



BLUE-Enterprise and Trade Statistics

BLUE-ETS

www.blue-ets.eu

SP1-Cooperation-Collaborative Project
Small or medium-scale focused research project
FP7-SSH-2009-A
Grant Agreement Number 244767
SSH-CT-2010-244767

Deliverable 8.4

Dissemination level: PU

Evaluation report on case studies

Edited by Miroslav Hudec

CBS: Deirdre Giesen, Piet Daas

INFOSTAT: Miroslav Hudec, Jana Juriová, Miroslav Klůčik

ISTAT: Grazia Di Bella, Daniela Bonardo, Fulvia Cerroni, Lorena Galiè, Valentina Talucci

SCB: Boris Lorenc, Andreas Persson, Stefan Berg, Klas Wibell, Thomas Laitila, Daniel Lennartsson, Richard Nilsson, Anders Wallgren, and Britt Wallgren

SSB: Frode Berglund, Gustav Haraldsen, Øyvin Kleven, and Kari-Anne Lund

UL: Mojca Bavdaž, Irena Bolko

DRAFT

28 February 2013

Deliverable 8.4

Evaluation report on case studies

Summary

The aim of this report is to provide an overall synopsis of realised case studies in fields of reduction of response burden, enhancing motivation, evaluation of various administrative data sources and evaluation of soft computing methods in improving imputation and reduction of non-response. Two main streams of data collection are surveys from respondents and re-use of already collected administrative data.

The first stream copes with issues of non-response, late response and inaccurate data due to the insufficient focusing on the perspective of the respondents to business surveys. Ignoring issues of motivation and response burden can cause a reduced quality of the survey data and can increase the costs for the national statistical institute (NSI) (e.g. demand for sending reminders and data imputation procedures). Nine case studies were conducted focusing on causes and effects of response burden and/or enhancing motivation in official business surveys.

In relation to the re-use of the administrative sources there are two main approaches: direct data production by an administrative source and use of administrative data to support statistical surveys. The collection and maintenance of administrative data sources is mostly beyond the control of the NSIs. It is the administrative data holder who manages these aspects. Therefore, NSIs should examine data quality and usability from these sources. In the part on evaluation of quality of administrative sources as substitute for data collection six case studies were realised.

A high rate of non-response is a significant issue because it causes bad quality of statistical data and increases the need for more sophisticated data imputation procedures, tailored reminders and motivation strategies supported by administrative data. These problems have two main aspects: large amount of data and statisticians' knowledge which cannot be expressed with crisp rules. These issues can be solved by soft computing methods as modern means to improve the data collection system. In this part six case studies were realised.

Although three working packages (three above paragraphs) have been researched independently, final evaluation reveals that researches are more or less intertwined. Our research has paved path for new research topics. Ideas and practical implementations are also discussed in this report.

Overview table of case studies

Case studies on reduction of response burden and enhancement of motivation

[Analysis of Swedish Data Providers Register](#)

[Analyses of Norwegian Data about causes and consequences of response burden](#)

[Analyses of Dutch Data about causes and consequences of response burden](#)

[Slovenian experiments to increase motivation](#)

[Swedish case study to increase motivation](#)

[Swedish case study into perception of use of feedback of statistics to increase motivation](#)

[Dutch case study on reducing response burden by questionnaire design](#)

[Norwegian case study on using web functionality to reduce response burden](#)

Case studies on QRCA for quality of administrative sources

[Evaluation of quality of two Dutch administrative sources](#)

[Measuring quality of the Italian social security data as input for the statistical production process](#)

[Evaluation of quality of three Swedish administrative sources](#)

Case studies on innovative tools and procedures developed for data collection and analysis

[Extracting rules from Intrastat database using fuzzy logic](#)

[Selecting respondents for tailored reminders](#)

[Flexible classification of respondents](#)

[Genetic programming classification](#)

[Genetic programming symbolic regression](#)

[Classification of Intrastat data by neural network approach](#)

Index

INTRODUCTION	6
1 Background	8
2 Evaluation of case studies	10
2.1 Case studies on reduction of response burden and enhancement of motivation (WP2, WP3)10	
2.1.1 Case study 8.1-1 – Analysis of Swedish Data Providers Register	10
2.1.2 Case study 8.1-2 – Analyses of Norwegian Data about causes and consequences of response burden	12
2.1.3 Case study 8.1-3 – Analyses of Dutch Data about causes and consequences of response burden	14
2.1.4 Case studies 8.1-4& 8.1.5 - Slovenian experiments to increase motivation	16
2.1.5 Case study 8.1.6 - Swedish case study to increase motivation	19
2.1.6 Case study 8.1.7 - Swedish case study into perception of use of feedback of statistics to increase motivation	21
2.1.7 Case study 8.1.8 - Dutch case study on reducing response burden by questionnaire design	23
2.1.8 Case study 8.1.9 -Norwegian case study on using web functionality to reduce response burden	25
2.2 Case studies on QRCA for quality of administrative sources (WP4)	27
2.2.1 Case study 8.2-1 – Evaluation of quality of two Dutch administrative sources	27
2.2.2 Case study 8.2-2 - Measuring quality of the Italian social security data as input for the statistical production process	31
2.2.3 Case study 8.2-3 - Evaluation of quality of three Swedish administrative sources	33
2.3 Case studies on innovative tools and procedures developed for data collection and analysis (WP5)	37
2.3.1 Case study 8.3-1 - Extracting rules from Intrastat database using fuzzy logic	37
2.3.2 Case study 8.3-2 - Selecting respondents for tailored reminders	39
2.3.3 Case study 8.3-3 – Flexible classification of respondents	41
2.3.4 Case study 8.3-4 – Genetic programming classification	43
2.3.5 Case study 8.3-5 – Genetic programming symbolic regression	45
2.3.6 Case study 8.3-6 – Classification of Intrastat data by neural network approach	47
2.3.7 Conclusion of soft computing	48
3 Implications for NSIs, Eurostat and ESSnet	50
3.1 Implications in response burden and motivation	50
3.2 Implications in measurement of quality of administrative sources.....	51
3.3 Implications on innovative tools and procedures developed for data collection and analysis	51
3.4 Implications for other projects and new integrations.....	52
4 Future research avenues and opportunities	54
4.1 Opportunities in reduction of response burden and motivation (WP2, WP3)	54
4.2 Opportunities in quality of administrative sources (WP4).....	54

4.3 Opportunities in soft computing (WP5).....	54
4.4 Opportunities related to synergy of our WPs.....	55
5 Conclusions.....	56
References.....	57

INTRODUCTION

The purpose of this introductory section is to provide the reader with a general view on the issues dealt within the BLUE-ETS project, gradually narrowing the view to specific methods developed and tested in small scale case studies related to modernization of data collection system in official statistics.

BLUE-ETS FRAMEWORK

Work Packages 2, 3, 4, 5 and 8 represent five of eleven work packages in the BLUE-ETS project, a collaborative project on official business statistics, funded by the European Community's Seventh Framework Programme under Grant agreement n. 244767. The project started in April 2010 and will end in March 2013 (see www.blue-ets.eu for more information).

The frame in which official statistical institutes (NSIs) collect data is limited by respondents' attitude to timely and accurately provide their data. Increasing reluctance of respondents causes decreasing quality of collected data. This issue could be mitigated by reducing the response burden (Work Package 2 – WP2) and increasing motivation of respondents to timely and accurately provide their own data and use data produced by NSIs (WP3). A high rate of non-response is a significant issue because it causes bad quality of statistical data and increases demand for more sophisticated data imputation procedures, tailored reminders and motivation strategies supported by administrative data (WP5). The second direction is using data that already exists – administrative data. However, the collection and maintenance of administrative data sources is mostly beyond the control of the NSIs. It is the administrative data holder who manages these aspects. The creation of the standardised tool for quality measurement of administrative data is a significant way not to measure quality only but to increase use of administrative data (WP4). Finally, developed methodologies and findings are tested on methodological case studies (WP8 – New vistas and potential of new methodologies). The framework of Work Package 8 has been anticipated by development of methodologies within other work packages (WP2, WP3, WP4 and WP5) and respective small scale case studies.

REPORT'S OBJECTIVE AND STRUCTURE

This report gives an overview of the case studies conducted in the course of WP8, case studies focusing on measuring and reducing respondents' burden; motivation; quality of administrative data and perspectives of soft computing. The wide variety of topics covered by these case studies reflects the fact that the data collection systems of NSIs are systems containing many different aspects. The quality and efficiency of these systems has significant impact on whole society, as the costs are carried by society and policy decision making highly depends on data collected and processed by official statistics' institutes.

In accordance with the Grant Agreement, promising approaches were empirically tested within Work Package 8. Case studies were usually conducted in the environment of the Dutch, Norwegian, Italian, Slovenian, Slovak, and Swedish NSIs participating in WP8.

The remainder of the paper is structured as follows. In chapter 1 we repeat main objectives and goals of the research project as well as areas which were selected to be evaluated on real data on case studies. Chapter 2 emphasizes main results of each case study. Each case study is represented by a structured description followed by a short evaluation. Implications for practical applications and development of fully applicable tools taking into account new trends are discussed in chapter 3. In chapter 4 ideas and opportunities for further research topics are briefly discussed. At the end, in chapter 5, a common conclusion is drawn.

Reports dealing with theoretical and development aspects as well as experiments are on the following website: <http://www.blue-ets.istat.it/index.php?id=7>.

ACKNOWLEDGEMENTS

The research, experiments and evaluation of results would not have been possible without the efforts of many individuals and institutions, whom we would like to acknowledge and thank here. In particular: Ulrika Ansanius, Caisa Bergman, Elisabeth Bolin and Brita Karlsson (SCB), Bjørn Are Holt, Kristin Foldal Haugen, Dag Gravem, Coen Hendriks, Ragnhild MælandLast, Tore Nøtnes, Mathias Revold, Marit Wilhelmsen, Li-Chun Zhang (SSB); Joep Burger, Astrea Camstra, Mattijn Morren, Saskia Ossen, Ger Snijkers, Martijn Tennekes (CBS), Saša Čuljak, Eva Belak, Dominika Lunder, Tomaž Mikelj, Iris Rošker, Rudi Seljak and Barbara Troha Ažbe (SORS), Magdaléna Bobulová, Emília Mladá (Statistical Office of the Slovak Republic), Karen Doherty (Statistics Canada), Erich Peter Klement (Johannes Kepler University, Linz), Ralf Münnich (University of Trier), Petr Novak (Czech statistical office), Dragan Radojević (Institute Mihailo Pupin), Tanja Švigelj (UL), Valentin Todorov (UNIDO), Vanessa Torres van Grinsven (CBS/Utrecht University), Steven Vale (UNECE), Reinhard Viertl (Vienna University of Technology), Mirko Vujošević (University of Belgrade), Dragan Vukmirović (Serbian statistical office).

In particular we would like to thank Mikkel Barslund and Jorgen Mortensen at CEPS for providing feedback and suggestions for improvement of the report.

DISCLAIMER

The views expressed in this paper are those of the authors and do not necessarily reflect those of the organisations they are affiliated with or the European Commission.

1 BACKGROUND

Blue-ETS is targeted to the MEETS Programme which is about re-focusing and adapting business statistics to new and ongoing change, new needs and concerns (including policy), by availing of new technologies to collect, integrate, manage, access and use statistical and administrative data, while reducing the response burden, improving the use of registers and enhancing business respondents and data users' motivation.

BLUE-ETS is a research project focused on modernization of official enterprise and trade statistics and, more specifically, providing high quality and robust statistical information for better policy and socio-economic research, and to support the Europe 2020 Strategy. Data collection is an indispensable part for achieving this goal.

Given the current huge pressures on productivity and rational use of resources in enterprises, they increasingly complain about the response burden and often refuse to participate in surveys (Bavdaz and Giesen, 2009). Even when they do participate, their attention to the survey task might not be sufficient, which leads to a more covert problem of poor quality of reported data. Experience and research have shown that this problem may be far from negligible (Bavdaz, 2010). When businesses refuse to cooperate, this may result not only in unit non-response (businesses refuse to reply on survey request), but also in item non-response (respondents do not fill all required fields in forms) and inaccuracy. The response burden in business surveys is only a very small part of the total administrative burden: 0.5% (EU Project on Baseline Measurement, 2009). However, response burden for (usually mandatory) business surveys is an issue for the business world. First of all, it has to be acknowledged that the burden caused by statistics is unevenly spread over businesses. Secondly, even a small time burden may cause a large irritation, due to the feeling that governmental organisations ask for data that respondents have already sent to the same or other governmental organisations or institutions or the feeling that the statistics are not useful. Therefore reduction of response burden and motivation of respondents were main tasks of Work Package 2 and 3 and their case studies.

The other direction is data collection from administrative and secondary or alternative sources (Balbi et al, 2013). The main issue is that NSIs are not involved into data collection of administrative and other sources so they are not able to make significant influence on data quality and especially the fluctuation of quality (Daas et al, 2011). Creation of measurement methods for quality indicators, their improvements and applying are significant tasks in this area. The development of a standardized approach that is able to quickly provide essential information on the data quality of administrative and other secondary sources was studied. The development of an approach that can be used to routinely determine the input quality of administrative data - for statistical purposes - has been the main focus of Workpackage 4. Such an approach enables the routine evaluation of data collected not under the control of the statistical organization. Since more and more statistical offices use administrative sources for official statistics, i.e. to reduce the response burden, there is an increasing need for this.

Databases of collected data contain missing values due to item or unit non-response. New approaches in official statistics like neural networks (Juriová, 2011), genetic programming (Klůčik, 2011) and fuzzy logic (Hudec, 2012) could cope with these issues more efficiently than traditional ones. Working package 5 and its case studies were focused on these problems. Additional data like data from interviews which reveals businesses image of NSIs and administrative data containing information like established reporting duty of respondents and values of realised trade are very helpful for estimation of missing values, reminders and motivation. Statisticians pose knowledge how to manage these data and obtain relevant information. Traditional tools are not able to catch statisticians' knowledge (usually expressed in linguistic terms, rules and quantifiers instead of crisp rules and numbers) and directly process them and reveal "information with meaning" in these large

data sets containing qualitative, quantitative and textual data. Applying soft computing methods could bring significant advantages.

The goals of project are not only theoretical research but also new methodological solutions and discuss them from practical perspective. This is reason why case studies are included into project and are under summarised evaluation in this report.

2 EVALUATION OF CASE STUDIES

This section encloses all case studies realized in the frame of WP8. Each case study is represented by a respective structured description. The structured description brings to the reader all relevant information on case study realisation such as: objectives, related field of statistics, data used in experiments and methodology. Structured descriptions are followed by a short evaluation (obtained results, strengths and weaknesses). Detailed descriptions of case studies are given in respective reports of WP8 (8.1 for response burden and motivation, 8.2. for measuring quality of administrative data and 8.3 for soft computing in data collection) and their theoretical research in working packages WP 2, WP 3, WP4 and WP5 respectively.

2.1 Case studies on reduction of response burden and enhancement of motivation (WP2, WP3)

The research and case studies were focused on two areas: 1) explore causes and consequences of response burden in official business surveys and 2) empirically test methods to reduce response burden and motivate businesses for better reporting in these surveys. Causes and consequences of response burden can be revealed by evaluation of relation between actual and perceived burden and response behaviour. In the second area case studies describe concrete actions aiming to influence respondents, namely to reduce respondents' actual and perceived burden and to increase motivation for survey participation and accurate and timely reporting.

2.1.1 Case study 8.1-1 – Analysis of Swedish Data Providers Register

Case study identification (title)	<i>Analysis of Swedish Data Providers Register</i>
Preceding work package	<i>WP2</i>
Institution proposing the case study	<i>SCB</i>
Case study leader	<i>Boris Lorenc, Boris.Lorenc@scb.se</i>
Other staff	<i>Linda Abrahamsson</i>
Case study objective	<i>To analyse patterns of response burden and related response behaviour.</i>
Data	<i>Data from the Register of Data Providers, established by Statistics Sweden. The register is, since 2009, populated with data on all business data collections carried out by Statistics Sweden on appropriation or on commission.</i>
Design	<i>Quantitative analyses (using already existing data)</i>
Related fields of official statistics	<i>General (all kinds of surveys).</i>
Related methodological topics	<i>Patterns of participation and nonparticipation in business surveys; estimating response burden on the business level across all the surveys of an NSI. Association of overall administrative burden and participation in surveys.</i>

Case study description	<i>Goals:</i> - identify patterns of non response and factors associated with it, on the survey level and the level of businesses, - based on the results, indicate points in an NSI's procedure where there might be a potential for improvement.
Description of methodology	- descriptive statistics, - multilevel analysis of a binary response, - other applicable methods.
Overall timing	August 2011 – June 2012
Steps with timeline	Step 1, extraction of data and a preparatory analysis. Step 2, single-level and multi-level analysis: November 2011 – April 2012. Step 3, Reporting: May 2012– June 2012.
Case study indicators	
Hardware, software	<i>R, especially the lme4 package</i>
Expected outcomes beyond final report	<i>Presentations, papers, feedback & recommendations for SCB and other beneficiaries.</i>

Regarding understanding of burden and response behaviour in business surveys the analysis carried out indicates that 1) too high levels of response burden of a business lead to business's reduced participation in new survey instances and possibly to other deteriorations in data quality and 2) businesses' response behaviour between the survey instances is not independent: in particular, businesses that are non-respondents in the first instance of a survey do have a large chance of remaining non-respondents and thus a particular attention needs to be paid to them.

Regarding design and implementation of surveys this case study showed that 1) there were some businesses in the data set that had a considerable amount of time set for participating in business surveys, on the order of person-weeks and person-months - this even if Statistics Sweden generally applies coordinated sampling for business surveys (thought, for business surveys, often positively coordinated sampling); as it was not the goal of this study to investigate which businesses these were and what specifically contributed to such high levels, this should be a task to be performed in the course of improving sampling design for business surveys: specifically, positively coordinated sampling ought to be evaluated against the high response burden that it creates and, on the basis of that, possibly implement some changes in the designs.; 2) pay, as per a preceding item, particular attention in designing contact strategies to businesses novel in a survey: this might include increased survey efforts for obtaining a response from a business new to a recurring survey as well as providing additional support to these businesses (explaining the data that are sought, providing help in using the data collection tool, and so on).

Regarding data collected into a register of data provision (RDP or one by another NSI built in the way similar to that the RDP is built) conclusions from this case study are 1) consider expanding with a measure of data collection effort, in order to gauge this factor's effect on response behaviour; 2) consider systematically collecting data providers' estimates of the time needed to participate in the survey instance, in order to gain an understanding of how this varies between the businesses and understand better the time's impact on response behaviour 3) consider including the variables on the mode by which the data were collected and whether the business was among the key data providers (contacts with such businesses are managed by a special unit of the data collection department); 4) consider including a more detailed outcome code (in particular for non-response).

2.1.2 Case study 8.1-2 – Analyses of Norwegian Data about causes and consequences of response burden

Case study identification (title)	<i>Causes and consequences of actual and perceived response burden based on Norwegian data</i>
Preceding work package	<i>WP2 and WP3</i>
Institution proposing the case study	<i>Statistics Norway</i>
Case study leader	<i>Gustav Haraldsen, gustav.haraldsen@ssb.no</i>
Other staff	<i>Øyvind Kleven, Frode Berglund</i>
Case study objective	<i>To analyse causes and consequences of perceived response burden.</i>
Data	<ul style="list-style-type: none"> • <i>Web survey on actual and perceived burden of respondents of the annual Structural Business Survey (SBS) respondents in 2010</i>
Design	<i>Quantitative analyses (combining already existing data)</i>
Related fields of official statistics	<i>SBS</i>
Related methodological topics	<i>Data collection</i>
Case study description	<i>The SBS questionnaire 2010 included several questions about actual and perceived response burden. This study uses these data and combines them with data from the data ordinary questions from the survey.</i>
Description of methodology	<i>Logistic regression analyses with as dependent variables dummies for perceived time burden, perceived cognitive burden and responding in time.</i>
Overall timing	<i>January 2012- February 2013</i>
Steps with timeline	<ol style="list-style-type: none"> <i>1. Inventory of available data</i> <i>2. Development of indicators</i> <i>3. Analyses</i> <i>4. Writing up first results for ICES paper (June 2012)</i> <i>5. Writing up final results for report.</i>
Case study indicators	<ul style="list-style-type: none"> - <i>perceived (time and cognitive) burden</i> - <i>type of structural business questionnaire</i> - <i>respondents' socio-demographics (single – or multi business, number of employees, type of industry)</i> - <i>response timing</i>
Hardware, software	<i>SPSS</i>
Expected outcomes beyond final report	<ul style="list-style-type: none"> - <i>presentation and paper for ICES IV (presentation together with BLUE-ETS partner Statistics Netherlands who performed similar analyses on Dutch data).</i> - <i>report for Statistics Norway</i> - <i>submitting article to a journal</i>
Additional comments	

From a survey quality perspective, there exists a theoretical argument that high response burden can lead to lower survey quality. This case study establishes that there is a correlation between actual and perceived response burden. Businesses who report that the questionnaire is burdensome actually use more time to collect the information needed and to fill in the questionnaire than businesses who claim that the questionnaire is not difficult to complete. Further, the results demonstrate that there is a correlation between both perceived and actual response burden on the one hand, and number of corrected values in the questionnaires on the other hand. There is also a correlation between the number of reminders sent to the businesses and response burden. This case study gives a clear indication that high response burden gives poorer data quality, although the effects found are not very strong. In recent literature, the hypothesis that there is a curvilinear relationship between perceived response burden and data quality has been advocated. The hypothesis receives little support from the results of this case study.

2.1.3 Case study 8.1-3 – Analyses of Dutch Data about causes and consequences of response burden

Case study identification (title)	<i>Causes and consequences of actual and perceived response burden based on Dutch data</i>
Preceding work package	<i>WP 2 & WP3</i>
Institution proposing the case study	<i>CBS</i>
Case study leader	<i>Deirdre Giesen, d.giesen@cbs.nl</i>
Other staff	<i>-</i>
Case study objective	<i>To analyse causes and consequences of perceived response burden.</i>
Data	<ul style="list-style-type: none"> • <i>Telephone survey on actual and perceived burden of respondents of the annual Structural Business Survey (SBS) respondents in 2006</i> • <i>Process data regarding the data collection of the SBS including the number of items of the questionnaire received by the respondents and the timing of the response</i>
Design	<i>Quantitative analyses (combining already existing data)</i>
Related fields of official statistics	<i>SBS</i>
Related methodological topics	<i>Data collection</i>
Case study description	<i>In the customer satisfaction survey for respondents to the SBS several questions are asked about the actual and perceived response burden. This study uses these data and combines them with data from the data collection process of the SBS to explore causes and consequences of perceived burden.</i>
Description of methodology	<i>Logistic regression analyses with as dependent variables dummies for perceived time burden, perceived cognitive burden and responding in time.</i>
Overall timing	<i>January 2012-October 2012</i>
Steps with timeline	<ol style="list-style-type: none"> <i>1. Inventory of available data</i> <i>2. Development of indicators</i> <i>3. Analyses</i> <i>4. Writing up first results for ICES paper (June 2012)</i> <i>5. Writing up final results for report.</i>
Case study indicators	<ul style="list-style-type: none"> <i>- perceived (time and cognitive) burden</i> <i>- actual burden (number of items on questionnaire and time taken)</i> <i>- overall evaluation of CBS (grade 1-10)</i> <i>- perceived usefulness of statistics</i> <i>- respondents' socio-demographics (size</i>

	<i>class, type of industry)</i> - <i>response timing</i>
Hardware, software	<i>SPSS</i>
Expected outcomes beyond final report	- <i>presentation and paper for ICES IV (presentation together with BLUE-ETS partner SSB who performed similar analyses on Norwegian data).</i> - <i>report for Statistics Netherlands</i>
Additional comments	

This case study sheds some light on factors affecting perceived response burden and the relationship between perceived response burden and response behaviour. The results show that both perceived cognitive burden and perceived time burden are related to actual burden, the attitude towards the statistical agency and the perceived usefulness of statistics for the own business. Also, the results show that perceived burden is related to the timeliness of the response.

These findings corroborate three important assumptions: 1) perceived burden is not only affected by actual burden, but also by attitudes (towards the statistical agency and statistics); 2) perceived burden affects response behaviour and through that 3) the costs of data collection costs (costs for sending out reminders). Also, based on these findings one can speculate that perceived response burden may affect the quality of resulting data, as late responses may affect the timeliness of the output.

These findings confirm that reducing perceived response burden should be a priority for NSIs. This study also offers some indications on how NSIs can reduce perceived burden: both by reducing actual burden and by influencing how respondents feel about the statistical agency and the usefulness of statistics for their own business. These findings suggest that - even for surveys where actual response burden cannot be reduced further - we may reduce perceived burden by improving the image of the statistical agency and the perceived usefulness of the resulting statistics for the businesses.

2.1.4 Case studies 8.1-4& 8.1.5 - Slovenian experiments to increase motivation

Case study identification (title)	<i>Two Slovenian experiments to increase motivation</i>
Preceding work package	<i>WP2 – NSIs’ practices concerning business burden and motivation, and WP3 – Business perspectives related to NSIs’ statistics</i>
Institution proposing the case study	<i>UL and SORS</i>
Case study leader	<i>Mojca Bavdaž, UL, mojca.bavdaz@ef.uni-lj.si</i>
Other staff	<i>Saša Čuljak, Eva Belak, Dominika Lunder, Tomaž Mikelj, Iris Rošker, Rudi Seljak and Barbara Troha Ažbe (at SORS); Irena Bolko and Tanja Švigelj (at UL)</i>
Case study objective	<i>Evaluate effectiveness of different kinds of NSI measures to improve motivation of business for reporting, and accurate and timely response.</i>
Data	<i>Primary data collection from businesses (post-hoc evaluation of NSI efforts); process data (response, timing, contacts).</i>
Design	<i>Experimental design</i>
Related fields of official statistics	<i>Industrial statistics, short-term statistics</i>
Related methodological topics	<i>Data collection, data editing, use of data by businesses</i>
Case study description	<p><i>NSI experimental treatments designed to address two sources of motivation:</i></p> <ul style="list-style-type: none"> • <i>Positive emotions (supposedly leading or contributing to an improved relationship) were expected to be triggered by:</i> <ul style="list-style-type: none"> ○ <i>A more personal and inviting text at the beginning of the questionnaire, and personalised help contacts and facilitated access to help support offered (with full name of the SORS staff, phone number and email contact given).</i> ○ <i>A telephone contact (SORS staff called the contact person named in the preceding questionnaire if available. SORS collects contact data in a standardised way, in the last part of the questionnaire, namely the respondent’s name, telephone number and e-mail address).</i> • <i>Worth arising from the survey task was expected to be recognised by:</i> <ul style="list-style-type: none"> ○ <i>Inclusion of short information containing some statistics suggesting value for businesses on the front page.</i>
Description of methodology	<p><i>All units were assigned to one of the four groups (of unequal size), each one receiving a different treatment:</i></p> <ul style="list-style-type: none"> • <i>Group 1: Max dose (treatment of both groups 2 & 3)</i>

	<ul style="list-style-type: none"> • <i>Group 2: Value for business</i> • <i>Group 3: Personal ties</i> • <i>Group 4: Control</i> <p><i>Treatments were applied one-off in an annual survey and four times in a monthly survey.</i></p>
Overall timing	<i>December 2011- December 2012</i>
Steps with timeline	<ol style="list-style-type: none"> 1. <i>Changes to two questionnaires</i> 2. <i>Development of protocols</i> 3. <i>Telephone contacts for relationship improvement</i> 4. <i>Collection and preparation of all relevant data</i> 5. <i>Post-hoc evaluation by phone</i> 6. <i>Data analysis</i>
Case study indicators	<p><i>Response rates</i></p> <p><i>Respondents' opinions about NSI efforts, burden and motivation</i></p>
Hardware, software	<i>SPSS, R</i>
Expected outcomes beyond final report	<i>Journal paper, presentation at an international conference.</i>
Additional comments	-

For these case studies two experiments were conducted addressing two sources of motivation aiming to provide and establish (more) personal contacts between the NSI and businesses, as well as improve perceived usefulness of NSI statistics for businesses by applying several treatments on two surveys, the Annual Industry Survey and the Monthly Survey on Turnover in Service Activities. The goal was to evaluate effectiveness of treatments in terms of improving motivation for response and accurate and timely data reporting. From initially included 400 unites in each survey, 281 questionnaires in the Annual and 358 questionnaires in the Monthly survey were received, and the number of successfully performed post-hoc evaluation calls was further reduced to 185 in the Annual survey and 102 in the Monthly survey, due to several reasons (i.e. declinations, unavailability, outsourcing of data reporting).

Results of evaluation phone calls show that in both surveys approximately a quarter of respondents that received the questionnaire with included statistics, reported to notice it. However, only a few respondents were also able to at least vaguely recall some content, but none of them found it useful for themselves or their work. On the other hand, when asked about hypothetical usefulness of this kind of statistics in general, more respondents were positive about the idea, especially if more concrete or tailored information was included. We can also notice that telephone conversation with the NSI was recalled by around a third of respondents in the Annual Survey and only some respondents from the Monthly survey. This difference could be explained by the difference in the lag between the NSI's and evaluation phone call in both surveys. Although help from the NSI was rarely requested, the difference between both surveys was also noted in respondents' attitude towards the NSI's offer for assistance – respondents in the Annual survey were considerably more in favour of it compared to the respondents in the Monthly survey who claimed that the questionnaire was short and unproblematic.

In the evaluation phone call we were also interested in respondents' opinion regarding burden and motivation. The questionnaire was based on seven factors (Image, Burden, Interest, Competences/Skills, Effort/Importance, and Relatedness) that were compared among the four experimental groups for both surveys, but results did not show any significant effect of treatments

on respondents' opinion. However, when analysing response rates across experimental groups a positive effect was found on response rates in the Annual survey, with control group having the lowest response rate, although it seems that both treatments did not add up. When timeliness of questionnaire submission was taken into consideration, results suggested that units with late response experienced more burden compared to units in time; and units that submitted with a slight delay reported to have invested more time into filling in the questionnaire compared to the units that submitted in time.

Although no significant effects were found for the treatments on respondents' opinion regarding burden and motivation, and the treatments were prevalently not recalled, it was possible to affect response rates in the more complex annual survey and get some interesting insights into data reporting activities from respondents' point of view, where importance of appropriate and efficient communication strategy was particularly stressed and business' need for more tailored statistics was noted. Future research should place attention both to designing and delivering the treatments.

The implementation of this research was hindered by some practical constraints and deviations from the planned protocol in addition to expected challenges of achieving co-operation from respondents. There was a considerable drop out of units during the course of both experiments resulting in a smaller number of conducted evaluation calls than expected. It was not always possible to ensure that follow-up evaluation calls would be conducted immediately after receiving the questionnaire suggesting that the results could (also) reflect more general attitudes and experiences. Due to lack of "new" respondents in the sample it was not possible to compare their performance and opinion with "old" respondents. It is thus for future research to investigate whether "old" and "new" respondents need a fundamentally different approach (e.g. because of differences with regard to routinization, perception of usefulness of statistics, requesting help). The evaluation phone call questionnaire was based on factors extracted from literature and adapted to the context of business surveys, but could not be empirically verified beforehand. However, these case studies can be regarded as some initial research steps towards establishing an efficient business surveys motivational inventory.

2.1.5 Case study 8.1.6 - Swedish case study to increase motivation

Case study identification (title)	<i>Swedish case study to increase motivation with feedback</i>
Preceding work package	<i>WP2/WP3</i>
Institution proposing the case study	<i>SCB</i>
Case study leader	<i>Boris Lorenc, Boris.Lorenc@scb.se</i>
Other staff	<i>Andreas Persson, Stefan Berg</i>
Case study objective	<i>To explore effects of making salient in a reminder letter different motives for participation</i>
Data	<i>Primary data collected by carrying out an experiment on top of the Swedish Structural Business Survey (SBS).</i>
Design	<i>Experiment</i>
Related fields of official statistics	<i>Structural business statistics.</i>
Related methodological topics	<i>Data collection – ways of soliciting an organisational response.</i>
Case study description	<p><i>Goals:</i></p> <ul style="list-style-type: none"> <i>- compare making salient benefit for business with benefit for society and with none of these salient.</i>
Description of methodology	<p><i>All the businesses that have not responded by a certain date received a certain reminder letter. The standard procedure is that the reminder introduces a promise of a feedback in the form of industry indices that enable comparison of the business in question with its industry and size class. This is assumed to make salient the benefit for the business. An alternative is to make salient the benefit for the society, by e.g. mentioning for what purposes the data collected are used (national accounts, etc).</i></p> <p><i>The businesses were allocated randomly (simple random sample within each SBS stratum) to one of the four experimental groups: 1) Useful for the business, standard; 2) Useful for the business, enhanced; 3) Useful for society; 4) Neither. This resulted in sizes of the four groups of 1,541, 1,541, 1,540, and 1,541 respectively.</i></p>
Overall timing	<i>November 2011 – February 2012</i>
Steps with timeline	<p><i>Step 1, Detailed study design (sampling, design of the reminders, planning the send-out): November 2011.</i></p> <p><i>Step 2, Data collection: November 2011.</i></p> <p><i>Step 3, Data analysis, reporting: December 2011– February 2012.</i></p>
Case study indicators	

Hardware, software	
Expected outcomes beyond final report	<i>Presentations, papers, feedback & recommendations for SCB and other statistical producers.</i>
Additional comments	<i>The case study is a continuation of the work reported in Hedlin, D, Lindkvist, H., Bäckström, H. & Erikson, J. (2008). An Experiment on Perceived Survey Response Burden Among Businesses. Journal of Official Statistics, 24, 301–318.</i>

In general, of the sampled businesses that have not provided data by the time that the experiment started, about 20 percent responded during the course of the experiment. There was not much variation in the response rates for the four treatments, which ranged from 0.23 to 0.27. The standard reminder looked as if perhaps performing somewhat better, however as the t-test resulted in a p-value clearly above the customary p-value of 0.05, it ought to be seen as not implying any difference. Thus, the reminder alternatives did not have a differential effect on the response rate in this experiment. Perhaps the main message from this experiment is that the reminder letter (in any of its variants) has likely been read less than completely by the recipients. Exactly what has been read or not is not possible to say based on this experiment. Thus, text additions as supplements to standardized reminders might not have as strong an effect on response rate as might have been expected or hoped.

The findings of this case study indicate that statistics producers should not presume that their reminder letters and other communication are read or understood fully. This not being a new result in itself, but still a reminder of the importance of need for more work on making content deemed important also salient in the communication, and also conducting tests to verify that this has been achieved. A corollary for the survey within which the experiment has been conducted might be to consider other, more perceivable ways of presenting the “Key Ratios” offer as a way of leveraging their usefulness for businesses, possibly in conjunction with other ways to enhance participation, aimed at those businesses – for instance the small ones – for whom “Key Ratios” are not a sufficient motivator.

2.1.6 Case study 8.1.7 - Swedish case study into perception of use of feedback of statistics to increase motivation

Case study identification (title)	<i>Swedish case study into perception of use of feedback of statistics to increase motivation</i>
Preceding work package	<i>WP2/WP3</i>
Institution proposing the case study	<i>SCB</i>
Case study leader	<i>Boris Lorenc, Boris.Lorenc@scb.se</i>
Other staff	<i>Andreas Persson, Stefan Berg, Klas Wibell</i>
Case study objective	<i>To evaluate a specific form of feedback</i>
Data	<i>Primary data collected by carrying out a study on top of the Swedish Structural Business Survey (SBS).</i>
Design	<i>Quantitative & qualitative analysis.</i>
Related fields of official statistics	<i>Structural business statistics.</i>
Related methodological topics	<i>Data collection – feedback as a motivational factor.</i>
Case study description	<p><i>Goals:</i></p> <ul style="list-style-type: none"> <i>- find out and evaluate how the feedback (consisting of branch indexes put in relation to businesses' own values on the same variables) is received and used by businesses,</i> <i>- indicate points in an NSI's procedure where there might be a potential for improvement.</i>
Description of methodology	<ul style="list-style-type: none"> <i>- stratified random sampling (SI within each SRS stratum)</i> <i>- a mixed-mode data collection, with email contacts in the first phase, and telephone interviews in the follow-up (which will be among non respondents + respondents with interesting responses)</i> <i>- sample size of 34 units, which resulted in 29 conducted interviews</i>
Overall timing	<i>December 2011 – March 2012</i>
Steps with timeline	<p><i>Step 1, Detailed study design (sampling, instrument design, interviewer recruiting and training): December 2011.</i></p> <p><i>Step 2, Data collection: January 2012.</i></p> <p><i>Step 3, Data analysis, reporting: February 2012– March 2012.</i></p>
Case study indicators	
Hardware, software	
Expected outcomes beyond final report	<i>Presentations, papers, feedback & recommendations for SCB and other beneficiaries.</i>
Additional comments	

Overall, effectiveness of making the feedback available to businesses is fairly low with the current setup: about 10% of the interviewed businesses did download the “Key ratios” report.

Following aspects of design and execution of providing the feedback might be re-evaluated:

- Timing of the feedback letter (collides with annual accounting work).
- The log-in step - if confidentiality is a concern - could be replaced by a registered letter containing the “Key ratios”.
 - enable process measure of whether “Key ratios” are downloaded;
 - Experiment with both forms (web and registered letter) to understand whether the increase in the number of downloads counterbalances the increased cost that the registered letter implies.
- Improve visibility and attractiveness of “Key ratios” and make the communication purpose clear; improve usability of the “Key ratios”.
- Attempt tailoring better the group that the business is compared with in the “Key ratios”.

2.1.7 Case study 8.1.8 - Dutch case study on reducing response burden by questionnaire design

Case study identification (title)	<i>Effects of questionnaire redesign on response burden</i>
Preceding work package	<i>WP2 & WP3</i>
Institution proposing the case study	<i>CBS</i>
Case study leader	<i>Deirdre Giesen, d.giesen@cbs.nl</i>
Other staff	
Case study objective	<i>To evaluate effect of redesign SBS questionnaire on response burden and response behaviour</i>
Data	<ul style="list-style-type: none"> • <i>Telephone survey on actual and perceived burden of respondents of the annual Structural Business Survey (SBS) respondents in 2006 (old design) and 2007 (new design (N=about 1500 for both 2006 and 2007)).</i> • <i>Process data regarding the data collection of the SBS such as unit response and timing of the response</i> • <i>Background data we have background information on size class, type of industry of businesses</i>
Design	<i>Quantitative analyses, combining already existing data.</i>
Related fields of official statistics	<i>SBS</i>
Related methodological topics	<i>Data collection, questionnaire design</i>
Case study description	<i>In 2007 a thorough revision of the Dutch SBS questionnaire was implemented. The revision aimed at making the questionnaire less burdensome for respondents. The revision included a reduction of the number of items on the questionnaire, a revision of text and layout to make the questionnaire more user-friendly and the introduction of an electronic version of the questionnaire. Data are available on the actual and perceived response burden from the old and the new design, but these so far have hardly been analysed. For this case study multivariate analyses per performed to test for significant effects on response behaviour and the actual and perceived burden.</i>
Description of methodology	<i>Evaluation study, with comparison of data before and after the introduction of the new questionnaire design.</i>
Overall timing	<i>July 2012-Dec 2012</i>
Steps with timeline	<ol style="list-style-type: none"> <i>1. Inventory of available data</i> <i>2. Development of indicators</i> <i>3. Analyses</i> <i>4. Writing of report</i>
Case study indicators	<ul style="list-style-type: none"> <i>- perceived burden</i> <i>- actual burden</i> <i>- overall evaluation of CBS (grade 1-10)</i> <i>- usefulness of statistics</i>

	<ul style="list-style-type: none"> - <i>Size class, type of industry</i> - <i>Response timing</i>
Hardware, software	<i>SPSS</i>
Expected outcomes beyond final report	<ul style="list-style-type: none"> - <i>Recommendations for questionnaire design</i> - <i>Journal paper(s)</i>
Additional comments	<p><i>It was also planned to conduct analyses of the effect of the redesign on the quality of the raw data. However, it proved to be too difficult to get access to these data within the time for this project.</i></p>

This case study shows that a redesign of a questionnaire can substantially reduce actual burden (from an average time spent of 4 hours to 3) and significantly reduce perceived burden. The findings show that is not just the time taken to complete a questionnaire that affects the burden caused by a given questionnaire design. This indicates that the new design makes responding not only quicker but also more pleasant (or less frustrating).

In the new design questionnaires were sent out later (about 6 weeks) and respondents were given less time to respond (about 60 days versus 77 days). Interestingly, this did not increase the percentage of late respondents. This is an important finding, as timely responses are important for meeting production deadlines and chasing respondents takes substantial resources. Probably several factors have affected the timing of the responses with the new design. First of all, less perceived burden seems to increase timely responses and the new design is perceived as less burdensome. Secondly, one can speculate what the effect is of the timing of sending out the questionnaires. It is known from help desk staff at CBS that businesses often complain that they receive the SBS questionnaires too early in the year, as their final financial statements are not ready yet. Maybe by sending out the questionnaires about 6 week later, more respondents were ready to provide data when they received the questionnaire. But maybe this finding also indicates that there is no need to give respondents 77 days to respond and they can do it in 60 days as well.

Overall, the results from these analyses indicate that questionnaire redesign can affect actual burden, perceived burden and through that response behaviour. Unfortunately, the data do not allow any estimates of how specific characteristics of the redesign (for example the electronic mode or the new wording) affect burden and response behaviour. For this type of knowledge experiments are necessary. Also, these analyse do not take into account the quality of the data. It would be interesting to see if and how the new design has affected this.

2.1.8 Case study 8.1.9 -Norwegian case study on using web functionality to reduce response burden

Case study identification (title)	<i>The usability of web functionality</i>
Preceding work package	<i>WP2 & WP3 – Business perspectives related to NSIs` statistics</i>
Institution proposing the case study	<i>Statistics Norway (SN)</i>
Case study leader	<i>Kari-Anne Lund, SN, Kari-Anne.Lund@ssb.no</i>
Other staff	<i>Frode Berglund, Øyvind Kleven, Gustav Haraldsen</i>
Case study objective	<i>Document if and how tailored instruction facilities in web business questionnaires affects response burden and motivation.</i>
Data	<i>Qualitative and quantitative data from experiments conducted in Jan/Feb 2013.</i>
Design	<i>Lab experiment (N=50) combining structured cognitive interviews with eye tracking data.</i>
Related fields of official statistics	<i>Business statistics, labour statistics, health statistics, sick-leave statistics,</i>
Related methodological topics	<i>Mixed methods, technology</i>
Case study description	<i>Experiment designed to give insight on attempts to reduce response burden and increase motivation in business surveys. Measure effects of introducing a tailored electronic calculator and an audio visual instruction.</i>
Description of methodology	<i>The mandatory survey of self-reported sick leave was chosen as a test case. The survey consists of 4 assumingly straight forward questions in addition to qualification questions. Two versions of the questionnaire were designed, with different instruction facilities offered to the respondents; written instruction initiated by a click on a question mark or a tailored Man-day Calculator and a video instruction.</i>
Overall timing	<i>December 2012- March 2013</i>
Steps with timeline	<ol style="list-style-type: none"> <i>1. Thesis outline</i> <i>2. Design experiment</i> <i>3. Programming of tool</i> <i>4. Conduct lab experiments</i> <i>5. Evaluation of results</i>
Case study indicators	<i>Use (1,0), preference (5 point scale), accuracy (verification rate)</i>
Hardware, software	<i>Altinn2 - a web portal common to all mandatory reporting from businesses to official institutions in Norway. Excel for data entry. SPSS for analysis.</i>

Expected outcomes beyond final report	<i>Paper on Quest 2013. Report for Statistics Norway Submitting article to a journal</i>
Additional comments	-

It seems as offering a calculator and using a video clip to explain how it should be used, did not add to the actual response burden and rather had a positive effect on the perceived response burden. Even more important, there are reasons to believe that the response quality increased. The general problem with respondents who don't seek instruction not readily presented to them, seems however to be even more so when the instructions are a video clip compared to a written text.

These results points towards several paths of further research. Taking the improved response quality and the reluctance to look up the audio clip as a starting point, the first question is if it is necessary to look at the video clip in order to master the calculator. That can be tested by taking it away. Other manipulations would be to take away the written instruction on how the calculator works given in the right pane of the screen or taking away both the written instruction and the video clip. If the video clip seems to be necessary in order to master the calculator and improve the response quality, we have to look for ways of making it more visible and attractive.

Another interesting result which could be followed up with more quantitative analysis is the notion that the information processing pattern is different when text and audiovisual features are mixed and when a calculation tool is offered. Eye tracking analysis' focusing on areas of interest (AOI-analysis) could give some more concrete answers to this question.

The motivation behind these experiments is the idea that more dynamic questionnaires will enhance a feeling of communication and collegiality between those surveyed and the surveyor, which may have a long term positive effect on the relationship between the respondents and the NSI. This is a long term effect that observations, like those we have made in our test, cannot answer. The way we see it, however, our tests are still a promising first step in looking into these matters.

2.2 Case studies on QRCA for quality of administrative sources (WP4)

In WP4 an approach was developed that enables a standardized evaluation of the input quality of administrative data from a statistics point of view. Apart from a Quality Report Card for Administrative sources, in which the findings are noted, a set of scripts was created to assure that the methods and framework created could actually be used. All scripts created are included in a data quality package for the R Statistical Computing Environment. In a separate report the results of the application of the above mentioned products on various administrative data sources at the NSI's involved and the R-package are described (Deliverable 8.2).

2.2.1 Case study 8.2-1 – Evaluation of quality of two Dutch administrative sources

Two Dutch administrative sources were studied from the input data quality point of view. Results were reported via the Quality Report Card for Administrative data (QRCA) developed in WP4 (deliverable 4.3). The sources studied are: the Insurance Policy record Administration (IPA) data and Value-added Tax (VAT) data.

Case study identification (title)	<i>WP8. 4.1 (CBS-VAT)</i>
Preceding work package	<i>WP4 Improve the use of administrative sources</i>
Institution proposing the case study	<i>Statistics Netherlands (CBS)</i>
Case study leader	<i>Piet Daas, CBS, pjh.daas@cbs.nl</i>
Other staff	<i>Saskia Ossen (CBS), Martijn Tennekes (CBS) and Joep Burger (CBS)</i>
Case study objective	<i>Evaluation of the measurement methods for the indicators and the Quality Report Card for Administrative data (QRCA) developed in WP4.</i>
Data	<i>Value Added Tax (VAT) data provided by the Dutch Tax office.</i>
Design	<i>Quantitative and qualitative quality evaluation, visual data inspection</i>
Related fields of official statistics	<i>Business statistics, Short Term Statistics, International Trade Statistics</i>
Related methodological topics	<i>Data and metadata quality evaluation</i>
Case study description	<i>VAT data is a secondary source of information that -at a first glance- seems excellently suited to i) replace the information currently obtained by the Short Term Statistics survey and as an ii) additional source of information for statistics on International Trade. However, more detailed studies revealed a considerable number of quality issues, such as the timely availability of the data, the selectivity of the data provided over time, and the authenticity of the data providers (to name a few). These quality issues make this data source an</i>

	<i>excellent case study to test the quality evaluation approach developed in WP4</i>
Description of methodology	<i>The approach developed in WP4 for the quality evaluation of administrative data will be applied (see Deliverable 4.3 for more details)</i>
Overall timing	<i>Case study is realized in 2012. From July till December 2012.</i>
Steps with timeline	<ol style="list-style-type: none"> <i>1. Script creation and testing</i> <i>2. Arranging access to data sources</i> <i>3. Evaluation studies</i> <i>4. Filling in QCRA</i> <i>5. Evaluation of overall findings</i>
Case study indicators	<i>The indicators developed in WP4 for the quality evaluation of administrative data will be applied (see Deliverable 4.2 Annex A for the complete list). These scripts are included in the data quality R-package that is part of Deliverable 4.3.</i>
Hardware, software	<i>Scripts were developed in the open source statistical software program R.</i>
Expected outcomes beyond final report	<i>For the VAT-data: results for the quality indicators and their measurement methods, including visual inspection results (Table plot) and report (see Deliverable 8.2). The results of this case study will be part of a paper and presentation at the ISI 2013.</i>
Additional comments	-

Case study identification (title)	<i>WP8. 4.2 (CBS-IPA)</i>
Preceding work package	<i>WP4 Improve the use of administrative sources</i>
Institution proposing the case study	<i>Statistics Netherlands (CBS)</i>
Case study leader	<i>Piet Daas, CBS, pjh.daas@cbs.nl</i>
Other staff	<i>Saskia Ossen (CBS), Martijn Tennekes (CBS)</i>
Case study objective	<i>Evaluation of the measurement methods for the indicators and the Quality Report Card for Administrative data (QRCA) developed in WP4.</i>
Data	<i>Insurance Policy record Administration (IPA) data collected by the Dutch Tax office and maintained by the Institute for Employee Benefit Schemes of the Netherlands.</i>
Design	<i>Quantitative and qualitative quality evaluation, visual data inspection</i>
Related fields of official statistics	<i>Business statistics, Social statistics, Enterprise Statistics, Regional Business Statistics, Labour statistics</i>

Related methodological topics	<i>Data and metadata quality evaluation</i>
Case study description	<i>IPA-data is a secondary source of information that provides detail information on employers (businesses), employees, and jobs. This makes IPA-data very suited as a general source of information for both business and social statistics. This and the fact that the amount of information provided is huge (each month a total of around 20 million records are received) makes IPA-data a very interesting case study to test the quality evaluation approach developed in WP4</i>
Description of methodology	<i>The approach developed in WP4 for the quality evaluation of administrative data was applied (see Deliverable 4.3 for details)</i>
Overall timing	<i>Case study is realized in 2012. From July till December 2012.</i>
Steps with timeline	<ul style="list-style-type: none"> <i>7. Script creation and testing</i> <i>8. Arranging access to data sources</i> <i>9. Evaluation studies</i> <i>10. Filling in QCRA</i> <i>11. Evaluation of overall findings</i>
Case study indicators	<i>The indicators developed in WP4 for the quality evaluation of administrative data will be applied (see Deliverable 4.2 Annex A for the complete list). These scripts are included in the data quality R-package that is part of Deliverable 4.3.</i>
Hardware, software	<i>Scripts were developed in the open source statistical software program R.</i>
Expected outcomes beyond final report	<i>For the IPA-data: results for the quality indicators and their measurement methods, including visual inspection results (Tableplot), and report (see Deliverable 8.2). The results of this case study will be presented at the NTTS 2013 conference in Brussels. The results of the visual inspection are included in the conference paper.</i>
Additional comments	-

The most important finding was that standardized evaluation of the quality of administrative data sources from a statistical point of view is difficult in practice. Not only is this caused by the two distinct points of view (see Deliverable 4.3) that can be distinguished for input quality. The actual evaluation of sources is also severely affected by the IT-environment of an NSI and hence the software available. This prompted a more flexible approach and focus less on the programming language in which the measurement methods were evaluated. In the Netherlands R was used but not in the other countries.

The QRCA demonstrated its use as a way of structurally noting down findings. The sequence used, however, needed to be changed. As a result -in accordance with the suggestion of ISTAT- the original sequence of the dimensions was altered. In the new version of the ARCA the dimensional

sequence is: Technical checks, Integrability, Accuracy, Completeness and Time-related dimension. The need to add additional findings (such as plots) to the report card also emerged. This was additionally included at the end of the report card. The updated version of the QCRA is included as an Appendix to Deliverable 8.2.

The time required to thoroughly evaluate administrative data is something that is a serious issue for many NSI's. It is difficult to image an NSI that routinely checks each quality aspect of the data for every delivery of an administrative source. Budget restraints would simply prevent this. Apart from automated checking of some aspects, which is -for instance- routinely done for quite some administrative data sources in the Netherlands, the visualization approach developed in WP4 of the BLUE-ETS project provides a reasonable alternative. It is fairly easy to create a tableplot of a data source and plot the 'profiles' of a selected number of variables. Looking an comparing such data 'pictures' for subsequent deliveries does not have to take much time and can be used to quickly identify any important data quality issues; see the paper of Tennekes et al. (2013) for more details.

The quality framework and approaches developed were presented at a special BLUE-ETS session at the European Conference on Quality in Official Statistics 2012 in Athens, Greece. Papers are available for download via the Q2012 website (www.q2012.gr, session 23) and described in Deliverable 4.3 and 8.2.

2.2.2 Case study 8.2-2 - Measuring quality of the Italian social security data as input for the statistical production process

Case study identification (title)	<i>WP8 - 8.2 Methodological case study for testing and evaluating WP4 input data quality indicators: Measuring quality of the Italian social security data as input for the statistical production process</i>
Preceding work package	<i>WP4 – Improve the use of administrative sources</i>
Institution proposing the case study	<i>ISTAT</i>
Case study leader	<i>Grazia Di Bella, Istat, dibella@istat.it</i>
Other staff	<i>Lorena Galiè, Daniela Bonardo, Fulvia Cerroni, Valentina Talucci</i>
Case study objective	<i>Testing data input quality indicators for evaluating and monitoring administrative data quality entering the production process. The main goal is give a contribute in order to define a new comprehensive quality-indicator instrument, a Quality Report Card for Administrative data (QRCA) that can be generally used by NSIs</i>
Data	<i>Administrative data set: Italian Social Security Data (SSD) produced by Inps (Italian Institute of Social Security) concerning monthly contribution declarations of employers for employees (May 2010); as reference list: Italian Business Register (named Asia – Archivio Statistico delle Imprese Attive) produced by Istat</i>
Design	<i>Methods for the quality indicators.</i>
Related fields of official statistics	<i>Business statistics, statistics on jobs.</i>
Related methodological topics	<i>Data quality measurements methods, record linkage, data analysis.</i>
Case study description	<i>Test the quality indicators - already produced in WP4 - on administrative data at disposal of the participating NSIs to verify their robustness.</i>
Description of methodology	<i>The measurement methods of the quality indicators, defined in Deliverable 4.2, are applied to the Italian Social Security Data (SSD). The choice of this archive for the indicators test derives from several elements: (i) it is a complex and big source including more statistical units connected to each other; (ii) its use can be extended to different types of statistical production process; (iii) Istat is particularly interested in it as it will be used to redefine the business register production and to improve its timeliness. To</i>

	<i>test quality indicators on SSD source we started by performing a preliminary analysis of the indicators and related measurement methods in order to evaluate their applicability to the data. Some comments for the quality QRCA implementation are given.</i>
Overall timing	<i>September 2012 - December 2012</i>
Steps with timeline	<ul style="list-style-type: none"> - <i>Administrative source selection</i> - <i>Administrative data set selection</i> - <i>Preliminary analysis of units and variables to which apply indicators and related measurement methods</i> - <i>Building of the Sas programme</i> - <i>Evaluation of results</i> - <i>Report for Deliverable 8.2</i>
Case study indicators	<i>Quality indicators defined in Deliverables 4.2</i>
Hardware, software	<i>SAS statistical package</i>
Expected outcomes beyond final report	<i>Blue Ets Workshop held in Rome, 11/2012</i>
Additional comments	-

The Istat case study concerns the quality evaluation of the administrative data set of the Italian Social Security Data (SSD) produced by Inps (Italian Institute of Social Security) using WP4 input indicators. It includes monthly contribution declarations of employers for employees (May 2010). To compare administrative data, the Italian Business Register (named Asia – Archivio Statistico delle Imprese Attive) produced by Istat, was used as reference list.

For the implementation we used SAS language because of the huge amount of SSD. For smaller sources, the use of the generalized R data quality package is desirable also for its effective graphics functions for synthesizing data. In cases where administrative sources include millions of records (as is the case of the sources refers to enterprises) an implementation plan for QRCA applied to each administrative data delivery as to be defined.

This source has been chosen to test the robustness of the QRCA and its indicators as it is a complex source comprising different types of units connected to each other (Linked Employer Employee Data).

The measurement methods have been successfully applied, so quality indicators reached the objectives. The quality framework defined in WP4 (see Deliverable 4.2) seems to include all quality aspects related to statistical use of administrative data.

It could be useful to highlight that quality dimensions are interrelated with each other and a low quality assessment of a dimension can affect the quality of another dimension.

The output of the Quality Report Card, reporting results of the measurement methods has a good efficiency in synthesizing and communicating the quality of administrative data. Case study full description is reported on Deliverable 8.2, Appendix B.

2.2.3 Case study 8.2-3 - Evaluation of quality of three Swedish administrative sources

In Sweden the quality of three data sources was studied according to the evaluation system described in Laitila et al. (2012). This broad perspective enables thorough evaluation of administrative data (registers) for statistics production. The sources studied are: the Income Statement register, the VAT-register and the Annual Company Reports (SRU) register, respectively.

Case study identification (title)	<i>WP8. 8.2-3 (SCB-IS)</i>
Preceding work package	<i>WP4 Improve the use of administrative sources</i>
Institution proposing the case study	<i>Statistics Sweden (SCB)</i>
Case study leader	<i>Thomas Laitila, SCB, thomas.laitila@oru.se</i>
Other staff	<i>Daniel Lennartsson, Richard Nilsson, Anders Wallgren, and Britt Wallgren (SCB),</i>
Case study objective	<i>Evaluation of the quality measurement procedure developed in WP4.</i>
Data	<i>Income Statement register (IS)</i>
Design	<i>Quality assessment procedure.</i>
Related fields of official statistics	<i>Business statistics, Social statistics</i>
Related methodological topics	<i>Data and metadata quality evaluation</i>
Case study description	<i>Income Statement register contains data on yearly wages and salaries, preliminary tax and benefits regarding each employee. About 60 % of total taxes in Sweden are collected with this system.</i>
Description of methodology	<i>The approach developed in WP4 for the quality evaluation of administrative data was applied (see Deliverable 4.3 Appendix B for details)</i>
Overall timing	<i>Case study is realized during the period July 2012 till January 2013.</i>
Steps with timeline	<ol style="list-style-type: none"> <i>1. Administrative source selection</i> <i>2. Metadata evaluation</i> <i>3. Accuracy evaluation</i> <i>4. Integration with a base register evaluation</i> <i>5. Integration with other surveys evaluation</i> <i>6. Note findings in report card</i> <i>7. Report for Deliverable 8.2</i>
Case study indicators	<i>The quality indicators developed in WP4 for the quality evaluation of administrative data was applied (see Deliverable 4.3 Appendix B for a complete list).</i>
Hardware, software	<i>Scripts were developed in the standard statistical software.</i>
Expected outcomes beyond final report	<i>Overall the system of indicators was found useful for register quality evaluation. By the construction of the indicators, the evaluation</i>

	<i>process did not only shed light on the quality of the registers evaluated, but also on the quality of the data sources used in the evaluation process.</i>
Additional comments	-

Case study identification (title)	<i>WP8. 8.2-3 (SCB-VAT)</i>
Preceding work package	<i>WP4 Improve the use of administrative sources</i>
Institution proposing the case study	<i>Statistics Sweden (SCB)</i>
Case study leader	<i>Thomas Laitila, SCB, thomas.laitila@oru.se</i>
Other staff	<i>Daniel Lennartsson, Richard Nilsson, Anders Wallgren, and Britt Wallgren (SCB),</i>
Case study objective	<i>Evaluation of the quality measurement procedure developed in WP4.</i>
Data	<i>Value Added Tax data (VAT)</i>
Design	<i>Quality assessment procedure.</i>
Related fields of official statistics	<i>Business statistics</i>
Related methodological topics	<i>Data and metadata quality evaluation</i>
Case study description	<i>The VAT-register data obtained from several sources, i) the monthly reported VAT by businesses to the Tax Board, ii) VAT reported to the Tax Board in the income-tax returns of smaller businesses and iii) information on VAT retrieved from the Customs each six months. All data is delivered to Statistics Sweden.</i>
Description of methodology	<i>The approach developed in WP4 for the quality evaluation of administrative data was applied (see Deliverable 4.3 Appendix B for details)</i>
Overall timing	<i>Case study is realized during the period July 2012 till January 2013.</i>
Steps with timeline	<ol style="list-style-type: none"> <i>1. Administrative source selection</i> <i>2. Metadata evaluation</i> <i>3. Accuracy evaluation</i> <i>4. Integration with a base register evaluation</i> <i>5. Integration with other surveys evaluation</i> <i>6. Note findings in report card</i> <i>7. Report for Deliverable 8.2</i>
Case study indicators	<i>The quality indicators developed in WP4 for the quality evaluation of administrative data was applied (see Deliverable 4.3 Appendix B for a complete list).</i>
Hardware, software	<i>Scripts were developed in the standard statistical software.</i>
Expected outcomes beyond final report	<i>Overall the system of indicators was found</i>

	<i>useful for register quality evaluation. By the construction of the indicators, the evaluation process did not only shed light on the quality of the registers evaluated, but also on the quality of the data sources used in the evaluation process.</i>
Additional comments	-

Case study identification (title)	<i>WP8. 8.2-3 (SCB-SRU)</i>
Preceding work package	<i>WP4 Improve the use of administrative sources</i>
Institution proposing the case study	<i>Statistics Sweden (SCB)</i>
Case study leader	<i>Thomas Laitila, SCB, thomas.laitila@oru.se</i>
Other staff	<i>Daniel Lennartsson, Richard Nilsson, Anders Wallgren, and Britt Wallgren (SCB),</i>
Case study objective	<i>Evaluation of the quality measurement procedure developed in WP4.</i>
Data	<i>Annual Company Reports register (SRU)</i>
Design	<i>Quality assessment procedure.</i>
Related fields of official statistics	<i>Structural business statistics</i>
Related methodological topics	<i>Data and metadata quality evaluation</i>
Case study description	<i>The SRU-register contains the yearly tax returns for Sole traders, Trading partnerships, Limited partnerships, Limited companies and Economic associations. They consist of three parts, balance sheet, profit and loss statement and tax adjustments. Limited companies give more information and sole traders give less detailed information. The SRU is the main source for the Structural Business Statistics survey.</i>
Description of methodology	<i>The approach developed in WP4 for the quality evaluation of administrative data was applied (see Deliverable 4.3 Appendix B for details)</i>
Overall timing	<i>Case study is realized during the period July 2012 till January 2013.</i>
Steps with timeline	<ol style="list-style-type: none"> <i>1. Administrative source selection</i> <i>2. Metadata evaluation</i> <i>3. Accuracy evaluation</i> <i>4. Integration with a base register evaluation</i> <i>5. Integration with other surveys evaluation</i> <i>6. Note findings in report card</i> <i>7. Report for Deliverable 8.2</i>
Case study indicators	<i>The quality indicators developed in WP4 for the quality evaluation of administrative data was applied (see Deliverable 4.3 Appendix B</i>

	<i>for a complete list).</i>
Hardware, software	<i>Scripts were developed in the standard statistical software.</i>
Expected outcomes beyond final report	<i>Overall the system of indicators was found useful for register quality evaluation. By the construction of the indicators, the evaluation process did not only shed light on the quality of the registers evaluated, but also on the quality of the data sources used in the evaluation process.</i>
Additional comments	-

In Sweden three data sources were evaluated by the approach described in Laitila et al. (2012). The indicators are grouped in sets which provides a working procedure involving the sequence of evaluating i) metadata, ii) accuracy, iii) integration with a base register, and iv) integration with other surveys. The definitions of the indicators are suggested with the view of implementing a data source (register) within statistical register systems as proposed by Wallgren and Wallgren (2007).

It is natural to first consider the (metadata) contents of an administrative data source and then evaluate the accuracy of the contents. For integration it is of interest to know if the source can be incorporated into the system by relating it to a base register. Performing these three first steps of the evaluation process does not only cast light on the new source evaluated. The experiences from the examples are that the evaluation process also provides insights on the sources themselves and other data sources used in the evaluation process. This is perhaps one of the more important results in the applications reported and an issue for future projects. To maintain a fully functional register system requires a continuous evaluation of the consistency with relevant external information. Here any kind of relevant information is of interest and not only administrative sources potentially useful for statistics production.

The fourth set of indicators addresses the issue of how an administrative data source can be utilized for improving surveys conducted at an NSI. Again the experience of the applications is that this step provides both information on the quality of the source evaluated and the data sources used for the evaluation. It was also found that this evaluation is a methodologically demanding task. This step will also involve subjective judgments since the considered potential usage of the new data sources depends on the experiences of those performing the evaluation.

2.3 Case studies on innovative tools and procedures developed for data collection and analysis (WP5)

Research in WP5 was focused on techniques to make data analysis and data collection more human oriented applying linguistic terms (statisticians possess knowledge how to deal with their tasks, but this knowledge cannot be always expressed by precise and crisp rules and numbers which are more restrictive than linguistic terms and quantifiers); biologically inspired systems which are based on remarkable human capability to perform a wide variety of tasks without precise measurements and computations (revealing dependencies and relations in large data sets) and mining textual data sets (revealing information from already published documents by business respondents). The work has focused on development of mathematical equations, programming methodologies and frameworks in textual data mining and soft computing and testing on real data from official statistics. In this section, case studies of soft computing approaches are presented. Case studies related to mining textual data sets can be found in Deliverable 5.1, Section 3.

2.3.1 Case study 8.3-1 - Extracting rules from Intrastat database using fuzzy logic

Case study identification (title)	<i>WP8 -8.3 Extracting rules from Intrastat database using fuzzy logic.</i>
Preceding work package	<i>WP5 – New Ways of Collecting and Analysing Information</i>
Institution proposing the case study	<i>INFOSTAT</i>
Case study leader	<i>Miroslav Hudec, hudec@infostat.sk</i>
Other staff	<i>-</i>
Case study objective	<i>Extracting information from Intrastat database by flexible rules. Applied for evaluation of imputed values by comparing their distribution with the distribution of surveyed data in the Intrastat database.</i>
Data	<i>Data used: Anonymised Slovak Intrastat data for the year 2009.</i>
Design	<i>Data modelling and quantitative analysis</i>
Related fields of official statistics	<i>Business statistics, administrative data, Intrastat</i>
Related methodological topics	<i>Data mining, Soft computing</i>
Case study description	<i>Test of current algorithms for data imputation by evaluation of similarities between estimated and collected data by fuzzy rules. The approach is able to reveal whether estimated values have more or less similar distribution such as data received from respondents.</i>
Description of methodology	<i>The approach developed in WP5 for the fuzzy rules evaluation has been applied (see Deliverable 8.3, Section 2.2 for more details). Pair of rules (one for surveyed data and one for estimated) described in natural language contains quantifiers (most, about half, few...) and linguistic terms (small, medium, high...).</i>

	<i>The degree of truth of a fuzzy rule is valued in the interval [0, 1] which in a more convenient way reveals similarity among data. The distinction between membership degrees of both rules reveals how efficient is the current algorithm for evaluating of missing values.</i>
Overall timing	<i>September 2011- September 2012</i>
Steps with timeline	<ul style="list-style-type: none"> • <i>Rule structure preparation</i> • <i>Intrastat database examination</i> • <i>Programming of tool</i> • <i>Evaluation of results</i>
Case study indicators	<i>Similarity between rules ([0, 1] interval)</i>
Hardware, software	<i>Relational database (MS Access), MS Visual Studio (code programmed in Visual Basic 2008).</i>
Expected outcomes beyond final report	<i>Invited paper in solicited session of SIS 2012, Rome. Paper in reviewing process for Springer book.</i>
Additional comments	-

The experiment of flexible rule evaluation by the SQL-like query has been applied for comparing whether distributions of estimated missing values have more or less similar distribution as the data received from respondents. Two main advantages are 1) rules are described by linguistic terms and quantifiers which are easily understandable, modifiable and directly applicable in evaluation processes; 2) a degree of truth is valued in the [0, 1] interval what means that entities having similar attributes values are always similarly treated.

The distinction between membership degrees of both rules (one on collected data and one on estimated data) reveals how efficient is the current algorithm for evaluating of missing values. If membership degrees of rules gravitate to each other than the current algorithm does not need improvements. This is significant information about the ability of current algorithms to properly estimate missing values.

Our approach has revealed that algorithm used in SO SR works pretty good for the estimation of countries of dispatch. In the case of estimation of number of items (goods) in each report, rule has revealed a significant difference and there is space for algorithm's improvement.

More sophisticated rules could be created through the support of administrative data. However, rules were not evaluated due to limitations caused by data anonymisation (obtained company ID was anonymised, so to link to the additional data across the company ID was impossible to realise). From technical point of view, the difference would only be in the connection to different data sources.

This approach could be extended to extract knowledge (rules) form other databases for different purposes e.g. support for decision making or dissemination. Offer broad audience not only data but mined information in easily and understandable way could improve image of NSIs. This kind of tool could be applied as a standalone program connected to the particular database.

2.3.2 Case study 8.3-2 - Selecting respondents for tailored reminders

Case study identification (title)	<i>WP8 -8.3 Selecting respondents for tailored reminders using fuzzy queries.</i>
Preceding work package	<i>WP5 – New Ways of Collecting and Analysing Information</i>
Institution proposing the case study	<i>INFOSTAT</i>
Case study leader	<i>Miroslav Hudec, hudec@infostat.sk</i>
Other staff	<i>-</i>
Case study objective	<i>Selection of respondents for different kinds of reminder letters supported by administrative data.</i>
Data	<i>Data used: Anonymised Slovak Intrastat data for the year 2008 provided with anonymised data from the registry on the month of first reporting duty for the same year.</i>
Design	<i>Data modelling and quantitative analysis</i>
Related fields of official statistics	<i>Business statistics, Intrastat, respondents motivation</i>
Related methodological topics	<i>Data mining, Survey system</i>
Case study description	<i>Attributes relevant for selection of relevant respondents are qualitative and quantitative and stored into several data sets. Fuzzy queries are able to integrate them into one condition and therefore select respondents for each group or reminder letters. In addition, query condition is described in terms of natural language.</i>
Description of methodology	<i>The approach developed in WP5 for the fuzzy querying has been applied (see Deliverable 8.3, Section 2.3 for more details). Fuzzy logic methodology for construction of membership functions able to cover both quantitative and qualitative data, and integrating them into one query condition by appropriate aggregation functions. If one attribute is more relevant than other preferences are covered by fuzzy implication. In this way all respondents with similar behaviour are selected into one class.</i>
Overall timing	<i>January 2012- October 2012</i>
Steps with timeline	<ul style="list-style-type: none"> • <i>Query structure preparation</i> • <i>Intrastat database and administrative registry examination and merging</i> • <i>Programming of tool</i> • <i>Evaluation of results</i>
Case study indicators	<i>Membership degree to query condition ([0, 1] interval)</i>
Hardware, software	<i>Relational database (MS Access), MS Visual Studio (code programmed in Visual Basic 2008).</i>

Expected outcomes beyond final report	<i>Paper for journal or conference under preparation. Discussion with Slovak NSI for further applicability.</i>
Additional comments	-

Statisticians apply the following reasoning when creating reminder letters:

If business is late 1-2 months, the delay is small, 3-4 significant, more than 4 high. If the task is totally new for the business, a delay of 3-4 months is considered as small and the letter should be less strict. Non-responded trade of a high trade value could make more significant bias than trade with small trade value.

Using crisp tools and several databases this is a highly laborious and low efficient work. Our experiments have provided two benefits: we have created methods for capturing statisticians reasoning, merging statistical data and data from registers. Fuzzy logic is able to easily merge three different types of values: crisp (usual numeric values), fuzzy discrete values, and continuous fuzzy numbers inside one query condition. The benefit of fuzzy logic is that it significantly reduces the complexity of the task.

We have proven that fuzzy queries are able to capture both crisp and flexible elementary conditions into the overall condition. But smooth membership degrees of selected respondents are not fully achieved due to the prevailing of qualitative attributes over quantitative. Qualitative attributes contain few elements each which imply that we can calculate limited number of membership degrees. Power of fuzzy logic becomes more obvious if quantitative elementary conditions prevail over qualitative.

This kind of tool could be applied as a standalone connected to the Intrastat database and used during the preparation of reminder letters. Selecting respondents for reminder letters depends on type of survey and associated administrative data. It means that this tool should be significantly changed to meet conditions of other kinds of surveys. We would like to emphasize that this does not concerns the core of method; management of fuzzy sets.

2.3.3 Case study 8.3-3 – Flexible classification of respondents

Case study identification (title)	<i>WP8 – 8.3 Classification of respondents using fuzzy logic.</i>
Preceding work package	<i>WP5 – New Ways of Collecting and Analysing Information</i>
Institution proposing the case study	<i>INFOSTAT</i>
Case study leader	<i>Miroslav Hudec, hudec@infostat.sk</i>
Other staff	<i>-</i>
Case study objective	<i>Flexible classification of business respondents as a support for tailored motivation.</i>
Data	<i>Data used: Business opinions that are collected through phone interviews realized by Slovenian NSI.</i>
Design	<i>Data modelling and quantitative and qualitative analysis</i>
Related fields of official statistics	<i>Business statistics, data collection, public relations</i>
Related methodological topics	<i>Data mining, Fuzzy classification</i>
Case study description	<i>Crisp classifications might causes that respondents with similar behaviour are classified into different classes and differently treated. Fuzzy approach improves classification by directly applying intensity of analyzed property which causes less complex classification space and causes that respondents with similar behaviour are always similarly treated.</i>
Description of methodology	<i>The approach developed in WP5 for the fuzzy classification has been applied (see Deliverable 8.3, Section 2.4 for more details). Fuzzy classification provide a resilient approach for data classification allowing the same data point to reside in multiple classes with different degree of membership and thus, improving the interpretation of the classified respondents. In this way all respondents with similar behaviour receives similar treatment. This way efficiently reveals which respondents (or data user) are key respondents (data users) for NSIs.</i>
Overall timing	<i>April 2012- October 2012</i>
Steps with timeline	<ul style="list-style-type: none"> • <i>Classification space preparation</i> • <i>Data from Slovenian NSI from businesses interviews examination</i> • <i>Programming of tool</i> • <i>Evaluation of results</i>
Case study indicators	<i>Number of flexible rules, rank of respondents by the [0, 1] interval</i>

Hardware, software	<i>Relational database (MS Access), MS Excel, MS Visual Studio (code programmed in Visual Basic 2008) and MatLab.</i>
Expected outcomes beyond final report	<i>The proposal for the synergy ERC call had been sent for evaluation to the EC. The synergy between fuzzy logic (results from WP5) and motivation of business respondents (results from WP3).</i>
Additional comments	-

The experiment of flexible classification evaluation by the SQL-like classifier has shown its ability to solve classification tasks. With the membership degree valued in the [0, 1] interval, only a small number of rules was necessary to solve the classification tasks. As the rules are described by linguistic terms, this ensures an easily readable, understandable and modifiable rule base, and faster processing in comparison of classical approaches where we have to include a significantly higher number of rules. Respondents are classified into several classes e.g. key respondents, relevant, irrelevant respondents. Depending of goals and/or resources available, NSIs could focus their attention on one or several classes of respondents.

Our experiments have proven the benefits of applying the fuzzy classification over the classical crisp classification in this task. The first experiment contained three qualitative indicators whereas the second experiments contained one qualitative and one quantitative indicator. We have observed that quantitative indicator has significant impact on this ranking. As number of quantitative attributes in the classification increases (e.g. classification according three attributes where two are quantitative) the advantages of fuzzy classification are more obvious. If we have only qualitative data in classification (values of membership degrees are reduced to several values), we not fully use the power of the fuzzy logic.

Our contribution is the creation of a classification space, which is capable of capturing statisticians' approximate reasoning and provides a solution for different classification tasks. It implies that there is field for further interdisciplinary research including motivation issues and classification methodologies.

This methodology does not share similar features as two case studies above (8.3-1 and 8.3-2). This approach is relatively independent from databases of official statistics and should work with data prepared for particular classification task e.g. interviewed data, administrative data and surveyed data by NSIs.

2.3.4 Case study 8.3-4 – Genetic programming classification

Case study identification (title)	<i>WP8 8.3 New ways of Intrastat data analysis by genetic programming approach – classification problem.</i>
Preceding work package	<i>WP5 – New Ways of Collecting and Analysing Information.</i>
Institution proposing the case study	<i>INFOSTAT</i>
Case study leader	<i>Miroslav Klúčik, klucik@rrz.sk</i>
Other staff	<i>-</i>
Case study objective	<i>Estimation of missing values in Intrastat data collection system, i.e. improving accuracy of estimates, reducing revisions, improving efficiency.</i>
Data	<i>Data used: Intrastat data 2008, 2009 anonymised, monthly original compiled data – Slovakia.</i>
Design	<i>Data modelling.</i>
Related fields of official statistics	<i>Intrastat.</i>
Related methodological topics	<i>Data mining, data editing, data analysis.</i>
Case study description	<i>Genetic programming on classification will be tested for estimation of missing data (late arrival data) and aggregate estimates.</i>
Description of methodology	<i>Genetic programming is a stochastic method that explores the space of potential solutions (computer programs) in order to find the global optimum. The algorithm searches for classification rules based on non-missing data. The work will be solely executed at INFOSTAT using Intrastat, data provided by the Statistical Office of the Slovak Republic.</i>
Overall timing	<i>December 2011- October 2012</i>
Steps with timeline	<ol style="list-style-type: none"> <i>1. Field description (text document), 12/2011-1/2012</i> <i>2. Data and tool description (text document), 1/2012-2/2012</i> <i>3. Models estimation, 1/2012-5/2012</i> <i>4. Evaluation of results (text document), 6/2012-10/2012</i>
Case study indicators	<i>Classification rate (%)</i>
Hardware, software	<i>Windows XP, EViews min. 5.0</i>
Expected outcomes beyond final report	<i>Scientific full paper, presentation at an international conference.</i>
Additional comments	<i>-</i>

The main objective of the case study was to test an artificial intelligence algorithm – genetic programming – as an alternative to human-based complex computation tasks relating to large databases. The genetic programming classification has been applied for imputation of values of Intrastat database according to discovered relationships in the database, i.e. classification rules able to assign a missing value according to combination of historic non-missing values. A tree structure based algorithm has been used (one tree = one classification rule) with two basic rules (AND, OR),

tournament based selection of best solutions, and standard genetic operations (crossover and mutation of rules according to assigned probabilities).

The achieved classification rate was very high, given the fact that the learning of the classification algorithm has been based only on small fraction of the whole database (10-25%). However, the complexity of the database was not very high; therefore, the achieved classification rates were not significantly different from a human-based search. The genetic programming search algorithm has identified the most occurring instances in the databases and constructed an effective classification rule.

The genetic programming algorithm, once used in a user-friendly application, has the ability for automatic search and more powerful computation abilities in comparison with human-based search. It is able to deal with large database of information without intervention from the side of the user during the computation process (after calibration of the tool).

Disadvantage of the current tested algorithm is that genetic programming prefers values from classes with higher number of instances, which discriminates other values (the database is imbalanced). For further research the selection method of classes shall be optional depending on the specific database. Secondly, the symmetric trees are an obstacle to produce more sophisticated rules (combinations of AND and OR), therefore other tree structure may be tested, e.g. with higher number of sub-trees (branches). There is a high potential in multi-class classification that would require development of a new algorithm.

This tool can be very useful in NSIs for imputation of missing data if offered in user-friendly non-licensed environment. The algorithm is highly case dependent and requires sensitive approach for each new database or task. The classification problem deals with time consuming calculation (classification rate computation) and sensitivity analysis, which can be solved by using powerful hardware. The research experiment was considerably restricted by anonymised data. The possibility of Intrastat, VAT and business statistics databases linkage would although bring more complexity, however, the application of genetic programming (the classification problem) on large datasets is expected to be more efficient with non-anonymised data. It is recommended to make use of the genetic programming tool in NSIs' environment where the issue of anonymity does not apply.

The case study was presented at BLUE-ETS workshops (February 2012, November 2012) and its detailed description is contained in the Deliverable 8.3.

2.3.5 Case study 8.3-5 – Genetic programming symbolic regression

Case study identification (title)	<i>WP8 8.3 New ways of Intrastat data analysis by genetic programming approach – symbolic regression problem.</i>
Preceding work package	<i>WP5 – New Ways of Collecting and Analysing Information.</i>
Institution proposing the case study	<i>INFOSTAT</i>
Case study leader	<i>Miroslav Klúčik, klucik@rrz.sk</i>
Other staff	<i>-</i>
Case study objective	<i>Estimation of aggregate indicators in Intrastat data collection system. Improving accuracy of estimates, reducing revisions, improving efficiency. Modernizing Intrastat system – improving quality of data.</i>
Data	<i>Data used: Intrastat aggregates, business statistics data – large companies, Extrastat data 2004 – 2010.</i>
Design	<i>Data modelling and quantitative analysis</i>
Related fields of official statistics	<i>Business statistics, short-term statistics, administrative data, Intrastat, Extrastat</i>
Related methodological topics	<i>Data mining, data editing, data analysis.</i>
Case study description	<i>Genetic programming software on symbolic regression will be tested for estimation of statistical aggregates.</i>
Description of methodology	<i>Genetic programming is a stochastic method that explores the space of potential solutions (computer programs) in order to find the global optimum. The algorithm searches for relationships between dependent variable and explanatory variables via symbolic regression. The work will be solely executed at INFOSTAT using Intrastat, Extrastat, business statistics and administrative data provided by the Statistical Office of the Slovak Republic.</i>
Overall timing	<i>December 2011- October 2012</i>
Steps with timeline	<ol style="list-style-type: none"> <i>1. Field description (text document), 12/2011-1/2012</i> <i>2. Data and tool description (text document), 1/2012-2/2012</i> <i>3. Models estimation, 1/2012-5/2012</i> <i>4. Evaluation of results (text document), 6/2012-10/2012</i>
Case study indicators	<i>RMSE</i>
Hardware, software	<i>Windows XP, EViews min. 5.0</i>
Expected outcomes beyond final report	<i>Scientific full paper, presentation at an international conference.</i>
Additional comments	<i>-</i>

In the second case study of genetic programming, the application of symbolic regression as an alternative to AR modelling for estimation of statistical aggregates was tested. The artificial intelligence algorithm's task was to discover linear or non-linear relationship between dependent and independent time series. A tree structure based algorithm has been used (one tree = one regression equation) with four basic operators (+, -, *, /), tournament based selection of best solutions, and standard genetic operations (crossover and mutation of rules according to assigned probabilities).

The genetic programming was able to compete with AR models, but has not provided significant increase of fitness. However, the genetic programming models are found automatically from large databases and also are not restricted by the form of the model. An additional output – count of explanatory variables within the best models – can be utilised for standard econometric tools at first hand. Application of genetic programming symbolic regression for modelling of statistical (macro economical) aggregates has significant advantages in comparison to standard regression techniques. The estimation tool does not rely on strict conditions applicable for OLS regression, such as length of time series or stationarity (although does not eliminate the false regression problem). Another substantial advantage is the ability of the genetic algorithm to deal with large number of time series which are analysed simultaneously. Both advantages apply strongly especially within the NSIs which are dealing with countless time series aggregates.

The main disadvantage is the absence of interpretation of the results. The complex nonlinear models give not a clear picture on the relationships between the variables in the model, therefore an alternative analysis of the results need to be provided (analysis of residuals, occurrence count).

The software applied for programming of genetic algorithm in the above case studies is difficult to apply in everyday use of NSIs since it is commercial software and requires licensing. It has been applied due to the author's familiarity with the code programming language (INFOSTAT – research institution also dealing with time series analysis for macroeconomic analyses and forecasts). Also the software is used mostly outside the NSIs particularly in time series analysis; therefore it would require also time- and resource-demanding training of NSIs' staff. The genetic algorithm proposed for classification and symbolic regression analysis is considered relatively easy comprehensible, therefore there are possibilities for implementing it in coding environment more familiar to the NSIs staff, such as R (Flasch et al., 2011).

Case study was presented at BLUE-ETS workshops (February 2012, November 2012) and the SIS conference in Rome in 2012. The detailed description of the algorithm and case study is contained in the Deliverable 8.3.

2.3.6 Case study 8.3-6 – Classification of Intrastat data by neural network approach

Case study identification (title)	<i>WP8 8.3 New ways of classification of Intrastat data by neural network approach</i>
Preceding work package	<i>WP5 – New Ways of Collecting and Analysing Information</i>
Institution proposing the case study	<i>INFOSTAT</i>
Case study leader	<i>Jana Juriová, juriova@infostat.sk</i>
Other staff	-
Case study objective	<i>The main objective is to use classification by means of neural networks for imputation of missing data in Intrastat data system.</i>
Data	<i>Data used: Intrastat data 2008 - anonymised monthly data provided by the Statistical Office of the Slovak Republic.</i>
Design	<i>Evaluation study</i>
Related fields of official statistics	<i>Intrastat</i>
Case study description	<i>The proposed neural network will be used for classification of surveyed Intrastat data and consequently for imputation of missing values.</i>
Description of methodology	<i>Neural network is a computational model from the category of soft computing methods, based on the abstraction of biological neural systems. In this case, the neural network is trained on the complete data file for classification. After reaching an acceptable degree of accuracy the network can be used for the classification of the rest of database with missing values. The neural network identifies the most similar class for each statistical unit and this enables the imputation of missing values.</i>
Overall timing	<i>September 2011- September 2012</i>
Steps with timeline	<i>1. Intrastat data examination 2. Programming of tool 3. Testing of tool on Intrastat database 4. Evaluation of results</i>
Case study indicators	<i>Class probabilities (%) – probabilities that units belong to certain class, RMSE</i>
Hardware, software	<i>Windows XP, Mat Lab</i>
Expected outcomes beyond final report	<i>Paper in contributed session at Conference SIS 2012, Rome Paper for conference under preparation</i>
Additional comments	-

The neural network (NN) case study was focused on solving a classification problem in business and trade statistics. The classification approach proposed under the WP5 was tested and verified on Intrastat database – data on dispatches for year 2008. The proposed neural network had to be

adjusted for the purpose of Intrastat database classification and trained for the classification of units according to the nature of transaction into 2 classes. The proposed neural network was trained to gain the highest probability of the inclusion into the right class and at the same time significantly higher than the inclusion into the other class. The goal was to identify the right class according to the available characteristics, i.e. surveyed items and to learn the network to distinguish between the two classes with the highest precision.

The best results were obtained with the neural network with 15 neurons and 1000 training cycles. The highest probability of inclusion into classes was with this neural network gained above 70% for both classes when the number of training cycles was 1000. Increasing the number of neurons in the hidden layer or increasing the number of training cycles did not result in this case in gaining better trained network.

After the network has been trained the best one was used for the classification of the original data to verify the proposed classifier. The probability of inclusion into the right class proved to be 76.8%. This confirmed the ability to use the trained network for suggesting the missing values.

Considering the gained results, we can see that the precision is still not sufficiently high, we would expect the probability around 90 per cent and there is also a problem with the searching algorithm, because results vary quite a lot with each run of the proposed network.

The proposed neural network could be tested also enlarged, i.e. for classification into more classes. In the case of classification into more than 2 classes the softmax function is more suitable to use (the softmax activation function is a biologically plausible approximation to the maximum operation).

To supplement, any introduction of new methods for the purpose of missing values imputation at the NSIs needs further research of variance estimation of proposed values. However, this approach has not been verified in this respect. When the proposed method will be improved, the variance estimation of values suggested by NNs should be computed and verified as well.

2.3.7 Conclusion of soft computing

The soft computing methodologies – fuzzy logic, genetic programming and neural networks, developed within the framework of WP5, were adopted to be able to solve some of the data collection and analysis issues of official statistics. Soft computing is rarely used in official statistics although it represents a significant potential. The aim was also to provide a proof for the EU statistical community of their usefulness. This aim is a consequence of the struggle of more sophisticated tools to replace or at least supplement the standard tooling of NSIs' staff (Saporta, 2012; EPROS, 2007).

Our research has met two goals. First one is described in the task description of WPs (MEETS). The second considered the conclusion of the European Commission (European Plan of Research in Official Statistics, 2007, p.17) where needs for research in soft computing were emphasized.

The development of fully functional software tools is time demanding and quite expensive. This could cause the NSIs to cope with additional burden, bearing the brunt of such steps in costs of products of official statistics. One possible answer to this problem is software sharing (Lehtinen and Gløersen, 2009). Generally speaking, NSIs share common mission and obligations. Sharing of software tools, through the limited open source approaches among statistical institutions could reduce the development effort inside NSIs. In order to deal with these issues UNECE, Eurostat and OECD on the Meetings of Management of Statistical Information Systems (MSIS) have created the Sharing Advisory Board. In order to meet this goal, we could advertise these approaches through the recently launched by the Sharing Advisory Board new software inventory to find way to continue

development of full functional tools and therefore offer them to NSIs. This software inventory is compiled in cooperation with the ESSnet project on a Common Reference Architecture (CORA), and already contains information and contact details for statistical software applications available for sharing.

3 IMPLICATIONS FOR NSIS, EUROSTAT AND ESSNET

Data collection is a dynamic and important stage of NSIs data production. Collecting data directly from businesses is costly for NSIs and creates a burden for respondents. These are two important reasons why NSIs use administrative and secondary data. One part of our research offered standardized tool for the quality measurement of these data sources.

However, not all data are available (or available in time) from administrative or secondary data. This means that business surveys remain a relevant part of data collection. Another part of our research focused on issues of motivation and response burden in order to improve quality of the survey data implying decrease of the costs for the NSI in the data collection stage. This part of research has offered quantitative methods for measuring response burden, exploring causes and consequences of response burden and has tested various methods to enhance motivation and reduce burden in business surveys.

NSIs have large amount of different kind of collected data or data from administrative and secondary sources (quantitative, qualitative and textual) describing the data collection stage. These amounts of data contain valuable information which cannot be effectively minded by traditional tools. Third part of our research offered methods for support motivation, reminders and data imputation based on soft computing.

3.1 Implications in response burden and motivation

The case studies exploring causes and consequences of response burden have some important implications for NSIs. First of all, these studies are some of the first to demonstrate that both actual and perceived response burden affect response behaviour and through this the costs and quality of data collection. This implicates that NSIs should monitor and manage both actual and perceived response burden. Moreover, the analyses from the Swedish register on response burden show that response burden should be managed not only on the level of individual survey, but also on the business level. Finally, this part of the research found indications that attitudes towards the statistical office and the perceived usefulness of statistics affect perceived response burden.

The case studies on actions to reduce burden and increase motivation clearly illustrated that questionnaire design matters. The experiments with various communication strategies show that it is not straightforward to affect motivation and ensuing response behaviour with these types of actions. The treatments (e.g. promising statistical feedback, illustrating the usefulness of statistics, enhancing personal ties and positive emotions) did produce some but not all of the expected effects. One important reason for this seems to be that responding to official business surveys is a routine behaviour for many respondents. They have become used to only quickly scan the materials sent to them for obligations and deadlines and have established attitudes towards official statistics which may be hard to change with a one-time action. So though it is important to improve the attitude towards the statistical agency and the usefulness of statistics among respondents to business surveys, accomplishing this is challenging; new attempts in this direction need careful pre-testing with the target group and are probably most successful if tailored to the needs of a specific business, focused on businesses or respondents who are new to the survey, and if coordinated over time and over surveys.

3.2 Implications in measurement of quality of administrative sources

NSIs use administrative data with various intensity but the common problem is quality measurement of these sources. NSIs have to cope also with the fluctuating quality of administrative sources. The significant shift is offered in our research. We offer them a standardized tool for measuring quality and a report card. The approach is immediately applicable by other NSIs or statistical oriented organizations if they are able to use R in their production environment. If the latter is not the case the alternative is to implement the methods in their programming language of choice. The QRCA can be generally used and allows comparison between data sources and NSIs.

3.3 Implications on innovative tools and procedures developed for data collection and analysis

Examination of statistical knowledge and tasks reveals that traditional tools cannot cope efficiently with all demands. We suggested tools based on the neural networks and genetic programming which help in the estimation of missing values by searching for similar patterns. Moreover, soft computing approaches can estimate attributes which are not in the focus of data collection but are relevant for national and international statistical institutes (Hudec and Juriová, 2013).

Fuzzy classification can help identifying key data users and respondents and reveal their potential and weaknesses so as to ensure that similar respondents/data providers are always similarly treated. Flexible classification is a significant support for tailored motivation strategies.

Fuzzy selection offers alternative way for tailored reminders. In dissemination the same tool is capable of giving answers to imprecise questions in the search for relevant data. It can solve more user demands and therefore improve image of NSIs as data provider.

Data dissemination by data portals of NSIs is a significant element for respondents' motivation either in positive or negative way. If a data portal is well designed and users can easily find relevant data and information they will be more motivated to provide their own data. Users want either raw data (large businesses) or extracted rules from data and information (small businesses) (Bavdaž, 2011b). It is assumed that, if we provide businesses not only with data but also „information with meaning“ by appropriate data mining tools they will be more satisfied and willing to provide their own data. Providing users tool capable of giving answers to their imprecise questions on websites could solve more users demands and therefore improve image of NSIs. Flexible querying provides not only list of selected entities but also ranked by degree of matching the condition. In case of territorial units this information could be directly processed to thematic maps having a colour gradient from a faint hue to deep hue of matching degree.

Modernization of the first and the last stage of data collection could create a chain reaction of improvements in data quality. Better data dissemination could motivate respondents to provide their own data timely and accurately and reduce the frequency of missing values implying more efficient imputation (less missing values and powerful soft computing tools). Finally, better and earlier data will be available for dissemination (websites) or exchange among institutes (e.g. by SDMX).

3.4 Implications for other projects and new integrations

ESSnet project on a Common Reference Architecture (CORA)

In the cooperation with the ESSnet project on a Common Reference Architecture (CORA), the Sharing Advisory Board (SAB) Inventory has been created. Since the Blue-ETS reports discuss developed methodologies and tools for use in national or international statistical institutes, it would be helpful to inform NSIs about these tools through the recently launched by the SAB new tools inventory. This inventory already contains information and contact details for over 50 statistical applications available for sharing.

For future development of tools and their improvement and implementation it could be useful to be in contact with this inventory to find partners for these activities. In this way NSIs could be focused on developing several tools and offer them through the Inventory. On the other side, they will be able to use tools developed by other teams. In this way resources for modernization of official statistics will be efficiently used (no duplicity in developing tools for the same purpose) and NSIs will obtain tools of high quality. The significant issue in NSIs is high fluctuation of skilled staff. If NSIs are developing many duplicity tools independently then there is a risk of non-finishing or finishing with significant constraints in quality.

Generic Statistical Business Model (GSBPM)

Recently the Generic Statistical Business Model (GSBPM) was proposed by the Joint UNECE/Eurostat/OECD Work Sessions on Statistical Metadata (METIS). The GSBPM (Vale, 2009) should be seen as a flexible tool to describe the set of business processes needed to produce official statistics. The use of this model can also be envisaged in other separate, but often related contexts such as harmonizing statistical computing infrastructures, facilitating the sharing of software components and providing a framework for process quality assessment.

The main goal of the GSBPM is to divide activities in statistical organizations into logical parts in a standardized way, because generally speaking, NSIs share common mission and obligations. This leads to exchange ideas, methodologies and tools which will reduce the software development efforts inside NSIs. Our results could be interesting for modernizing and standardizing sub steps of data collection.

ESSNet on SDMX

In area of exchange of aggregated data between statistical institutes the Statistical Data and Metadata eXchange (SDMX) standard appears. Another standard, DDI (Data Documentation Initiative) is also under interest of Eurostat. Gradually they penetrate into data collection.

Storing of questionnaires structure together with supporting metadata in standard format meets two goals: they can be easily understandable and modifiable for similar surveys and new staff will be able to easily obtain relevant knowledge of realized surveys. It means that calculations of dependent variables, validations and interactive support could be added in standardized way in all kinds of online questionnaires.

ESSnet on Administrative data

The ESSnet on the use of administrative and accounts data for business statistics aims to find common ways for use of administrative data for business statistics are share the practices and methods to all interested member states. The quality studies in this ESSnet are performed in several work packages. Main quality focus is the effect of the quality of administrative data on the output of NSIs. Also attention is paid to the fact that information once collected from enterprises should be used for producing business statistics to a possible extend, aiming to reduce the reporting burden on businesses and to improve quality of statistical information. During the BLUE-ETS project the WP4 –leader kept in touch with the project leaders of the various work packages in this ESSnet to assure knowledge and experiences on the determination of the quality of administrative data sources was exchanged and was available to all involved.

4 FUTURE RESEARCH AVENUES AND OPPORTUNITIES

In the frame of the Blue-ETS project we were focused on several aspects on data collection in order to improve the collection and the quality of data for business and trade statistics. Our WPs have been running more or less independently. In the second part of our research, gradually, we become more and more aware of the potential of synergy of our results. This section firstly discusses opportunities for each WP and then moves attention to the opportunities for synergy.

4.1 Opportunities in reduction of response burden and motivation (WP2, WP3)

Even though a wealth of paradata is collected as part of the statistical production process, these data are often not complete, sometimes difficult to access, and not systematically analysed in depth. Comparison of paradata among NSIs and their evaluation for reduction of response burden and motivation enhancement could be an interesting topic for further research which could tell us which additional data are interesting to measure and how motivation differs among countries. Admittedly, paradata do constitute “golden mines” and should be further mined.

Rapid development of new technologies should be used for further improvement of data collecting from businesses in order to make providing data easier and more rewarding for respondents and more efficient for NSIs. For example, NSIs could easier generate online questionnaires from their documents (supported by standards SDMX and DDI and their related tools). Respondent can benefit from interactive and easy to use questionnaires that also provide them with relevant statistical feedback. Also, to increase the overall image of NSIs and official statistics in the business world, it seems important to bring official statistics in everyday day business use by developing some statistical products addressing specific business needs based on official statistics and the latest ICT knowledge.

4.2 Opportunities in quality of administrative sources (WP4)

Goal of WP4 work was develop a general applicable approach to study the input quality of administrative data sources. For this purpose a quality framework, measurement methods and scripts were developed. All effort was put into enable application of the findings by others. Interestingly, the quality indicators developed are not limited to administrative data sources but could be applied to all data sources that potentially provide valuable information for official statistics. As such, new data sources, e.g. so-called Big Data, can be evaluated by them. This certainly opens up opportunities to extend this work into areas that focus on the use of new data sources for statistics.

4.3 Opportunities in soft computing (WP5)

The aim of our WP was to develop methodologies and tools based on soft computing and mining textual data sets in order to modernize data collection system and improve data quality. This way gives us valuable knowledge for further research on hybridization of methodologies in order to utilise strengths and mitigate weaknesses. The new opportunities are as follows:

1. Improving mining and analysis textual data sources by fuzzy logic. Co-clustering of documents and words and storing imprecise terms in fuzzy databases are main areas where we expect new results. This opportunity had been developed to the proposal for the new FP7 project. The proposal

is focused on supervised and unsupervised evaluation of available textual and numeric data related in field of health service and statistics. Textual data sets are relevant source of information and should not be neglected by NSIs.

2. Integration of neural networks and fuzzy logic leads to a more sophisticated tool for estimation of missing values. Briefly speaking, in first step neural network will estimate missing values and in second step fuzzy logic will evaluate estimated values. If required fuzzy logic will provide feedback for improvement of estimation and neural network itself.

3. Motivation of respondents is also influenced by the ways how NSIs have developed data portals for dissemination. It is proposed that, if users (often they are respondents) can find relevant data and information easily, they will be more willing to provide their own data. Providing users tool capable of giving answers to their imprecise questions on websites could solve more users demands and therefore improve image of NSIs. Equations and relations in the field of fuzzy logic developed in this work package could be used in data dissemination without significant modifications.

4.4 Opportunities related to synergy of our WPs

Researches in WPs although separate at beginning have revealed further promising synergic topics.

The significant result is the new project proposed for the ERC Synergy grant. Adaptive survey designs (ASD) bring considerable advantage to the field of designing surveys by taking into consideration the fact that the impact of various design features varies significantly over respondents (Calinescu and Schouten, 2012). When auxiliary information is available (registry, paradata, previous behaviour of respondents) designs may be tailored to optimize response rates through strategies designed to enhance motivation. Although adaptive survey designs are promising approaches, they cope with limitations (business or respondents although having similar attribute values, may be classified into different classes and therefore receive different treatments) which could be solved by fuzzy logic.

WP2, WP3 and WP5 examined significance of reminders. In WP5 tool for flexible selection of respondents for different types of reminders has been created using limited administrative data. In WP2 and WP3 this problem has been analysed from respondents' perspective. Merging knowledge from WP5, WP2 and WP3 and using rich paradata sets could create more tailored reminders.

Motivation of respondents by offering a set of statistics (Key ratios) to businesses who respond on time had been examined. Fuzzy classification has created classification space ensuring that similar entities are similarly treated. Integration of these two approaches: classification space capable to cope with paradata and data obtained from interviews and offering different amount of reward (key ratios + some discount when business send request for new data) according to behaviour in responding could support strategy for motivation of each group of respondents.

Administrative data and new data sources, e.g. so-called Big Data contains large amounts of data and information. Identification of identical or similar establishments in a business record could be very useful, since it is one of issues in building a business registers and use of data. Fuzzy functional dependencies and neural networks could reveal similar establishments and also reduce size of database without compression. Database could be reduced and fully functional without decompression.

5 CONCLUSIONS

Data collection is the first and very important part of statistical data production. High quality data obtained either from administrative sources or from respondents are crucial for high quality of NSIs' outputs reached by reasonable costs. Even with maximal effort some data cannot be collected and they have to be estimated. Case studies presented in this evaluation report deals with all above mentioned areas. We have offered methodologies and tools for improvement all mentioned aspects of data collection.

This process should be continuously improved because society, respondents, administrative and secondary sources are changing and we should be prepared. We have discussed new opportunities which implied form our work. Some of recognised opportunities had been further developed to the proposal for new projects.

REFERENCES

- Balbi S., Hudec M., Juriová J., Klůčik M., Stawinoga A., Triunfo N. (2013). Report on analysis of existing practices in the data collection field. BLUE-ETS Project Report (deliverable 5.2).
- Bavdaž, M. (Ed.) (2011a). Business use of NSIs' statistics based on external sources (NSIs, publications, expert opinions). BLUE-ETS Project Report (deliverable 3.1).
- Bavdaž, M. (Ed.) (2011b) Final report integrating findings on business perspectives related to NSIs' statistics. BLUE-ETS Project Report (deliverable 3.2).
- Bavdaž, M., Giesen. D. (2009) Description of work for work packages 2 and 3. In P. Roberti: *BLUE-Enterprise and Trade Statistics: Description of work. Annex to Seventh Framework Programme Grant Agreement 24476* (pp. 14-16) Brussels: European Commission.
- Bavdaž, M. (2010). Sources of Measurement Errors in Business Surveys. *Journal of Official Statistics*, 26 (1), 24-42.
- Calinescu, M., Schouten, B. (2012). Adaptive survey designs that minimize nonresponse and measurement risk. Statistics Netherlands
- Daas, P., Ossen, S., Tennekes, M., Zhang, L-C., Hendriks, C., Foldal Haugen, K., Laitila, T., Wallgren, A., Wallgren, B., Bernardi, A., Cerroni, F. (2012b) Quality Report Card for Administrative data sources including guidelines and prototype of an automated version. Third deliverable of workpackage 4 of the BLUE Enterprise and Trade Statistics project, version 1.1 final
- Daas, P., Tennekes, M., Ossen, S., Di Bella, G., Galiè, L. , Laitila, T., Wallgren, A., Wallgren, B., Hendriks, C., Zhang, L-C., Kristin Foldal Haugen (2013) Guidelines on the usage of the prototype of the computerized version of QRCA, and Report on the overall evaluation results. Second deliverable of workpackage 8 of the BLUE Enterprise and Trade Statistics project, version 04 draft.
- Eu Project on Baseline Measurement and Reduction of Administrative Costs (2009). *Final Report. Measurement data and analysis as specified in the specific contract 5&6 on Modules 3&4 under the Framework Contract No ENTR/06/61. Report on the Statistics Priority Area*, March 5 2009.
- European Plan of Research in Official Statistics (EPROS) Main conclusions from the activities in the 5th Framework Programme (2007). Luxembourg: Office for Official Publications of the European Communities.
- http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-07-003/EN/KS-RA-07-003-EN.PDF
- Flasch, O, Mersmann, O, Bartz-Beielstein, T, Stork, J (2011) rgp: R genetic programming framework. <http://CRAN.R-project.org/package=rgp>, rpackage version 0.2-4
- Giesen, D.& Raymond-Blaess,V. (Eds). (2011). Response burden measurement and reduction in official business statistics. A literature review of national statistical institutes practices and experiences. BLUE-ETS Project Report (deliverable 2.1).
- Giesen, D. (Ed.) (2011). Response Burden in Official Business Surveys: Measurement and Reduction Practices of National Statistical Institutes. BLUE-ETS Project Report (deliverable 2.2).
- Giesen, D., Bavda, M., Bolko, I. (2013) Comparative report on integration of case study results related to reduction of response burden and motivation of businesses for accurate reporting. BLUE-ETS Project Report (deliverable 8.1).

- Hudec, M., Juriová J. (2013) Soft computing in official statistics data collection. IADIS International Conference on Information Systems (Accpeted for publication).
- Hudec, M. (2012). Dynamically modelling of fuzzy sets for flexible data retrieval. 46th scientific meeting of the Italian statistical society (SIS 2012), Rome.
- Juriová, J. (2011) Use of Neural Networks for Data Mining in Official Statistics, EUROSTAT and European Commission International Conference (NTTS 2011), Brussels.
- Křůčík, M., Hudec, M., Juriová, J. (2012) Final report on the case study results on usage of IT tools and procedures developed for data collection (Soft computing tools for Official Statistics). BLUE-ETS Project Report (deliverable 8.3).
- Křůčík, M. (2011) Introducing New Tool for Official Statistics: Genetic Programming, Paper for the 2011 European NTTS conference, Brussels.
- Lehtinen H., Gløersen R. (2009). Cooperation in development of open source software. UNECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS 2009), Oslo.
- Saporta, G. (2012) Mining Official Data: What's New?, 2012 SIAM International Conference on Data Mining, Anaheim, California (Invited paper, Data Mining in Official Statistics Workshop).
- Tennekes, M., de Jonge, E., Daas, P.J.H. (2013) Visualizing and Inspecting Large Datasets with Tableplots. Journal of Data Science, 11 (1), 43-58.
- Vale, S. (2009). DEVELOPING A GENERIC STATISTICAL BUSINESS PROCESS MODEL. UNECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS 2009), Oslo.