Issues in Automated Financial Data Collection in The Netherlands

Gert Buiten[1], Ran van den Boom[1], Marko Roos[2]
Ger Snijkers[2]
[1]Statistics Netherlands, Henri Faasdreef 312,
Den Haag, 2492 JP, The Netherlands
[2] Statistics Netherlands, CBS-weg 11,
Heerlen, 6412 EX, The Netherlands

Abstract
This paper describes Statistics Netherlands' long-term vision on automated financial data collection. The basis for this vision is to become an integral part of the chain of financial administrative processes in such a way that in the long run an almost complete micro data set with financial data of all Dutch companies can be constructed, with automated data deliveries for Small and Medium Enterprises (SME's) on a voluntary basis and the delivery of statistical feedback data in return. After an overview of the general context and goals, the paper discusses the importance of standardisation and the solution chosen in The Netherlands. Next, the current status is summarized. Finally, the major issues encountered so far are discussed, focusing on statistical issues, implementation issues, feedback data issues and organisational issues.

Key Words: SBR/XBRL, automated data collection, Reference Classification System of Financial Information, RCSFI, Standardised Reference Chart of Accounts

1. Statistics Netherlands' Approach to Automated Financial Data Collection

Information about businesses and their activities is generated in a chain of administrative processes. This chain is gradually being automated, opening up the possibility for automated data collection. This chapter describes Statistics Netherlands' approach to integrating and automating the steps in the chain.

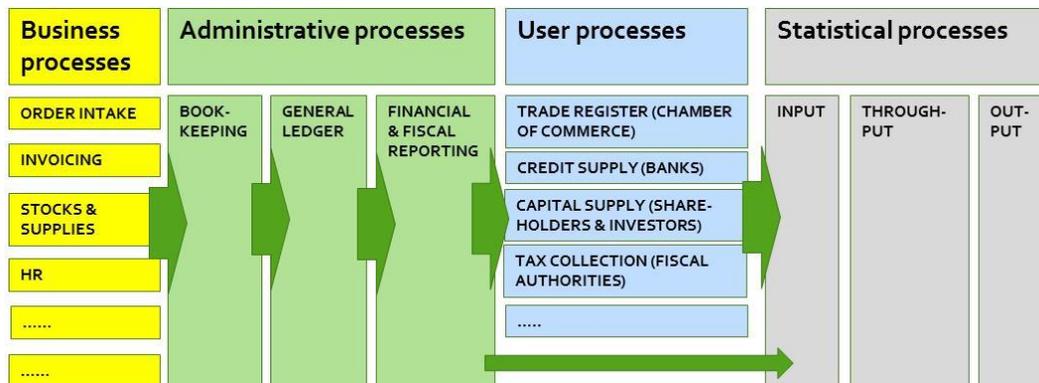### 1.1 The Chain of Administrative Processes
The input data for business statistics is generated by a chain of administrative processes. More and more of these processes are supported by or even performed with software, increasing the amount of digitally available data. Figure 1 shows a simplified picture of this chain.
Businesses activities are carried out in business processes that usually generate some kind of recording of these activities. Examples are order intake, invoicing and billing, stock keeping en human resources (e.g. wage administration).
Business processes feed data into a series of administrative processes by which the business translates the data on activities into dashboards, data on financial results, financial reports, tax declarations, etcetera. The first process is bookkeeping, in which data from various business processes are fed into one central system and coded

accordingly. The second process is the general ledger, in which the codes from the book keeping system are classified into one or more sets of profit and loss accounts and balance sheets. The third process is financial and fiscal reporting, in which selections of data from the general ledger are used to compile reports, tax declarations etcetera for specific users inside and outside the business.

Different users have their own processes in which they use the business reports or data for their own purposes, such as tax collection, credit provision, capital provision etcetera. Finally, the statistical processes obtain their data from both the administrative processes of companies and the user processes.

Figure 1: The chain of administrative processes

1.2 Statistics Netherlands' View on Data Collection in the Chain

Currently, data collection by Statistics Netherlands and most other National Statistical Institutes (NSI's) focuses 1) on the re-use of administrative data gathered by other users like the Tax authority and the Chamber of Commerce and 2) on the use of survey data from the reporting processes of companies (Snijkers et al., 2011).

In addition to this, Statistics Netherlands aims to use digital data that can be collected without additional response burden on businesses. This data comes in two types: 1) data obtained directly from the software systems used in the primary business processes, i.e. (big) data on order intakes, payments, receipts and so on and 2) data generated by the financial administrative processes within the companies, especially the general ledger.

By using these two types of data, Statistics Netherlands aims to fully integrate data collection with the chain of administrative processes. Using this approach, in the long run a practically complete, high quality micro data set can be obtained from SME's. This vision is visualised in figure 2.

For large and/or complex companies, data collection will continue to be mainly based on a survey approach using samples and questionnaires, in which companies will have to make an effort to provide data according to the definitions, classifications and concepts as provided by Statistics Netherlands. One important aspect is that a number of financial variables are required on a consolidated basis, with intra-company transactions eliminated. We are studying this subpopulation in order to reduce the survey approach as much as possible.

For small and medium enterprises, input data will focus on the use of register data from users, (big) data captured from the business processes and financial data collected from the administrative processes. In the latter case, data is preferably collected from administrative intermediaries such as bookkeeping companies, accountant firms and

providers of online accounting software. In each of these cases the burden on the individual business is minimal or completely absent.

Only for variables that cannot be obtained in any of these three ways, additional surveying on a sample basis is envisaged.
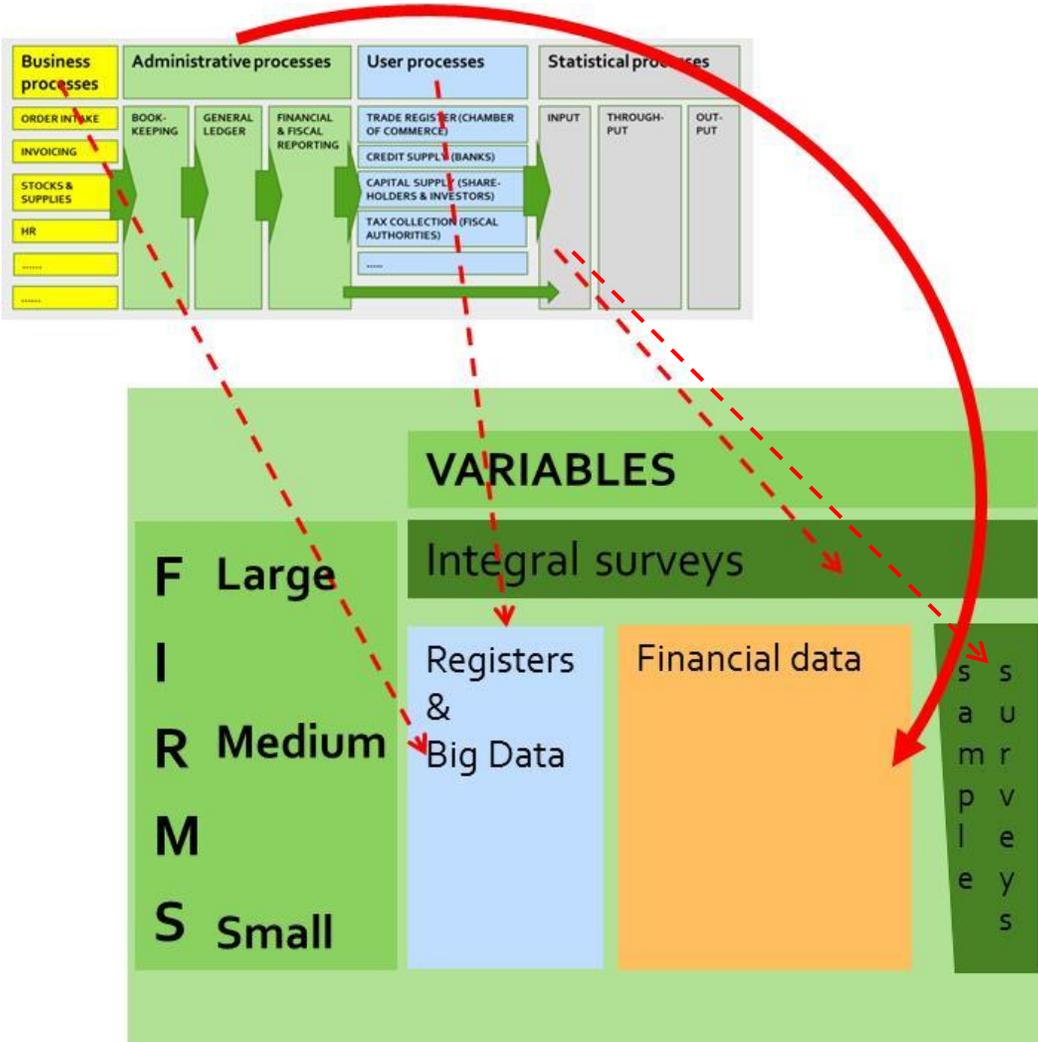
**Figure 2:** Statistical data collection in the chain of administrative processes

1.3 Major Pre-conditions for Automated Data Collection

There are two major pre-conditions that have to be fulfilled before this vision can be realised.

Firstly, businesses must be willing to deliver digital data on a voluntary basis. Otherwise data delivery will simply result in an increase in the response burden on businesses. Therefore, an essential part in this strategy is that Statistics Netherlands not only collects data, but also provides relevant feedback data to businesses – either directly to individual firms or indirectly via administrative intermediaries such as accounting firms and software providers.

Secondly, the information in the chain needs to be standardised. Standardisation is needed for two reasons. On the one hand, meta data have to be standardised to make data from different businesses comparable. This is currently not the case in The Netherlands: each company uses its own chart of accounts, its own balance sheet format etcetera. On the other hand, the meta data must be machine readable in order to be processed automatically.

Both types of standardisation can be achieved by the use of a 'Reference Classification System of Financial Information' or in other words a 'Standardised Reference Chart of Accounts'.
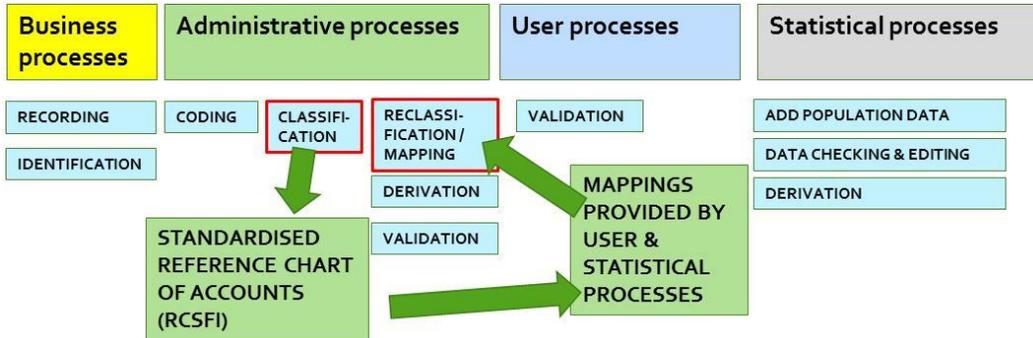
Figure 3: Main statistical actions in the chain of administrative processes

Such a Reference Chart of Accounts can play a crucial role in improving communication between steps within the chain and thus creates possibilities for a stronger integration of the chain. This can be explained by examining the actions in the chain of administrative processes that facilitate (statistical) reportnig. These actions are highlighted in Figure 3. In the business processes, data are recorded and identified in one way or another – e.g. by product codes, shipment numbers, bar codes, customer ID's and the like (see e.g. Zhang, 2011). In the administrative processes, these data are first of all coded into an overall coding system (that might be the same as the general ledger classification, but this is not necessarily always the case). Next, they are classified using the general ledger classification that can be seen as a coherent and (usually) hierarchical classification in the form of a chart of accounts. Based on this classification, various reports are generated by means of reclassification or 'mapping'.

Figure 4 provides a simplified example of a mapping. Data from the general ledger are selected and assigned to the variables required for the report. In this process, some data from the general ledger may be aggregated. Also, new variables may be derived from the general ledger data, such as ratio's, balancing items, etcetera.
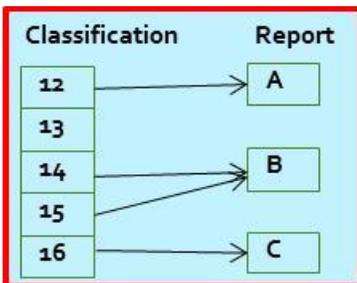
Figure 4: Example of a mapping between general ledger and a report

The use of machine readable mappings creates two essential possibilities. First of all, it allows this process step to be fully automated. Secondly, it allows the user processes to supply to the administrative processes all the necessary information for creating a report. Thus, an important step can be taken towards further integration and automation of the chain of administrative processes.

For this purpose, a group of private-sector bodies, software companies and government institutes have developed a Reference Classification System of Financial Information (RCSFI). Statistics Netherlands was one of the parties involved. In the Netherlands such a reference standard is important, since businesses are legally allowed to use their own formats and codes for bookkeeping, general ledger, profit and loss accounts and balance sheets. Legally prescribed templates only apply to specific reports, like tax declarations.
Table 1 gives an example of a selection of the RCSFI and a mapping. The first column of the table shows the unique alphabetical code that is assigned to a specific item in the RCSFI. Using these unique codes, software developers can automate data processing, importing and exporting routines. The first column also shows the hierarchical nature of the RCSFI, in which items are broken down into underlying details. Businesses are able to choose the level of detail at which they wish to record their data, depending on their specific situation like types of costs and revenue.

The RCSFI is explicitly positioned as a reference classification and not as a mandatory standard. Any company wanting to enjoy its benefits is free to use its own formats, codes and the like, as long as these can be mapped onto the RCSFI. This flexibility makes the reference classification a useful tool for data processing, importing and exporting with various types of administrative software packages. Therefore, a number of software developers have also joined the group that developed the RCSFI.

Table 1: Example of the Reference Classification System of Financial
Information (RCSFI)

Reference code	Description	Mapping
BVrd	Stocks	12400
BVrdHan	Trading goods	12410
BVrdHanVoo	Stocks of trading goods	12411
BVrdHanHW	Re-valuation of stocks of trading goods	12412

The technical format for automated data transmission between various software packages and between different parties used in the Netherlands is the XBRL-format in the context of Standard Business Reporting (SBR). Providing digital reports by SBR has been made mandatory by the Tax authority and the Chamber of Commerce. Hence it also became the standard for a number of other uses, like financial reports that are required by private banks..

2. Current Status

In realising the vision of automated administrative data collection, considerable progress has been made but there is still a long way to go.
Statistics Netherlands is firmly embedded in the network of parties that operate the chain of administrative processes: software providers, accounting companies, bookkeeping firms, the tax authorities, banks and so forth. This network has been formalised into a

private/public governance structure centered around SBR. Both SBR/XBRL and the RCSFI are included in this governance.

Within the governance structure, a roadmap has been developed for the implementation of SBR and XBRL. This includes a calendar showing the dates as of which specific types of tax declarations and deposits of balance sheets can only be delivered via XBRL-messages. This applies to all accountants and other intermediairies that file on behalf of companies. Companies by themselves are still allowed to manually fill out the digital reports on the Tax Authority portal.
The largest private banks have implemented a policy by which XBRL is actively promoted as the main channel for credit reporting to banks.
Statistics Netherlands itself has opened a channel for XBRL-messages for a number of short term statistics. The extension to other statistics is underway.

Version 1 of the RCSFI was released in 2014. Its use is not mandatory, so dissemination and actual use are done on a voluntary basis. At the moment, some of the largest software providers support the use of RCFSI in their software. An increasing number of smaller and more recent entrants into the market are also supporting RCFSI. The next step is that accounting firms and individual companies implement the RCSFI in their administrative processes. This requires an investment, since for each business a mapping must be made from the current classification used to the RCSFI. This additional investment is lowest for new businesses that have to set up their administration from scratch. For the others, various initiatives are being taken to reduce the required investment. Recently, the first commercially available tool supporting this mapping process was released.This software automatically recognises commonly used general ledger accounts and assigns them a RCFSI code.

A number of software companies are using the RCSFI to facilitate the transfer of data between software packages. For example, providers of dashboard and management information software are creating import tools based on RCFSI. This will facilitate the first time use of their software for new customers. This in turn creates an added value to implementing RCSFI from the point of view of businesses (especially administrative intermediaries).

The first XBRL messages based on RCSFI have been received by Statistics Netherlands. To speed up the implementation of RCSFI, a 'leading group' has been formed with a number of innovative software providers and accounting companies. This group is also developing various ways through which Statistics Netherlands can provide feedback data to companies. The first pilot projects are underway.

Standardisation is the key to success, but this is not only in the hands of Statistics Netherland – nor in the hands of any other individual party. It is for the market to pick this up and make it spread like an oil slick. What Statistics Netherlands can do is stimulate that process.

3. Major Issues

The effort to get Statistics Netherlands' vision on automated data collection up and running is facing a number of issues. The most important issues will be discussed below,

grouped by statistical issues, implementation issues, feedback data issues and organisational issues.

3.1 Statistical Issues

Four major statistical issues have been encountered so far. First of all, non-financial variables can usually not be provided directly from general ledger. Examples are retail area in square meters and hours worked. Possible solutions to this are the use of survey data from large companies for imputations for SME's and the use of alternative sources.

A second type of issue involves non-recorded detail, such as exports to Eurozone and non-Eurozone countries. At best, businesses keep track of the country in which their clients reside. In the short term, research focusses on the use of survey data from large firms as a basis for imputations. In the long run, efforts focus on including these data in administrative software.

A third issue arises from the freedom that businesses have to choose their own levels of aggregation for the individual types of costs, income and the like. In general, companies will keep detailed records of items that are important for running their business processes or that are mandatory for legal reasons. Less relevant items are only coded at higher levels of aggregation (e.g. property tax is directly coded as housing costs). What is and is not important may differ between various branches of industry as well as between individual companies within an industry. The approach taken to this problem is to consider this as a type of item-non response and use standing imputation techniques to standardise on the required level of aggregation.

A fourth group of issues involves the possibility of differences between statistical and administrative units. For most small enterprises there usually is no difference, but for a number of medium enterprises the statistical units may contain several administrative units. An example is a manufacturing company with a separate sales unit that also sells products from other companies. For statistical purposes, the transactions within the statistical unit have to be consolidated. In the short term, this problem is circumvented by focusing on the 1-on-1 cases. In the medium long term, an approach could be to use units that consolidate towards other administratieve units, e.g. the unit for the Tax authority. In the long run, the solution is to increase the consolidation functionalities in administrative software. A coherent implementation of the RCSFI in the administration of the companies concerned makes it much easier to perform this type of consolidation.

Discussions are also underway with users of statistical data on the extent to which variables are actually required for SME's separately. The basic proposition in these discussions is that at times it may be necessary to sacrifice details (for specific variables) in order to gain in other areas (data on practically all SME's, providing much more detail by e.g. activity, region).

Where data are truly necessary and if all others approaches fail, a last resort is to include these variables in an additional survey. Such a survey might be conducted on an annual basis, or on a multi-annual frequency.

3.2 Implementation Issues

A number of implementation issues of a technical nature linked to SBR/XBRL and RCSFI have been encountered. A first issue is that mapping a detailed set of codes in individual administrations to the RCSFI can be a considerable effort. One solution is to provide mapping examples and best practices to support this effort. These examples are disseminated both through the RCSFI website as well as in bilateral contacts with e.g.

branch associations. As mentioned above, one software developer recently released a RCSFI-mapping support tool which will facilitate this process.

A second issue is the hierarchical nature of the RCSFI in combination with freedom of level of detail. This may lead to mapping errors as well as to ambiguity about the content of the 'Other …..' items (as in 'Other housing costs'). Mapping errors arise when both a total and one or more of the underlying items are included, resulting in double counting. Conversely, including only part of the underlying items may lead to omissions. Ambiguity in the nature of 'Other….' items arises if a total is mapped and only part of the underlying items. Recalculating the 'Other ….' items implicitly includes one or more of the missing items. These issues are tackled by providing a set of guidelines (like using RCSFI codes at the lowest level of detail as possible) and conformance suites for software suppliers.

A final implementation issue is that currently the nature of mutations as well as a number of consolidation items are missing from the RCSFI. The way forward here is to investigate – together with software developers and accounting companies – where in the administrative chain this information originates and how it can be included in the RCSFI.

3.3 Feedback Data Issues

The idea of delivering relevant feedback data to companies also leads to issues, that center around three questions: what information, in which format and through which channel? These issues are being investigated in co-operation with a leading group of innovative software and accounting companies. Suggestions for the kinds of feedback information on so far include: variables from the surveys involved, additional detail by e.g. region and activity and quarterly data on otherwise annual figures. Also, benchmarks data could be provided that allow companies to assess how they are performing in comparison with other relevant groups of firms. Other possibilities are data that help in evaluating the business model of a firm. For example, do companies that outsource certain activities have better financial results?

Regarding the format and channel of feedback data, two approaches have been identified. First of all, a 'push' approach that includes specific types of publications or products made by Statistics Netherlands and subsequently disseminated to businesses and/or their administrative intermediaries. Examples include PDF-publications and e-massages. A pilot in this area is underway, which hopefully also helps to further improve the business case for implementation of the RCSFI for businesses. Secondly, a 'pull' approach centers around disseminating feedback data in the form of open data that is fed directly into systems used by accountants or companies via webservices and API's (Application Programming Interface). This requires efforts by software providers, but opens up the possibility of including feedback data directly into dashboards that accountants provide to their customers. A pilot is now underway that builds on Statistics Netherlands' existing Open Data-platform.

A more general issue with regard to feedback information is that Statistics Netherlands must not become a competitor to private businesses that earn money by providing data. This requires that we operate carefully and in a co-operative way.

3.4 Organisational Issues

Finally, there are a number of organisational issues. Some of these are direct consequences of the approach adopted by Statistics Netherlands, others have arisen while elaborating the approach and trying to put things in practice. These issues center around the ultimate goal of integrating statistical data collection with the financial-administrative

information chain. To achieve this, Statistics Netherlands must be an accepted partner and player in this chain.

First of all, Statistics Netherlands must establish a different relationship with enterprises than is currently the case: one that is based on co-operation and mutual benefits, as opposed to a one-sided relationship based on mandatory data deliveries.

Secondly, part of this new relationship is the close co-operation between Statistics Netherlands and software providers and bookkeeping and accounting companies as intermediate providers of enterprise data and as disseminators of feedback data. In the long run, this opens up possibilities for collecting data for statistical purposes without burdening individual firms. An interesting example has come up recently in a pilot with one branch association of accountancy companies. This association has established a benchmark platform for its members were the data are automatically uploaded when the status of the underlying data changes into 'draft version'. In other words: once the required setting has been made in the software program, data delivery is done *without* the click of a button.

Advocating the new approach and a widespread use of RCSFI has to be done in close co-operation with these partners. Businesses have to be convinced that investing in the implementation of RCSFI has a positive business case, making reporting easier, lowering response burden and providing valuable feedback data. This requires a co-ordinated and sustained Public Relations – and takes time.

A third aspect of this issue is that Statistics Netherlands must learn to speak the language of businesses – instead of approaching businesses with statistical concepts, definitions and the like. This means that statisticians become responsible for working out how statistical concepts can be derived from business data instead of expecting each individual firm to figure this out by themselves for every survey questionnaire and technical explanations. The RCSFI and the mappings based on this is an important tool in bridging this gap between business and statistics.

Fourthly, this new approach requires that Statistics Netherlands invests time and effort in building up and maintaining a network of contacts in the private sector.

Fifthly, sufficient knowledge of SBR/XBRL and of the RCSFI have to be built up and maintained within Statistics Netherlands.

Sixthly, taking automatically generated data into statistical production requires a stepwise implementation rather than a Big Bang approach. Statisticians have to build up experience with the quality of the data to avoid quality loss or unnecessary breaks in time series.

Finally, a number of internal processes and structures need to be adapted in order to reap the full benefits of this new approach. The chain approach opens up the possibility of influencing 'at the source' the way data is recorded, processed and delivered as a statistical input. This requires organizing internal processes to better make this connection, such as by using the RCSFI internally and providing and maintaining the necessary mappings. Some statistical checks within our own processes will become less important, whereas new ones need to be added.

4. Conclusion

This paper describes Statistics Netherlands' long-term vision on automated financial data collection. The basic idea is to become an integral part of the chain of financial administrative processes in such a way that an almost complete micro dataset with financial data of all Dutch companies can be constructed, with automated SME data deliveries on a voluntary basis and the delivery of statistical feedback data in return. This is a long term investment, of which only the first steps have now been taken. Until now, the reactions of accounting firms, software providers and individual business have been

positive and promising. However, there is still a very long way to go. We firmly believe that this is the future for statistical business data collection.

References

The RCSFI website http://www.referentiegrootboekschema.nl/home
'Roadmap SBR op weg naar 2020'
http://www.sbr-nl.nl/fileadmin/SBR/Werken_met_SBR_Roadmap/Herijkte_Roadmap_SBR_op_weg_naar_2020_definitieve_versie_2016-07-19.pdf

Statistics Netherlands, Press release https://www.cbs.nl/nl-nl/uitgelicht/cbs-ontvangt-eerste-berichten-via-sbr

Snijkers, G., Göttgens, R., and Hermans, H. (2011a), Data collection and data sharing at Statistics Netherlands: Yesterday, today, tomorrow. Paper presented at the 59th Plenary Session of the Conference of European Statisticians (CES), United Nations Economic Commission for Europe (UNECE), Geneva, June 14–16.

Zhang, L.-C. (2011), Topics of statistical theory for register-based statistics and data integration, Statistica Neerlandica 66(1): 41–63.