DE GRUYTER
OPEN

Journal of Official Statistics, Vol. 34, No. 4, 2018, pp. 797–809, http://dx.doi.org/10.2478/JOS-2018-0040

Preface

1. New Techniques and Technologies for Official Statistics

Official statistics are steadily moving away from the traditional "one survey – one product" paradigm, with growing efforts to integrate the data sources at hand and include new data sources, be it administrative data or altogether new "big data", such as web scraped data or mobile phone data. Moreover, there is an increased interest in and awareness of the need for reliable processes and systems underpinning an agile production of official statistics and the importance of both producing official statistics and effectively disseminating them with methods that ensure that they reach their intended users.

Research in official statistics, being close to the cutting edge, is a leading indicator of this trend. In recent years, interest in the conference series on New Techniques and Technologies for Statistics (NTTS), organised since 1992, has increased considerably, and NTTS has emerged as a major official statistics research forum. As can be seen in Figure 1, the most recent conference, NTTS 2017, which took place in March 2017 in Brussels, saw a record level of participation. Figure 2 illustrates the diversity of stakeholders brought together by the conference, which included official statisticians (half of the participants represented national statistical institutes, regional statistical institutes, and Eurostat), other parts of the public sector (one-fourth of the participants represented European Union Institutions and national authorities), academia (15% of participants) and the private sector (8%).

Building on the successful experience of the Journal of Official Statistics (JOS) special issue with articles from NTTS 2013 (see Karlberg et al. 2015), the NTTS 2017 Scientific Committee reached out to the JOS editorial board in 2016 to explore the possibility of a JOS special issue based on articles from NTTS 2017. This special issue of the JOS, in which eight articles are presented, represents the final outcome of a highly selective screening and peer review process. First, we would like to thank all authors, including authors of the numerous articles that were unsuccessful in the reviewing process, for considering the JOS as a platform for their work. We would also like to emphasise that this special issue would not have been possible without the 60 referees who kindly agreed to review one (or frequently more versions) of the manuscripts. The manuscripts all benefitted from these constructive reviews, improving greatly in quality.

In this context, we would like to express our deep sorrow at learning that Rein Ahas, Professor in Human Geography at the University of Tartu, passed away unexpectedly on 18 February 2018 at the age of 51. Rein served with us on the NTTS 2017 Scientific Committee, and the very last service that he rendered in this regard was a review of one of the manuscripts for this special issue. In the review, he emphatically insisted on the proper application of scientific standards, constructively providing concrete advice on

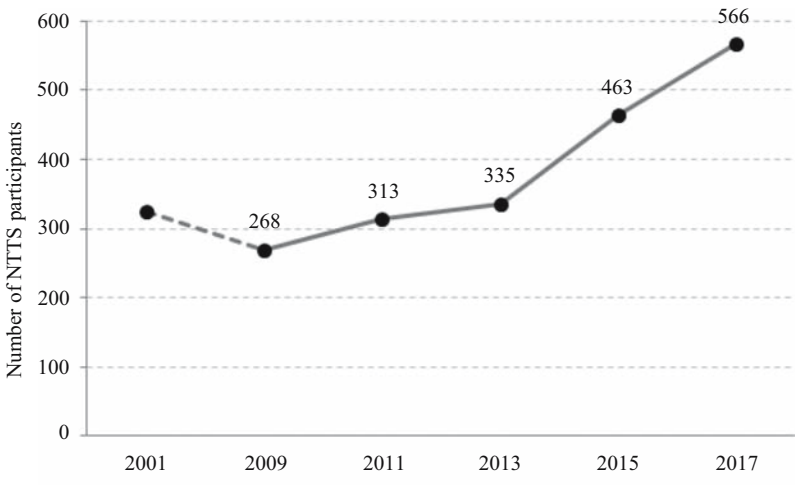

Fig. 1. Number of effectively attending NTTS participants (registrations adjusted for no-shows). Source for 2009-2017: the registration system of each respective conference. For 1992-2001, the sole data source identified to date is a conference report on NTTS 2001.

how the authors could do so in practice. With this, we would like to acknowledge the exceptional service that Rein has given to the NTTS conference series, and offer our deepest condolences to his family.

Appropriate management of core statistical, methodological and IT processes is a challenge that many statistical offices face. Section 2 presents the first two articles of this special issue, which both deal with innovative solutions tackling this fundamental aspect.

In Section 3, we present three articles that all deal with the integration of multiple sources in official statistics, with an emphasis on administrative and registry data, challenges regarding reconciliation (matching and linkage) and the quality control that this entails.

While regular official statistics production based on new big data sources is still in its infancy, there are many experiments underway to investigate its feasibility. Two articles quite far apart on the exploration/application scale are presented in Section 4. One of them has a specific official statistics concept in mind, whereas the other article is of a more exploratory nature.

The eighth and final article of this special issue, presented in Section 5, deals with an innovative approach to dissemination, by means of natural language, underpinned by fuzzy logic. In Section 6, we offer some concluding remarks, focusing on features common to the articles of this special issue.

2. Agile Processes for Statistical Offices

In the first article of this special issue, Salgado, Esteban, Novás, Saldaña and Sanguiao advocate an agile approach to data organisation and process design based on functional modularity. They underline that while it is common practice to see these principles applied in (a) the design of software for statistical production, they are rarely applied consistently when it comes to two other aspects of statistical production, (b) statistical production

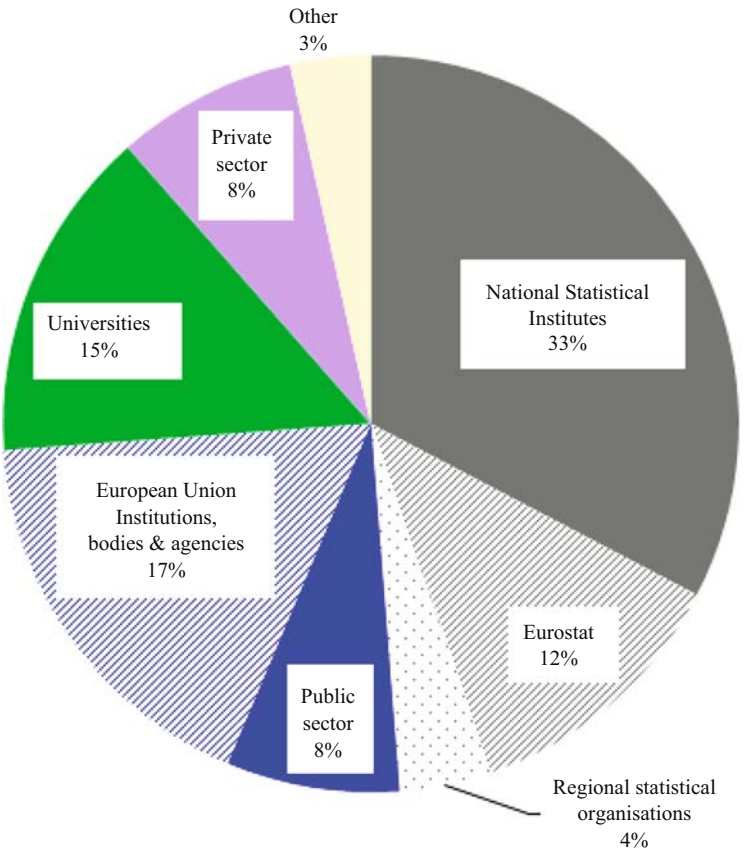

Fig. 2. Participants at NTTS 2017 by type of organisation. Source: registration database (n = 566). Participant categorisation was done manually.

metadata, and (c) statistical methodology. This unfortunate state of affairs is by no means unique to official statistics; in many organisations, stakeholders tend to ignore non-IT aspects of processes, and (implicitly) delegate or outsource key process design decisions to IT services, instead of requiring the business process owners to set up their processes. This results in a lopsided setup, that is, well-defined IT processes that might provide "the right answer to the wrong question", as they may be supporting vaguely or suboptimally defined business processes. Salgado and coauthors argue that the same principles must be applied to *all* of the three aforementioned aspects (a), (b) and (c). This is especially the case, since (as they convincingly argue) official statistics production is a *complex system* (Saltzer and Kaashoek 2009), composed of "(i) a large number of components, (ii) a large number of interconnections between these components, (iii) many irregularities in these interconnections, since the lack of regularity is indeed the rule rather than the exception, (iv) a long description of the system and its related management [. . .] and (v) a team of designers, implementers, and/or maintainers to handle the system".

Their proposed approach is based on object-oriented and functional computation paradigms. "The former comprises a standardised key-value pair abstract data model, where keys are constructed by means of the structural statistical metadata of the

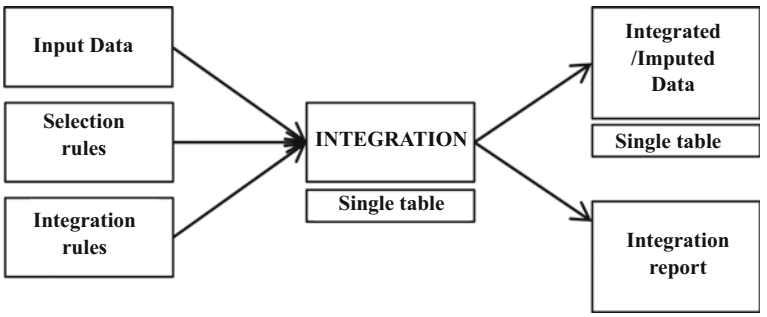

Fig. 3. Abstract information model of a general integration process. Source: Cesaro and Tininini.

production system", whereas the latter relies heavily on "the principles of functional modularity (modularity, data abstraction, hierarchy, and layering) to design production steps." Salgado and coauthors conduct a proof of concept (for the editing process step) in the field of Short-Term Business Statistics, and draw several valuable conclusions on various aspects seeking to emphasise lessons learned that could be transferrable to other statistical offices. In view of their experience, they advocate a non-linear "spiral approach" to software development, and make their case for a "change of mindset to conceive software as constantly evolving instead of as a closed definitive tool [as this] is necessary to industrialise and modernise the statistical production". The authors give practical hints, such as "wrapping" in various ways (developing a simple Excel file to serve as the "front end" for generating XML code; complying with a "SAS only" office policy by running SAS macros that execute R scripts in batch). Importantly, they make all code available in the public domain via GitHub (Esteban et al. 2017).

In the second article of this special issue, Cesaro and Tininini propose a service-oriented architecture (SOA) following the style of lightweight basic integration, and describe how this has been successfully deployed for validation (as well as for prioritisation among multiple sources) for the Italian Statistical Business Registers (SBR). They thoroughly investigate key design choices that drive performance.

While these two special issue articles have different focus, they share many features. Both of them rely on, and discuss, core artefacts developed by the official statistics modernisation community, such as the Generic Statistical Business Process Model (GSBPM), the Generic Statistical Information Model (GSIM) and the Common Statistical Production Architecture (CSPA). They are in favour of high-granularity processes/services (although the level of granularity might differ between the two articles) that are stateless (depending only on the input they receive). They follow the *active metadata* paradigm, letting metadata (including rules, as illustrated in Figure 3) be a part of the input to processes rather than hard-coded into the processes themselves. In short, both articles present solutions that are conducive to shareability and agility.

3. Consolidating and Reconciling Multiple Sourses

In the words of Vâju and Meszaros (2018), statistical authorities "need to produce data faster in a cost effective way, to become more responsive to users' demands, while at the

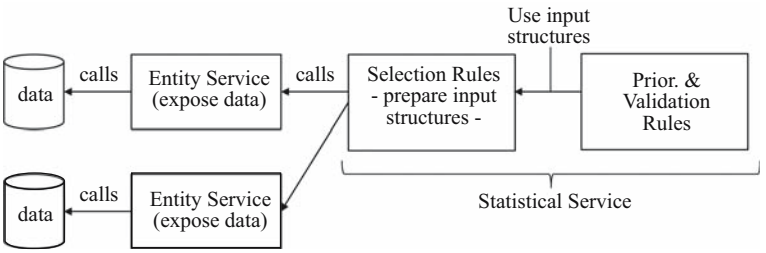

Fig. 4. Logical SOA architecture with Entity services and statistical services: selection rules adapt input data for subsequent processing for prioritization and validation purposes. Source: Cesaro and Tininini.

same time providing high quality output. One way to fulfil this is to make more use of already available data sources, and in particular administrative sources, most typically used in combination with other sources." A key objective of the service developed by Cesaro and Tininini is to support activities fulfilling this objective, by integrating multiple sources, as illustrated in Figure 4. The consolidation and reconciliation of multiple sources is also the topic of three other articles of this special issue, and we will present those in this section.

When one speaks of a set of "administrative data", the typical case that springs to mind is that of a single, monolithic, governmental authority at national level with one common set of rules for data collection. However, in many cases, administrative data are provided by decentralised autonomous administrations (for instance, municipalities that collect data on their inhabitants). Thus, consolidation of multiple sources (each reporting entity being one source) has to take place in order to arrive at one (e.g. national) data set for a certain domain. In the third article of this special issue, van Delden, van der Laan and Prins address this situation – more specifically, how to deal with the heterogeneity in reporting that this can entail; they present a method (illustrated in Figure 5) to detect under- and overreporting by data suppliers for "decentralised administrative data" for the case of change estimates. The method is successfully applied to a case study with administrative hospital data, and the authors conclude by setting out a number of steps concerning adaptations and extensions needed to deploy the methods in official statistics production. While the authors specifically treat the case of decentralised autonomous administrations, the scope of applicability might be even wider. For instance, local offices of a national administration might also have *de facto* developed their own administrative traditions, even if there is a common set of rules nominally applicable to the entire authority.

A reconciliation of multiple sources along another dimension takes place when multiple lists for (largely) the same set of units are brought together. This is the topic of the fourth article of this special issue, wherein Di Consiglio and Tuoto propose a method that is applicable in cases where the goal is to measure the size of a population (partially) enumerated in different lists. Their multiple lists linkage procedure tackles the problem of adjusting population estimates in the presence of linkage errors. With a distribution of estimates close to the one that could be expected without any linkage errors, their proposed class of estimators performs better than the alternatives investigated.

Whereas Di Consiglio and Tuoto consider the reconciliation of more or less complete data sets, each normally covering a very large part of a target population under study,

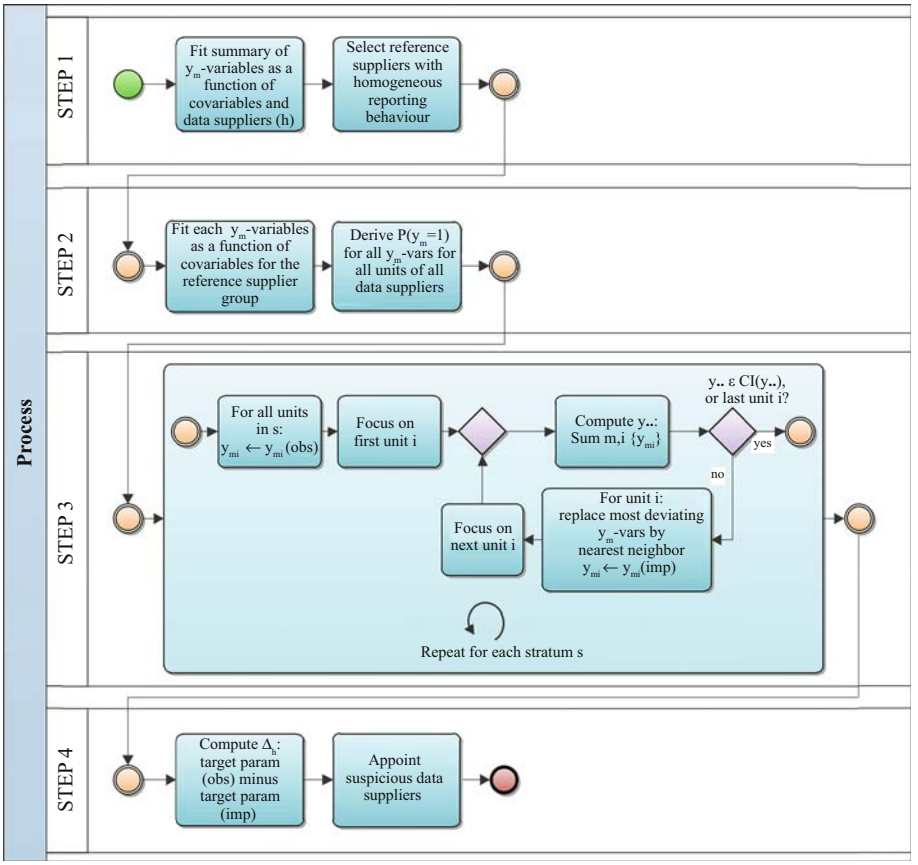

Fig. 5. Flow chart of the four steps of the methodology proposed by van Delden and coauthors.

record linkage can also be conducted for the purpose of adding value to a survey, by bringing in data from administrative records. In the fifth article of this special issue, Gessendorfer, Beste, Drechsler, and Sakshaug note that "Surveys that perform record linkage to administrative records are often required to obtain informed consent from respondents prior to linkage. A major concern is that nonconsent could introduce biases in analyses based on the linked data." This missingness due to nonconsent is illustrated in Figure 6.

To remedy this missing data problem created by nonconsent, they propose that nonconsenters be matched with statistically similar units in the target administrative

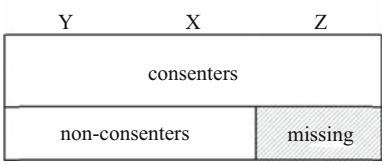

Fig. 6. The missing data situation when a survey data set has been combined with administrative data by means of record linkage (X being variables common to both data sets, Y being variables present in the survey and Z being variables present in the administrative data set). Source: Gessendorfer and coauthors.

database. In an empirical study, they assess the effectiveness of statistical matching for this purpose using data from two German panel surveys that have been linked to an administrative database of the German Federal Employment Agency. Their findings are mixed: the method can be effective in reducing nonconsent biases in marginal distributions, but biases in multivariate estimates can sometimes be worsened. This finding is valuable in itself; it is important that scientific journals do not limit their presentations to "success stories", dooming researchers to re-conduct studies of approaches already found to be inappropriate by others. (As noted by Karlberg and Radermacher (2014), "a lopsided evidence base, with the range of outcomes truncated to exclude failures, is a good example of the classical 'file drawer problem' [Rosenthal 1979]".)

4. Nontraditional Data Sources

While the administrative data sources discussed in the previous section demonstrably come with their own set of challenges, they are comparatively well-structured and well-defined. In contrast, "big data" are much harder to incorporate in statistical production, owing not only to their sheer volume, but also to issues such as their lack of structure and the fact that they are frequently held by third parties. Already five years ago (DGINS 2013), it was acknowledged that big data "represent new opportunities and challenges for Official Statistics" and the European Statistical System and its partners were encouraged to effectively examine the potential of big data sources in that regard. For the better part of the last decade, the official statistics community has been grappling with the challenge of integrating the new big data sources into official statistics production.

At the European level, a multi-pronged action plan and roadmap was established, and within the ESS Vision 2020 (Eurostat 2015), the Big Data (BIGD) project was tasked with implementing it. This has generated several outcomes, including methodological studies (such as the overview by Beręsewicz et al. (2018) of methods for treating selectivity in big data studies), analyses regarding aspects such as legal issues, ethics, quality and IT requirements, and a number of pilot studies (BDES 2018) involving new sources such as web scraping, smart meters, vessel tracking and mobile phone data.

On this latter data source, Vanhoof, Reis, Ploetz and Smoreda note in the sixth article of this special issue that "Mobile phone data are an interesting new data source for official statistics. However, multiple problems and uncertainties need to be solved before these data can inform, support or even become an integral part of statistical production processes." They then proceed to analyse the performance of five home detection algorithms (HDAs), based on mobile phone data characteristics such as amount of activities, amount of distinct days of activities, time of day constraints and space constraints. However, in their study, based on French Call Data Record (CDR) data, it turns out that no matter which HDA is applied, the dissimilarity between ground truth data and the estimated home location is large. While it is unsurprising that this is the case for holiday months (see Figure 7 for an illustration of the situation in August), this dissimilarity remains at a high level for all calendar months. The authors propose remedial actions at three levels: (i) studies at the individual level allowing the simultaneous observation of ground truth data and the related CDR data, (ii) reconciliation at the national level to compensate for the differences in local market shares between operators,

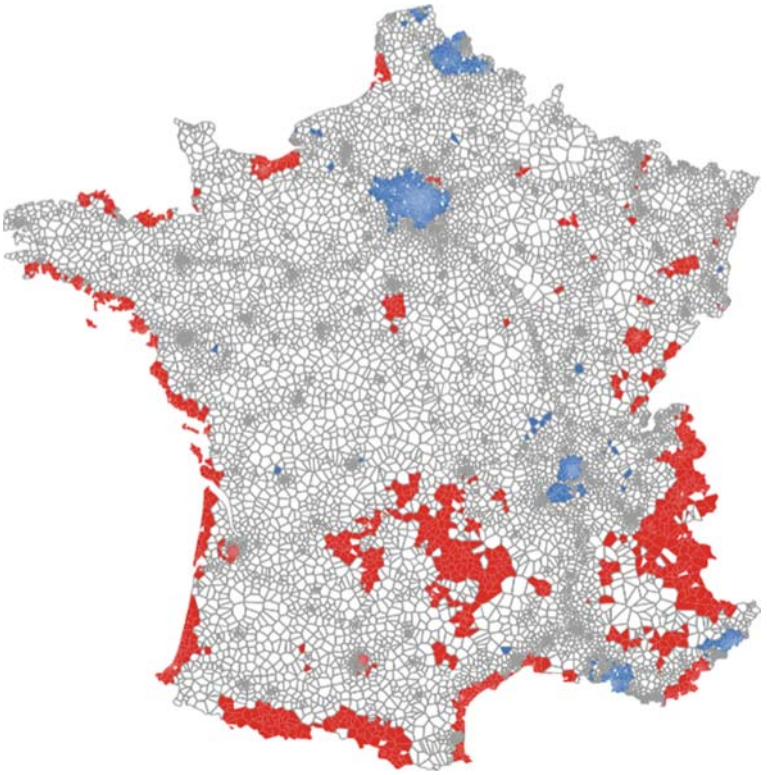

Fig. 7. French hotspots (red) and coldspots (blue) in August based on the ratio between the (mobile phone data based) activity-based home location and the population counts of a validation dataset. Note the effects of the reduced activity level in the capital region and the increased activity level in many typical holidaymaking areas. Source: Vanhoof and coauthors.

and (iii) testing at the international level to ascertain robustness of methods – across countries and over time.

To encourage experimentation with big data sources, one might try approaches that involve a competitive element. Two initiatives in this regard, both launched in 2016, were presented at NTTS 2017. The Big Data for Official Statistics Competition (BDCOMP) was the first official statistics nowcasting competition at EU level with a big data focus (Kovachev et al. 2017), requiring participants to submit a nowcast before the publication of official statistics. The BDCOMP could be said to follow a "marathon" approach, with participating teams delivering monthly submissions over roughly one year's time, whereas the EU Big Data Hackathon (Eurostat 2017a) was more of a "sprint". Over a period of just two days (and nights), participants had to develop solutions to a policy question ("*How would you support the design of policies for reducing mismatch between jobs and skills at regional level in the EU through the use of data?*"). A total of 22 teams from European National Statistical Institutes competed to develop a data analytics tool to support this, and then, on the third day, "pitch" their solutions, each with a time slot of just 6 minutes, to the evaluators. Two independent panels of evaluators (one statistical panel and one industry panel) assessed the contributions according to the criteria of relevance,

methodological soundness, communication, innovativity and replicability. The 1st, 2nd and 3rd prizes were awarded to the teams from Croatia, France and Estonia, respectively (Eurostat 2017b).

Opik, Kirt, and Liiv describe, in the seventh article of this special issue, the 3rd prize winning contribution to the EU Big Data Hackathon. They "present a visual method for representing the complex labour market internal structure from the perspective of similar occupations based on shared skills". Their method, based on graph theory (West 2001), is designed to enable adding extra layers of external information. Moreover, they offer a prototype for a tool allowing users to interact with the visualisation.

To demonstrate their methods and tools, Opik and coauthors conducted a case study in which they analyse the labour market together with the megatrend of automation and computerisation of jobs. Starting out by integrating data sets on job vacancies, they arrive at a graph depicting 2,950 occupations, with links between them based on their occupational similarity, and nodes annotated with megatrend (susceptibility to automation/computerisation) and supply data.

They proceeded to build a user interface that visualises the graph in a way that renders a zoomable and scrollable scalable vector graphics (SVG) document for browsing the graph online (supported by most modern web browsers). It has both a "move and zoom" mode, allowing for exploration, and a "query" mode (illustrated in Figure 8). The source code for the prototype (as well as the prototype itself) is made available on GitHub.

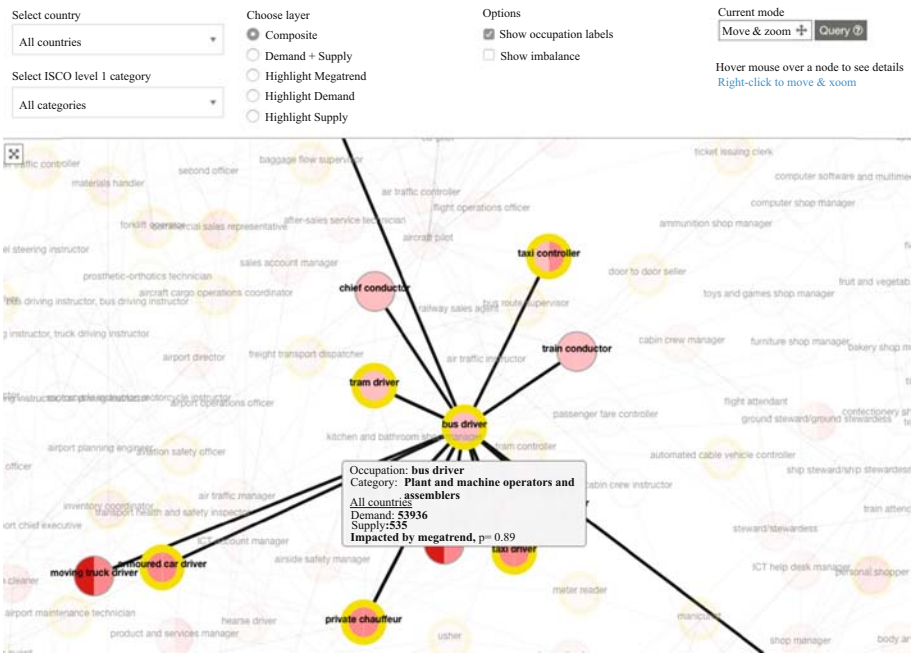

Fig. 8. Screenshot illustrating the query mode activated for the visualisation prototype. When an analyst moves the mouse cursor over a node, a small tooltip with demand and supply numbers is displayed. Hovering also highlights connected jobs and fades out the rest of the graph. Source: Opik and coauthors.

5. New Ways of Disseminating Official Statistics

The official statistics community has become increasingly aware of the importance not only of how official statistics are produced, but also of how they are presented. As an example, two (out of five) key areas of the ESS Vision 2020 (Eurostat 2015) – *focus on users* and *improve dissemination and communication* – concern how to understand user needs, and how to improve user communication. To tackle these issues, the ESS Vision 2020 implementation project *Digital communication, User analytics and Innovative products* (DIGICOM; see Kormann 2016) was launched. Given the simultaneous focus on the use of new big data sources and the communication and visualisation of their outcome, the EU Big Data Hackathon (2017a) was carried out jointly by the BIGD and DIGICOM projects.

While the DIGICOM project has a rich and varied portfolio of activities concerning, for example, user analysis, visualisation, open data dissemination, statistical literacy and gamification (Kormann et al. 2018), there are numerous other innovative activities underway at any given moment in the official statistics research community. A manifestation of this is the final article of this special issue, in which Hudec, Bednárová and Holzinger propose a method for disseminating statistics verbally, by means of natural language. Moreover, and in order to reflect the elasticity of many verbal quantifiers, the authors apply fuzzy logic to allow a "sliding scale" definition of linguistic terms; for attributes related to the values of a unit of the variable under study (such as "high pollution" and "low pollution") as well as quantifiers related to the relative frequencies of units possessing these attributes ("few enterprises", "about half of the enterprises", "most enterprises"). They demonstrate their concept using a test interface interpreting summaries from real municipal statistics data.

6. Outlook for Continued Research and Innovation in Official Statistics

A number of recent trends in statistical research and innovation have been manifested in the articles of this special issue. As pointed out in the article by Salgado and coauthors, the modernisation of statistical production "is to be accomplished under the high pressure of product release calendars within the traditional stove-pipe production model and a decreasing amount of budgeted resources". In Section 2, we have seen approaches that would allow official statistics to innovate in an agile way while under such resource constraints.

A recurrent theme in this special issue is that of learning through sharing. For instance, both Salgado and coauthors and Opik and coauthors embrace this principle by making their source code freely and publicly available on GitHub. Moreover, sharing is also a matter of letting colleagues know what does not work, so that they are already aware of the weaknesses associated with (and the need to improve) certain approaches. In this respect, Gessendorfer and coauthors, in demonstrating the limitations of statistical matching, as well as Vanhoof and coauthors, who point to the weaknesses of home detection algorithms, render valuable services to the official statistics research community.

"Improving the numerical and statistical literacy of citizens, journalists and policymakers will help to increase their awareness and ability to critically assess news, including fake news, and their participation in the democratic process" as noted by the

Power from Statistics panel on statistics in the digital era (Eurostat 2018). As Hudec and coauthors emphasise, the objective of the method that they propose is not to replace existing dissemination with verbal summaries, but rather to provide an alternative way of dissemination, which might be useful to certain categories of users. For instance, this might be a useful complement (to numbers) for persons that are innumerate, or have a low level of statistical literacy. Moreover, visualisation is not the ideal way to disseminate to everyone – visually impaired users, but also users for whom the interpretation of diagrams does not come naturally may be unable to grasp what is being communicated visually. "Reading graphs and charts is far from intuitive", as pointed out by Cairo (2018).

When new, unforeseen political and societal challenges emerge, policymakers and other stakeholders require timely evidence, produced with short lead times. In this context, the article of Opik and coauthors demonstrates the capacity for rapid innovation in the official statistics community. Based on the effectiveness of the "hackathon" approach, we were glad to learn that NTTS 2019, which will take place in Brussels from 12 to 14 March 2019, will also include a hackathon. This is just one of the many elements that contribute to the relevance of the NTTS series of conferences in advancing the research frontiers of official statistics.

Martin Karlberg
Guest Editor

Silvia Biffignandi
Piet J.H. Daas
Loredana Di Consiglio
Anders Holmberg
Risto Lehtonen
Ralf T. Münnich
Boro Nikic
Marianne Paasi
Natalie Shlomo
Roxane Silberman
Ineke Stoop
Guest Associate Editors

7. References

BDES. 2018. *Minutes*. Minutes from the 2018 conference on Big Data for European Statistics (BDES 2018). Available at: https://webgate.ec.europa.eu/fpfis/mwikis/ess-netbigdata/images/c/cf/BDES_2018_05_14-15_Sofia_Minutes.pdf (accessed August 2018).

Beręsewicz, M., R. Lehtonen, F. Reis, L. Di Consiglio, and M. Karlberg. 2018. *An Overview of Methods for Treating Selectivity in Big Data Sources*, Eurostat Statistical Working Paper. Doi: https://doi.org/10.2785/312232.

Cairo, A. 2018. "Uncertainty and graphicacy: How should statisticians, journalists, and designers reveal uncertainty in graphics for public consumption?" In *Power from*

Statistics: data, information and knowledge – Outlook Report 91-102. Publications Office of the European Union. Doi: https://doi.org/10.2785/721672.

DGINS (*Directeurs Généraux des Instituts Nationaux de Statistique*). 2013. *Scheveningen Memorandum – Big Data and Official Statistics*. Adopted by the European Statistical System Committee on 27 September 2013. Available at: https://ec.europa.eu/eurostat/documents/7330775/7339365/Scheveningen-memorandum-27-09-13/2e730cdc-862f-4f27-bb43-2486c30298b6 (accessed August 2018).

Esteban, E., S. Saldaña, and D. Salgado. 2017. *Software Implementation of Optimization-based Selective Editing Techniques at Statistics Spain (INE)*. UNECE Work Session on Statistical Data Editing. The Hague, 24-26 April 2017. Available at: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2017/mtg2/Paper_19_Stat-Spain.pdf (accessed August 2018).

Eurostat. 2015. *ESS Vision 2020 – Building the Future of European Statistics*. Luxembourg: Publications Office of the European Union. Doi: https://doi.org/10.2785/078607.

Eurostat. 2017a. *European Big Data Hackathon*. Available at: https://ec.europa.eu/eurostat/cros/EU-BD-Hackathon (accessed August 2018).

Eurostat. 2017b. *Winners of the EU Big Data Hackathon awarded*. Eurostat news release published 17 March 2017. Available at: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/CDN-20170317-1 (accessed August 2018).

Eurostat. 2018. *Power from Statistics: Data, Information and Knowledge – Guidance Report*. Publications Office of the European Union. Doi: https://doi.org/10.2785/302736.

Karlberg, M., S. Biffignandi, P.J.H. Daas, A. Holmberg, B. Hulliger, P. Jacques, R. Lehtonen, R.T. Münnich, N. Shlomo, R. Silberman, and I. Stoop. 2015. "Preface." *Journal of Official Statistics* 31(2) . Doi: https://doi.org/10.1515/jos-2015-0011.

Karlberg, M. and W. Radermacher. 2014. "Discussion: Methodology Architecture – an area under construction." *Statistical Journal of the IAOS* 30(4): 393–398. Doi: https://doi.org/10.3233/SJI-140855.

Kormann, C. 2016. *Innovation for Dissemination in the European Statistical System – the Approach of the DIGICOM Project*. Paper presented at the 2016 UNECE workshop on statistical data dissemination and communication. Available at: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.45/2016/Christine_Kormann_paper.pdf (accessed June 2018).

Kormann, C., S. Klasinc, M.J. Vinuesa Angulo, N. Tsiligkaki, A. Ilkova, P. Campos, M. Jankowska, P. Collesi, X. Caruso, and S. Luhmann. 2018. *A European Effort to Explore Games and the Gamification of Official Statistics*. Paper presented at the 2018 Conference of European Statistics Stakeholders. Available at: https://coms.events/cess2018/data/abstracts/en/abstract_0014.html (accessed August 2018).

Kovachev, B., M. Karlberg, B. Nikic, B. Oancea, and P. Righi. 2018. *The Big Data for Official Statistics Competition – Results and Lessons Learned*. Paper presented at NTTS 2017. Available at: https://ec.europa.eu/eurostat/cros/ntts2017programme/data/abstracts/abstract_295.html (accessed August 2018).

Rosenthal, R. 1979. "The file drawer problem and tolerance for null results." *Psychological Bulletin* 86: 638–641. Doi: http://dx.doi.org/10.1037/0033-2909.86.3.638.

Saltzer, J.H., and M.F. Kaashoek. 2009. *Principles of Computer System Design: An Introduction*. MIT Press.

Vâju, S., and M. Meszaros. 2018. *Administrative Data and Quality – Guidelines Towards Better Quality of Administrative Data*. Paper presented at the 2018 European Conference on Quality in Official Statistics (Q2018). Available at: https://www.q2018.pl/wp-content/uploads/Sessions/Session%2037/M%C3%A1ty%C3%A1s%20M%C3%A9sz%C3%A1ros/Session%2037_Matyas%20Meszaros.docx (accessed August 2018).

West, D.B. 2001. *Introduction to Graph Theory*. Prentice Hall.