

ESSnet on Statistical Disclosure Control

Mid-term evaluation

The objective of this mid-term evaluation is to study the work delivered in the first year of the ESSnet on SDC and to give recommendations for the remaining part of the work.

The evaluation is carried out by the Evaluation Panel composed of Eurostat representatives involved in confidentiality issues and interested in the results of the ESSnet.

I. General comments

The Evaluation Panel appreciates the work done in the first year of the ESSnet on SDC. The deliverables provided are of a great value; however the most important, practical part of the work remains still to be done in 2009. Based on the deliverables received, the panel understands that the theoretical development have predominated in the first year. It is essential that the implementation side of the project is developed appropriately in the second year of the ESSnet. For instance, we expect very much that testing and implementation of the proposed methodology and eventually formulation of suggestions and guidelines to be followed are developed. The general objective of the ESSnet is to share the knowledge and methods that can be useful and applicable for the ESS. It is therefore critical that the theoretical work delivered so far is turned into a more practical knowledge and tools available for all.

The deliverables received are presented as isolated pieces of work. At mid term, it would have been beneficial to present an appraisal of the contribution of the work to the overarching objective of the ESSnet and to benchmark the deliverables against the problems raised and collected in the Inventory. Final report should be more elaborated in this respect.

II. Specific comments

Task 1 Dissemination of CENEX results

1a Improvement of the Handbook

The improvements brought to the CENEX version of the Handbook by the current ESSnet could not be appreciated fully by the panel because they are not documented. The comments received from the Member States are only available for members of the project (or users

having access to the internal website protected by password). A fair evaluation of the results would have been easier if the new version of the Handbook had been provided with tracked changes, highlighting the modifications made under the ESSnet project. It is expected that the final version of the Handbook is completed by case studies and recommendations, which the panel considers as major outcomes of the ESSnet work as far as the improvement of the Handbook is concerned (inventory of comments and improvements should be enclosed to the final report).

It would also be interesting to have an executive summary of the Handbook, i.e. a shorter version of it and limited to major definitions and methods with examples. This could be very useful for those readers who are not very interested in details but who would like to have a general overview of the SDC methods.

1b Follow up of the CENEX inventory

The course organized in Ljubljana is a successful initiative and it seems that participants are very satisfied with this training. Nevertheless the other tasks of the item 1b1 "Needs of less experienced countries" should have received more attention. The objective here was to contact these countries directly and to address their problems through for example on site visits and consultations.

The report on CENEX inventory is rather schematic and refers mainly to numbers. This is according to the opinion of the Evaluation Panel just a starting point for more detailed analysis with more recommendations for solving various problems of NSIs with regard to SDC.

It is essential in this context to mention that the results of the inventory should be considered in almost all ESSnet tasks allowing for identification of major problems and structural approach to them (grouping of countries according to different aspects of SDC for example and addressing some solutions to those blocks of countries).

Task 2 Making SDC-tools better usable by NSIs

Firstly, the tasks under this part of ESSnet should provide an answer on how to make SDC better usable in NSIs. The 3 tasks under this heading have been defined in order to find the solutions in those cases where MS did refrain from using Argus software. What is expected here is the inventory of problems NSIs and Eurostat have with SDC tools and proposals on how these problems could be solved. We hope to receive this as an output to these tasks at the end of the project.

Secondly, in order to widen the use of the software developed/improved during the project it would be important to have an overview of the problems encountered by the partner NSIs while testing the software in their environment. It is important to know more what problems occurred and how they can be tackled during the timeframe of the project. We consider this a important step in the implementation of the project.

2a Widening the scope of input and output data formats

The SPSS format was already included as the input format to ARGUS in the CENEX project. Please clarify what is the exact contribution of this project to that task.

2b Problems with IT infrastructure

The results of the testing done are essential here. The description of the testing with main findings and results will be needed at the end of the project.

2c Sustainability

The report is very interesting and provides very good ground for the future development of software solutions.

Given the deliverable is already available, it is important that the various options envisaged are assessed by the consortium rapidly so that the practical solutions to be developed in the second half of the project are in line with a new strategy for sustainable IT tools.

On the technical side, despite open source solutions are appealing the dependency on Xpress or CPLEX license needs to be addressed in this context and developed in the final report.

It would be very much appreciated to see the results of the reflection indicating clearly the way to follow in order to achieve the sustainability of Argus software and thus to make it fully usable by NSIs.

Task 3 Output checking

Guidelines

The results of the questionnaire and guidelines for output checking to be applied by MS are expected at the end of year 2.

Automatic output checking

Solutions for automatic protection of outputs are highly desirable. The jackknife method proposed seems interesting. It is important that the potentialities of the method is assessed by the whole consortium in parallel with the testing expected in year 2 and thus, if testing proves to be positive, the method is integrated in the set of standard tools to be made available at the end of the project.

Task 4 Communication, web and FAQ

The website is user friendly and constitutes a rich source of information on SDC.

The FAQ section is expected to be more developed in the second year. The valuable comments of users of Argus (available at internal website) and comments to the Handbook might be included here.

Task 5 Improvement of software for microdata

The list of improvements to be applied in the software is expected at the end of the project. For each improvement the following aspects should be covered: link to the inventory (explaining the need for improvement), synopsis of the methodological approach proposed (highlighting new features) with the results of its testing and finally its detailed technical specification.

5a Big surveys

5a1 Standardized anonymisation of microdata sets

No deliverables mentioned here.

5a2 Blocking methods

The deliverable provided is interesting. We would like to see the practical implementation integrated in standard tools in the second year if results are confirmed.

5b Alternative risk models or Microdata dissemination strategies

The deliverables provided here constitute an interesting theoretical basis for developing anonymisation of microdata. Practical guidelines would be welcome.

A general approach to different microdata dissemination strategies (PUF, research license, restricted license, ...) are also expected here. Specifically, a comparative study of different type of release of microdata (with the emphasis on data utility) with recommendation for respective anonymisation strategies would be a high value for the ESS.

5c Record linkage

The software for distance-based record linkage was already produced within the CASC project and the contribution of this project to this task is not yet clear.

Task 6 Improvement of software for tabular data

6a Protection of frequency tables and tables with negative values

The problem of negative values, together with problem of linked tables, constitute the essential problems to be solved in order to make the SDC tools better usable by NSIs and Eurostat. It is thus essential to propose clear and adequate solutions to these problems. These solutions should be based on thorough testing by project partners (the description of this testing should be also available at the end of the project).

The proposed improvements should be clearly defined and new software features described in details and implemented. The documentation for the implementation of the proposed solutions should be adapted to end users.

III. Conclusions

The objective of this mid-term evaluation was to study the work delivered in the first year of the ESSnet on SDC and to give recommendations for the remaining part of the work.

As mentioned above the work done in the first year of the ESSnet on SDC is appreciated. The SDC project has contributed to emulation of SDC development activities in the ESS and to raise more widely the awareness on SDC issues. However the second year should be more concentrated on practical implementation. The testing of the solutions proposed by ESSnet partners in their working environments is essential to check the adequacy and viability of those solutions and to allow for future implementation.