



Evaluation and looking back

1. Introduction

Statistical Disclosure Control is a topic of growing importance in Europe. The need of enhancing our methods on SDC is a common need in many countries in Europe. However the number of researchers active in this field is rather small, down to one or two per country. So several researchers are working in very isolated places throughout Europe. Each working with only limited resources. Although they are hard working people, this is not a very efficient way of working. This awareness has been the basis of the cooperation in Europe.

2. Starting point

The CASC project was not the first project in this field. In the 4th Framework project we have seen already the SDC project, a consortium of partners in The Netherlands, United Kingdom and Italy. This project aimed more at the research side of Statistical Disclosure Control but already the first versions of the ARGUS software emerged from this project. Positive side effects of this project were the growing awareness of some major project partners that European cooperation was a very good way. This led to long and intense discussions on the proposal of a new project in the 5th Framework. It was a common feeling between the partners that we should aim at practical results, leading to methods that could be easily used in the daily practice in the statistical institutes and similar institutes. So the software developments i.e. μ -ARGUS and τ -ARGUS became the key issue on this project. We accepted only research that could lead to practical extension on ARGUS. Eventually the CASC-project was submitted and after some discussions with Eurostat and a successful negotiation phase this project was approved. January 1st 2001 became the official starting point for the CASC-project, a three year project, which in the end was extended with six month only to allow us to organise a final meeting/conference under nicer circumstances in Barcelona.

3. Major events

During the project a few major events can be memorised. During the first UN-ECE worksession on Statistical Confidentiality in Thessaloniki, March 1999, a few potential project partners came together and discussed very prematurely the setup of a project. But the spirit was born there already.

Two years later at the second UN-ECE worksession on Statistical Confidentiality in Skopje we were happy to organise already our first CASC project meeting. Besides that several CASC-partners presented an outline of their plans in the CASC project. Despite of the instable political situation in Skopje we can say that we gave birth to a good project there.

After Skopje we were able to present through several presentations at the Eurostat organised NTTS conference in Crete. The CASC-project filled complete session there. Throughout the whole CASC-project we have used these conferences as an option to meet each other and have project meetings.

We also participated in the AMRADS project, aiming at the dissemination of results of the various framework projects. As a result of this Josep Domingo-Ferrer organised a workshop "SDC: From Theory to Practice" in Luxembourg in December 2001. This also led to the Springer publication "SDC: From Theory to Practice", thanks to the work of Josep Domingo-Ferrer.

In 2002 we organised an 'internal' CASC-conference in Plymouth. The aim of this meeting was to bring the large number of partners a bit more together and to discuss internally all the aspects of the projects. One afternoon of this meeting was devoted to a workshop on mathematical models in optimisation by JJ Salazar. Very useful for the more statistical orientated partners in this project. An other good reason for organising this meeting was the lack of other meetings/conferences to use as a 'meeting points in this project.

In 2003 there was a new worksession on SDC, this time a joint effort of the UN/ECE and Eurostat. This worksession took place in Luxembourg and led to a publication in the Eurostat series "Monographs in Official Statistics". As usual many CASC-related papers were presented there.

In August 2003 the two-yearly ISI-session took place in Berlin. At this major event we organised two complete sessions to give all the CASC-partners the opportunity to present to the whole statistical world the results of the CASC project.

Finally in May 2004 we organised an Adieu-CASC conference in Barcelona. Thanks to the efforts of the local organiser (Josep Domingo-Ferrer and Vicenc Torra) we again succeeded in publishing a Springer Volume ("Privacy in Statistical databases 2004").

4. Results

We will not elaborate on all the result achieved during the CASC-project, as they have been discussed at other levels already. We will only summarise the major results here. As the CASC project has been designed to achieve practical results the

most visible outcome of this project is the twin ARGUS packages: μ -ARGUS for microdata and τ -ARGUS for tabular data. 'Hidden' in these packages contributions of many CASC-partners can be found.

For *microdata* we have investigated methods like micro-aggregation, rank swapping, Sullivan's making method, qualitative microaggregation and Post-Randomisation. Besides this we have developed further the Franconi-Benedetti risk models to assess the disclosure risk per record and per household. This all has been build in μ -ARGUS in addition to the Dutch (simple) risk approach. At a certain stage in the project we have drawn the conclusion that Sullivan's masking method was theoretically interesting, but an efficient implementation that also could be used in daily practice was too much. So we discontinued the research here and devoted the remaining part of the project to the start of the development of record-linkage software that eventually can be an extension of μ -ARGUS.

On the *tabular* side we have improved the German Hypercube method and paid a lot of attention to optimisation methods. A serious draw-back of the first version of τ -ARGUS was the lack of options for hierarchical tables. Many statistical tables show hierarchies in the spanning variables. For the secondary cell suppression this complicates the solution very much. The resulting optimisation methods become more and more complex. The classical optimisation models have been extended, enabling us now to protect hierarchical tables. As it was to be expected that for large tables these methods might become computationally very demanding, we also have implemented a modular approximation of these optimisation models.

Network based solutions have also drawn attention for secondary cell suppression. Up till now these networks could be use only for low-dimensional non-hierarchical tables. We have shown that some extensions are possible and a very efficient implementation for 2-dim tables with one hierarchy has been included in τ -ARGUS.

Also auditing software has been developed to test the quality of suppression patterns.

Testing is always in important job when developing software. Experiences from previous projects have learned that you have to take this matter seriously. You cannot only rely on good friends, who will do some testing for you on a rainy afternoon. So a special work package testing had been added to the project and several partners were only participating in the project for the testing. This approach gave us much more control over the test-efforts and this has proven to be a serious advantage.

All the results of CASC can be found on the CASC website:

<http://neon.vb.cbs.nl/casc/default.htm>

A significant step forwards has been made in the research and development of Statistical Disclosure Control, giving Europe a leading role in this field. But the world is changing continuously, leading to new challenges in the future. If possible the CASC team will try to keep together and search for new opportunities to bring solutions for the new challenges in this field.