from stove-pipes to a coordinated system of data collection activities. This resulted in the establishment of a centralized Data Collection Division in 2011, even though we are still in the midst of establishing this coordinated system. These developments caused a number of shifts in the organization as a whole.

39. These developments affected the span of control and the focus of local statistics managers. Instead of running individual surveys, local managers become managers of integrated statistics and look where to get the information they need. Thus, they become content matter experts. Consequently, their span of control shifts from managing individual surveys to managing data flows.

40. Another shift we see is from local decision making to corporate decision making. Local statistics managers and managers of data collection processes are subject to corporate strategies on data collection, in which harmonized systems, tools, and procedures are used. Nowadays, it would be quite unimaginable that within Statistics Netherlands, each statistics department would have to deal with multi-source/mixed-mode designs on its own.

41. In this process of harmonization we have adopted a process-and-knowledge driven approach, in which processes that require the same systems, tools and knowledge are combined. As we have illustrated, staff can be flexibly allocated, since they have the same knowledge and practices, and since they use the same systems and tools.

42. In our approach, both social and business surveys are treated in the same way, instead of allocating them to separate departments. Basically, they follow the same processes. In this way, these surveys can be managed and monitored effectively, using the same systems and procedures. This is a third shift: the traditional differences in treatment of social and business surveys disappear. An aspect we still have to deal with, however, is differences in culture for staff that was focused on either of these populations. This still prevents staff from effectively learning from each others practices, skills and experiences.

43. An issue that is discussed among our staff is the effects of this organizational model on survey data quality. They argue that social surveys and business surveys need to be designed differently, and even some local statistics managers still argue that e.g. manufacturing industry surveys need to be designed differently than construction industry surveys. And they are right! But the arguments they use have to do with methodological issues, and not with organizational and process management issues. Harmonisation of processes does not mean that you cannot tailor the survey design to special populations. It does mean that a survey design is restricted by the coordinated system of data collection. An issue that is of importance, however, is that this coordinated system should meet methodological standards, in order to get good survey data collected with reduced compliance costs (i.e. response burden).

VI. References

Al, P., and B. Bakker (2000). Re-engineering social statistics by micro-integration of different sources; an introduction. Netherlands Official Statistics (Special Issue on Social Statistical Database), Vol. 15, No. Summer, pp. 4-6.

Bakker, B. (2002). Statistics Netherlands' Approach to Social Statistics: The Social Statistical Dataset. OECD Statistics Newsletter, Vol. 2002, No. 11, Pp. 4-6.

Daas, P., M. Roos, Ch. de Blois, R. Hoekstra, O. ten Bosch, and Y. Ma (2011). New data sources for statistics: Experiences at Statistics Netherlands. Paper presented at the 5th European Conference on New Techniques and Technologies for Statistics, 22-24 February 2011, Brussels, Belgium.

Conrad, F.G., en M.F. Schober (2008). Envisioning the Survey Interview of the Future. Wiley, Hoboken.

-
- Göttgens, R., D.J. Beukenhorst, G. van Dam, G. Snijkers, and M. Verbruggen (2005). CBS Data Collection Strategy (in Dutch: CBS-brede waarneemstrategie), Report Statistics Netherlands, Heerlen/Voorburg.
- Hermans, H., G. Snijkers, and M. Roos (2011). CBS Data Collection Strategy, version 2011: “Collection and delivery” (in Dutch: CBS-brede waarneemstrategie, versie 2011: “Halen en brengen”). Report Statistics Netherlands, Heerlen/The Hague.
- Keller, W., and A. Willeboordse (2000). New Methods for Statistical Processing in a new organization environment. Proceedings of the 2nd International Conference on Establishment Surveys, American Statistical Association, pp. 403-412.
- Program Data Collection (2009). Annual Plan Data Collection 2009 (in Dutch: Jaarplan Data Verzameling 2009). Report Statistics Netherlands, Heerlen/The Hague.
- Rivière, P., (2002). What makes Business Statistics special? International Statistical Review, Vol. 70, No. 1, pp. 145-159.
- Roos, M. (2010). Using XBRL in a Statistical Context. The Case of the Dutch Taxonomy Project, Journal of Official Statistics, Vol. 26, No. 3, pp. 559–575.
- Snijkers, G. (2008). Getting Data for Business Statistics: A Response Model. Proceedings of the 4th European Conference on Quality in Official Statistics, Rome, Italy.
- Snijkers, G., (2009). Getting Data for (Business) Statistics: What’s new? What’s next? Paper presented at the 2009 European Conference for New Techniques and Technologies for Statistics (NTTS), 18-20 February 2009, Brussels.
- Snijkers, G., and M. Bavdaz (2011). Business Surveys. In: M. Lovric (ed.): International Encyclopedia of Statistical Science. Springer Verlag, Berlin.
- UNECE (2009). Generic Statistical Business Process Model. Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS): www.unece.org/stats/gsbpm.
-