

# ESSNet S D C

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

## **Minutes of the ESSNet project meeting in Rome 3-4 November 2009**

Present: Anco Hundepool (Netherlands),  
Luisa Franconi, Daniela Ichim, Gianni Seri, Maurizio Lucarelli (Italy)  
Josep Domingo-Ferrer (Catalonia, URV)  
Sarah Giessing, Christopher Gürke (Germany)  
Johan Heldal (Norway)  
Michael Carlson (Sweden)  
Pedro Campos (Portugal)  
Mathias Templ (Austria)  
Georges Zangerlé (Luxembourg)  
Kathalin Gajdos (Hungary)  
Felix Richie, Keith Spicer (UK)  
Jean-Marc Museux (Eurostat)

### **Tuesday**

Meeting starts at 9.30

#### *Guidelines for output checking:*

Presentation by Felix Ritchie of the guidelines for output checking. Lively discussion follows on rules for output checking and strategies to follow in this area. Sarah suggests to apply perturbation to output in order to preserve confidentiality of results given to researchers. However this view is not shared by others in the group. Museux suggests that maybe assumption is that users are more happy with real data; also it is good having a living document however how would it be possible to preserve sustainability of this idea? German delegate suggests the creation of a Wiki book where staff of RDC can contribute.

#### *Administrative issues:*

Time sheets by the end of the year: there is a deadline for sending all administrative papers: send them by the end of January to Anco.

There has been a change in the contract: University of Rome is helping in the topic of simulated data. Contract has to be signed now. Final report of the project with contribution by all.

#### *Synthetic data:*

Presentation by Josep Domingo-Ferrer, review of work done so far, and a presentation by Istat, description of a new method and presentation of the results on test data. A paper will be presented at the Bilbao Worksession and report with the results on enterprise data will be produced by Istat and University of Rome (prof. Brunero Liseo).

It follows a lively discussion on the use of simulated data. Destatis is investigating the use of simulated data for data structure file. Istat sees different setting in US and Europe; here we can use synthetic or perturbed data for PUF. Josep wishes it would be possible to find a method to reproduce perfectly the data without being unsafe. Felix suggests that synthetic data can be used as test data set for RDC and they would be useful for users of the RDC to prepare programs and test methods. Museux see them useful for PUF if we can increase information content. Matthias Temple suggests that maybe we need to create different synthetic data sets for different users.

FAQ

These are far from complete. Each of us should think of two FAQs and provide answers to these.

No restriction on the subject of the FAQ.

*Web Site:*

CBS has put .hrc files for the new NACE classifications. This was created for the SBS-statistics and the SBS-course. Luxembourg asks whether also other classifications for SBS could be introduced. Anco thinks that maybe this is outside the scope of the web page. ONS has a web page on the institutional web site with FAQ on SDC. Anco could make a link to this page.

*SDC courses:*

Two courses have been held during the project. The second one was in Vilnius. Positive reactions to both courses. Next May there should be another training in Luxembourg as part of the ESTP programme. Also a special SBS-course is given in Luxembourg

*$\mu$ -ARGUS*

New features: reads and writes SPSS file, blocking options in micro-aggregation and other, inclusion of R-synthetic data procedure, record linkage, new manual.

Luisa Franconi suggests that as more and more methods are being developed, it would be useful to move to open source. Eurostat agrees on this point: we should move to open source as it was done with Demetra software for season adjustment. Anco agrees that good solutions should be available to all otherwise there is a waste of resources. Eurostat sees the interface as a big issue; there is also a data cruncher and a library of routines that can be used by all and directly into the production system. Matthias briefly reviews the new development of sdc-micro. Presentation by Portugal on comparison between Argus and sdc-micro. Discussion follows on what still needs to be developed in R as it is not complete.

*$\tau$ -ARGUS :*

New features: Reads SPSS, new table presenter, generator a-priori/history files, linked tables (modular is ready, now the interface will be built), new manual, weighting, distance function for secondary suppression in modular, negative values in modular. Discussion for free solver: till now it was not possible but now it seems there is hope to go in this direction.

Luxemburg has used Argus for the SBS-tables: still difficulties in managing data set, still lack of communication of problems and error messages. Sarah made comments on features that could be extended. Output file containing audit information. Luxemburg asks for the introduction of waivers. Luxembourg intends to use waivers to increase the number of published cells.

*Sas and  $\tau$ -ARGUS:*

Presentations by Norway and Sweden: macros from SAS to  $\tau$ -ARGUS. Documents will be made available in the ESSNet web site with a summary of the work done. Destatis has developed a series of SAS macro that can manage linked tables problems (exporting files, export a priori reads back the results) a report has been written and will be presented at the Bilbao worksession, it will be put in ESSNet web page. There is the proposal to share programs and macros in SAS and make them available through the web site.

Meeting closes at 4.30 pm.

**Wednesday**

Meeting starts at 9.40

*Sustainability:*

Istat presentation on sustainability of Argus. It follows a lively discussion on direction of open Argus.

*Handbook:*

Comments on the handbook and decision on which direction to go: live document or printed book? Keith commented that it was quite improved compared with previous version. Needs continuous assessment and improvements.

Museux believes that have achieved something: we need a publication. If we want to keep it alive we need to change mode: Wikipedia type. Eurostat has experiences here and is willing to help.

Pedro explains how the previous version was not toward NSI; still is oriented to methods not to problems. Maybe not good for internal training. Sarah says that practical case studies are meant to solve this. Future: it may have a wiki structure. Luxembourg agrees that it has improved, but the introduction and regulation are weak. Introduction too fast and regulation needs update. Jean-Marc will send updates. Cross references are in Dutch and need to be corrected.

Eric has contact with Wiley. Comparisons with already published book: this should be more for NSI people for understanding the issues. Clarify who is the target. The conclusion is that Anco will find the printer and try to publish it. Minor changes will be made.

Guidelines on output checking: what to do there? Felix thinks that we need a reference to the work done in the handbook but the guidelines should be kept different.

Methodology in frequency tables: still quite challenging to apply methods. We still need guidelines on when a table can present a problem.

As for the Wiki Eurostat sees different types of wiki: on Platform and contents. For the second type Eurostat has already Statistics explained in Eurostat web site.

Meeting closes at 13.00