



Georges ZANGERLE  
C3 - Structural Business Statistics  
Tel.: +352 247-84242  
georges.zangerle@statec.etat.lu

## **ESSNet on SDC, June 5-6 2008, Wiesbaden**

In ESSNet on SDC, Luxembourg has been assigned to the following tasks:

- the reviewing/commenting of the handbook on SDC ;
- testing of the ( $\tau$ -Argus) software (with respect to usability and IT-infrastructure).

This work has been performed from a user's perspective, more particularly from the perspective of the Structural Business Statistics (SBS) confidentiality framework.

### **Comments on $\tau$ -Argus**

*(software version 3.2.0)*

So far, our tests have been conducted in order to check the automation qualities of  $\tau$ -Argus.

#### **Recoding a hierarchical spanning variable and then applying an a-priori file**

Currently, there is a problem when applying the a-priori file information on a manually (i.e. interactive mode) recoded table. Such information will not be processed correctly by  $\tau$ -Argus in case of a hierarchical spanning variable. The procedure that produces this error would first consist in recoding the hierarchical spanning variable, given that the level of detail to be delivered is not homogenous, and then to apply the a-priori information before calculating secondary suppressions. As a result, the a-priori information would be taken into account for some recoded cells but not for others, even if the latter were not recoded.

Our workaround to this problem has been to perform the recoding directly in the input file. However, this has not completely solved the problem, as we still wanted to preserve the benefits of working with a hierarchical spanning variable in  $\tau$ -Argus. Consequently, additional protection flags were added to the a-priori file in order to avoid  $\tau$ -Argus from suppressing cells which would otherwise be suppressed via secondary suppression but which would have no protecting value because they were not deliverable in the framework of the transmission programme.

#### **Handling authorization waivers – missing feature ?**

While  $\tau$ -Argus offers the possibility to handle the rare case where respondents ask for protection in the framework of foreign trade (cf. request rule), there seems to be no possibility for directly addressing the opposite case where respondents authorize publication of an aggregate even if this publication might cause a disclosure risk.

In the SDC handbook, reformulating the concentration rule is mentioned as a solution. However, it is not clear how such an individual (and manual ?) treatment could be put into practice with  $\tau$ -Argus, in particular when dealing with a relatively large number of authorization waivers.

### Handling periodical data

Prior year suppression patterns are a major constraint in the statistical disclosure control process for SBS data.

According to the guidance on periodical data on page 152 of the SDC handbook, an option might be to *'copy' the suppression pattern of the previous period and add some (primary and secondary) suppressions for cells which become sensitive in the current period*. We wonder how this might be put into practice with  $\tau$ -Argus.

In the handbook, a second option is mentioned, i.e. *assigning low costs to prior-period secondary suppressions when assigning secondary suppressions to the current tables*. For this, the manual also mentions the a-priori file, which may be helpful :

*[...] Specifying all this information manually during a  $\tau$ -ARGUS-run for a large table can be rather cumbersome and error prone. In the batch-version of  $\tau$ -ARGUS it is even impossible to do this. **So the a-priory file has been introduced**. In this file the user can specify for each cell (by giving the values of the spanning variables) :*

*[...]*

*- specify a new value of the cost-function for a cell; this can be either high or low. This will influence the chance of a cell as a candidate for secondary suppression. (cf. pp.151-152)*

*Also the cost-function can be changed here for a cell. This will make the cell more likely to become secondary cell suppression, when the value is low, or less likely when the value is high. (cf.  $\tau$ -Argus manual version 3.20 p.32)*

The problem is that "generating" the a-priori file itself can be cumbersome and error prone depending on the size of the table.

Furthermore, the modification of the cost function in the a-priori file is not documented in an easily understandable manner. More guidance or examples on the choice of a decent value, i.e. having both a protecting "value" and a high probability for secondary suppression, would be useful. Another workaround for this may be to perform a first confidentiality pattern calculation with  $\tau$ -Argus and then to compare the output with the confidentiality flags applied for the previous year. Then, on a case by case basis, the a-priory file is fed with the necessary protection flags to influence the secondary suppression pattern of the current year. However, this is also a cumbersome procedure, while not being optimal as well.

## Preliminary comments on the Handbook on Statistical Disclosure Control

(handbook version 1.01)

Luxembourg's contribution schedule on this project is set for the year 2009. However, for the discussions during the project meeting in Wiesbaden, a preliminary review of chapters 1 (Introduction) and 4 (Magnitude Tabular Data) has been performed.

### The handbook's readership

*'This handbook aims to provide **technical** guidance on statistical disclosure control **for NSIs** on how to approach this problem of balancing the need to provide users with statistical outputs and the need to protect the confidentiality of respondents.'* (cf. p.6)

Sometimes, it is not clear whether or not the handbook is addressed to a readership that possesses a fundamental knowledge of mathematical concepts. There are some sections in the handbook (e.g. 'How to process linked tables' pp.150-151) where discussions and problems are mainly described in a mathematical manner. If the readership is supposed to be specialized in mathematics, then such a way of describing SDC can naturally be deemed adequate.

However, if the target readership should also include people who have a less formal education in mathematics, this way of describing SDC problems might not easily contribute to understandability and readability. Obviously, one cannot really avoid mathematics in the context of SDC, in particular not if an in-depth understanding is needed, nevertheless, it would be recommendable :

- either to "warn" the reader in advance whenever the language of a section becomes deeply mathematical and when that kind of formal discussion is not entirely necessary for the overall understanding of a basic idea - such a section could be preceded by a distinctive symbol or by the words "formal discussion" ;
- or to summarize the basic ideas for the general readership at the end of each section.

A combination of both may also be appropriate. In the end, it all depends on the target readership, which we should define more explicitly at the beginning of the handbook. That said, the handbook is all in all very well written but if it is to be used as a reference guide, some sections might have to be reconsidered.

### The $\tau$ -Argus manual and the handbook on SDC

It may appear confusing to some readers that a lot of elements normally belonging to the scope of the  $\tau$ -Argus manual actually made it into the SDC handbook (e.g. 'A-priory file' pp.151-152, etc.). The purpose of this is certainly to illustrate how theory can be put into practice with  $\tau$ -Argus and thus to promote the latter as a solution.

Nevertheless, one should be careful not to mix up the  $\tau$ -Argus manual with the SDC handbook. This unnecessarily increases the volume of the SDC handbook and it generates redundancies with the  $\tau$ -Argus manual. Furthermore, similar redundancies with the SDC handbook can be found in the  $\tau$ -Argus manual as well, which makes it difficult to work with both the SDC handbook and the  $\tau$ -Argus manual, for it is not always clear where the main pieces of information are supposed to be. Eventually, one ends up consulting both manuals to get the complete picture of a given SDC problem or a feature of the software.

Therefore, it should be determined if, when and how such interferences with the  $\tau$ -Argus manual can happen. To increase readability and to stress the purpose of the SDC handbook, specific information relating to  $\tau$ -Argus could be transferred into devoted e.g. "How to do this with  $\tau$ -Argus"-boxes/sections below the actual information or it could be marked with a distinctive symbol next to the paragraphs where applicable.