

Working Paper
ENGLISH ONLY

**UNITED NATIONS ECONOMIC COMMISSION
FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN
UNION (EUROSTAT)**

Joint UNECE/Eurostat work session on statistical data confidentiality
(Ottawa, Canada, 28-30 October 2013)

Topic (vii): Software developments and demonstrations

Open source software Argus

Prepared by Peter-Paul de Wolf, Statistics Netherlands

Open source software ARGUS

Peter-Paul de Wolf*

* Statistics Netherlands, P.O. Box 24500, 2490 HA The Hague, The Netherlands, pp.dewolf@cbs.nl¹

Abstract: In recent years, the ARGUS-software has become a valuable tool to apply Statistical Disclosure Control methods. Many European statistical institutes use the current Windows based version of the software. Recently a European project was started to port the current version to an Open Source version that can also be used on other platforms. In this paper we will describe the planning of this porting-project and show some preliminary results.

Keywords: Statistical Disclosure Control, τ -ARGUS, μ -ARGUS, software, Open Source

1 Introduction

The main task of statistical offices is to produce and publish statistical information about society. Publishing a vast amount of detailed information has the risk of disclosing sensitive individual information both on individual persons as well as on economic entities. The growth of the internet has made people aware of consequences of the concept ‘big brother is watching you’ and hence privacy issues have become more and more a concern. The Statistical Disclosure Control theory is used to solve the problem of how to publish and release as much detail in these data as possible without disclosing individual information (Hundepool et al, 2012).

Statistical offices need to find the right balance between the need to inform society as well as possible, on the one hand, and the need to safeguard the privacy of the respondents on the other. Traditionally, tailor-made software has been developed by several statistical offices as an aid to this task.

Combining efforts of several European institutes and being partially financed by Eurostat, general purpose SDC software was developed since the late 1990’s: the ARGUS software. Several European projects helped to finance the development of the software: e.g., Computational Aspects of Statistical Confidentiality (CASC), Centre of Excellence SDC (CENEX-SDC), ESSnet-SDC and ESSnet-harmonisation.

The main developers and maintainers of the ARGUS software are/were Statistics Netherlands’ employees. Recently, one of them had to retire and he was not replaced by another employee. This is one of the reasons to decide to launch a project, mainly

¹ The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

funded by Eurostat², to port the ARGUS software to an Open Source version. That way, ‘the Open Source community’ can more easily take part in the further development of this software and thus reduce the maintenance burden of Statistics Netherlands. At the same time, using platform independent code, at the end of the project it will be possible to compile the software for Windows as well as Unix/Linux environments.

In section 2 we will summarise the project and in section 3 we will show some preliminary results. During the UNECE/Eurostat work session of October 2013 we will demonstrate the at that time available version of the Open Source ARGUS software.

2 The project

The ARGUS software consists of two separate programs: μ -ARGUS to be used to protect microdata and τ -ARGUS to be used to protect tabular data. In each interface it is possible to apply various SDC methods. These SDC methods are developed by several European institutes: NSIs as well as universities.

The current version of the ARGUS software can be found at <http://neon.vb.cbs.nl/casc>. At this website, the user’s manuals of μ -ARGUS (Hundepool et al., 2010) and τ -ARGUS (Hundepool et al., 2011) can be found as well.

The interfaces are compiled using Microsoft Visual Studio and can be used from within a Windows environment. Some of the SDC methods that can be called from these interfaces were provided as windows binaries/executables by the original programmers directly. SDC methods were implemented in C/C++, R, Fortran and Delphi.

Even though the current ARGUS software has always been provided as binaries/executables, most source code is available upon request. However, the new project to port ARGUS to Open Source, should make it more easy to obtain the source code and to contribute to future developments.

2.1 Outline of the project

The original ARGUS software consists of a graphical user interface (GUI), written in Visual Basic, that connects several SDC methods and controls the flow of the program. The different SDC methods that can be used from within the GUI, are mostly provided as shared libraries (DLL’s in Windows) or separate executables. These additional SDC methods are written in C/C++, Fortran, Pascal and R.

² The contents of the current paper are the sole responsibility of the author. The European Commission is not responsible for any use that may be made of the information contained in the current paper.

In the Open Source project, the main objective is to port the current version of ARGUS while maintaining the current functionality as much as possible. The GUI will be ported to ‘standard’³ Java and some of the additional shared libraries with SDC methods will be rewritten to ‘standard’ C/C++.

Moreover, in the Open Source version it should also be possible to use free Open Source LP-solvers in τ -ARGUS, next to the commercial LP-solvers Xpress and CPLEX. To that end, the cell suppression code (written by J.J. Salazar’s team) underlying the Modular and the Optimal approaches, had to be rewritten such that Open Source LP-solvers can be used. The revised code, using SCIP with Clp as LP-solver, was written as part of another European project called Data without Boundaries⁴. As a consequence, the Modular and Optimal approaches have to be rewritten to ‘standard’ C/C++ as well.

As far as μ -ARGUS is concerned, many new SDC methods for the protection of microdata files as Public Use Files (PUFs) or Microdata Under Contract (MUCs) are currently being developed in R. Actually, another software package to apply SDC methods to microdata (called sdcMicro) is developed in R by M. Templ, see <http://cran.r-project.org/web/packages/sdcMicro/index.html>. The same group has provided a GUI to use their sdcMicro routines from within R as well. During the Open Source ARGUS project we will try to collaborate with Matthias Templ and develop a μ -ARGUS version that can also use the R-implementation of some of the SDC methods available in sdcMicro.

When we have the ARGUS software as Open Source, it does not necessarily mean that the Open Source community will automatically ensure the future existence and further development of the ARGUS software. Moreover, since we want this software to be the standard software for statistical disclosure control at National Statistical Offices (NSIs), the governance structure should be defined properly.

That is, there should be a ‘board’ of experts that will decide which developments should go into the official releases of ARGUS. This board will most likely be a subgroup of the Eurostat Expert Group on Statistical Disclosure Control and act under the EUPL (European Union Public Licence) model for Open Source projects. However, setting up such a governance structure is not part of this project.

Additionally, for future development and maintenance of the ARGUS software, financing is most likely needed. Hopefully financing will come from NSIs as well as Eurostat, just like the previous European projects that have led to the current status of the software.

³ With ‘standard’ code we mean platform independent code, i.e., code that can be compiled under e.g., Windows as well as Unix/Linux.

⁴ DwB (Data without Boundaries) is a 4-year FP7 project funded by the EU under the Grant Agreement N° 262608.

2.2 Planning of the project

The project started in December 2012 and lasts until December 2014. By that time a first Open Source version of the ARGUS software will become available to the SDC community. As τ -ARGUS is the most used part of the software, this part will be ported in the first place.

In June 2014, a beta-version of the Open Source version of τ -ARGUS will be made available to be tested by any interested NSI. During the UNECE/Eurostat work session on SDC of October 2013 we will show the current state of software.

The current status of the software *at the time of writing this paper* is that it is possible from within the new GUI to import microdata along with the associated metadata, change the metadata if necessary and save the changes, specify the table to be protected and display the table under consideration. It is not yet possible to apply any suppression method.

A first ‘standardised C/C++’ version of the Modular and Optimal approaches for cell suppression is working for Xpress and CPLEX. However, this has not yet been tested from within the new GUI.

3 Tentative preview

Note: The figures we refer to in this section are put together at the end of this paper. The ‘old’ version of τ -ARGUS will be referred to as τ -ARGUS, whereas the preliminary Open Source version will be referred to as pOS τ -ARGUS.

Up until now the project has been concentrating on the GUI of the τ -ARGUS software. Therefore, in this section we will only show some preliminary screenshots of the Open Source τ -ARGUS. Note that these screenshots are indeed preliminary, in the sense that they may still undergo some changes before the project has finished.

We tried to stay as closely as possible to the look and feel of the ‘old’ τ -ARGUS. However, whenever we felt that the GUI would benefit from certain changes, we adjusted the GUI accordingly.

The Specify Table window has not really changed, save some minor lay-out issues. See Figure 1 for an example of a Specify Table in the new Open Source version.

The biggest change in the GUI is the ‘main screen’ of τ -ARGUS. See Figure 2 for an example of the ‘old’ situation: whenever a table had been specified, the main screen would display the number of unsafe combinations in every dimension. Since this information is hardly used, we will no longer support this information in the Open Source version.

In the ‘old’ version, the table that needed SDC methods to be applied, was displayed in a separate window. In the new Open Source version, the table under consideration will

be displayed in the main screen. See Figure 3 and Figure 4. Note that in the new Open Source version, the Holding-checkbox and the Request-field seem to be absent. This is because the microdata that was used to make this particular table did not include holding info nor a request variable and the new GUI does not show any options that cannot be used with the table at hand.

References

- Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte Nordholt, E., Spicer, K. and de Wolf, P.P. (2012), *Statistical Disclosure Control*. Wiley Series in Survey Methodology, Wiley, Chichester, UK.
- Hundepool, A., van de Wetering, A., Ramaswamy, R., Franconi, L., Polettini, S., Capobianchi, A., de Wolf, P.P., Domingo, J., Torra, V., and Giessing S. (2010), *μ -ARGUS, user's manual, version 4.3*, The Hague, The Netherlands: Statistics Netherlands.
- Hundepool, A., van de Wetering, A., Ramaswamy, R., de Wolf, P.P., Giessing, S., Fischetti, M., Salazar, J., Castro, J. and Lowthian, P. (2011), *τ -ARGUS, user's manual, version 3.5*, The Hague, The Netherlands: Statistics Netherlands.

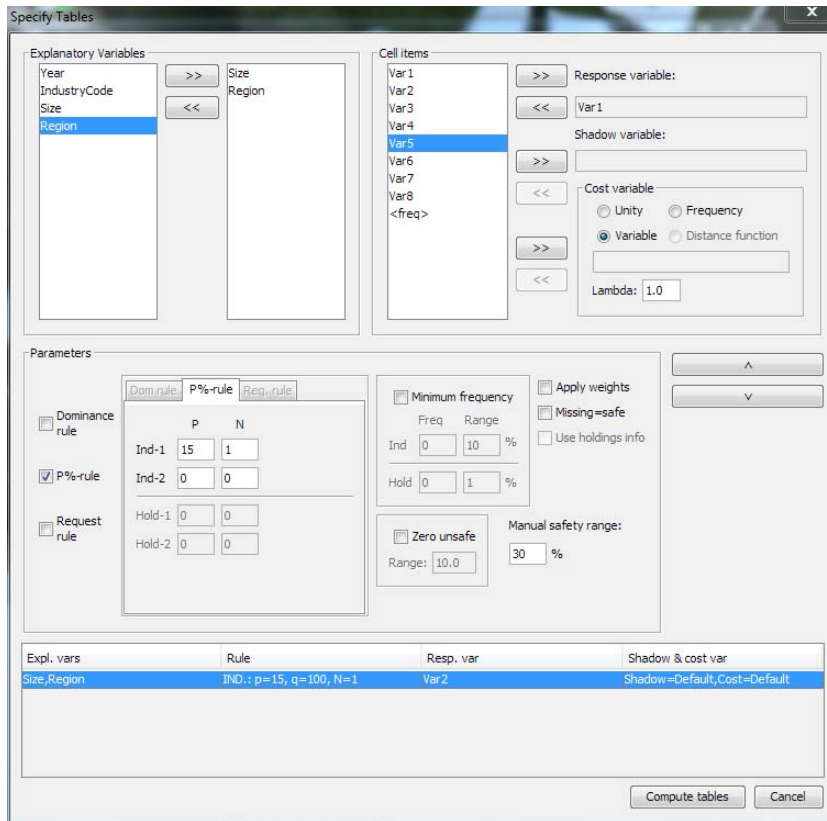


Figure 1: Specify Tables window of pOS τ -ARGUS.

Variable	dim 1	dim 2
Size	0	12
Region	0	12

Code	Label	Freq	dim 1	dim 2
	Total	42723	0	0
-2		9	0	5
-4		5	0	6
-5		20002	0	0
-6		8831	0	0
-7		5498	0	0
-8		4594	0	0
-9		3779	0	1
99		5	0	0

Figure 2: Main screen of τ -ARGUS

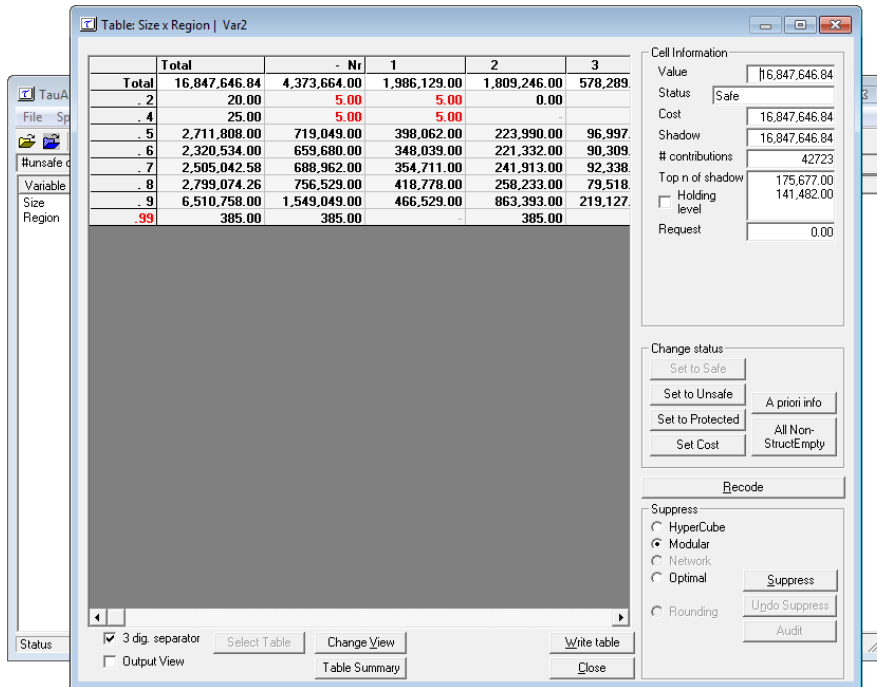


Figure 3: Table display window of τ -ARGUS.

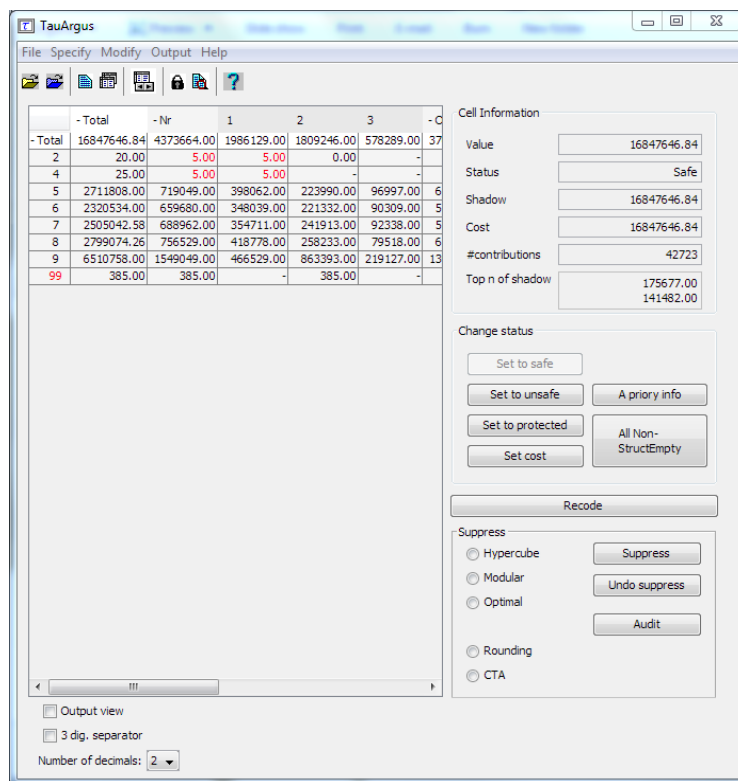


Figure 4: Main screen with table display of pOS τ -ARGUS.