

ePub^{WU} Institutional Repository

Elmar Wilhelm Fürst and Peter Oberhofer and Christian Vogelauner and
Rudolf Bauer and David Martin Herold

Innovative methods in European road freight transport statistics: A pilot study

Article (Draft)

Original Citation:

Fürst, Elmar Wilhelm and Oberhofer, Peter and Vogelauner, Christian and Bauer, Rudolf and Herold, David Martin (2019) Innovative methods in European road freight transport statistics: A pilot study. *Journal of Statistics and Management Systems*. pp. 1-22. ISSN 2169-0014

This version is available at: <http://epub.wu.ac.at/7067/>

Available in ePub^{WU}: July 2019

ePub^{WU}, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

This document is an early version circulated as work in progress. There are minor differences between this and the publisher version which could however affect a citation.

Innovative methods in European road freight transport statistics: a pilot study

Abstract

By using innovative methods, such as the automated transfer of corporate electronic data to National Statistical Institutions, official transport data can be significantly improved in terms of reliability, costs and the burden on respondents. In this paper, we show that the automated compilation of statistical reports is possible and feasible. Based on previous findings, a new method and tool were developed in cooperation with two business partners from the logistics sector in Austria. The results show that the prototype could successfully be implemented at the partner companies. Improved data quality can lead to more reliable analyses in various fields. Compared to actual volumes of investments into transport, the costs of transport statistics are limited. By using the new and innovative data collection techniques, these costs can even be reduced in the long run; at the same time, the risk of bad investments and wrong decisions caused by analyses relying on poor data quality can be reduced. This results in a substantial value for business, research, the economy and the society.

Keywords

Road freight transport, transport statistics, innovative methods, automated data collection, data quality, policy improvement

1 Introduction

Transport is one of the most important mass phenomena, as everyone is affected by it either directly (commuting to work, leisure travel, etc.) or indirectly (purchase of goods which were transported to the location of sale, people paying visits to others, etc.). Many decisions, be they personal, political or economic, are influenced by the availability of transport infrastructure and services and in turn have a substantial impact on transport demand and the transport system. Among the various transport modes, road freight is often labelled as the most important in terms of volume and performance [1]. However, due to its negative impacts on society and the environment, it is also considered critical [2-4]. Many political and corporate decisions, modelling and forecasting require a broad knowledge of transport activities. Therefore, numerous institutions and organizations rely on high-quality transport data. Considering the consequences, costs and irreversibility of wrong decisions (e.g. infrastructure misinvestment, unfocused policy agendas), it becomes evident that the costs of compiling reliable data are rather low compared to the consequences of poor project planning [5-7].

In particular, the domain of road freight transport statistics shows considerable room for improvement as the majority of the National Statistical Institutes (NSI) of European countries still rely predominantly on paper-pencil questionnaires (which imply serious quality concerns)

[8] and (partly) web-based questionnaires rather than on automated systems to compile data. Moreover, surveys and methods are developed at member-state level and are not fully coordinated across European borders. By using innovative data collection methods, statistics could be significantly improved [9]. The maintenance and running costs of the software would be relatively low. Additionally, the costs of submitting statistical reports (on the part of the companies) and of data input, processing and analysis (on the part of the statistical office) could be substantially reduced. Without additional efforts, sample sizes could be enlarged and respondents' burden lifted [10]. As a result, data quality could be improved significantly and data collection could be more easily harmonized across national borders [11].

The purpose of this paper is to carve out the relevance of innovative methods for data collection, focusing on the domain of road freight transport. Moreover, we have designed and tested a prototype application demonstrating that the automated compilation of statistical reports is possible and feasible and we suggest how such a system should be designed to be usable at a European level. Based on the findings of previous studies, which reveal that the relevant data is electronically available in most companies, we conduct an experimental development to prove that such a system can be operated successfully. Finally, we discuss the positive effects of such innovative methods for various stakeholders, such as policy-makers at different national and international levels and in a variety of companies.

The rest of this paper is structured as follows. In section 2 we provide the theoretical background and define the research gap. Section 3 describes the research concept and the methodological approach to develop and implement an innovative data collection scheme. Section 4 presents the results of the experimental pilot study which are then critically discussed in section 5. Last but not least, we draw conclusions and describe the study's limitations as well as the implications for further research.

2 Theoretical background

In many ways, statistics are in the public interest as they are used for making fundamental decisions. This is particularly true for official statistics which aim to monitor changing economic, social and cultural aspects of society and which are central to effective government [12]. Thus, there is a need for developing suitable organizational structures and legal frameworks [13]. Statistical data can generally be seen as a product with producers and consumers, and therefore also associated costs and value [5]. In this respect, data quality is an important factor which can be described as the usefulness of the statistics for a particular goal. It is multidimensional and related to 'fitness for use' [10,14]. Being used by a broad range of stakeholders, data quality is a primary issue for official statistics [15,16]. It requires commitment and implies accountability on the producers' side. Central criteria for assessing data quality include recency, accuracy, availability and relevance [17]. Within the European Statistical System, a harmonized Code of Practice ensures the quality of official statistics. It defines principles for the institutional environment (professional independence, adequacy of resources, commitment to quality, statistical confidentiality, impartiality and objectivity),

statistical processes (sound methodology, appropriate statistical procedures, non-excessive burden on respondents and cost effectiveness) and statistical output (relevance, accuracy and reliability, timeliness and punctuality, coherence and comparability, accessibility and clarity) [16]. The high quality of (corporate) input data is an important prerequisite for the superior quality of official statistics data [14]. Any (theoretical) decision based on the data as well as decisions about the data should include the costs involved to achieve (a certain level of) data quality; the benefits obtained from it; and the costs of poor data quality [5]. Moreover, public confidence in the quality of official statistics is crucial. If users suspect that data is not correct, the reputation and credibility of the NSI are at risk [5,10,18]. Finally, data quality also depends on the mode of collection [5].

Official transport statistics are relevant for several specific areas, such as infrastructure investments, land-use and transport planning, transport system inventory, transport system performance or travel patterns [19]). However, transport statistics are often insufficient, e.g. for analyzing supply chains [20,21]. Moreover, current methods have repeatedly been criticized as being incoherent and inadequate. This is particularly true for road freight transport [22], where the majority of NSIs in European countries still predominantly rely on paper-pencil questionnaires (which imply serious quality concerns) and the complex manual processing of data content [8]. Consequently, several authors have already called for alternative sources of data [10] and particularly for automated survey techniques in order to produce official national statistics [23-25]. By using innovative data collection methods, statistics could be significantly improved [9]. Furthermore, both reporting costs (on the part of the companies) and costs for data input, processing and analysis (on the part of the statistical office) could be substantially reduced [10]. While academic research has started to evaluate new technologies, it is crucial for NSIs to keep track of innovative developments as well by enabling function-specific processes and to redesign surveys using information management tools [26] to increase quality as required by the Eurostat's Code of Practice mentioned above [16]. Although there is a broad body of literature on innovative methods of conducting transport surveys on travel behavior and patterns of passengers [27-31], only little academic research has been performed on innovative methods of collecting official freight transport statistics.

Due to different methodologies currently applied in the different reporting states (e.g. thresholds, samplings, exceptions, data-processing, grossing-up procedures, imputations) and missing information on vehicles registered in third countries [32], there is a lot of room for improvement in the consistency of European road freight transport statistics. Currently, these are conceived as sample surveys with diverging coverage of the lorry population using mostly paper-pencil questionnaires. Attempts to render statistical data more valid and reliable include the use of other survey methods, particularly innovative and automated methods, to raise sample sizes and minimize error-proneness as well as costs in the collection process [22,33]. Possible pathways towards a sustainable and long-term oriented approach could include the use of GPS-navigation systems [34-36] and the extraction of data from corporate management systems as well as combinations of these. However, such approaches require the willingness to change traditional structures. Several features need to be considered when designing a new

system, which include maintainability, efficiency, reliability, durability, scalability, vulnerability, flexibility, adaptability, and extensibility [37].

From a theoretical perspective, there is therefore every indication that automated data collection and transfer would significantly improve transport statistics. This research paper now asks whether this can be transferred to an experimental practical setting in which obstacles and hindrances occur, and whether there is evidence that such solutions could be broadly applied in the industry.

3 Research concept and methodological approach

The research for the underlying project was conducted in several steps. Based on the theoretical background, a quantitative and qualitative pre-study identified prerequisites, proving that companies generally have the required data available in (diverging) electronic formats and systems. The pre-studies also identified a general level of acceptance for innovative methods of generating transport statistics [12,33].

Based on these findings, a large-scale research project was initiated for the official Austrian road freight transport statistics, which was funded by the Austrian Research Promotion Agency. Following the concept of 'Consortium Research' [38], both scientific institutions and practitioners were part of the research team. An experimental pilot study was conducted which aimed to prove the practical applicability by developing (building and evaluating) an artifact [39]. The pilot study was oriented towards the Cooperative Experimental System Development Approach (CESD) [40]. Following this approach, we developed a general but tailor-made software product. The innovative artifact was then implemented at two partner companies, thus providing information about 'proof of concept', 'proof of value' and 'proof of use' [41]. The interface used considered the particularities of one large and one medium-sized partner company from the transport sector as well as the requirements of the Austrian NSI for electronic reporting. During the testing phase of several weeks it was repeatedly adapted and further developed. Finally, results were discussed and evaluated in several meetings which were attended by a steering committee from the Ministry for Transport, Innovation and Technology and Statistics Austria. Finally, the findings are being disseminated.

4 A pilot study of innovative methods in Austria

4.1 The problem statement

'The continuous improvement in transport survey data methods, procedures, and tools is an imperative, not a luxury' [28]. To address this issue, we aimed to show how automated data collection can be used in the domain of road freight transport in Austria to improve official transport statistics.

The Austrian NSI (Statistics Austria) is part of the European Statistical System (ESS). The ESS puts particular emphasis on the creation and development of transport survey methods and

technologies. According to EU Regulation 70/2012 [42], the re-codification of the former Regulation 1172/98, official road freight transport statistics cover all transport of cargo performed by lorries or articulated vehicles and registered in the reporting countries of the European Statistical System (the *nationality principle*). The regulation provides a general legal and methodological framework for the different national surveys and stipulates which statistical units have to provide data and which exceptions apply.

Austrian law assigns the responsibility for the production of these road freight statistics to Statistics Austria. Drawing on the vehicle registry, which is updated monthly using data provided by the Austrian Association of Insurance Companies, a stratified sample considering load capacity, place of registration, and cumulated load capacity of all registered lorries is prepared quarterly. This sample, which meets EU quality requirements, comprises 6,500 vehicles per quarter out of a total of about 72,000 registered vehicles. Each local unit operating one or more lorries or articulated vehicles within the sample is required to report all journeys during the observation period of one week for the vehicles drawn. For this purpose mostly paper-pencil questionnaires are currently used. In addition, an online questionnaire making it possible to submit responses electronically was released in April 2008. Its adoption rate of slightly above 10% [43] is rather low (presumably due to the complex structure of the required data). The former sample size of 180,000 vehicle weeks per year has been reduced to the current level of 26,000 vehicles weeks per year in 2006, corresponding to about 14% of the original data volume. The preliminary aim of this reduction was to alleviate the administrative burden for respondents. As a consequence, however, the accuracy of the data, and thus its usability, have suffered. Analyses at a more detailed (geographical) level are subject to significantly higher sample errors as grossing-up procedures are based on lower vehicle counts [44,45].

Although extensive efforts are made by Statistics Austria's staff to verify and correct the reported data manually and despite automated plausibility tests, a number of specific quality concerns remain, as follows [22,46-49]:

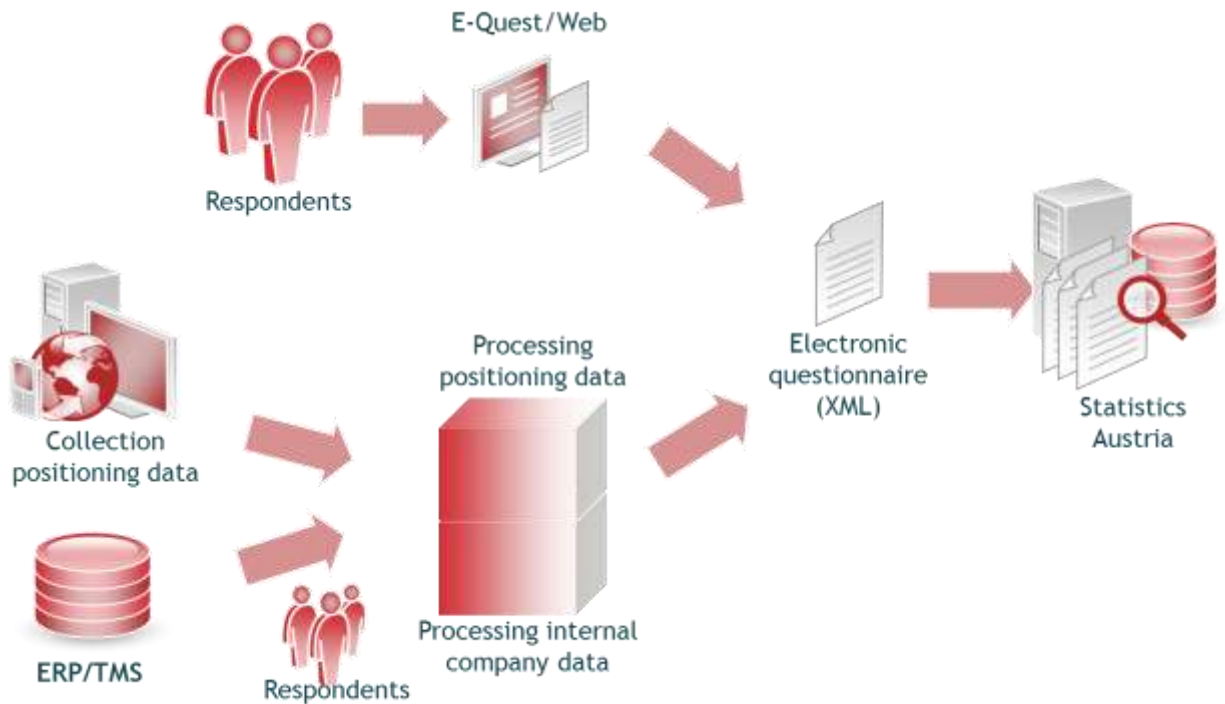
- Insufficient precision for analysing the connections between journeys and cargo types
- Insufficient precision of reported tonne-km as a result of automated imputation of the distance covered between the origin and destination of a journey using distance matrices, disregarding potential detours in between
- Possible over-estimation of empty trips as a result of the implemented imputation procedure assuming that distances between places of unloading and subsequent places of loading are empty journeys (again using distance matrixes)
- Assumed under-reporting of trips as respondents might be tempted to minimize their efforts
- Limited accuracy of cargo types as many respondents prefer to use collective categories (mixed cargo) instead of reporting individual cargo types

4.2 Innovative methods to improve data quality: opportunities and challenges to overcome

We hypothesize that innovative methods are able to tackle the data quality concerns listed above by largely eliminating or avoiding the sources of respective errors through the use of more authentic data by automatically transferring data from corporate IT systems (e.g. Enterprise Resource Planning (ERP) and Transport Management (TM) systems) to the National Statistical Institute. Introducing such automated data collection methods would improve data quality for various reasons: it would (1) make it possible to avoid the elaborate process of extracting data by hand from electronic data sources, manual data processing, and typing onto paper or into a web-interface (on the respondents' part); (2) ease the companies' burden and allow for the early assessment of reported data; (3) remove the need to further process and check the data and to reenter them (again mainly manually) into an electronic system (on the part of the NSI).

The introduction of automated data collection methods, however, faces two main challenges regarding the seamless data transfer from a technological perspective: the heterogeneous data structures of ERP (Enterprise Resource Planning) and Transport Management (TM) systems and consequently the missing, unclear and incoherent data in these operational systems. The systems are often not compatible, either with each other or with the interface of the NSI. To overcome these challenges, XML (Extensible Markup Language) seems to be a promising basis as it enables the development of interfaces for a generic data transfer application. As a consequence, extracted data from various ERP and TM systems can easily be transferred to the NSI. Providing a data format that is able to meet all potential data requests on the one hand and to complete the generic format with the corresponding data on the other are the main challenges for developing such an XML structure. The compatibility of these different conceptual views and physical representations requires the development of a consolidated data model as well as the specification of interfaces. Following these theoretical thoughts, the research project developed an experimental prototype to map and seamlessly transfer data contained in respondents' IT systems to the 'XML-gateway' of Statistics Austria, the Austrian NSI. The target process is illustrated in figure 1 and comprises two main paths.

Figure 1. Target process



Based on the findings from the pre-studies, automated data collection methods are particularly beneficial for larger companies. Firstly, their reported data volume is significantly higher due to their larger vehicle fleets and greater reporting obligations – in general, large companies have to report four times a year for one week. Secondly, their IT systems are usually more complex, which results in higher expenses for manually extracting data from the IT systems and typing them into the online questionnaire or filling out the paper-pencil forms. The cost-saving potential is therefore considerably higher for large companies. They are also more likely to have the financial sources to invest in methods with a long-term strategic goal and longer amortisation times. Furthermore, innovative methods could be used for a variety of other controlling purposes. Consequently, by using innovative methods to directly transfer data to the NSI automatically they can save time and resources, and, as a result, enhance their competitiveness. For small and medium enterprises, on the other hand, the use of the available online questionnaire should be further encouraged as they often lack financial resources and their reporting obligations are limited.

5 Results

The presentation of the results of this pilot study contains several stages. In a first step, the data to be reported is identified and integrated into a high-level data model. Next, the technical components of the architecture and the embedment of the prototype into the overall statistical reporting process in the countries are shown. The features of the prototype are then described in detail. Finally, experiences and lessons learned from its implementation at the partner companies are given.

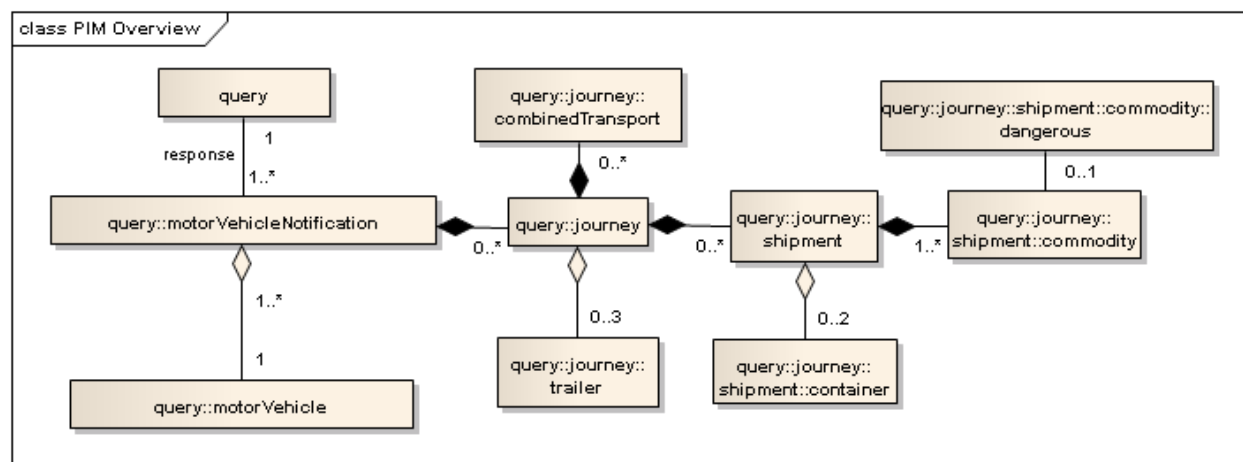
5.1 Development of a high-level data model

In order to develop the interface, the type of corporate data which must be reported to the European Statistical System by the NSI had to be identified. This resulted in four different semantic data types and related sources in the ERP and TM systems, as follows:

- *Data on companies (and local units)* – also used to distinguish transport for hire or reward and transport on own account; usually a by-product of the NSI sampling.
- *Data on vehicles* – uniquely identified through their license number. Further technical information, such as weight, load capacity, age etc. is contained in administrative registers or in Fleet Management Systems.
- *Data on journeys* – including origin and destination, intermediate stops for loading and unloading, routes and distances. These represent the core information of TMS; additional data can be obtained from Position Data Services software.
- *Data on goods* – such as type of cargo, weight, place of loading and unloading, type of dangerous goods (if applicable). Can be obtained from the ERP or TM systems.

A high-level data model was developed in order to integrate these data types hierarchically. Figure 2 illustrates this model, including the main entities and the relationships between them, and describes the data that is collected as part of the road freight transport statistics.

Figure 2. High-level information model of main entities



Conditions for and relationships within such a collection of data information on vehicles that have to be reported by the selected respondent are:

- 0 or more journeys performed during the observation period per vehicle can be reported.

- 0 to 3 trailers per journey can likewise be reported, as well as 0 or more different shipments per journey.
- 0 or more combined transports per journey can be reported.
- 0 or more countries (international journeys/transit) per journey can be reported (not shown in the diagram).
- 0 or more shipments per journey may be reported; each shipment can be associated with 0-2 containers.
- A shipment can comprise 1 or more commodities; each commodity can be classified as dangerous goods if applicable.

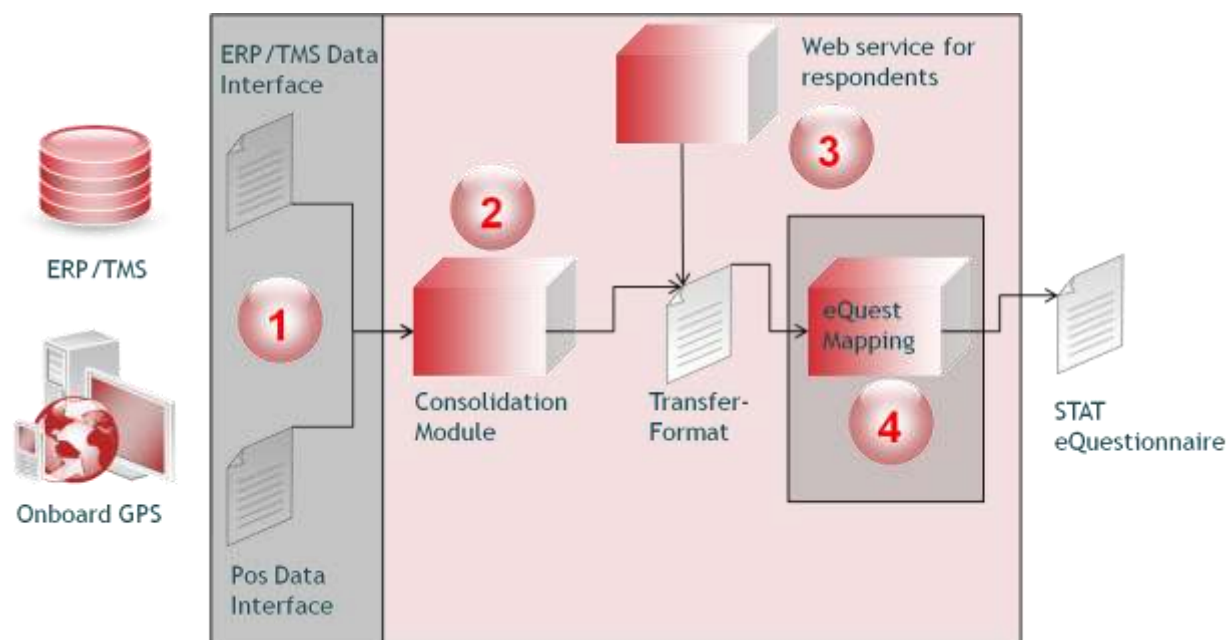
Such a high-level information level diagram has subsequently been expanded and specified at the detailed attribute level.

Differentiation between a generic transfer format and a country-specific electronic questionnaire was introduced in order to enable the applicability of the methodology in all reporting countries. The transfer format therefore covers all data of presumable interest to national statistics and Eurostat. Documentation will be in the public domain and will serve as a proposal for an interface standard of European road freight transport. The correctness and the completeness of the interface and its usability have been successfully tested at the partner companies.

5.2 Technical and business architecture

The figure below illustrates both the technical components of the architecture and the embedment of the prototype into the country's overall statistical reporting process.

Figure 3. Technical components of the architecture and the embedment



Different applications provide access to relevant data by means of the respective interfaces. The ERP and TM systems are the main data source; positioning data is used to complement these (1). Data is combined in a consolidation module (2). In order to assist the respondent in completing the report, services for reference data collection, consolidation and verification, automated goods classification and event/route detection are included (3) (see also table 1). The generic transfer format serves to map data for a specific electronic questionnaire (e.g. in Austria, the 'eQuest'-structure of Statistics Austria) (4), and the completed questionnaire can eventually be dispatched to the NSI.

Table 1. Innovative service features of the prototype

Service	Feature description
Reference data collection	The NSI's electronic questionnaire already contains reference data on the selected company and the vehicles which will automatically be transferred to the newly compiled report.
Consolidation and verification	Extracted data is consolidated using the semantic rules of the data model; for quality control, a set of pre-defined rules is applied and flawed records are flagged and presented for correction.
Automated goods classification	To improve the quality of coding goods according to goods classifications, machine-learning (adaptive) algorithms were implemented; text information was used for coding purposes.

Event / route detection	Using positioning data (e.g. GPS), the actual route travelled including distances and stops is identified by using a set of heuristics to detect potential loading and unloading stops. Recorded data is matched with extracted data to increase precision.
-------------------------	---

5.3 Prototype implementation at the partner companies

At both partner companies there was already a high degree of technical infrastructure at all stages of the transports and other business operations. This means that practically all data required for the statistical report was available in electronic form. However, data in different systems within the company might be updated at different dates and therefore these datasets were not necessarily accurate and coherent at all times. Moreover, for the actual data transfer to the NSI, all data needs to be in the appropriate format. At present, relevant datasets can be extracted from partner companies' information systems only if suitable interfaces are provided to recode data into the reporting format. Based on these practical considerations, such solutions were designed and implemented at the companies in order to enable automated report generation and data transfer. In this respect, tailor-made solutions are necessary as there is no standard product available on the market which could fulfill this task. Currently, a unique approach must be designed for every individual company. This also represents the main obstacle for the partner companies as the intended implementation of the automated reporting systems will usually involve substantial initial investments. Large providers of ERP and transport management systems have an incentive to produce a standard solution in the long run if there is sufficient demand; however, there may still need to be some adaption to individual requirements.

At the two partners, the standard interface was successfully implemented. The experts at the companies particularly appreciated that the systems allow for editing the reports and associated corporate data before dispatching the final report and thus leaving them complete governance of their own data. The implementation into the existing software environments took three and seven man-days, respectively. Although these efforts seem to be reasonable, the partner companies felt that – due to the low respondents' burden following the sharp cut of sample size in Austria in 2006 – the benefits do not justify these efforts.

Does this lead to the conclusion that a system like the prototype is generally cost ineffective? Or is this rather a subjective impression? Due to a lack of detailed data from Statistics Austria we present a general estimation. Particularly large companies with high reporting obligations could quickly profit from automated data generation and transfer. On average, reporting for one vehicle-week takes the companies 27 minutes [50]. Large companies possess fleets of several hundred lorries and have to report four times a year for a certain sample of their lorries (plus trailers). Before 2006, they had to prepare a report for all their lorries. Following the cut of the sample size to 14% of the original amount in 2006 means that a company with about 450

lorries now has to report 63 vehicle-weeks on average per quarter and therefore 252 vehicle-weeks per year, resulting in about 14 man-days (based on the calculation that reporting for one vehicle week takes 27min). Thus – even if one allows for the wages of programming staff to be double the wages of the person in charge of preparing the reports, the implementation of the prototype at the company would already pay off after the first year where seven man-days were required, as in our case. This calculation does of course not include the positive added value of having this data available at the company in a structured form. Additionally, various stakeholder benefit due to quality improvements which are not easily quantifiable but certainly significant.

6 Discussion and Conclusion

In this paper we argue for the relevance of innovative methods for data collection in general, and in the domain of road freight transport in particular. Under current circumstances, there is a three way trade-off between data quality, the respondents' burden and costs for the NSI. By designing a prototype that directly connects corporate ERP and TM systems (complemented by positioning data) to the NSI, we show how automated methods can improve statistics by eliminating this trade-off. We test this prototype for the Austrian market and explain how such a system should be designed to be applicable on a European level. The study follows the principle of Consortium Research [38] and applies the Cooperative Experimental System Development Approach (CESD) [40] to a domain of corporate and public interest. The resources and knowledge of both academics and practitioners are combined and an experimental experiment is delivered. This has been proven to work appropriately in the participating partner companies. Due to its generic design it can easily be implemented in other companies in the transport sector as well.

The findings of the pilot study in Austria identified several possible positive effects: considerably improved data quality due to greater sample sizes; the elimination of error sources; improved accuracy and detail; and the faster availability of results. Consequently, the implementation of innovative methods ensures compliance with Eurostat's principles from the 'Code of Conduct' by addressing the adequacy of resources and the commitment to quality, improving methodology, accuracy, timeliness and cost effectiveness, using appropriate statistical procedures, and relieving the burden on respondents [16].

However, companies are reluctant to implement the solution as initial investment costs and efforts are regarded as too high in relation to the current burden of reporting Austrian road freight transport statistics. Furthermore, they do not perceive the benefit of improved data quality and availability. There is, however, considerable potential for companies to change their perspective. On the individual corporate level, the scarce data currently collected for statistical reports of freight transport – being input for large-scale grossing-up only – does not exhibit any benefit. By using innovative data collection methods leading to a broad and accurate database, several side effects can be generated. These include controlling purposes, routing and consignment disposition, internal benchmarking, carbon footprint calculations, energy and fuel

saving calculations, etc. At a cross-company level, data could also be used for benchmarking purposes concerning various topics, such as strategic or environmental topics, although a third party would be needed to manage these data in a reliable and trustworthy manner [51]. However, the NSI, which meets all requirements in this respect, could act as such a platform manager due to several convincing advantages in a mutually beneficial situation: new incomes could be generated; provision of data would no longer be only a burden but also a direct pay-off for the companies; the quality of the overall statistics would improve; self-reinforcing processes towards (environmental and economic) efficiency and effectiveness would be induced at both corporate and co-operational (supply chain) levels; and, most important, the quality (accuracy, timeliness, sample size etc.) of official transport statistics could be boosted. Automated data collection and electronic data processing is capable of substantially increasing data quality, thus achieving all the goals mentioned above. Taking into account these additional benefits of improved data quality, the (initial) investment in automated collection tools becomes substantially more attractive.

It is therefore necessary to promote the idea of using innovative methods in companies to improve the process of transport data reporting, reduce redundant resources and consequently raise the quality of official data. Regarding the data reporting process, a long-term target is to completely replace paper-pencil questionnaires with electronic alternatives. On the basis of an analysis of individual companies' transport activities, a new sampling model could be developed, which would rely on automatically collected data from larger companies and possibly reduce reporting obligations for smaller companies. This in turn could support the survival of smaller and medium-sized companies – which are the backbone of many western industrialized economies – in a competitive environment driven by price pressure.

At a national level, the effects lead to the conclusion that the generated positive external effects of improved data quality (e.g. avoiding misinvestment in infrastructure projects) can be internalized by assigning public grants to support companies in the implementation of innovative methods. As a prerequisite, a new legal basis is needed which would allow for the further development of innovative data collection techniques. At the European level, all reporting countries could equally benefit from the developed technology as it is designed for ubiquitous use and can therefore contribute to more coherent road freight transport statistics within the European Statistical System. This design would also motivate software providers of ERPs and TM systems to include road freight transport data reporting as a feature in their packages. Taken to a general level, automated technology is capable of improving corporate and official processes and the underlying solution could serve as a role model and be transferred to other fields in further research.

As demonstrated in the course of this paper, there is a multitude of beneficiaries of improved official data quality at various levels and in different fields, and authors have repeatedly argued the need for it [10,23-25]. In a final step, we use the field of environmental sustainability in the domain of transport as an example to demonstrate the need for improved transport data and highlight the advantages for sustainability analyses. Transport is the fastest growing sector in

terms of the consumption of energy and the production of greenhouse gases in the European Union [51,52]. Among the various sectors, transport accounts for 13.1% of global GHG emissions and 24% in EU-27 [16]. Particularly researchers in the field of sustainability have repeatedly called for methods to improve data quality, availability and reliability in order to strengthen measurements related to environmental and social impacts [2,53,54]. Diverging data from different official sources and unreliable statistics are frustrating, erode confidence and make it difficult for policy-makers to develop sustainability strategies (e.g. CO₂ reduction goals) [2,55]. By illustrating the case of carbon footprinting in road freight transport in the UK, McKinnon and Piecyk [2] demonstrate serious problems and show how calculations based on different data sources lead to varying results. Innovative data collection methods can thus enlarge the data base and result in more detailed and accurate data.

In this way, a broad statistical body of data can be developed, eliminating the need to use a combination of inconsistent and incoherent data sources and reducing the amount of estimations that have to be made due to missing information. As a consequence, sustainability analysis in particular, and a variety of other analytical purposes in general, can be improved. Moreover, due to more efficient processes through automation and the electronic use of data transfer, various resources can be saved, resulting in ecological and economic benefits. Although one should not neglect IT-induced CO₂ emissions, such as energy consumption or hardware production [56], intelligent processes, environmental applications and responsible resource consumption are enablers for CO₂ emission reductions and capable of satisfying specific sustainability needs for internal and external stakeholders [57]. Data with improved quality therefore become more valuable and can be used for a multitude of (sustainability) purposes. Indeed, there is a broad variety of potential applications which companies can benefit from. These could be benchmarking with competitors in general or sustainability benchmarking (SBM) in particular [52], making it easier to comply with environmental obligations and to reduce liabilities. This might refer to e.g. emission calculations for products or services, or for companies, institutions on a regional, national and international level [59]. Consequently, the strengths and weaknesses of a company can be identified on a transparent and credible basis for all stakeholders. Besides comparative analyses at the corporate level, the measurement of environmental effects (e.g. carbon footprinting) on a cross-company basis can be improved.

Building on theoretical insights and practical experiences, the findings represent a first step in revolutionizing statistical data reporting in Europe. However, there is still a long road to go. Based on our findings, the next step should focus on further developments aimed at a 'marketable' product as well as spreading education about and raising awareness of the benefits of the approach. Moreover, a detailed analysis of the (Austrian) transport market, focusing on transport movements and the performance of companies, appears to be necessary to align and improve sampling. Future research should also be focused on extending the research across European borders. Transferring the pilot study to other countries will make it possible to test its feasibility and acceptability among other companies. Moreover, future research should analyse the acceptability among and willingness of NSIs towards the technology as well as evaluating its added value in companies.

Of course, some limitations apply due to the prototype character of the project, leading to limited generalizability. The opportunities and threats of broad connectedness need to be further evaluated [58] as the increasing availability of electronic data goes hand in hand with the risk of easy and unauthorized access and the (ab)use of information [60]. However, it is hoped that the considerable benefits resulting from improving the quality and availability of transport data will far outweigh these challenges for researchers, practitioners and public authorities.

Acknowledgments

This paper depicts the results of the project "InnoRFDat-X" which was funded by the Austrian Research Promotion Agency under the program "IV2Splus". Special thanks go out to the other project partners (AIT, Petschl Transporte and Gebrüder Weiss) for their support.

References

1. European Commission. *Eu transport in figures*; Publication Office of the European Union: Luxembourg, 2017.
2. McKinnon, A.C.; Piecyk, M.I. Measurement of co2 emissions from road freight transport: A review of uk experience. *Energ Policy* **2009**, *37*, 3733–3742.
3. Herold, D. M., Farr-Wharton, B., Lee, K. H., Groschopf, W. The interaction between institutional and stakeholder pressures: Advancing a framework for categorising carbon disclosure strategies. *Business Strategy & Development* **2018**.
4. IPCC. *Climate change 2007: Synthesis report*; IPCC: Geneva, 2007.
5. Karr, A.F.; Sanil, A.P.; Banks, D.L. *Data quality: A statistical perspective*; National Institute of Statistical Sciences: Research Triangle Park, North Carolina, 2005.
6. Knudsen, M.A.; Richn, J. Ex post socio-economic assessment of the oresund bridge. *Transport Policy* **2013**, *27*, 53-65.
7. Ülengin, F.; Özeydin, Ö.; Ülengin, B.; Kopp, A.; Önsel, Ş.; Kabak, Ö.; Aktaş, E. Are road transportation investments in line with demand projections? A gravity-based analysis for turkey. *Transport Policy* **2013**, *29*, 227-235.
8. Ferrer, S.; Ruiz, T. Enhancing survey methods and instruments to better capture personal travel behaviour. In *Proceedings of the NTTS - Conferences on New Techniques and Technologies for Statistics*, Brussels, 2013; pp 608-616.
9. Madre, J.-L.; Nguyen, T.-T.; Marchal, P. What short additional information should be collected to improve and validate the post-processing of gps mobility data? In *Proceedings of the NTTS - Conferences on New Techniques and Technologies for Statistics*, Brussels, 2013; pp 593-598.
10. Brackstone, G. *Managing data quality in a statistical agency*; Statistics Canada: 1999; pp 139-150.
11. Seyb, A.; McKenzie, R.; Skerrett, A. Innovative production systems at statistics new zealand: Overcoming the design and build bottleneck. *Journal of Official Statistics* **2013**, *29*, 73-97.
12. Hand, D.J. Modern statistics: The myth and the magic. *Journal of the Royal Statistical Society* **2009**, *172*, 287–306.

13. Pullinger, J. Statistics making an impact. *Journal of the Royal Statistical Society* **2013**, 176, 819–840.
14. Otto, B. Data governance. *Business and Information Systems Engineering* **2011**, 3, 241–244.
15. United Nations. *Fundamental principles of official statistics*; United Nations: New York, 2013.
16. Eurostat. European statistics code of practice for the national and community statistical authorities. Eurostat: Luxembourg, 2011.
17. Kenett, R.S.; Shmueli, G. On information quality. *Journal of the Royal Statistical Society* **2014**, 177, 3–38.
18. Holt, D.T. Official statistics, public policy and public trust. *Journal of the Royal Statistical Society* **2008**, 171, 323–346.
19. Richardson, A.J.; Ampt, E.S.; Meyburg, A.H. *Survey methods for transport planning*. 1 ed.; Eucalyptus Press: Melbourne, Australia, 1995.
20. McKinnon, A.C. Logistics and the environment. In *Handbook of transport and the environment*; Hensher, D.A.; Button, K.J., Eds. Elsevier Science: Oxford, 2003; pp 665–685.
21. Herold, D. M., Lee, K. H. The influence of internal and external pressures on carbon management practices and disclosure strategies. *Australasian Journal of Environmental Management* **2019**, 26, 63–81.
22. Roider, O.; Sammer, G. In *Development of an innovative model concept to improve the austrian transport statistics*, Reliability and Statistics in Transportation and Communication, Riga, Latvia, 2012; Transport and Telecommunication Institute: Riga, Latvia, pp 152–159.
23. Madre, J.-L.; Kalinowska, D. Vehicle-based surveys: Toward more accurate and reliable data collection methods. In *8th International Conference on Survey Methods in Transport*, Annecy, France, 2008.
24. Daas, P.; Roos, M.; De Blois, C.; Hoekstra, R.; Ten Bosch, O.; Ma, Y. New data sources for statistics: Experiences at statistics netherlands. In *2011 European New Technique and Technologies for Statistics Conference*, Brussels, 2011.
25. Nordbotten, S. The use of administrative data in official statistics - past, present, and future. In *Official statistics - methodology and applications: In honour of daniel thorburn*, Carlson, M.; Nyquist, H.; Villani, M., Eds. Stockholm, 2010; pp 205–223.
26. Cook, L. Editorial: The national statistician: 2 years on. *Journal of the Royal Statistical Society* **2003**, 166, 1–4.
27. Bonnel, P.; Lee-Gosselin, M.; Zmud, J.; Madre, J.-L. Transport survey methods. Keeping up with a changing world. In *Keeping up with a changing world: Challenges in the design of transport survey methods*, Bonnel, P.; Lee-Gosselin, M.; Zmud, J.; Madre, J.-L., Eds. Emerald Group Publishing Limited: Bingley, UK, 2009; pp 3–15.
28. Zmud, J.; Lee-Gosselin, M.; Munizaga, M.; Carrasco, J.-A. Transport surveys: Considerations for decision makers and decision making. In *Transport survey methods: Best practice for decision making*, 1 ed.; Zmud, J.; Lee-Gosselin, M.; Munizaga, M.; Carrasco, J.-A., Eds. 2013; pp 3–21.
29. Stopher, P.R.; Prasad, C.; Wargelin, L.; Minser, J. Conducting a gps-only household travel survey. In *Travel survey methods: Quality and future directions*, Stopher, P.R.; Stecher, C., Eds. Elsevier: 2006; pp 91–114.
30. Carvalho, D.; Rodrigues, A.; Jones, S. *Mediate - methodology for describing the accessibility of transport in europe*, 2011.

31. Wolf, J.; Wilhelm, J.; Casas, J.; Sen, S. A case study: Multiple data collection methods and the ny/nj/ct regional travel survey. In *Travel survey methods: Quality and future directions*, Stopher, P.R.; Stecher, C., Eds. Elsevier: 2006; pp 71-90.
32. Eurostat. *Methodologies used in surveys of road freight transport in member states and candidate countries*. Eurostat Methodologies and Working papers: 2011.
33. Fürst, E. The future of transport surveys: Users' requirements for official transport statistics. *Austrian J Stat* **2009**, *38*, 265-275.
34. Arentze, T.; Feng, T.; Robroeks, J.; Van Brakel, M.; Huibers, R. Compliance with and influence of a new in-car navigation system for trucks: Results of a field test. *Transport Policy* **2012**, *23*, 42-49.
35. Herold, D.M.; Lee, K.-H. The influence of the sustainability logic on carbon disclosure in the global logistics industry: The case of dhl, fdx and ups. *Sustainability* **2017**, *9*, 601.
36. Herold, D. M.; Lee, K. H. Carbon disclosure strategies in the global logistics industry: Similarities and differences in carbon measurement and reporting. In *Pathways to a Sustainable Economy*. Springer, Cham: 2018, pp 87-101.
37. Eltinge, J.L.; Biemer, P.P.; Holmberg, A. A potential framework for integration of architecture and methodology to improve statistical production systems. *J Official Stat* **2013**, *29*, 125-145.
38. Österle, H.; Otto, B. Consortium research. *Business and Information Systems Engineering* **2010**, *2*, 283-293.
39. March, S.T.; Smith, G.F. Design and natural science research on information technology. *Decision Support Systems* **1995**, *15*, 251-266.
40. Grønbaek, K.; Kyng, M.; Mogensen, P.H. Towards a cooperative experimental system development approach. In *Computers and design in context*, Kyng, M.; Mathiassen, L., Eds. MIT press: Cambridge, MA, 1997; pp 201-238.
41. Winter, R. Interview with jay f. Nunamaker, jr. On "toward a broader vision of is research". *Business and Information Systems Engineering* **2010**, *2*, 321-329
42. European Union. Regulation (eu) no. 70/2012 of the european parliament and of the council of 18 january 2012 on statistical returns in respect of the carriage of goods by road. European Union: 2012.
43. Statistics Austria. *Standard-dokumentation metainformation (definitionen, erläuterungen, methoden, qualität) zur statistik des straßengüterverkehrs*; Wien, 2010.
44. Oberhofer, P.; Fürst, E. Sustainable development in the transport sector: Influencing environmental behaviour and performance. *Bus Strateg Environ* **2013**, *22*, 374-389.
45. Fürst, E. Das neue stichprobenkonzept für die straßengüterverkehrsstatistik. *Statistische Nachrichten* **2008**, *63*, 409-417.
46. Fürst, E.; Oberhofer, P.; Kummer, S. Innovative datenerfassung in der straßengüterverkehrsstatistik: Analyse der rahmenbedingungen und ergebnisse eines forschungsprojektes. *Internationales Verkehrswesen* **2013**, *65*, 27-29.
47. Kummer, S.; Kotnik, S.O.; Novakov, S.; Popsky, D. Innovative erhebungsmethoden und datengrundlagen im bereich verkehr. WU - Vienna University of Economics and Business: Vienna, 2008.
48. Paradigma Unternehmensberatung Gmbh. *Innorfdat-x. Endbericht für das forschungsprojekt zu innovativen methoden in der straßengüterstatistik*; 2012.
49. Statistics Austria. *Straßengüterverkehr - entwicklung des meldeaufwands 2001-2012*; Statistics Austria: Vienna, 2013.
50. Fürst, E.; Herold, D. Fare Evasion and Ticket Forgery in Public Transport: Insights from Germany, Austria and Switzerland. *Societies* **2019**, *8*, 98.

51. Koslowski, T.; Strüker, J. Erp on demand platform. *Business and Information Systems Engineering* **2011**, *3*, 359-367.
52. European Environmental Agency. *Final energy consumption by sector (csi 027/ener 016)*; European Environment Agency: 2010.
53. Eurostat. Transport energy consumption and emissions.
http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Transport_energy_consumption_and_emissions
54. Wiedmann, T. Carbon footprint and input-output analysis - an introduction (editorial). *Economic Systems Research* **2009**, *21*, 175–186.
55. Dobrovnik, M.; Herold, D.; Fürst, E.; Kummer, S. Blockchain for and in Logistics: What to Adopt and Where to Start. *Logistics* **2018**, *2*, 18.
56. Herold, D.M. Has carbon disclosure become more transparent in the global logistics industry? An investigation of corporate carbon disclosure strategies between 2010 and 2015. *Logistics* **2018**, *2*, 13.
57. Buhl, H.U.; Laartz, J.; Löffler, M.; Röglinger, M. Green it reicht nicht aus! *Wirtschaftsinformatik & Management* **2009**, *1*, 54-58.
58. Schmidt, N.-H.; Ereke, K.; Kolbe, L.M.; Zarnekow, R. Sustainable information systems management. *Business and Information Systems Engineering* **2009**, *1*, 400-402.
59. Peters, G.P. Carbon footprints and embodied carbon at multiple scales. *Current Opinion on Environmental Sustainability* **2010**, *2*, 245–250.
60. Buhl, H.U.; Müller, G. The “transparent citizen” in web 2.0. *Business and Information Systems Engineering* **2010**, *2*, 203-206.