DE GRUYTER OPEN

Journal of Official Statistics, Vol. 34, No. 2, 2018, pp. 419–443, http://dx.doi.org/10.2478/JOS-2018-0019

# Business Data Collection: Toward Electronic Data Interchange. Experiences in Portugal, Canada, Sweden, and the Netherlands with EDI

*Gert Buiten*[1], *Ger Snijkers*[2], *Paulo Saraiva*[3], *Johan Erikson*[4], *Anna-Greta Erikson*[4], *and Alice Born*[5]

This article discusses the experience and the ideas of National Statistical Institutes from four countries – Portugal, Sweden, Canada, and the Netherlands – in order to build a fully automated data collection system, to provide a system-to-system (S2S) data exchange or Electronic Data Interchange (EDI) between all stakeholders in the production chain. This joint work is a summary of an invited session at the Fifth International Conference on Establishment Surveys, which was devoted to 'the future of business data collection'. Taken together, the four presentations provide an overview of recent experiences with S2S/EDI data collection for financial business data. The basis for such a system is an integrated unbroken digital information chain that runs from the recording of financial data in computerised administrative systems of individual businesses all the way to publishing economic statistics – the Business Information Chain. This chain can be 'closed' and made into a cycle by including a feedback loop, for example by providing benchmark data to businesses. However, to make it happen, technical standardisation, vertical and horizontal conceptual harmonisation between all partners in the chain, and positive business cases for all partners are needed. The article starts by putting EDI developments in historical perspective.

*Key words:* Economic statistics; business information chain; reference chart of accounts; integrated data system; standardisation, harmonisation; S2S; XBRL/SBR.

## 1. Introduction

In the field of business data collection by National Statistical Institutes (NSIs), an evolution can be seen from traditional paper collection as the dominant mode to a

[1] Statistics Netherlands, PO BOX 24500, 2490 HA, The Hague, The Netherlands. Email: g.buiten@cbs.nl
[2] Statistics Netherlands, PO BOX 4481, 6401 CZ, Heerlen, The Netherlands. Email: g.snijkers@cbs.nl (corresponding author)
[3] Statistics Portugal, Av. António José de Almeida, 1000-043 Lisboa, Portugal. Email: paulo.saraiva@ine.pt
[4] Statistics Sweden, PO BOX 24300, 104 51 Stockholm, Sweden. Emails: johan.erikson@scb.se and annagreta.erikson@scb.se
[5] Statistics Canada, 150 Tunney's Pasture Driveway, Ottawa, Ontario, K1A 0T6, Canada. Email: alice.born@canada.ca
**Acknowledgments:** This article is a fruitful collaboration between the authors of four countries of an invited session at the Fifth International Conference on Establishment Surveys, which was devoted to 'the future of business data collection'. It represents the results of an extensive network of senior experts of Statistics Portugal, Statistics Netherlands, Statistics Canada, and Statistics Sweden. Gratitude is owed to Ger Snijkers, Statistics Netherlands, for recognizing the value of those experiences and ideas on modernising official statistics production as building blocks of an overall approach to automated data collection for financial business data, and bringing together the essential elements and motivations. We would like to thank the reviewers and the guest editors for their helpful comments and suggestions. **Disclaimer**: The views expressed in this article are those of the authors and do not necessarily reflect the official policy of their National Statistical Institute.

multi-source/mixed-mode approach, in which register data is used in combination with computer-assisted (i.e., human-to-system, H2S) data collection modes, such as web questionnaires. Because of this development, Zhang (2012, 21) states that "the 20th century witnessed the birth and maturing of sample surveys; the 21st century will be the age of data integration." This integration not only applies to the integration of data sources like registers and surveys, but also the integration of micro data within administrative business processes, as well as the integration of these micro data into data collection processes.

In addition to H2S data collection modes, system-to-system (S2S) data collection, also called Electronic Data Interchange (EDI), has been emerging as result of a further computerisation of administrative business processes (Bharosa et al. 2015; Couper and Nicolls 1998). Even though the start of S2S data collection dates back to the 1980s, its implementation continues to be challenging, despite the obvious advantages, such as a lesser administrative burden for businesses and more (timely) primary data for NSIs. This article discusses EDI/S2S business data collection and its implementation, and quality challenges. We focus on EDI for the collection of financial business data. This EDI methodology builds on an integrated, unbroken digital chain of financial administrative processes with as little human intervention as possible, and ideally runs from the recording of single financial data in digital administrative business systems all the way to data analysis and the dissemination of economic statistics. In its most recent views, Eurostat (2017), would consider such a fully automated chain to be a system of smart statistics. This chain can be 'closed' and made into a cycle by including a feedback loop, for example by providing benchmark data to providers, which would result in a continuous intelligent business learning system, using smart statistics.

This article is a summary of an invited session on the future of business data collection at the Fifth International Conference on Establishment Surveys, with contributions from four countries: Portugal (Saraiva 2016a), Canada (Born 2016), Sweden (Erikson et al. 2016), and the Netherlands (Buiten et al. 2016). Each of these countries has experiences with EDI/S2S data collection. These four countries are not representative, but their experiences and future plans on EDI/S2S data collection give insights into implementation approaches and challenges.

Before discussing EDI/S2S, in Section 2 we present a brief overview of the evolution of business data collection to put this method into historical perspective. Section 3 discusses the Business Information Chain as the core of an EDI system. Following this chain, EDI/S2S is discussed in Section 4, including necessary requirements to get Business Information Chain integration, that is, an efficiently working EDI data collection system. This is illustrated with examples from the four countries represented in this article. Section 5 focusses on quality issues for EDI data collection, and Section 6 concludes the article.

## 2. The Evolution of Business Data Collection: A Brief Overview

The evolution of computerisation in primary business data collection is summarised in Figure 1, starting with H2H, to H2S, and finally to S2S data collection. Today, all methods are still being used. The second and third steps have their origins in the 1980s, while the first step dates from before that time (for an overview see Couper et al. 1998; Groves 2011).

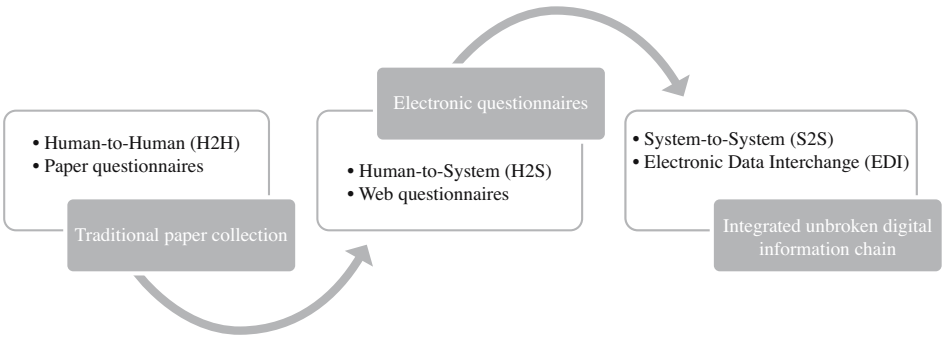


*Fig. 1. Evolution of computerisation in primary business data collection (adapted from Erikson et al. 2016).*

At the same time as electronic questionnaires came into use, various ways of direct S2S data exchange emerged – between businesses as well as between businesses and government authorities, such as the Tax Office and NSIs (Couper et al. 1998; Bharosa et al. 2015). This is the third step in Figure 1, and goes by a number of names: 'system-to-system' data collection/data exchange (S2S) or 'machine-to-machine' communication (M2M), in addition to EDI or automated data capture or data exchange. The basic idea is that S2S data exchange is set-up only once and then runs automatically over time without human involvement. Such solutions are also requested by businesses, simply because they have the potential to significantly reduce response burden. Bharosa et al. (2015, 6) in follow-up to Hansen and Hill (1989, 405) define EDI, as "the movement of business documents electronically between or within firms (including their agents or intermediaries) in a structured, machine-retrievable data format that permits data to be transferred, without re-keying, from a business application in one location to a business application in another location." This definition also applies to EDI in the context of statistical data collection (Ypma et al. 1997). In this article we use S2S and EDI interchangeably.

The first EDI attempts date back to the 1980s, when information and administrative processes within businesses were computerised (Bharosa et al. 2015). In 1995, for example, the U.S. Bureau of Labor Statistics (BLS) implemented EDI for the purpose of transmitting large datasets from businesses to BLS for the Current Employment Statistics (CES) survey (Johnson 2016). The BLS experiences from that time indicate that firms had large start-up costs (Clayton et al. 2000); it took six to nine months to get the system working and prepare a first test file. On average, it took three test files to obtain a file that matched the data requirements. In addition, maintenance costs were considerable. The high costs were reasons for firms not to participate. It turned out that the EDI solution was only efficient for bigger firms, with enough worksites, which was about five percent of US firms. Furthermore, the upfront costs for BLS were high. However, in the end, the EDI method was of huge value, as Johnson (2016) concludes: "more than 40 percent of CES worksites are processed with relatively limited burden and low costs for BLS."

In 1996 in Europe, a consortium of European NSIs, which included the Netherlands, Sweden, Italy, Germany, with Portugal and Finland as associate countries, started a project to investigate the feasibility of EDI for collecting financial business data (De Bolster and

Metz 1997). As part of this project, Statistics Netherlands introduced a system called EDIsent for their Structural Business Survey (SBS). This system was aimed at the automated transmission of financial data coming directly from business records. Couper and Nicholls (1998, 16) stated that if this EDI strategy was successful, "it may make many traditional business surveys obsolete". However, the system required large upfront costs and investments, as well as continuous maintenance and annual updating of the mappings with the individual business records, which was carried out by Statistics Netherlands' business field staff (Lunter 1997). These efforts out-weighed efficiency gains, and in the early 2000s, the system was abandoned; this was at the same time that the use of e-questionnaires was expanded (e.g., with the SBS e-questionnaire). With regard to this project, Ypma et al. (1997, 7) stated that "the very elimination of the paper questionnaire was one of the driving forces behind the introduction of EDI in statistical data collection." Looking back, we might conclude that the EDI initiative came too early; at that time no standards were in place. Consequently, questionnaires have not been eliminated, but rather, paper was replaced by electronic questionnaires. But today, maybe more than ever, the EDI data collection concept is worthwhile exploring. Because of the rapid expanding use of the internet since the mid-1990s, the electronic transfer of large data sets became more feasible and less expensive (Bharosa et al. 2015).

In addition to primary data collection, in the 1970s, secondary data collection emerged; registers and administrative data were being used in the production of statistics to replace surveys, supplement survey data, or for data validation or imputation (UNECE 2007, 2011; Wallgren and Wallgren 2007). As a result of the computerisation of registers, the exchange of large data files became feasible, and NSIs developed multi-source approaches to combine primary and secondary data, a process that is ongoing (Bakker and Daas 2012; Smith and Phipps 2014). All four countries have a long history of using alternative data sources in the production of national statistics (Erikson and Nordberg 2000; Snijkers et al. 2011; Born 2016; Saraiva 2016b).

Figures 2 and 3 describe the current situation and the future of statistics production. In Figure 2, statistical output is produced along stove pipe or silo processes, based on one or a

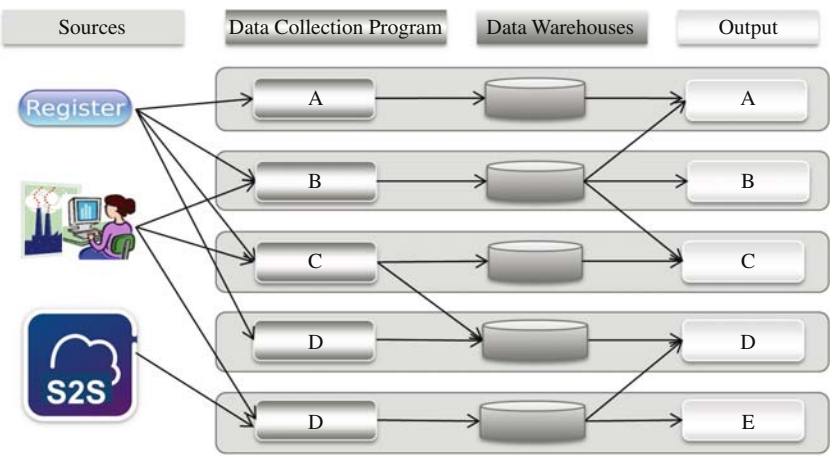


*Fig. 2. Coordinating data collection and dissemination – the current situation (adapted from Erikson et al. 2016).*

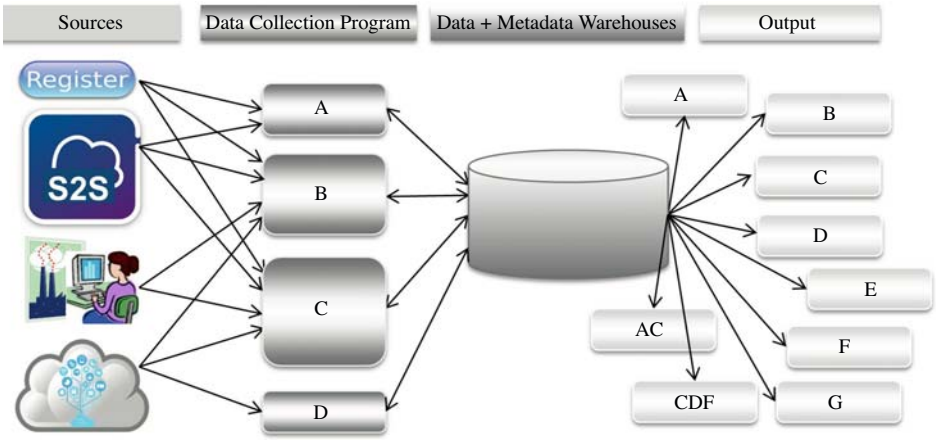


*Fig. 3. Coordinating data collection and dissemination – the future (adapted from Eriksen et al. 2016).*

combination of sources, that is, registers and surveys. In Figure 3, statistical output is produced independently of the data sources. Administrative registers are the basic source for statistics. In addition, data could be collected from business accounting systems through a S2S solution, whereas another collection process could collect data not readily available in business systems that requires manual entry by respondents, using a web questionnaire (H2S solutions). A fourth data source is big data that are available in the 'cloud' (Groves 2013; Daas et al. 2015; Di Consiglio et al. 2016).

These figures illustrate the developments for all four countries represented in this article. Figure 2 illustrates the multi-source/mixed-mode approach, as discussed above, with the production of various statistics still taking place in isolated production processes (silos). Figure 3 illustrates the integration of all available data sources in a data warehouse (sometimes also referred to as a 'data lake'), which serves as the one data repository for all statistics. Comparable processes in the various data collection programs are harmonised, as discussed in Subsection 4.2.

Today, many NSIs are working towards both the extended use of secondary sources (including big data) and the implementation of web surveys, which can be concluded from the large number of papers presented at conferences addressing these issues, see e.g., the programs of the 3rd, 4th, and 5th ICES conferences: www.amstat.org.ASA/Meetings/ICES.aspx (Smith and Phipps 2014), as well as e.g., the European Conferences on New Techniques and Technologies (NTTS: https://ec.europa.eu/eurostat/cros/content/ntts-conferences_en), and the European Conference on Quality in Official Statistics (http://ec.europa.eu/eurostat/web/quality/ess-practices/quality-conferences). A data collection method that has been discussed to a lesser degree is EDI or S2S data exchange.

## 3. The Business Information Chain

Before discussing EDI approaches and challenges, we discuss the Business Information Chain. Figure 4 shows a simplified picture of this chain (based on Bharosa et al. 2015). This chain is not only the basis for EDI, but also the origin of both register and survey data.

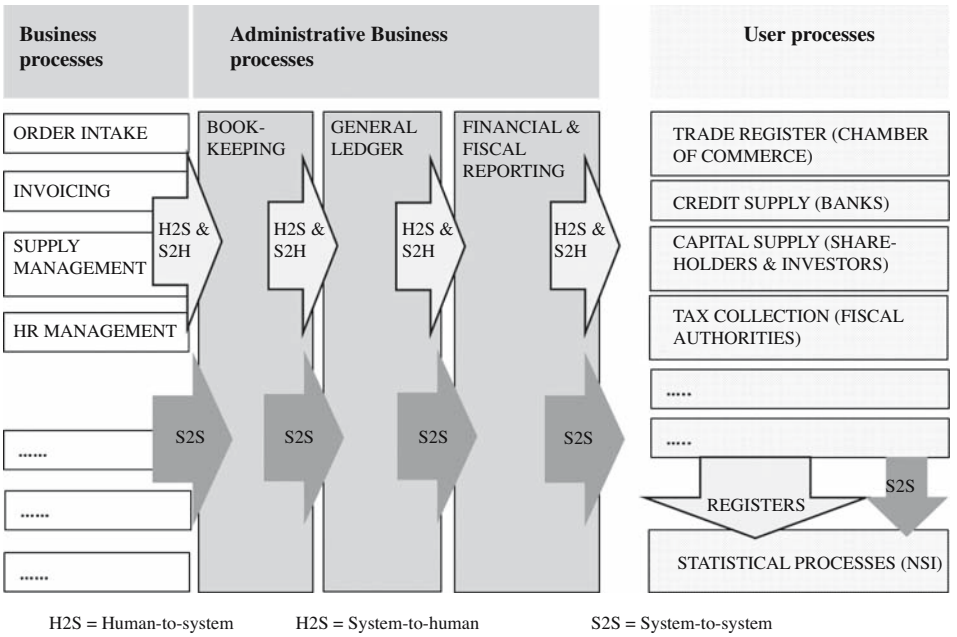


*Fig. 4. The Business Information Chain: A functional view of the current chain of financial administrative processes (adapted from Buiten et al. 2016).*

Consequently, knowledge about administrative processes in businesses is already important in business surveys to adequately tailor survey designs to the business context (see e.g., Snijkers et al. 2013; Bavdaz 2010). Information about this chain is also important for assessing the quality of administrative data (Daas et al. 2011; Reid et al. 2017). In the context of EDI starting with this chain is inevitable.

Business activities are carried out in business processes. With regard to financial information, these activities include order intake, invoicing (related to e.g., trading of goods and services and buying materials for the production process), stock keeping, human resources (HR), and payment of employees. These activities generate records at transaction levels such as product codes, shipment numbers, bar codes, customer IDs and similar. Depending on the nature of these activities and the way in which individual business processes (as shown in Figure 4) are organised, these records may be stored in one or more subsidiary ledgers such as customer administration, stock administration, wage administration, and so on, which can reside in different business departments. These transaction data feed into a series of administrative processes through which the business translates data into dashboards, financial reports, tax declarations, and others. (Bharosa et al. 2015; Zhang 2012). In this process, data can flow from one department to another.

The first administrative process is bookkeeping, in which data from various business processes feed into one central system. At this stage in the chain, business process data on individual transactions or events are coded, clustered and integrated in accordance with bookkeeping rules. The second process concerns the general ledger, in which bookkeeping data are classified into one or more sets of profit-and-loss accounts and balance sheets: the charts of accounts. Now, the data in the chain represent business (sub-)units, depending on

the required level of detail. The bookkeeping data are clustered and integrated in such a way that they result in general ledger information about the units, required by the business. In practice, businesses with integrated systems may be able to drill down from the general ledger to detailed information in subsidiary ledgers (going backward in Figure 2). The third administrative process is financial and fiscal reporting, in which selections of data from the general ledger are used to compile financial reports, tax declarations, and other reports for specific users both inside and outside the business. Again, the data need to be transformed in such a way that the requested information about the pre-defined units can be reported. These three processes take place within businesses. Next, we have the user processes that take place outside the businesses. Each user has its own process in which they use the business reports/data for their own purposes, such as tax collection, credit provision, and capital provision. At the end of this chain is the NSI, with its statistical processes, having both secondary and survey data as input.

This chain may be highly fragmented. Data is transferred within this chain by various modes, applying H2S, S2H, and S2S data communication. In data exchange requiring human intervention, data stored in one system need to be re-entered manually in another system and several employees may be involved, working with other systems at other locations. This is cumbersome, expensive, and error-prone. The emergence of the internet resulted in a huge increase in H2S chain integration, making it possible for business employees, accountants, and others to directly log into (parts of) this chain, and transfer large data sets. This stimulated S2S chain integration and standardisation, which is discussed in the next section. In the four countries represented in this article, initiatives have been taken to further integrate this chain.

## 4. Business Information Chain Integration

In order to achieve an efficiently working Business Information Chain, the share of S2S data flows should be as high as possible, resulting in a fully integrated unbroken digital information chain. Extensive data communication, both forward and backward in the chain, can only be attained when all major stakeholders participate and adopt the same standards. This requires (Bharosa et al. 2015):

- technical standardisation (‘how?’),
- conceptual harmonisation (‘what?’), both vertically and horizontally, as well as
- positive business cases for all stakeholders (‘why?’).

### *4.1. Technical Standardisation*

Assuring the interconnectivity and interoperability between all stakeholders through all steps in the Business Information Chain means that technical standards have to be put in place that are easily and widely accepted. Several competing standards are available, including XML, XBRL, and SAF-T.

Extensible Mark-up Language (XML) was developed at the end of the 1990s by the World Wide Web Consortium (W3C: www.w3c.org/XML; Bharosa et al. 2015). Its purpose was to send and store structured data, represented as plain text, via the internet. XML is a mark-up language that defines a set of rules for encoding documents in a format

that is both human- and machine-readable. XML was designed to be simple and flexible, with general use across the internet. Its generic nature, however, is also its problem. Thus, applications for specific purposes have been developed by various institutes, that is, XML-based languages designed for exchanging data defined by specific taxonomies. For the exchange of financial business information, these languages include XBRL, SAF-T, and SAF-P.

XBRL (eXtensible Business Reporting Language) is a freely available and global standard for exchanging business information, especially financial information, such as financial statements (Bharosa et al. 2015). The XBRL specification is developed and published by XBRL International, Inc. (www.xbrl.org), and has been adopted by institutions in more than 60 countries around the world, including Canada, Sweden, the Netherlands, United States, Germany, United Kingdom, Spain, Denmark, and Belgium. (See www.xbrl.org/xbrl-around-the-world). It is supported by many software packages to facilitate structured financial data reporting. XBRL has some advantages over XML, such as the inclusion of validation and business rules in the taxonomy-making data validation at an early stage in the chain. These rules are part of the data definitions, the metadata, which means that in an XBRL-based chain, not only the data but also these rules can be reused by stakeholders.

SAF-T (Standard Audit File for Tax) is an international standard defined by the Organisation for Economic Co-operation and Development (OECD) for electronic exchange of accounting data from organisations to a national tax authority or external auditors (see www.oecd.org/tax/administration/45045602.pdf.). SAF-T was first adopted in the EU in 2008 by Portugal and has since spread to other European countries, including Luxembourg, Austria, Germany, and France. This language has the standardised SAF-P extension, which enables payroll accounting software to create a file containing reliable transaction data from payroll systems.

In the four countries considered, XBRL-based solutions have been chosen or are being considered, but experiences with these standards differ. Sweden opts for XBRL to move toward a fully automated S2S solution, but is still starting up. In the Netherlands, the government as a whole chose XBRL as the standard in a nationwide SBR-approach – Standard Business Reporting. Statistics Canada considered using XBRL and the SBR approach as well. Statistics Portugal successfully adopted a SAF-T approach in joint cooperation with other government agencies, which is implemented in the legally established Simplified Business Information (IES) system.

In Sweden (Erikson et al. 2016), an open standardised data format was developed in the 1990s for exchanging accounting information between businesses and systems using different software packages. This standard, called SIE (Standard Import Export: www.sie.se), allows for S2S communication within a business and H2S for financial reporting. SIE is used to transmit the chart of accounts data, based on the Swedish standard Chart of Accounts (BAS, see Subsection 4.2). Statistics Sweden (Erikson et al. 2016) recently initiated a process to change statistical processes that coordinate data collection by clustering variables to fewer data collection programs, which is an example of moving from Figure 2 to Figure 3. Statistics Sweden has also adopted XBRL as its technical standard, with the intention that an enterprise should only have to report information once, and this information is then reused by other authorities; this comes

close to the Portuguese solution that we discuss below. However, businesses in Sweden have been slow to adopt XBRL, possibly in part because the SIE format is well-established, although the attitude appears to have been shifting since last year. A government decision in 2016 will allow businesses to provide annual reports in XBRL format to the Swedish Companies Registration Office, which might stimulate businesses to take up XBRL.

In the Netherlands (Buiten et al. 2016), Standard Business Reporting (SBR: www.sbr-nl.nl/english-site) was developed and introduced in 2013 as the standard for electronic data exchange between businesses and government agencies. The technical standard used in SBR is XBRL. In addition, in 2014 a reference chart of accounts, the Reference Classification System of Financial Information (RCSFI) was introduced, facilitating SBR reporting of chart of account data (see Subsection 4.2).

Statistics Netherlands is one of the government agencies involved in the Dutch SBR-community that aims at further standardising and harmonising the exchange of business data. The Tax Authorities (tax declarations), the Dutch Chamber of Commerce (annual business reports), and the largest Dutch commercial banks (credit reports) already receive vast amounts of SBR-data, since SBR is the required channel. The uptake of SBR in general is not regulated by law, but for tax declarations, SBR is the mandatory sole channel. Since 2012, the required use of SBR has been extended to more and more tax reports. From 2016, Chambers of Commerce also require small businesses to submit their annual reports electronically using SBR. In 2017, this requirement started applying to mediumsized businesses, and will be effective from 2019 onward for large businesses. Commercial banks have set SBR as the preferred channel for SMEs by fining other ways of data transmission; if SBR is not used, a fine of EUR 250 must be paid.

Within this framework, Statistics Netherlands has developed an extension to the Netherlands' SBR Taxonomy, including data definitions for Short Term Statistics (STS) and the Annual SBS surveys. The Netherlands' Taxonomy is the official Dutch 'SBR dictionary', defining the terms used in XBRL-messages. So, for additional statistical applications, the dictionary had to be extended. In addition, the infrastructure for receiving XBRL messages and feeding them into the statistical systems is operational. A good relationship with software providers and the accounting industry has been established within the SBR community. As in Sweden and Canada, the uptake of SBR by businesses for statistical reporting is still very limited.

Statistics Canada (Born 2016) examined the use of XBRL as an exchange standard for business reporting as a way to reduce response burden on businesses in the collection of financial data. The first approach was to continue using tax data for small simple businesses, and develop a taxonomy for financial reporting (Canadian Chart of Accounts, COA), which was generated in XBRL format for exchange. However, the uptake of XBRL by Canadian businesses is limited, and Statistics Canada expanded the use of e-questionnaires with standardised modules for financial information – revenues and expenses – for 60 business surveys. More recently, Statistics Canada has been collaborating with other federal departments and the private sector in determining the feasibility of a government-wide taxonomy for business reporting to government agencies, similar to SBR approaches used by Statistics Netherlands and other statistical agencies. As part of the study, the focus has been on developing a taxonomy using the financial items

reported to the Canada's Tax Agency – Canadian Revenue Agency, in XBRL format. More accurate mapping between the business financial information and this taxonomy will lead to better quality financial information, which is used by Statistics Canada.

In Sweden, the Netherlands, and Canada, the adoption of a technical standard is not regulated by law; Portugal opted for another solution. In 2007, through a legal act, Portugal established the mandatory nation-wide Simplified Business Information system (IES, www.ies.gov.pt/site_IES/site/home.htm) to enable S2S financial reporting between businesses and government agencies (Pereira 2011; Saraiva 2016a). With IES, four legally required business reports have been integrated: the annual declaration of accounting and tax information to the Portuguese Tax and Customs Authority, the registration of the accounts within the Portuguese framework of commercial registry legislation, business data for statistical purposes to Statistics Portugal, and annual accounting data for statistical purposes to the Bank of Portugal. In these reports, businesses had to report identical information in different ways to four government agencies. The first steps toward this system were made in early 2006 by defining a common data set to be collected. In a joint collaboration, the data needs of all four agencies were analysed in detail and decisions were made to integrate variables and reduce redundancy, resulting in a harmonised Chart of Accounts taxonomy (which is an example of vertical harmonisation as discussed in Subsection 4.2.1). The final EDI approach features SAF-T files containing business billing and accounting data that can be transmitted through the Tax Authority Portal. By law, the business chart of accounts had to be organised according to this information system, making it possible to produce a SAF-T file whenever requested by the Tax Authority.

In addition to the IES system and web questionnaires, businesses may also use another EDI data collection system solely provided by Statistics Portugal – Automated Data Transmission (TAD, (Saraiva 2016a, http://swhupload.ine.pt). TAD is a S2S approach involving the upload of XML files via the TAD portal. TAD allows businesses respondents to prepare an XML envelope, which can contain variables from one or more businesses and surveys, thus avoiding the completion of online questionnaires. After uploading the file, the system integrates the variables accordingly, and can be accessed by the respondent or Statistic Portugal's data collection technical staff. This solution is currently available for 14 surveys. Businesses have to apply to get access to this system. Even though this approach requires businesses to invest in the system, there has been an increasing acceptance of this system by respondents who indicate a significant reduction in the resources involved, as well as increased quality. The next step is to provide fully automated data exchange using web services with two-way communications.

In Portugal and the Netherlands, XBRL solutions are implemented in large parts of the chain. The use of a common technical standard by the most important government agencies is an important factor, as well as close cooperation with the business community. In the case of Portugal, this is regulated by law, whereas in the Netherlands a public-private approach including commercial banks has been adopted. In Europe the essence of standardisation is recognised by the European Commission (2016) by actively working towards "European standards for the 21st century". Technical standardisation is the first step, but there is more; it is not only about the 'how', but also about the 'what'.

### *4.2. Vertical and Horizontal Harmonisation*

To establish a fully computerised information chain, technical standardisation in itself is not enough. Harmonisation of concepts is the other requirement needed to make an automated statistical production system work. Based on the cases in the four countries, it is clear that technical standardisation goes hand in hand with conceptual harmonisation, that is, metadata and taxonomies (UNECE 1995, 2000). In an integrated statistical process (as shown in Figure 3), metadata support the entire statistical production, defining concepts, classifications, variables, data sources, and statistical outputs. They are just as important as the data themselves, and an essential tool for an NSI. Taxonomies, in general, are metadata dictionaries that define individual reporting elements in relation to each other. The application of taxonomies ensures harmonised data exchange, both with regard to content and format.

Harmonisation within the Business Information Chain includes harmonisation of metadata and its taxonomies, as well as relationships between data elements and other rules, in order to assure basic governance principles, such as accountability and comparability. We identify two levels of content harmonisation, vertical and horizontal harmonisation (Bharosa et al. 2015), which go hand in hand (see Figure 4). Vertical harmonisation refers to external harmonisation between all stakeholders. In order to achieve this, horizontal harmonisation is also required, that is, internal harmonisation for all steps in the chain within an organisation. For NSIs, this means that all surveys and statistics dealing with chart of accounts data need to be harmonised.

#### 4.2.1. Vertical Conceptual Harmonisation

Vertical or external harmonisation of data used by multiple parties is an essential precondition to get the Business Information Chain working. As Figure 5 illustrates,

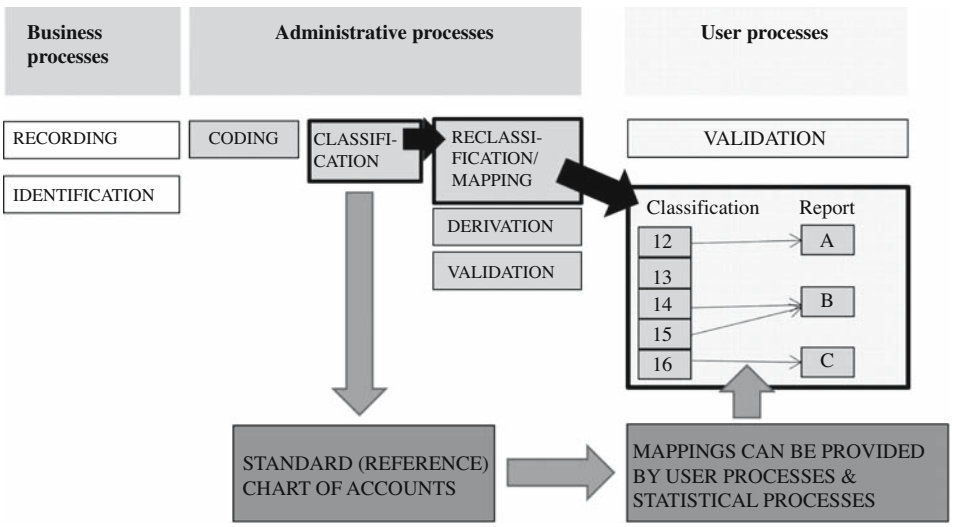


*Fig. 5. The main harmonisation actions in the Business Information Chain (adapted from Buiten et al. 2016).*

(meta)data that are harmonised across various data users and suppliers increase the usability and comparability between data from individual businesses. The use of a standard Reference Chart of Accounts by businesses and various data users implies that these chart of accounts data can be directly mapped to the various user processes (the grey arrow path in Figure 5) without needing to recode or reclassify the data (the black arrow path). Based on such a standard taxonomy, various reports can be generated that directly feed into the user processes. They even offer the possibility for data users to provide examples of mappings or mapping guidelines to facilitate the mapping process in businesses. This would increase quality and reduce administrative burden at the same time. Standards for harmonisation may be established within the business community itself and between businesses and government agencies such as the Tax Office and the Chamber of Commerce.

A major challenge in vertical harmonisation, in order to achieve interconnectivity and interoperability between all stakeholders, is to choose a set of metadata that serves multiple purposes and is applicable to all stakeholders, that is, the same metadata can be used in the statistical production system by the Tax Office and can also be adopted by businesses for their financial administrative processes (or easily converted to match these processes).

When it comes to applying an efficient S2S method for financial reporting, a harmonised Chart of Accounts is required. In Subsection 4.1, we saw that in Portugal the establishment of the IES system started with vertical harmonisation. In Section 2, we saw that the EDIsent system in the Netherlands (in the 1990s) did not work because for each and every business the system needed to be mapped to the business' chart of accounts model (the black route in Figure 5). More recently, Statistics Netherlands was one of the participants in a joint public-private partnership that developed the Reference Classification System of Financial Information (RCSFI: www.referentiegroot-boekschema/english; Buiten et al. 2016). This classification system became available in 2014. Such an initiative was necessary, since legally, companies are free to design and apply their own chart of accounts. The implementation of the RCSFI is now underway on a voluntary basis and is expected to take several years to be fully accepted by businesses and the accountancy sector. Sweden has a non-mandatory Chart of Accounts (BAS, Erikson et al. 2016) used by approximately 95 percent of all businesses. Statistics Sweden collaborates with other government agencies to harmonise terms and concepts, and aims to create an integrated hierarchy of taxonomies that use both common and agency-specific terms. Statistics Canada (Statistics Canada 2015, Born 2016) has adopted a Reference Chart of Accounts (COA) as the reference taxonomy for organising business financial information (e.g., revenue, expenses, assets, and liabilities). The COA is used for statistical purposes only, and is mapped to tax data and business survey variables.

#### 4.2.2. Horizontal Harmonisation

Horizontal or internal harmonisation is another important precondition for NSIs to participate in any broader, external harmonisation initiative. As the country examples illustrate, an integrated metadata system is a cornerstone of an automated production system. Interconnecting metadata between the various internal production systems and processes is a necessary requirement in order to produce integrated statistics, making the

connection between various sectors and topics, as is shown in Figure 3. However, this is a difficult harmonisation process, as illustrated by the Portuguese, Canadian, and Dutch cases. All three examples show the same route: first, the establishment of conceptual harmonisation across data collection programs, followed by process integration. Apart from NSIs, these harmonisation actions have to take place within all businesses and data users that take part in an EDI initiative.

As we have seen, in Portugal, vertical harmonisation in accordance with the IES System was regulated by law in 2007. To make the system work within Statistics Portugal, a central data collection department was established in 2005, followed by an Integrated Survey Management System (called SIGINQ) (Saraiva 2016c) in 2007. At first, SIGINQ covered the business surveys, and then also the household surveys. Integration in this system is based on four pillars: processes, metadata, registers, and variable-oriented. Following the principle of maximising the reuse of components, the system combines generic functions with survey-specific features. Some components are shared by all surveys (business or household), and others are specific for a single survey.

The central data collection department was based on an integrated Metadata System (see http://smi.ine.pt/?LANG=EN), developed in 2003 by Statistics Portugal. The Portuguese Metadata System consists of a repository of concepts, classifications, variables, data collection instruments and methodological documentation relating to statistical activities carried out and disseminated data on Statistics Portugal's Portal (which relates to Figure 3). The components of this system are integrated; as a consequence, its management is subject to strict rules of harmonisation and integration. The main goals of this repository are twofold: 1) to support survey design; and 2) support data dissemination, documenting indicators disseminated through the dissemination database. This system constitutes an instrument of coordination and harmonisation within the Portuguese Statistical System.

Through standardisation and harmonisation, uniform processes across the various production units could be created, eliminating differences between common processes. Using economies of scale, productivity was increased and the risk of failures reduced. As a result, resources were freed up and could be used to invest in new innovations and develop new capabilities. On the other side, a harmonised system resulted in loss of flexibility. A challenge that remains is dealing with differences in unit definitions across statistical programs, and establishing a unified unit register. Another challenge is the establishment of a fully integrated data warehouse. The basic structure is still survey-oriented.

Saraiva (2016c) states that establishing an integrated data collection organisation and infrastructure is clearly not a trivial matter. These are more management aspects than technical challenges that must be known in advance and dealt with. Critical success factors include strong top management sponsorship, the involvement of a multidisciplinary team, an extensive requirements design, and internal change management.

In 2010, Statistics Canada launched the Integrated Business Statistics Program (IBSP, Statistics Canada 2015) to provide a more efficient model for producing economic statistics. The main objective was to enhance the economic statistics program so that it remains as robust and flexible as possible, while reducing the burden on business respondents. The IBSP encompasses around 60 surveys covering three major sectors – manufacturing, wholesale and retail trade, and services and one overall topic: capital

expenditures. The goal is that by 2019/2020, the IBSP will include approximately 150 business surveys covering all sectors of economic statistics.

The IBSP uses the multi-source/mixed-mode approach (as discussed in Section 2) as the standardised approach for economic data collections conducted at Statistics Canada, including:

- maximising the use of administrative information (such as tax information) to reduce business response burden,
- using electronic questionnaires as the principal mode of primary data collection, and
- harmonising concepts and questionnaire content.

All IBSP surveys must apply statistical standards including:

- North American Industrial Classification System (NAICS) to classify the target population by industry,
- North American Product Classification System (NAPCS) to categorise and collect business input and output data, and a
- Chart of Accounts (COA) as the reference taxonomy for organising business financial information (e.g., revenue, expenses, assets, and liabilities).

There are a number of financial variables that are common across many economic surveys. By harmonising the definitions of these variables and systematically applying standards, common content has been developed and implemented across statistical programs.

One key issue that had to be resolved in developing financial data content was ensuring that the conceptual needs of the Canadian macroeconomic accounts are met through the use of administrative and survey data. The COA bridges the two sets of concepts. As part of developing the IBSP content model, the COA was reviewed and revised to ensure that COA variables, which are directly linked to tax concepts, meet the information requirements of national accountants. A new version of the taxonomy, COA 2010, was released in March 2011, in which items in the Income Statement were changed to align directly with the items in the General Index of Financial Information (GIFI) taxonomy (Martineau 2012). New items were added, some were amalgamated and others were split. One limitation to expanding service product detail in the national accounts is the limited capacity to collect consistent and reliable NAPCS-based information on service inputs, notably purchased services used by firms in the production process. Collection of data on services inputs can be complex and can impose a heavy burden on respondents. In order to address this issue, adjustments were made to the COA 2010 to more closely align with the tax data variables.

The last couple of years of Statistics Canada's Integrated Business Survey Program have proven both challenging and rewarding. The model works well for over 70 business surveys, but a lot of unplanned work was required to get it to that point. The flexibility of the modular approach as well as the standardisation of processes, content and methods have already yielded early benefits when the programs had to be adapted to respond to user needs. Some key lessons learned are (Ravindra 2016) that transition work cannot be underestimated, test now or pay the price later, there needs to be a balance between complexity and quality, and one size does not always fit all.

The third example on horizontal harmonisation comes from the Netherlands. Since the 1990s, Statistics Netherlands has been moving away from survey silos to integrated survey systems (moving from Figure 2 to Figure 3, see Snijkers et al. 2011). In 2000, a redesign program, IMPECT (IMPlementation EConomic Transformation process), was started, aimed at harmonising the Annual SBS surveys and Monthly STS surveys. To accomplish this goal for the SBS, common key questions related to the chart of accounts data were identified for all SBS questionnaires in all main sectors in the economy. These questions were the basis for a core questionnaire, planned to be sent to all businesses in the entire sample.

During the harmonisation process of the key questions, Statistics Netherlands focused on the key concepts that the questions were supposed to measure. After identifying these concepts, the original data needs and the key concepts were reviewed to evaluate whether questions were needed at all (and could be dropped), and to work toward a harmonised questionnaire design, that is, a set of business survey questions (like in the Canadian case). After agreeing on the concepts to be measured, industry-specific questions were identified and included in a sub-questionnaire of the core questionnaire. In the end, Statistics Netherlands produced one core questionnaire, composed of 58 tailored questionnaires: two for each NACE main groups, one for the large and crucial businesses (asking for more details), and a shorter one for small businesses. As a result of the IMPECT project, a harmonised process for the SBS and STS statistics was established, with a 40 percent reduction in staffing for these surveys. To make the IMPECT program a success, the same critical success factors as for Statistics Portugal were applied.

In addition to conceptual harmonisation, organisational harmonisation at Statistics Netherlands was taken one step further in 2007, when all data collection activities were centralised in a general Data Collection Division (Snijkers et al. 2011). In hindsight, all the harmonisation steps made it possible to move forward toward fully automated data collection using the SBR/RCSFI standards.

### *4.3. Positive Business Cases for All Partners: EDI Drivers*

A critical factor to achieve vertical harmonisation is the business case for each of the partners, if not regulated by law. It must be worthwhile for businesses, their service providers (accountants) and software developers to invest in EDI, and consequently in internal business horizontal harmonisation. One solution to get everyone on the same level is legal regulations, which was achieved in Portugal. With the establishment of the IES businesses, software developers and accountants had to adopt this mandatory system.

Software developers play an important role in this chain, as they provide the technological implementation; if the chain is not supported by software, the S2S system will not work. Because of the legally established system, software developers in Portugal have a good business case to invest in this solution. If a software house offers a ‘statistical report compliance’ with their products, they are differentiated from their competitors, as they only need to make the technical requirements broadly available and easy to access by businesses and their service providers.

However, in the cases of Canada, Sweden and the Netherlands, such a legal driver is lacking. All data users need to support and use a common taxonomy on a voluntary basis

(such as the COA in Canada, the BAS in Sweden, and the RCSFI in the Netherlands). Establishing such a standard taxonomy is a big step in itself. Common arguments in favour include cost reduction (including response burden) and efficiency gains in statistical reporting, albeit at the cost of independence.

The main drivers for businesses, accountants, and software developers to invest in an automated statistical information chain are very weak. Businesses are sampled, unless they are crucial businesses in which case they are multi-surveyed. In addition, the system needs to be stable over time. There is no business case for investing in a system that is uncertain to be used or yield some internal benefits, or is unstable. The latest developments in the Netherlands indicate that since SBR has been required for tax reporting and reporting to the Chambers of Commerce, and commercial banks issue fines if SBR is not used, requests for SBR for statistical reporting are increasing. Here, other parties are paving the road for statistical EDI reporting.

In Sweden, we observe a similar movement (Erikson et al. 2016). In July 2016, a government decision was taken to task the Swedish Companies Registration Office to prepare a solution for receiving annual reports in XBRL format. This work will be carried out in formal cooperation between the authorities involved (including Statistics Sweden and the Swedish Tax Agency). The decision also states that this work should prepare for the common concepts and taxonomies needed in this particular area (annual reports) to allow for an expansion of computerised data provision to other financial information, and also to facilitate data sharing and reuse between authorities. This means the road to computerised data provision to several authorities using common concepts and technical solutions has now formally been opened by the Government, a starting point for a successful SBR program in Sweden.

Adapting to conceptual standards throughout the information chain within businesses may also have some internal benefits; it may increase cost-efficiency for businesses, not only to submit data, but also to prepare internal reports and adjust systems and processes if new user requirements arise. Horizontal harmonisation within businesses aligns the internal business and administrative processes (see Figures 1 and 5), making the chain more efficient. In addition, the use of standards creates the possibility of using metadata throughout the chain, making it easier to do business analysis across the chain (Srinivasan 2017; Thomas and McSharry 2015).

Even though these rational cost-efficiency arguments may be clear, it is still hard to convince businesses to standardise the chain solely for statistical purposes. Providing feedback and benchmark data is considered to be an incentive for businesses to participate in surveys, even if the surveys are mandatory by law (Snijkers and Jones 2013; Torres van Grinsven et al. 2014), and may also serve as a driver for businesses to adopt EDI. A 2013 study carried out by Statistics Portugal showed that the majority of businesses consider official statistics useful or very useful to society (75%), but less useful for their own company (45%) (Saraiva and Moreira 2016; see also Willimack and Snijkers 2013; Haraldsen et al. 2013). As a result of this conclusion, one of the actions taken was to create regular and personalised feedback for use by companies. The initiative began with the release of statistical PDF reports. The production of these reports is sourced from the dissemination data warehouse, where confidentiality of indicators is already assured. The personalised reports can only be accessed by the top level respondent, who has a master-

key provided directly to the company administration by postal mail. Today, this service covers more than 50,000 different companies. It includes a generic type of report, as well as several versions of personalised reports, containing specific information about each company. Figure 6 shows a partial example of personalised feedback to a company, including some valued information, such as the share of the company in a certain sector of activity, and its position relative to competitors. Statistics Portugal significantly improved its relationship with data providers by designing a regular service of customised feedback reports.

Accountants and intermediary service providers also need to invest in an EDI system. Their current business case is making money by assisting businesses in their reporting obligations. With the introduction of an EDI system, this business case will disappear. By adopting EDI, accountants undermine their own business case, and a new one has to be established. Assisting businesses in analysing their data, doing business and market analyses, and 'dash boarding', thus providing better and even new services to their customers could be a new business case. In these analyses, statistics could be used for benchmarking, and as such, for closing the Business Information Chain. In this case, the data itself leads to an adapted business case, becoming a production factor for businesses, and especially service-oriented accountants. In the Netherlands, for example, recent contacts with accountancy firms and software developers show a willingness to invest in EDI reporting if benchmark data and other useful statistical output can be incorporated. Software developers no longer develop traditional accounting software, but rather include


*Fig. 6. Example of Personalised Feedback from Statistics Portugal to a company (Saraiva 2016a).*

methods for dashboarding where real-time business performance assessment is presented in a dashboard, based on both the business and statistical data, which might include the Statistics Portugal example (Figure 6).

## 5. Quality Considerations

In the 20th century sampling was an efficient way to collect data from a population of units (Groves 2011; Zhang 2012). However, the question is whether sampling remains efficient within an EDI approach. A census, having an enumeration of the entire population, may be much more efficient, considering the investments required by businesses, software developers, and others. This applies to the Portuguese IES system, which includes all businesses in Portugal that by law must report their financial statements to the government. Other methods to consider may be non-probability sampling (Baker et al. 2013), or an opt-in approach, in which businesses themselves decide to provide data (and set up an automated data exchange process following the information chain). No matter what the final solution will be, similar to traditional sampling, key quality issues such as precision, representativeness and bias remain important aspects for understanding the accuracy of a statistical estimate. The total survey error framework still serves as a quality framework, which can be renamed 'Total Data Error Framework'. (A detailed overview of error sources for EDI is discussed by Snijkers et al. 2014, see also Snijkers 2016; Haraldsen 2013; Zhang 2012; Groves et al. 2009). To secure data quality, the relevant question is whether we get the data from the predefined unit. The relevant question about the measurement is whether we get the correct measurements for predefined variables.

Constructing a predefined unit for the data to be exchanged between businesses, NSIs and other parties in the chain is essential (UNECE 2011). The problem is that the definition of units used by several parties does not directly comply with the required statistical units. Reconciling definitions of units (tax units, legal units, etc.) to statistical units (enterprise group, establishments, local units, etc.) can be a huge conceptual challenge. This issue is known from the use of registers, but also applies to EDI; in an EDI system, data collection becomes dependent on the unit as defined in the business records. At the end of the process, data is acquired, but it may be unclear for what unit; Snijkers (2016, in the context of business surveys) refers to this as unit errors. This may require post-field unit reconstruction. For most small businesses there are usually no differences (1-on-1), but for medium-sized and large businesses, the relationship between administrative units and statistical units may be complex. One approach to reducing unit errors is 'profiling', which is typically a task related to the statistical business register. Profiling (Eurostat 2010) is aimed at understanding the legal, operational and accounting structure of businesses, and making sure that this structure is reflected correctly in the business register as the basis for statistical data collection and processing. Profiling has the potential to offer a better understanding of complex unit structures. Nevertheless, it faces reluctance from NSIs, because it is resource-consuming and requires specially trained staff. This approach is applied in the Netherlands and in Portugal. In surveys, complex businesses in the Netherlands must report consolidated data, which means that, in a questionnaire, the unit definitions have to be presented (as part of the metadata). With regard to the implementation of SBR, it was decided to first focus on small (1-on-1) SMEs, because of

this issue. In Portugal, within the IES system, the unit issue still remains a problem of great concern. The profiling process of converting administrative units into statistical units can be quite difficult conceptually and often involves some form of modelling, including information from other data sources. In other countries, such as Sweden, the legal unit is a stable type of unit that most reporting relies upon. Next to a harmonised Chart of Accounts, these three countries have also established a harmonised unit definition.

In the short term, the unit problem may be circumvented by focusing on the 1-on-1 cases. In the medium term, an approach could be to use units that consolidate towards other administrative units, such as tax units. In the long term, the solution is to include a consolidation function in administrative software. Unit rules would then be included in a harmonised reference Chart of Accounts, making it possible to perform unit consolidation, and to validate consolidated numbers for target and auxiliary variables.

The second side of quality issues relates to the measurements. In the case of EDI, the data that NSIs receive are the data stored in the business records. In case a standard taxonomy is applied by all parties in the chain, this taxonomy defines the required data. A first quality issue is that the taxonomy needs to be complete; all required variables (including the required answer formats) need to be included. If the taxonomy is incomplete, variables will be lacking; if answer formats are not detailed enough, the data may not be usable. Therefore, the vertical harmonisation process may take some time, as the final result needs to serve all users, including businesses.

When variables are not specified as data needs in a conceptual taxonomy, we could speak of specifications errors for the reference Chart of Accounts, leading to item nonresponse. In parallel to specification errors in surveys (Biemer and Lyberg 2003; Snijkers 2016), where a concept is translated into one or more questions on a questionnaire, we have one or more predefined data items that make up a variable. If items in the taxonomy are missing, we have a specification error, like missing questions on a questionnaire, and this would require post-field data validation and imputation.

After the data items in a taxonomy have been specified correctly (i.e., the reference Chart of Accounts including all data items, at the required level of detail), this chart of accounts will be used in practice by mapping items in the business administration to these data items (as illustrated in Figure 5). In the mapping process errors can occur, when administrative items are mapped to wrong data items in the reference chart, leading to measurement errors or item nonresponse (if data items are not mapped to administration items) (Zhang 2012; Reid et al. 2017). When no standard taxonomy is available, mapping business data to a taxonomy is quite cumbersome, time-consuming and error-prone, as can be concluded from the experiences with the EDIsent approach in the Netherlands in the 1990s (Lunter 1997; Ypma et al. 1997). Differences may exist between the two systems regarding names and codes of data items, levels of aggregation (a statistical item may be composed of different accounting items), existence of data (a statistical item may have no accounting counterpart). However, even with a reference Chart of Accounts in place, the mapping process can be a considerable effort, and error-prone. One solution here is to provide examples, guidelines, and best practices to support the mapping process. Furthermore, validation rules can be applied; in the Portuguese IES system 2000, validation rules are included to check the data.

A final quality concern that needs to be discussed here is stability over time. This concerns stability on the business side regarding the content of the business administration and the units the information refers to, but also stability regarding the harmonised Chart of Accounts and the information required in reports. Changes made here will certainly require changes being made in the mappings. As we have seen in the EDIsent example in the Netherlands, updates were required annually. A way to deal with this would be to investigate where to map in the Business Information Chain. And again, it shows that vertical harmonisation across all stakeholders is essential for a stable EDI system.

## 6. Conclusions

Over the last decades, data collection for economic statistics has become quite complex, having to integrate various secondary data sources, and dealing with various modes of data collection. The dominant mode today is electronic questionnaires, and will be for the coming years. Yet the options for successfully implementing EDI solutions seem to increase, even if they are not made mandatory by law.

Over time, NSIs also became more and more dependent on others. With the general use of secondary data sources, they became dependent on stable delivery of these registers. With EDI, they even became dependent on others who facilitate primary data collection, such as software developers. As a consequence, NSIs may lose their independence and flexibility with regard to data collection.

Technical standardisation, and both vertical and horizontal conceptual harmonisation within the entire Business Information Chain are hurdles that have to be tackled. Based on the experiences in the four countries described in this article, we can conclude that getting all stakeholders in the chain on the same level is hard to achieve on a voluntary basis. It seems especially hard to establish a business case for businesses, software developers and accountants. It is possible that selection procedures other than sampling, which make NSIs a small player in the field, are needed. In addition, NSIs may work together on initiatives undertaken by other stakeholders, such as the Tax Offices. Another option discussed in this paper is presenting feedback data, and making the data a product in itself, and thereby closing the chain. These options seem promising. A solution of a different order is presented by Portugal, where the EDI system was made mandatory by law.

Regarding quality, both unit and measurement issues continue to require attention. Experiences in Portugal indicate that the construction of statistical units is a serious issue. A stratified approach, making a difference between 1-on-1 and more complex businesses, may be considered in implementing EDI. In addition, quality issues related to the data received may need to be monitored, like mapping errors. Including validation rules in the system is a good practice to deal with this issue. An additional approach would be to retain human intervention in the system, by displaying the data for a respondent (as in a completed e-questionnaire) to check, complete and sign off on before sending (Haraldsen et al. 2011). Dynamics make maintenance and constant quality monitoring important. These issues require study to ensure that unbiased and precise statistical output is achieved over time.

At the beginning of this article we quoted Zhang (2012) who stated that the 21st century will be the age of data integration. We have seen that an *integrated* Business Information

Chain is at the heart of a successful EDI system. However, it is striking to note that today, for successful implementation of EDI, the same hurdles have to be tackled as in the late 1990s (Ypma et al. 1997). Still, progress has been made – technical standards have been developed, and considerable progress has been made with regard to conceptual and organisational harmonisation. With the ever growing computerisation since the start of the 21st century, EDI has become more feasible and less expensive.

Although some hurdles still remain, we believe that EDI will be the future for business data collection. Financial business data will either be collected by other organisations, feeding into registers such as tax registers, or directly by an NSI. An EDI system will be used in all data exchanges. As a consequence, when moving to EDI, the differences between primary and secondary data collection will be blurred. Quality issues related to EDI resemble secondary data quality issues (Reid et al. 2017).

In the future, we believe these data will feed into a system of 'smart statistics'. Eurostat (2017) thinks of smart statistics as being "the future of official statistics where data capturing, processing and analysis will be embedded in the system itself, starting with the digital footprints of an activity." A fully integrated Business Information Chain with S2S data exchange matches this description. By making the chain into a loop, real-time business statistics will be input for a data-driven intelligent business learning system, by continuously monitoring and benchmarking business performance (Srinivasan 2017; Thomas and McSharry 2015).

## 7. References

Baker, R., J.M. Brick, N.A. Bates, M. Battaglia, M.P. Couper, J.A. Dever, K.J. Gile, and R. Tourangeau 2013. "Summary Report of the AAPOR Task Force on Non-probability Sampling." *Journal of Survey Statistics and Methodology* 2013(I): 90–143. Doi: https://doi.org/10.1093/jssam/smt008

Bakker, B.F.M. and P.J.H. Daas 2012. "Methodological Challenges in Register-Based Research." *Statistica Neerlandica* (special issue) 66(1): 2–7. Doi: http://dx.doi.org/10.1111/j.1467-9574.2011.00505.x

Bavdaz, M. 2010. "Sources of Measurement Errors in Business Surveys." *Journal of Official Statistics* 16(1): 25–42.

Bharosa N., R. van Wijk, N. de Winne, and M. Janssen (eds) 2015. *Challenging the Chain. Governing the automated exchange and processing of business information*. Amsterdam: IOS Press, Available at: www.iospress.nl/book/challenging-the-chain/ (accessed 19 April 2018).

Biemer, P.P. and L.E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken, NJ: Wiley.

Born, A. 2016. "Harmonizing Financial Information from Businesses at Statistics Canada." In Proceedings of the Fifth International Conference of Establishment Surveys: Statistics (ICESV) Switzerland, 20-23 June 2016, Geneva, Switzerland. Statistics Switzerland. Available at: http://ww2.amstat.org/meetings/ices/2016/proceedings/146_ices15Final00226.pdf (accessed 19 April 2018).

Buiten, G., R. Boom, M. Roos, and G. Snijkers 2016. "Issues in Automated Financial Data Collection in The Netherlands". In Proceedings of the Fifth International Conference of Establishment Surveys: Statistics (ICESV) Switzerland, 20-23 June 2016, Geneva,

Switzerland. Statistics Switzerland. Available at: http://ww2.amstat.org/meetings/ices/2016/proceedings/148_ices15Final00209.pdf (accessed 19 April 2018).

Clayton, R.L., M.A. Searson, and Ch. D. Manning. 2000. "Electronic Data Collection in Selected BLS Establishment Programs." In Proceedings of the Second International Conference of Establishment Surveys (ICESII), Montreal, 18-21 June 2000, 439–448. Alexandria, VA: American Statistical Association. Available at: http://ww2.amstat.org/meetings/ices/2000/proceedings/S08.pdf (accessed 19 April 2018).

Couper, M.P., R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II, and J.M. O'Reilly. 1998. *Computer Assisted Survey Information Collection*. New York: Wiley.

Couper, M.P. and W.L. Nicholls II. 1998. The History and Development of Computer Assisted Survey Information Collection Methods. *Computer Assisted Survey Information Collection*, edited by M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clarck, J. Martin, W.L. Nicholls II, and J.M. O'Reilly, 1–21. New York: Wiley.

Daas, P., M. Puts, B. Buelens, and P. van den Hurk. 2015. "Big Data as a Source for Official Statistics." *Journal of Official Statistics* 31(2): 249–262. Doi: https://doi.org/10.1515/jos-2015-0016.

Daas, P., S. Ossen, M. Tennekes, L.-C. Zhang, C. Hendriks, K. Foldal Haugen, A. Bernardi, F. Cerroni, T. Laitila, A. Wallgren, and B. Wallgren 2011. List of Quality Groups and Indicators Identified for Administrative Data Sources. Luxembourg: Eurostat (BLUE-ETS deliverable 4.1).

De Bolster, G.W. and K.J. Metz. 1997. The TELER-EDISENT project. *Netherlands Official Statistics* (Special issue on EDI: The State of the Dutch Art): 12 (autumn), 51–59. Voorburg/Heerlen: Statistics Netherlands.

Di Consiglio, L., M. Karlberg, M. Skaliotis, and I. Xirouchakis. 2016. "Overview of Big-Data Research in European Statistics Agencies." In Proceedings of the Fifth International Conference of Establishment Surveys: Statistics (ICESV) Switzerland, 20-23 June 2016, Geneva, Switzerland. Statistics Switzerland. Available at: http://ww2.amstat.org/meetings/ices/2016/proceedings/187_ices15Final00346.pdf (accessed 19 April 2018).

Erikson, A.-G., J. Erikson, and C. Hertzman 2016. "Automated Data Collection and Reuse of Concepts in Order to Minimise the Burden." In Proceedings of the Fifth International Conference of Establishment Surveys: Statistics (ICESV) Switzerland, 20-23 June, Geneva, Switzerland. Statistics Switzerland. Available at: http://ww2.amstat.org/meetings/ices/2016/proceedings/147_ices15Final00184.pdf (accessed 19 April 2018).

Erikson, J. and L. Nordberg. 2000. "Use of Administrative Data as Substitutes for Survey Data for Small Enterprises in the Swedish Annual Structural Business Statistics." In Proceedings of the Second International Conference of Establishment Surveys (ICESII), Buffalo, NY, 17-21 June 2000, 813–820. Alexandria, VA: American Statistical Association. Available at: http://ww2.amstat.org/meetings/ices/2000/proceedings/S28.pdf (accessed 19 April 2018).

European Commission 2016. European Standards for the 21st Century. Brussels: European Commission. Available at: http://ec.europa.eu/DocsRoom/documents/16980 (accessed 19 April 2018).

Eurostat. 2010. Business registers: Recommendations manual. Luxembourg: Eurostat. Available at: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf.

Eurostat, 2017. Smart Statistics. Paper discussed at the Joint Dime/ITDG plenary sessions, 14-15 February 2017. Luxembourg: Eurostat.

Groves, R.M. 2011. "Three Eras of Survey Research." *Public Opinion Quarterly*, 75(5): 861–871. Doi: https://doi.org/10.1093/poq/nfr057.

Groves, R.M. 2013. *Official Statistics and "Big Data"*. Paper presented at the 2013 European Conference for New Techniques and Technologies for Statistics (NTTS), Brussels, 5-7 March 2013. Luxembourg: Eurostat.

Groves, R.M., F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau. 2009. *Survey Methodology*, 2nd edition. Hoboken, NJ: Wiley.

Hansen, J.V. and N.C. Hill. 1989. "Control and Audit of Electronic Data Interchange." *MIS Quarterly* 13(4): 403–414. Available at: http://aisel.aisnet.org/misq/vol13/iss4/2/ (accessed 19 April 2018).

Haraldsen, G. 2013. "Quality Issues in Business Surveys." In *Designing and Conducting Business Surveys*, edited by G. Snijkers, G. Haraldsen, J. Jones, and D.K. Willimack, 83–125. Hoboken, NJ: Wiley.

Haraldsen, G., J. Jones, D. Giesen, and L.-Ch. Zhang 2013. "Understanding and Coping with Response Burden." In *Designing and Conducting Business Surveys*, edited by G. Snijkers, G. Haraldsen, J. Jones, and D.K. Willimack, 219–252. Hoboken, NJ: Wiley.

Haraldsen, G., G. Snijkers, M. Roos, A. Sundvoll, T. Vik, and H.-P. Stax. 2011. *Utilizing Web Technology in Business Data Collection: Some Norwegian, Dutch and Danish Experiences*, Paper presented at the 2011 European Conference for New Techniques and Technologies for Statistics (NTTS), Brussels, 22-24 February 2011. Luxembourg: Eurostat.

Johnson, N. 2016. "One hundred years of Current Employment Statistics Data collection." *Monthly Labor Review*: January 2016. Washington DC: US Bureau of Labor Statistics.

Lunter. L. 1997. EDIsent after the pilot phase (in Dutch: EDIsent na de pilot fase). Report. Voorburg: Statistics Netherlands.

Martineau, P. 2012. *Use of the Chart of Accounts in determining the content of Statistics Canada business surveys*. Internal document. Ottawa: Statistics Canada.

Pereira, H.J. 2011. "Simplified Business Information (IES) – Is coordination between public entities really possible." In Proceedings of the International BLUE-ETS Conference on Burden and Motivation in Official Business Surveys Heerlen, 22-23 March 2011, 177–188. Heerlen: Statistics Netherlands. Available at: https://www.blue-ets.istat.it/index.php?id=78&tx_ttnews%5Btt_news%5D=25&cHash=75a34e2a1c989b918ff346a3911c90fc (accessed 19 April 2018).

Ravindra, D. 2016. "Challenges and Benefits of Producing Business Statistics within a Highly Centralized Model." In Proceedings of the Fifth International Conference of Establishment Surveys (ICESV), 20-23 June 2016. Geneva, Switzerland. Statistics Switzerland. Available at: http://ww2.amstat.org/meetings/ices/2016/proceedings/153_ices15Final00050.pdf (accessed 19 April 2018).

Reid, G., F. Zabala, and A. Holmberg 2017. "Extending TSE to Administrative Data: A Quality Framework and Case Study from Stats NZ." *Journal of Official Statistics* 33(2): 477–511, Doi: https://doi.org/10.1515/jos-2017-0023

Saraiva, P. 2016a. *Integrated Survey Management System: Statistics Portugal Experience*. Presentation at the Fifth International Conference of Establishment Surveys (ICESV), 20-23 June 2016, Geneva, Switzerland. Statistics Switzerland.

Saraiva, P. 2016b. *The Experience of Statistics Portugal in the Appropriation of Administrative Data for Statistical Purposes*. Invited presentation (in Portuguese) at the 3rd National Conference of Producers and users of Statistics, Geographical and Environmental Information (INFOPLAN). IBGE, December 5-9, 2016, Rio de Janeiro, Brazil.

Saraiva, P. 2016c. *Integrating Data Collection: Wins and Challenges*. Presentation at the UNECE Workshop on Statistical Data Collection, 'Visions on Future Surveying', Statistics Netherlands, The Hague, 1-5 October 2016. Geneva: United Nations Economic Commission for Europe (UNECE).

Saraiva, P. and A. Moreira. 2016. *Motivating Respondents: the Importance of Personalised Feedback*. Presentation at the UNECE Workshop on Statistical Data Collection, 'Visions on Future Surveying', Statistics Netherlands, The Hague, 1-5 October 2016. Geneva: United Nations Economic Commission for Europe (UNECE).

Smith, P. and P. Phipps. 2014. Preface. *Journal of Official Statistics* (Special Issue on Establishment Surveys) 30(4): 575–577. Doi: https://doi.org/10.2478/jos-2014-0038

Snijkers, G. 2016. "Achieving Quality in Organizational Surveys: An Holistic Approach." In *Methodische Probleme in der empirischen Organisationsforschung*, edited by S. Liebig and W. Matiaske, 33–59. Wiesbaden: Springer.

Snijkers, G., R. Göttgens, and H. Hermans. 2011. *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. Geneva: UNECE Available at: www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2011/20.e.pdf.

Snijkers, G., G. Haraldsen, J. Jones, and D.K. Willimack. 2013. *Designing and Conducting Business Surveys*. Hoboken, NJ: Wiley.

Snijkers, G. and J. Jones. 2013. "Business Survey Communication." *Designing and Conducting Business Surveys*, edited by G. Snijkers, G. Haraldsen, J. Jones, and D.K. Willimack, 359–430. Hoboken, NJ: Wiley.

Snijkers, G., M. Roos, T. Hooijmans, Th. van Kasteren, M. Storms, V. de Haan, and G. Buiten. 2014. Reference Chart of Accounts (RCSFI): a case study – toward a quality framework (in Dutch: Referentie Grootboekschema (RGS): een casestudie – aanzet tot een kwaliteitskader). Report. Heerlen: Statistics Netherlands.

Srinivasan, V. 2017. *The Intelligent Enterprise in the Era of Big Data*. Hoboken, NJ: Wiley.

Statistics Canada 2015. *Integrated Business Statistics Program Overview*. Catalogue No. 68-515-X. Ottawa: Statistics Canada, Available at: http://www.statcan.gc.ca/pub/68-515-x/2015001/mi-rs-eng.htm.

Thomas, R. and P. McSharry 2015. *Big Data Revolution: What Farmers, Doctors and Insurance Agents Tell Us about Discovering Big Data Patterns*. Hoboken, NJ: Wiley.

Torres van Grinsven, V., I. Bolko, and M. Bavdaz. 2014. "In Search of Motivation for Business Survey Response Task." *Journal of Official Statistics*, 30(4): 579–606. Doi: https://doi.org/10.2478/jos-2014-0039

UNECE 1995. *Guidelines for the Modelling of Statistical Data and Metadata*. Geneva: United Nations Economic Commission for Europe (UNECE), Available at: https://www.unece.org/fileadmin/DAM/stats/publications/metadatamodeling.pdf (accessed 19 April 2018).

UNECE. 2000. *Terminology on Statistical Metadata*. Geneva: United Nations Economic Commission for Europe (UNECE). Available at: http://ec.europa.eu/eurostat/ramon/coded_files/UNECE_TERMINOLOGY_STAT_METADATA_2000_EN.pdf

UNECE. 2007. *Register-based statistics in the Nordic countries: review of best practices with focus on population and social statistics*. Geneva: United Nations Economic Commission for Europe (UNECE). Available at: http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf.

UNECE. 2011. *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices*. Geneva: United Nations Economic Commission for Europe. Available at: https://www.unece.org/fileadmin/DAM/stats/publications/Using_Administrative_Sources_Final_for_web.pdf (accessed 19 April 2018).

Wallgren, A. and B. Wallgren 2007. *Register-Based Statistics – Administrative Data for Statistical Purposes*. Hoboken, NJ: Wiley.

Willimack, D.K. and G. Snijkers. 2013. "The Business Context and its Implications for the Business Response Process." In *Designing and Conducting Business Surveys*, edited by G. Snijkers, G. Haraldsen, J. Jones, and D.K. Willimack, 39–82. Hoboken, NJ: Wiley.

Ypma, W.F.H., A.J. Willeboordse, and W.J. Keller. 1997, EDI in the collection of statistical data: an introduction. *Netherlands Official Statistics* (Special issue on EDI: The State of the Dutch Art) 12 (autumn): 7–15. Voorburg/Heerlen: Statistics Netherlands.

Zhang. L.-C. 2012. "Topics of Statistical Theory for Register-Based Statistics and Data Integration." *Statistica Neerlandica*, 66(1): 41–63. Doi: http://dx.doi.org/10.1111/j.1467-9574.2011.00508.x

Received November 2016
Revised December 2017
Accepted January 2018