

# ESSNet S D C

A Network of Excellence in the European Statistical System in the field of  
**Statistical Disclosure Control**

## **The first year** (Anco Hundepool)

### **General**

The ESSNet on Statistical Disclosure Control is a two-year project, aiming at the cooperation between NSIs in Europe in the field of Statistical Disclosure Control. The full project description can be found in the technical proposal.

In this report we will highlight the work of the ESSNet in the first year. Several reports ready for full publication can be found at the official website, while certain drafts and internal documents are available at the internal project website.

### **Meetings**

The major project meeting took place 5 and 6 June in Wiesbaden. The major topics during this meeting were the discussions on the results of the testing efforts by all the minor partners. Separate minutes of this meeting are available at the website.

A second meeting took place during the PSD'2008 conference in Istanbul. The major partners, the steering committee, discussed the progress of work and made arrangements for the work during the remainder of 2008.

## **Overview of the work per topic**

### ***Dissemination of CENEX results***

#### **1.a Improvement of the handbook**

##### **1.a.1. Consultation of NSIs.**

Many partners in the ESSNet have contributed to the discussions. Their findings have also been laid down in their reports available at the ESSNet project website. This was the starting point for many improvements in the handbook and the software.

##### **1.a.2. Enhance the handbook: Examples and recommendations**

Many lessons have been learned from the consultations mentioned above. An improved version (1.1) of the handbook is now available from the website.

We have also started section on Case studies in order to illustrate the methods described in the handbook and make the theory more easily understandable. We plan to further expand this section in 2009.

## **1.b. Follow up of the CENEX inventory**

### **1.b.1. Needs of less experienced countries**

The CENEX questionnaire has taught us many aspects of the situation with respect to SDC in almost all European countries. A report of the lessons learned from this questionnaire has been produced and is available from the website.

### **1.b.2. Tailored courses and follow up visits.**

Teaching courses has always been a good way of spreading the knowledge on SDC. This was already recognised in the proposal. This time we have decided to concentrate on a smaller group of countries. This should improve the ‘critical mass’ of knowledge. The training course this year (9-12 December), hosted by the Slovenian Statistical Office, was attended by participants mainly from the south-eastern part of Europe: Hungary (4), Romania (4), Bulgaria (2), Slovakia (2), Czech Republic (2), Italy(1) and Slovenia (6 as the host country). The main part of the costs for the participants was reimbursed by the ESSNet (70%-rule), but for the remainder a grant from the Unesco Chair in Data Privacy (Universitat Rovira i Virgili, Tarragona) was gratefully acknowledged.

New items in this course were sessions where the participants applied ARGUS on their own familiar datasets. Participants were asked to bring their own data sets to get familiar to the ARGUS-software. Almost all participants brought their own datasets and this proved to be a great success. A more detailed report will be written in January.

For next year’s training course we will concentrate on the north-eastern part of Europe.

A continuous help-desk function has been working illustrated by the many requests for help received by the partners.

Daniela Ichim (IStat) has visited ONS in February in order to assist with the SDC-work.

## ***Making SDC-tools better usable by NSIs***

### **2.a. Widening the scope of input and output data formats**

Apart from minor improvements in ARGUS we have introduced SPSS as a new data format. Both  $\mu$ -ARGUS and  $\tau$ -ARGUS can now read and ( $\mu$ -ARGUS) write SPSS system files. This will facilitate the use of ARGUS in many NSIs as many NSIs use this format as a standard file format to store final datasets.

## **2.b. Problems with the IT-infrastructure**

The test-reports by the testers have paid attention to this issue. Improvements in ARGUS will be considered next year.

## **2.c. Sustainability**

Although only planned for next year some internal reports have been written already. Issues of open source/R are of course addressed. Also the initiative by the UN-ECE/Eurostat MSIS group will be taken into account. Contacts have been made already. Internal reports are available at the project site.

## ***Guidelines for output checking***

For the guidelines for output checking a separate sub-group has been established. After a first meeting in Voorburg it was decided to prepare a questionnaire and send it to all NSIs in Europe. The response was very encouraging. The results are being processed now. The questionnaire is available at the project website. After this work a second meeting is planned in March 2009 to draw the conclusions from this survey and decide on the remaining work in this part of the ESSNet.

With respect to the automatic output checking a report on the jack-knife has been prepared and published on the website.

## ***Communication; Web and FAQ***

We decided not to build a complete new website for this project, but add a new section to the familiar CASC-website. This website is known in the SDC-world and this will attract more visitors. The whole website has been restyled, giving it more modern face. New results from this project are and will be made available there.

A start has been made with the FAQ.

## ***Improvement of software for micro data***

### **5.a. Big surveys**

For Very large surveys and for census data, certain methods in  $\mu$ -ARGUS might be too computer intensive/ time consuming. However it is not always really necessary to apply the methods on the whole data file in one big run. Certain partitioning might be very effective. The deliverable reports on this and paves the way for certain extensions of  $\mu$ -ARGUS next year.

### **5.b. Alternative risk models or Microdata dissemination strategies.**

Two reports have been produced for this topic.

- Dissemination of Microdata Files Stemming from Social Surveys; Improvements of the Estimation of the Re-identification Risk Measures Based on Log-Linear Models
- Community Innovation Survey: a Flexible Approach to the Dissemination of Microdata Files for Research

The first report presents a new way to estimate the risk of disclosure for social microdata. More tests and assessment will be performed next year. The second report investigates possible dissemination strategies for enterprise microdata (in particular data stemming from the Community Innovation Survey). Both reports will give a background for producing safe micro data in the respective area of interest (social and business). Daniela Ichim visited ONS in February. She worked together with ONS on a test to apply the Italian made routine in R for risk estimation and disclosure limitation on the British sample of the Community Innovation Survey microdata. The test was successful and ONS and the data custodian are evaluating the results.

### **5.c. Record linkage**

The record linkage software is now available to the ESSNet-team and will be built in  $\mu$ -ARGUS next year.

## ***Improvement of software for tabular data***

### **6.a Protection of frequency tables and table with negative values**

#### **6.b. Inclusion of audit-routines**

The first version of an audit routine had been produced during the CASC-project by the Ilmenau-partners. This version however proved to be rather slow and very memory inefficient. We have improved this version. The speed has been improved drastically and also it now works with both CPLEX and XPress as solvers. It is now part of  $\tau$ -ARGUS. It is not very visible in  $\tau$ -ARGUS as it replaces the previous audit-routine, which was not always reliable.

### ***Synthetic data files***

The project team is already ahead of schedule. A report has been written and is available at the project website. A missing part on development of new methods and test on real surveys will be the topic of the foreseen subcontract with the University of Rome and further work with permanent staff at Istat. This latter part will be the topic of a specific report in 2009. Also a first version of the software for synthetic data has been written in R. This has been produced already ahead of schedule. These routines will be included in  $\mu$ -ARGUS next year. Joining R and

ARGUS could also be a test-case and an opening to open source for the ARGUS software.

### ***Analysis of problems on linked tables.***

A report on possible strategies for linked tables has been written and has been presented at the PSD'2008 conference in Istanbul. An updated version is available at the website as deliverable.

### ***Management***

The Steering committee has met during the Wiesbaden meeting and also during the PSD'2008 conference in Istanbul. This was the most efficient way to manage the project.

In general we can say that the project progresses more or less as planned, apart from the two issues mentioned in the beginning. But beyond a slight delay this will not be a threat whatsoever to the successful completion of the ESSNet project.

### ***Personnel***

The ESSNet was happy to bring together a great project team, consisting of the major SDC-specialists of the NSIs in Europe, together with a few specialists from the academia. The major participants formed also the steering group of this project. The steering group consisted of Anco Hundepool (CBS) as the project leader and Luisa Franconi (IStat), Sarah Giessing (Destatis), Keith Spicer (ONS) and Josep Domingo-Ferrer (University Rovira i Virgili, Taragona). This group controlled the main activities of the project. Besides these partners many other NSIs and a few others participated. The smaller NSIs mainly played the role of tester. Experiences from the past have shown, that it is essential to bring these testers in the project and not rely on some voluntarily testing. That will lack the necessary quality.

Apart from the positive feelings a few manageable problems have to be reported.

During the first year Slovenia could unfortunately not fulfill its required tasks. The original partner (Nina Jukic) has left the Slovenian Statistical Office and her replacement (Andreja Smukavec), a very good replacement, was unable to work during this year due to serious sad personal reasons. But as she will return to the office in spring 2009 we still expect a great contribution from her.

In Italy we have seen problems with the hiring of the planned temporary staff. IStat has banned the hiring of all temporary staff after signing the contract. IStat has also reported that to Eurostat. This left us with the problem of finding a solution for this. However in December we have reached a conclusion how to solve this. The outcome will be that IStat

will do the majority of the tasks with the permanent staff thanks to an extension of the SDC-unit. For the remainder task on synthetic data we propose to sign a subcontract with the University of Rome. A good qualified professor has been found there. We expect that, although there will be some delays, we will be able to fulfil all our tasks during the project period.

Finally in Portugal Teresa Crespo has a serious illness and she has been replaced by Pedro Campos. He has also fully taken over Teresa's tasks.

### ***Conclusions***

In general we can say that we have been able to do almost all the work as planned. In such a large project we can always expect some deviations from the original plans. The smaller delays due to some staffing problems are still fully under control and manageable, while on some other topics we are already ahead of schedule, like work on the sustainability and the synthetic data (software produced already).

We are confident and look forward to the work in the second part of the ESSNet project.