Preface

1. Emerging Trends and Challenges for Official Statistics

As always, official statistics faces a wide range of challenges, and many of those are associated with trends that have been long in the making. On the one hand, challenges are linked to developments that are mostly under the control of statistical institutes; with ever more complex sample designs, paired with an ambition to compute and present statistics and indicators of increasing complexity, even point estimation is not without its challenges. When precision estimates are to be calculated under such circumstances, practitioners often resort to replication methods.

On the other hand, there are trends that are only partially influenced by statistical institutes, such as the exploration of new data sources. While it could be debated whether administrative data is indeed a new source, or whether it rather represents a renaissance, returning to the presampling era of producing official statistics, increased use of such data is a current trend, driven by various factors, such as cost and response burden reduction, and including (sometimes) legal obligations of a statistical office to use all available data before resorting to (sample) surveys. A possible exception to these steadily advancing trends is Big Data, which constitute a rapidly emerging type of source, of great apparent potential, but as of yet with few actual official statistics applications. The well-known problem of the nonrepresentativity of Big Data, and the difficulty of linking sample units with the observational units of Big Data, still constitute major obstacles to their usefulness for official statistics. However, a promising application, where the impact of this could be seen as less severe, is that of Small Area Estimation (SAE).

Regardless of the source used (sample survey, census, administrative or Big Data), it is well known to official statisticians that merely "providing the figures" without any context may end up doing a disservice to policymakers and, in the worst case, result in inappropriately founded policy decisions. There is a clear trend towards a more reflective approach, with an emphasis not only on producing high-quality statistics, but also on rendering explicit details on exactly how this is being achieved. More precisely, there is a need to understand how high the quality is and where there is room for improvement.

2. New Techniques and Technologies Tackling the New Challenges

The present activities of the official statistics research community reflect (and sometimes shape) current trends and attempts to tackle challenges such as those outlined in Section 1. One of the fora exploring the state of the art regarding official statistics research is New Techniques and Technologies for Statistics (NTTS), which is an international scientific

conference series. The NTTS conferences, which are organised by Eurostat, the statistical office of the European Union, cover new techniques and methods for official statistics and the impact of new technologies on statistical collection, production, and dissemination systems.

The purpose of the NTTS conferences is both to allow the presentation of results from currently ongoing research and innovation projects in official statistics and to stimulate and facilitate the preparation of new innovative projects related to research in statistics within the European Framework Programmes for Research and Development.

From 1992-2001, the first four NTTS conferences were organised triennially. After an eight-year gap, the series resumed in 2009, and has been organised biennially in Brussels ever since. With NTTS being one of the major European scientific conferences with a specific focus on official statistics, the Scientific Committee of NTTS 2013 and the JOS Editorial Board agreed to explore the potential for a special issue of JOS based solely on papers from NTTS 2013 (www.NTTS2013.eu). Following a call for papers, a large number of contributions were received, screened and peer reviewed, and the present JOS special issue represents the positive outcome of this undertaking.

While the articles cover a variety of domains, they could be said to represent three major areas of new developments, upon which we have based this issue's structure. Section 3 presents two of the articles which tackle challenges linked to the traditional source of sample surveys; they both address estimation under complex sampling designs, but in very different ways.

The largest group of articles of this special issue deals with various aspects of new data sources. In Section 4, we present four articles (two on administrative data and two on Big Data) which fall into this category.

Finally, this special issue also includes three articles that deal with post-survey topics which we denote as metadata issues. Section 5 is devoted to these articles which include aspects of quality, respondent attitudes, and statistical disclosure control.

No JOS article is complete without some concluding remarks and future outlook, which we present in Section 6.

3.   Estimation

As illustrated by Oguz Alper and Berger in the first article of our special issue, an analytical (in this case regression-based) approach to variance estimation is still possible, even when one is faced with the triple challenge of a complex survey design, a complex function, and longitudinal effects.

Moving on from this in-depth treatment of a specific application, in the second article Zardetto proposes a general estimation system that easily manages complex sample survey estimators and their precision assessment. This illustrates a number of emerging trends: the move to Free and Open Source Software taking place in many statistical offices and supported by the ESS Vision 2020 (http://ec.europa.eu/eurostat/documents/10186/756730/ESS-Vision-2020.pdf/), as well as a systems approach, with general applications being developed for the benefit of the entire statistical system, according to a shared services philosophy.

4.   New Data Sources in Official Statistics

4.1.   *Administrative Data*

Administrative data have many uses, but coincidentally, both of the articles related to administrative data presented here concern imputation. Again, however, just as for the estimation articles of Section 3, this section's articles offer the same type of variety on the application/system scale. On the one hand, Fleishman, Gubman, and Tur-Sinai discuss a very specific application of imputation, where dwelling prices are used as proxies for the socioeconomic level of an area. On the other hand, Schnetzer and coauthors present a set of general measures to evaluate imputation quality, and embed this assessment in a broader quality framework.

4.2.   *Big Data*

Given the major recent developments in the field of Big Data, with for instance the High-Level Group for the Modernisation of Statistical Production and Services (HLG) sponsoring the project *The Role of Big Data in the Modernisation of Statistical Production* (HLG 2015; this project was the subject of a highly attended satellite workshop at NTTS 2015) and the ESS Big Data Action Plan and Roadmap 1.0 (Eurostat 2014), it is interesting to recall that at the time of NTTS 2013, Big Data in official statistics was still in its infancy. Apart from the keynote address by Robert M. Groves, the presentations dealing with Big Data were few and far between. One of the notable exceptions was the paper of Daas, Puts, Beulens, and van den Hurk, which also forms part of this special issue, in which the exploration of both opportunities and challenges associated with the application of Big Data for official statistics is discussed. Refreshingly, the article not only expresses itself in generalities, but also presents concrete applications to Dutch traffic loop data and Dutch social media messages.

Under the Fay-Herriot model (Fay and Herriot 1979), Big Data can be used as covariates at area level, without the need for linking at unit level, a requirement for which Big Data frequently prove insufficient, in spite of their size. As many Big Data sources (GPS, GSM, satellite images, traffic loops) have a geographical dimension, this is a promising application. Moreover, SAE provides an excellent framework for using Big Data (albeit mainly in the role of area-level covariates), as the tools developed for other types of arealevel covariates are applicable to Big Data as well. In the special issue article by Giusti and coauthors, mobility data are combined with survey data to explore the potential of SAE in this regard.

5.   Metadata

The article by Schnetzer and coauthors discussed in Section 4.1 does in fact deal with measuring (imputation) quality rather than proposing methods for the production of official statistics; in other words, there is a certain overlap between this section and the previous one.

5.1. Respondent Attitude

In the seventh article of this special issue, Torres van Grinsven and Snijkers use social media data sources to gauge the sentiment of business survey respondents, with a view to improving the statistical bureau's understanding of this group of respondents. This understanding could then lead to better strategies for increasing business survey response rates.

While the authors use a novel data source, they do not do so for the direct purpose of producing official statistics. Interestingly, the actual number of social media messages relevant to the analysis of Torres van Grinsven and Snijkers is rather small (a few hundred). Nonetheless, this could in a sense be said to be a Big Data application, because unless there had been an enormous database to begin with (three million new entries each day) to scan, the few messages related to business survey respondents could not have been extracted. Thus, the high granularity of Big Data (rendering it possible to extract information even on a very small subpopulation) is what has rendered this particular analysis possible.

5.2. Statistical Disclosure Control

With an increased demand from policymakers and researchers for specialised frequency tables, many statistical agencies are considering the development of web-based software platforms where users can generate tables of interest.

However, this requires a particular type of metadata to be generated (and then immediately used) "on the fly": the disclosure risk has to be assessed and, if needed, a statistical disclosure control method has to be applied prior to the near realtime release of the table whilst preserving the utility of generated tables to the users to the greatest possible extent. The perennial struggle to strike a balance between utility to data users and data subjects' privacy is thus further complicated by the time pressure introduced by dynamic "on-the-fly" table generation. To remedy this, Shlomo, Antal, and Elliot propose, in the eighth article of the special issue, a new disclosure risk measure and a data utility measure that can be defined at the level of the generated output table and calculated "on the fly".

5.3. Comprehensive Metadata

In the final article, Signore, Brancato, and Scanu argue for going beyond the production chain, and integrating all relevant metadata into one system. Most notably, this extension includes business-related metadata concerning planning, execution, and assessment. The business metadata are needed for the planning phase concerns, e.g., strategic goals and high-level decisions, programme planning (including multiannual plans), and plans for training as well as methodological and IT support. In the execution phase, service-level agreements with stakeholders (administrative data suppliers, outsourcing contractors) are at the core of the business-related metadata. Finally, the authors mention the various business-related metadata available in the assessment phase (user satisfaction surveys, statistics on data access, spontaneous feedback, staff satisfaction surveys, etc.).

6. Advancing the State of the Art of Official Statistics Research

There are several indispensable ingredients in a special issue such as this one. First and foremost, there must be a large body of high-quality articles. Therefore, we thank all the authors who have kindly submitted their conference papers for our consideration. In addition, we would like to thank the referees for taking the time to review all manuscripts, sometimes several versions, and providing constructive comments, leading to considerable improvements in relation to the original manuscripts. Well over 70 referees have contributed to this special issue.

With conference papers as the point of departure for this endeavour, there is a clear challenge in transforming these papers into articles to be included into a scientific journal such as JOS. The purpose and audience are sometimes very different, and mere project presentations, which might be perfectly legitimate at a conference, need to undergo considerable reworking if they are to make it into a scientific journal. On the side of the NTTS Scientific Committee, a step in this direction has been taken through the more stringent requirements for NTTS 2015 conference abstracts. The articles that were ultimately included in this special issue of JOS follow the remit of the JOS editorial board for research results about "survey methods, quality, applications, policy issues and other aspects of production of official statistics" and each article contains relevant sections including an overview of the state of the art, the contribution of the article, a critical discussion of alternatives, a summary discussion with conclusions and future areas of research.

Through this special issue, as with the NTTS series of conferences, we hope to contribute to, and disseminate, some of the richness of the state of the art in official statistics research.

Martin Karlberg
Guest Editor

Silvia Biffignandi
Piet J.H. Daas
Anders Holmberg
Beat Hulliger
Pascal Jacques
Risto Lehtonen
Ralf T. Münnich
Natalie Shlomo
Roxane Silberman
Ineke Stoop

Guest Associate Editors

7. References

Eurostat. 2014. "ESS Big Data Action Plan and Roadmap 1.0." Approved by the 22nd Meeting of the European Statistical System Committee. Available at: http://www.cros-portal.eu/content/ess-big-data-action-plan-and-roadmap-10 (accessed May 7, 2015).

Fay, R. and R. Herriot. 1979. "Estimates of Income for Small Places: an Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 74: 269–277.

HLG. 2015. "The Role of Big Data in the Modernisation of Statistical Production." Available at: http://www1.unece.org/stat/platform/display/bigdata/2014+Project (accessed 23 March 2015).