SAVOIR POUR AGIR

statec
LUXEMBOURG

Luxembourg, October 29, 2009

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

ESSNet on SDC 2008-2009 – LU final report 2009

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

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

This work has been performed from a user's perspective.

COMMENTS ON τ-ARGUS 2

TAU ITEM 1. [SOLVED] RECODING A HIERARCHICAL SPANNING VARIABLE AND A PRIORI FILE 2
TAU ITEM 2. HANDLING WAIVERS – MISSING FEATURE ? 3
TAU ITEM 3. [PLEASE REFER TO HB ITEM 12] HANDLING PERIODICAL DATA 3
TAU ITEM 4. [SOLVED] RUN-TIME ERRORS WITH THE VB INSTALLATION PROCEDURE FOR τ-ARGUS 3
TAU ITEM 5. [SOLVED] "SPECIFY TABLES" SUPPRESSES ANY PREVIOUSLY CALCULATED TABLES 4
TAU ITEM 6. [SOLVED] SAMPLING WEIGHTS 4
TAU ITEM 7. [ABANDONED] PROTECTION REQUESTS AS A SEPARATE FILE 5
TAU ITEM 8. RELATIVE SENSITIVITY VALUES AS SUPPLEMENTARY CELL INFORMATION 5
TAU ITEM 9. ABNORMAL TERMINATION IN BATCH MODE 6
TAU ITEM 10. [SOLVED] MISLEADING BATCH COMMAND SYNTAX DESCRIPTION FOR THE COST FUNCTION 6
TAU ITEM 11. ERRONEOUS τ-ARGUS GENERATED BATCH FILE CONTAINING A LAMBDA 7
TAU ITEM 12. [SOLVED] A PRIORI BATCH COMMAND NAME INCONSISTENCY 7
TAU ITEM 13. [SOLVED] ISSUES RELATING TO THE WRITETABLE BATCH COMMAND 8
TAU ITEM 14. LOG FILE LINE NUMBERS 8
TAU ITEM 15. [SOLVED] LOGBOOK BATCH COMMAND NOT MENTIONED IN THE τ-ARGUS MANUAL 9

COMMENTS ON THE HANDBOOK ON STATISTICAL DISCLOSURE CONTROL 10

HB ITEM 1. [SOLVED] THE HANDBOOK'S READERSHIP 10
HB ITEM 2. THE τ-ARGUS MANUAL AND THE HANDBOOK ON SDC 11
HB ITEM 3. [PLEASE REFER TO HB ITEM 10] WAIVERS 11
HB ITEM 4. CHAPTER 1 – MINOR ISSUES 12
HB ITEM 5. CHAPTER 2 – ETHICAL CODES – CONTENT-RELATED ISSUES 12
HB ITEM 6. CHAPTER 2 – LAWS – CONTENT-RELATED ISSUES 13
HB ITEM 7. CHAPTER 2 – MINOR ISSUES 14
HB ITEM 8. CHAPTER 4 – DISCUSSION ON THE (N,K)-DOMINANCE RULE MAY BE INCOMPLETE 14
HB ITEM 9. CHAPTER 4 – HOLDINGS AND THEIR EXTENT IN STRUCTURAL BUSINESS STATISTICS 15
HB ITEM 10. CHAPTER 4 – WAIVERS 16
HB ITEM 11. CHAPTER 4 – 'HOW TO PROCESS LINKED TABLES' NOT UNDERSTANDABLE ANYMORE 18
HB ITEM 12. CHAPTER 4 – HANDLING PERIODICAL DATA 18
HB ITEM 13. CHAPTER 4 – MINOR ISSUES 20

Comments on τ-Argus

We have tested τ-Argus and its corresponding manual, from a perspective of SBS, ICT and CIS. Additionally, some areas have been tested independently from that context. The final report 2009 covers the software version 3.3.0 build 17, which is significantly more stable than 3.3.0 build 2. Whenever a previously reported issue has been solved, we have tagged the title of the relevant item.

As an overall remark, we can say that many issues or feature requests previously reported have been satisfactorily addressed in the most recent build.

tau Item 1. [solved] Recoding a hierarchical spanning variable and a priori file

Issue reference(s)	τ-Argus software versions 3.3 build 2 and 3.2 build 7
Reported	June 2008
Updates	October 2009, July 2009, November 2008

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 τ-Argus in case of a hierarchical spanning variable. The procedure that produces this error first consists 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 is taken into account for some recoded cells but not for others, even if the latter have not been 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 τ-Argus. Consequently, additional protection flags were added to the a priori file in order to avoid τ-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.

Update in November 2008: This could potentially contribute to more serious issues in the case of complex tables, such as the SBS tables by employment size classes, as the protection flags strongly limit τ-Argus in its capacity to find a feasible solution.

Update in July 2009: The issue still applies to the τ-Argus software version 3.3. build 2. For SBS, this is a critical issue which needs to be addressed. Alternatively, a decent workaround needs to be implemented in the software or at least be described in the manual.

Update in October 2009: The issue has been solved in τ-Argus software version 3.3. build 17.

tau Item 2. Handling waivers – missing feature ?

Issue reference(s)	τ-Argus software versions 3.3 build 2 and 3.2 build 7
Reported	June 2008
Updates	July 2009, November 2008

While τ-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 authorise the publication of an aggregate even if this publication might present a disclosure risk for them.

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

Update in November 2008: Please also refer to the related HB Item 3.

Update in July 2009: The issue still applies to the τ-Argus software version 3.3. build 2.

tau Item 3. [please refer to HB Item 12] Handling periodical data

It seems that this issue cannot be handled only by the software. Therefore, this item has been moved to the handbook section of this report.

tau Item 4. [solved] Run-time errors with the VB installation procedure for τ-Argus

Issue reference(s)	τ-Argus software version 3.2 build 7
Reported	November 2008
Updates	October 2009, July 2009

During the meeting in Wiesbaden in June 2008, some τ-Argus installation problems were reported. When installing build 7 this year [i.e. 2008], our IT department reported that the Visual Basic installation procedure failed due to run-time errors, while the installation went smoothly when using the alternative installation procedure. We can therefore confirm the reports by other participants during the ESSNet on SDC meeting in Wiesbaden 2008. It should be considered to put only the second installation procedure on the SDC website or at least to make it the primary download option.

Update in July 2009: In order to avoid any installation trouble beforehand, we have used the alternative installation procedure. The website still suggests using the VB procedure first.

Update in October 2009: All updates that followed so far only included the alternative installation procedure, which never posed any problems in our IT environment. The issue can therefore be deemed as solved.

tau Item 5. [solved] "Specify tables" suppresses any previously calculated tables

Issue reference(s)	τ-Argus software versions 3.3 build 2 and 3.2 build 7
Reported	November 2008
Updates	October 2009, July 2009

If the user selects "Specify tables" after computing other tables in the same session and then decides to press the "cancel" button in the "Specify tables" dialog box, any previously calculated tables can no longer be accessed by "View tables".

In order to avoid any annoying situations where a user would have to let τ-Argus re-perform the potentially lengthy calculations, we recommend that :

- either there should be a question dialog box of the type "Warning ! This will erase any previously calculated tables, are you sure to continue ?" when "Specify tables" is selected after prior calculation in the same session ;
- or pressing the "cancel" button in the "Specify table"-menu should no longer suppress access to any previously calculated tables.

Update in July 2009: The issue still applies to the τ-Argus software version 3.3. build 2.

Update in October 2009: The issue has been solved in the τ-Argus software version 3.3. build 17.

tau Item 6. [solved] Sampling weights

Issue reference(s)	τ-Argus software versions 3.3 build 2 and 3.2 build 7
Reported	November 2008
Updates	October 2009, July 2009

The sampling weight usually indicates the number of units in the population that are represented by the sampled unit. τ-Argus offers two options for handling sampling weights :

- apply weights : applies the sampling weights for computing the cell totals ;
- apply weights in safety rule : safety rules take into account the sampling weights.

We observed an unusual behaviour with the "apply weights" option in a very precise scenario : a given minimum frequency rule (e.g. 4), less contributors in the sample for a given cell than required to satisfy the minimum frequency rule (e.g. 3), a total sampling weight above the minimum frequency (e.g. 7). In this scenario, the option "apply weights" produces opposite results than when it is not selected and the same results than when "apply weights in safety rule" is selected. In other scenarios, we did not observe this kind of unusual behaviour.

In addition, for increased convenience in interactive mode, it would be useful to actually see the total sampling weight as supplementary cell information in the "View Table" dialog. The possibility to export this piece of information could also be interesting.

Update in July 2009: The issue still applies to the τ-Argus software version 3.3. build 2. The requested feature is also not available.

Update in October 2009: The borderline case issue seems to be solved in Argus software version 3.3. build 17. The feature request has not been addressed but can be deemed as minor, even though it would be nice to have.

tau Item 7. [abandoned] Protection requests as a separate file

Issue reference(s)	τ-Argus software versions 3.3 build 2 and 3.2 build 7
Reported	November 2008
Updates	October 2009, July 2009

When discussing the possibility of including the waivers feature into τ-Argus in June 2008, it was suggested that the list of waivers should be included in another file than the microdata. Given the fact that the number of available waivers is usually minor compared to the total number of respondents, this truly makes sense.

We wonder why this feature should not also be available for the protection requests processed by the request rule. The number of protection requests may be minor in some countries. Consequently, an option to include the protection requests via a separate file would make sense. Of course, such data can be included with the microdata before importing the latter into τ-Argus, but this is also true for data on waivers.

Update in July 2009: The requested feature is not available in the τ-Argus software version 3.3. build 2.

Update in October 2009: After discussions with Anco Hundepol, we understand that this feature request is not feasible. Given the fact that this information could also be included in the microdata file, we have decided to abandon this issue.

tau Item 8. Relative sensitivity values as supplementary cell information

Issue reference(s)	τ-Argus software versions 3.3 build 2 and 3.2 build 7
Reported	November 2008
Updates	July 2009

It would be useful to see the relative cell sensitivity value expressed in terms of the chosen rule, e.g. the actual dominance rate or the actual p% rate, in the "View table" dialog. While not being an essential feature, it actually makes the interactive mode more comfortable and informative.

Update in July 2009: The requested feature is not available in the τ-Argus software version 3.3. build 2.

tau Item 9. Abnormal termination in batch mode

Issue reference(s)	τ-Argus software version 3.3 build 17, 3.3 build 2 and 3.2 build 7
Reported	November 2008
Updates	October 2009, July 2009

When executing a batch file from interactive mode and after encountering an error, τ-Argus terminates. The log file usually reports an error together with the remark "Program abnormally terminated". When a similar error is generated in interactive mode or in case of an error with the <LOGBOOK> batch command[1], a proper dialog box is shown.

We understand that the batch mode is mainly meant to be used as a production mode. Nonetheless, it can also be precious in a preparation phase[2] or when testing the software in the framework of this project, so that using the batch mode in interactive mode becomes inevitable. In such a case, it can be quite annoying and initially confusing when τ-Argus terminates without reporting any errors through an alert dialog box.

To avoid losing the benefits from a fully automated batch mode as it is currently implemented, it might be a good idea to implement a batch command <VERBOSEMODE>. This command, when encountered by the program in the batch file, would tell τ-Argus to show the alert dialog boxes and not to terminate in case of encountering errors. At the end of the batch, the mode could be reset to default, i.e. non verbose mode. Similarly to the <GOINTERACTIVE> command, this command should be ignored when the batch mode is executed from command-line.

Update in July 2009: The issue still applies to the τ-Argus software version 3.3 build 2.

Update in October 2009: The issue still applies to the τ-Argus software version 3.3 build 17.

tau Item 10. [solved] Misleading batch command syntax description for the cost function

Issue reference(s)	τ-Argus manual versions 3.3 (pp. 48 and 49) and 3.2 (pp. 47 and 49)
Reported	November 2008
Updates	July 2009, October 2009

According to the τ-Argus manual [3.2] on page 47, *"if the cost variable is specified either a numerical variable is specified or '1' is chosen for frequency or '2' for unity."* In the example on page 49 [of the manual 3.2], the syntax is explained as follows : //Exp|resp|shadow|cost –1= unit –2 = freq.

[1] in which case τ-Argus also terminates after confirming the dialog "Run-time error 76: Path not found".

[2] e.g. testing the metadata file, setting up a table, testing the batch file prior to use in production, etc.

The second description corresponds to what actually works with the software, i.e. -1 for unity and -2 for frequency. Writing only 1 or 2 will lead τ-Argus to terminate with an error. Consequently, the sentence on page 47 of the manual [3.2] should be adapted.

Update in July 2009: The issue still applies to the τ-Argus manual 3.3.

Update in October 2009: The issue has been solved in the τ-Argus manual 3.3 build 14.

tau Item 11. Erroneous τ-Argus generated batch file containing a lambda

Issue reference(s)	τ-Argus software versions 3.3 build 17, 3.3 build 2, 3.2 build 7
Reported	November 2008
Updates	October 2009, July 2009

When τ-Argus is asked to write a batch file, the lambda is added to the <SAFETYRULE> command. When the same batch file is executed, τ-Argus will terminate with an error. As described in the τ-Argus manual [3.2] on page 47, the lambda should be added to the <SPECIFYTABLE> command.

Update in July 2009: The issue still applies to the τ-Argus software version 3.3 build 2.

Update in October 2009: The issue still applies to the τ-Argus software version 3.3 build 17, even though the lambda is now added to the <SPECIFYTABLE> command. τ-Argus abnormally terminates when fed with the previously written batch file. We consider this as a minor issue.

tau Item 12. [solved] A priori batch command name inconsistency

Issue reference(s)	τ-Argus software version 3.2 build 7 τ-Argus manual 3.2, page 48
Reported	November 2008
Updates	July 2009, October 2009

In the τ-Argus manual [3.2] on page 48, the a priori batch command is referred to as <APRIORI> whereas in the software it is named <APRIORY>. This should be fixed either in the manual or in the software.

Update in July 2009: The issue has been solved in the software version 3.3 build 2.

tau Item 13. [solved] Issues relating to the Writetable batch command

Issue reference(s)	τ-Argus software versions 3.3 build 2 and 3.2 build 7 τ-Argus manual versions 3.3 (p.49) and 3.2 (p.48)
Reported	November 2008
Updates	July 2009, October 2009

The batch command <WRITETABLE> is documented in the manual by explaining the meaning of its parameters and through several examples. However, there is no longer any explicit description of its syntax. The latter used to be available in version 3.1 of the manual (on page 48) : (TabNo,P1,P2,Filename). This description should be put back into the manual.

Furthermore, it would be useful to know whether or not a default value for P2, i.e. the second parameter, has to be provided if the latter is not used in the context of the chosen format.

Finally, the SBS format does not seem to have been implemented in the batch command. τ-Argus in its current version produces an "illegal batch command" error for this option.

Update in July 2009:

- <WRITETABLE> documentation (syntax, default value for P2): the issue still applies to the τ-Argus manual 3.3;
- SBS format implementation in the batch command <WRITETABLE>: when attempting to use the batch command WRITETABLE (1,4,1,"<path>") or (1,4,"<path>") or (1,4,,"<path>"), the software version 3.3 build 2 still produces the "illegal batch command" error. The documentation in the τ-Argus manual 3.3 has not changed. We therefore assume that this issue has not been addressed.

Update in October 2009: The issue has been solved in τ-Argus software version 3.3 build 17 as well as in the manual.

tau Item 14. Log file line numbers

Issue reference(s)	τ-Argus software versions 3.3 build 2 and 3.2 build 7
Reported	November 2008
Updates	July 2009, October 2009

In batch mode, errors are referenced as follows : e.g. "ExploreFile went wrong in line 0" or "Illegal batch command () in line 31". Given the fact that the log file can be cumulative, i.e. including the log of prior sessions, it would be useful to have line numbers which match the indications. While not being an essential feature, it may in some cases speed up the debugging process in reaction to such errors.

Update in July 2009: The requested feature is not available in the software version 3.3 build 2.

Update in October 2009: It has not been fixed but it can be considered as a minor issue.

tau Item 15. [solved] Logbook batch command not mentioned in the τ-Argus manual

Issue reference(s)	τ-Argus manual versions 3.3 and 3.2
Reported	November 2008
Updates	July 2009, October 2009

Even though it is a fully working batch command, <LOGBOOK> is not mentioned anywhere in the τ-Argus manual.

Update in July 2009: The issue still applies to the manual version 3.3.

Update in October 2009: The issue has been solved in the τ-Argus manual version 3.3 build 14.

Comments on the Handbook on Statistical Disclosure Control

In 2008, some items were reported in the context of a preliminary review of chapters 1 (Introduction) and 4 (Magnitude Tabular Data) of the SDC handbook (version 1.01).

In 2009, an in-depth review of the chapters 1 (Introduction), 2 (Regulations, a general Background), 4 (Magnitude tabular data) and 5 (Frequency tables) of the SDC handbook version 1.10 was performed. Moreover, we have checked whether or not the items reported in 2008 have been addressed – whenever an issue was solved, we have tagged the title of the relevant item.

As an overall remark, it should be mentioned that since version 1.01 many sections as well as the handbook as a whole [version 1.10] have significantly improved and become more readable. Our remarks still apply to version 1.20 of the handbook, therefore we have not updated any of the issues reported in our draft report.

HB Item 1. [solved] The handbook's readership

Issue reference(s)	HB versions 1.10 (p.5) and 1.01 (p.6), Chapter 1 Introduction, 1.2 Introduction
Reported	June 2008
Updates	June 2009

*'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 [version 1.01] (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 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.

Update in June 2009: Concerning the chapters under review, i.e. 1, 2 and 4, it can be said that thanks to additional comments, explanations and revisions, the handbook's readability has vastly improved. As far as Chapter 4 is concerned, there is one remaining readability issue – please refer to HB Item 11 for further details.

HB Item 2. The τ-Argus manual and the handbook on SDC

Issue reference(s)	HB versions 1.10 and 1.01 (Chapter 4) τ-Argus manual versions 3.3 and 3.2
Reported	June 2008
Updates	June 2009

It may appear confusing to some readers that a lot of elements normally belonging to the scope of the τ-Argus manual actually made it into the SDC handbook [version 1.01] (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 τ-Argus and thus to promote the latter as a solution.

Nevertheless, one should be careful not to mix up the τ-Argus manual with the SDC handbook. This unnecessarily increases the volume of the SDC handbook and it generates redundancies with the τ-Argus manual. Furthermore, similar redundancies with the SDC handbook can be found in the τ-Argus manual as well, which makes it difficult to work with both the SDC handbook and the τ-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 τ-Argus manual can happen. To increase readability and to stress the purpose of the SDC handbook, specific information relating to τ-Argus could be transferred into devoted e.g. "How to do this within τ-Argus"-boxes/sections below the actual information or it could be marked with a distinctive symbol next to the paragraphs where applicable.

Update in June 2009: The following pages in the handbook version 1.10 contain details regarding τ-Argus which should only form part of the τ-Argus manual because they are not essential for the understanding of the handbook:

- p.136: "The software package τ-Argus … for frequency tables." ;
- p.137: Figure 4-1: Overview of the τ-Argus architecture ;
- p.144: "Note that in the current implementation … (BS2000)." ;

HB Item 3. [please refer to HB Item 10] Waivers

This item has been moved to allow for a more readable structure of this report.

HB Item 4. Chapter 1 – minor issues

Issue reference(s)	HB version 1.20, Chapter 1 Introduction, HB version 1.10, Chapter 1 Introduction, pages 4-8
Reported	June 2009
Updates	October 2009

The chapter *1 Introduction* of the SDC handbook provides a very good overview of the subject. There are however a few minor points which, if addressed, would allow increasing the readability of this part of the book.

- The subchapter *1.1 Preface* should form an isolated chapter of the handbook, i.e. before the chapter *1 Introduction*. While the information contained in this section needs to have its place in the handbook, the introduction should rather keep the reader focused on the topic of disclosure risk and statistical disclosure control.

- The following statement on page 5 under the point *Tabular data protection* does not seem to be easily understandable: “In the majority of cases, confidentiality protection is achieved only by statistical tools due to the absence of legal and IT restrictions.”

- Some parts of subchapter *1.3 Concepts and definitions* (e.g. the description of the various types of outputs) appear to be redundant or complementary with subchapter *1.2 Introduction*. It should be considered to consolidate these two sections. The consolidated subchapter could then be named *1.2 Statistical output types and disclosure risk*. The title *Concepts and definitions* might also be confused with the *Glossary*.

- The section *Risk and Utility* should be renamed *Disclosure risk and data utility*.

- Update regarding HB version 1.2: Additionally, there are some cross-reference errors.

HB Item 5. Chapter 2 – Ethical codes – content-related issues

Issue reference(s)	HB version 1.10, Chapter 2 Regulations, a general Background, 2.1 Ethical codes, pages 12-17
Reported	June 2009
Updates	/

- ISI Declaration on Professional Ethics: It would be nice to obtain a quick overview of its background: what were its objectives, how was it received by the NSIs, is it still applicable, how has it evolved, etc. ;

- European Statistics Code of Practice: Given the degree of detail of the section relating to the Conference of European Statistics and for the sake of completeness, the detailed indicators provided in principle 5 of the European Statistics Code of Practice should also be quoted in the SDC handbook. In addition, it would be very useful to provide similar comments and/or historical background information for the European Code of Practice than for the CES section.

- It would be interesting to know more about the links between these ethical codes (provided there are any).

- This subchapter focuses a tad too much on microdata access for scientific purposes. While this is necessary for a better understanding of chapter 3 on microdata protection, the ethical background of tabular data protection merits to be mentioned here as well.

HB Item 6. Chapter 2 – Laws – content-related issues

Issue reference(s)	HB version 1.10, Chapter 2 Regulations, a general Background, 2.1 Ethical codes, pages 17-18
Reported	June 2009
Updates	/

- This section should be updated with regards to the recent adoption of the Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics. Given that this regulation repeals the regulations 1101/2008 (which itself codified the regulation 1588/90 and its amendments) and 322/97, the latter should only be briefly mentioned for historical purposes.

- Like the subchapter 2.1 on ethical codes, this subchapter also focuses a tad too much on microdata access for scientific purposes rather than on the general need to control data for disclosure risk. Therefore, it would be more relevant to focus on the regulation 223/2009 rather than on the Commission regulation no 831/2002, which implements the general principles outlined in the former. For anyone working in statistical disclosure control, a sound knowledge of this legal framework is of crucial importance because it provides key definitions and supersedes any national laws regarding statistical confidentiality :

    - definition of confidential data and other important concepts such as e.g. direct and indirect identification (article 3) ;
    - public use files (article 19) ;
    - protection of confidential data: exclusive use of confidential data for statistical purposes, exceptions to that rule, physical and logical protection (article 20) ;
    - transmission of confidential data within the ESS and between ESS and ESCB exclusively for statistical purposes (article 21) ;
    - protection of confidential data in the Commission (article 22) ;
    - access to confidential data for scientific purposes (article 23) ;
    - data from public sources (article 25) ;
    - violation of statistical confidentiality (article 26).

HB Item 7. Chapter 2 – minor issues

Issue reference(s)	HB version 1.10, Chapter 2 Regulations, a general Background, pages 12-14, 18
Reported	June 2009
Updates	/

In an attempt to increase readability, the following points could be considered:

- The section *International Statistical Institute* (cf. page 12) should be renamed *ISI Declaration on Professional Ethics*.

- pages 12 bottom - 13 top: The two sentences "The International Statistical Institute … written consultations." should be reformulated as follows: "After an intense preparation process taking place from 1979 to 1985, the International Statistical Institute (ISI) adopted the ISI Declaration on Professional Ethics in 1985."

- The following "titles" on page 13 should not be formatted with bold characters because technically they do not contribute to the structure of the subchapter which they form part of. The bold character formatting may mislead the reader :

    - 4.5 Maintaining confidentiality of records ;
    - 4.6 Inhibiting disclosure of identities ;
    - Principle 5: Statistical confidentiality.

- The section *Conference of European Statisticians* (cf. page 13) should be renamed *UNECE Principles and Guidelines of Good Practice*.

- Hyperlinks to any external reports should be put into section *2.4 References* and be cross-referenced in the text if deemed necessary. Consequently, the links on the bottom of page 12 and on the top of page 14 should be replaced by "(see 2.4 References)". Moreover, the link to the UNECE document is no longer working.

HB Item 8. Chapter 4 – Discussion on the (n,k)-dominance rule may be incomplete

Issue reference(s)	HB versions 1.01 and 1.10, Chapter 4 Magnitude tabular data, 4.2.1 Sensitive Cells in Magnitude Tables, paragraph "Holdings", pages 123-128
Reported	June 2009
Updates	/

On page 123, it is mentioned that the minimum frequency hardly makes any sense if n is greater than 3. This is justified by the very low likelihood that 2 or more units (which are not linked) would form an intruder coalition.

One may wonder whether it makes sense to use a (n, k)-dominance rule with n equalling 3 or higher. The actual disclosure risk in this case also seems to depend on the likelihood that 2 or more units form an intruder coalition. The n parameter of the (n,k)-dominance rule does not seem to be have been discussed in this way in the handbook.

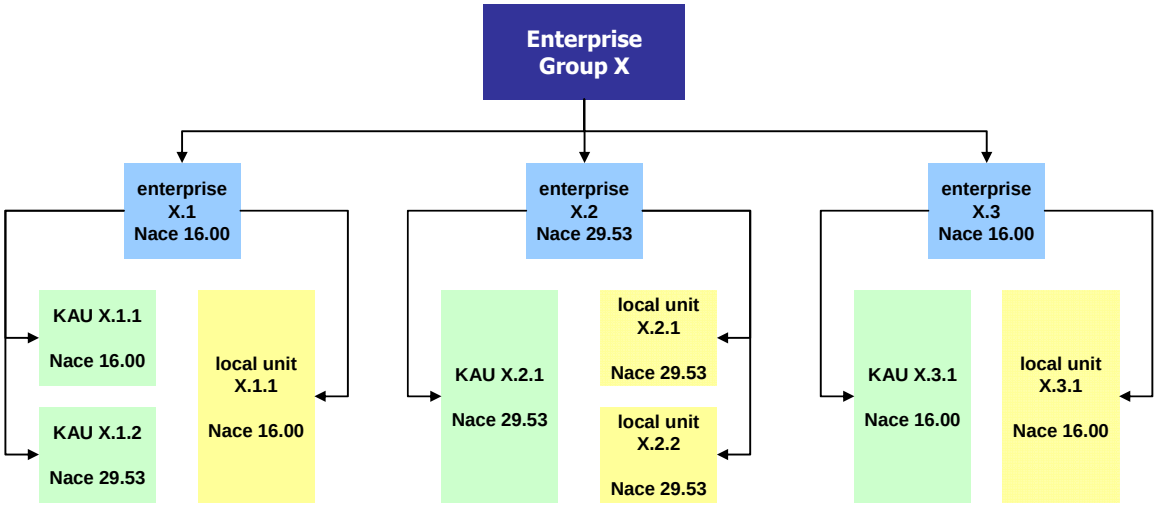
HB Item 9. Chapter 4 – Holdings and their extent in Structural Business Statistics

Issue reference(s)	HB versions 1.01 and 1.10, Chapter 4 Magnitude tabular data, 4.2.1 Sensitive Cells in Magnitude Tables, paragraph “Holdings”, page 129
Reported	June 2009
Updates	/

“We need to group the contributions from one company together to one contribution before we can apply the sensitivity rules.”

From a Structural Business Statistics (SBS) point of view, this statement has wider consequences than those described in the manual.

In SBS, data is available for several types of statistical units: enterprises, kind-of-activity units (KAU), local units. Some enterprises act as enterprise groups, so this needs to be considered. The relation between the statistical units can be illustrated by the example here below.



Consequently, based on what is said in the handbook and given the knowledge about statistical units, we would have to apply the following SDC workflow in SBS:

- First, we need to execute SDC procedures for data based on the enterprise concept. When doing this we must ensure that within one cell, enterprises which are held by the same enterprise groups in that same cell are grouped together. Thus, (truncated) enterprise groups may have an impact on the confidentiality pattern of the enterprise data set. The assumption would be that enterprises belonging to the same (truncated) enterprise group may form an intruder coalition within a given cell.

- In the second stage, we would execute SDC procedures for data based on KAUs or local units, using the pattern of the enterprise data set as a constraint where the cell totals are similar to address the risks resulting from linked tables. However, this would not completely eliminate the impact from both enterprises and enterprise groups because KAUs or local units could be grouped together using either of the two concepts to form an intruder coalition within a given cell. Given that the enterprise group level is the topmost

level, using only the enterprise identifier as a holding indicator is not enough for lower level statistical units such as local units and KAUs.

The current section *Holdings* in the SDC handbook implicitly suggests some of what precedes but the actual extent is not fully clear. How far should we go when testing the risk of intruder coalitions within cells and what should be the reference holding indicator ? Maybe some of this could be discussed more in-depth in the handbook.

Furthermore, the risk of coalitions formed by units in a cell also (but not only) depends on the links between those units. It is more likely that two members of the same enterprise group form a coalition than in the case of two competitors. This could be mentioned more explicitly in the handbook.

HB Item 10. Chapter 4 – Waivers

Issue reference(s)	HB versions 1.01 and 1.10, Chapter 4 Magnitude tabular data, 4.2.1 Sensitive Cells in Magnitude Tables, paragraph “Waivers”, page 128
Reported	November 2008
Updates	June 2009

On page 128 of the SDC handbook, waivers and their impact on the sensitivity rules are described. It concludes that :

“[…] we will be able to deal with such a situation properly, if we reformulate the concentration rules as […] $x_r + x_s > k / 100 \cdot X$*, for the (2,k) dominance rule.”*

This may suggest that the adaptation to the dominance rule would not cause any trouble. According to the above formula, the coalition of the largest respondent with waiver and the largest respondent without waiver must not dominate in case of a (2,k)-rule. At first glance, it may seem paradoxal why the largest respondent with waiver should still be protected. Yet, one should not forget that there is also the risk that the largest respondent with waiver may reveal the largest respondent without waiver. The dominance rule aims at addressing both risks simultaneously and therefore most often leads to overprotection. One way to solve this problem in the context of a (2,k)-rule would be to have at least two waivers, more specifically one for each of the two largest respondents in a cell.

Nevertheless, this is rarely an option and one would therefore be better off to switch to an adapted p%-rule for the cells where there are waivers. The adapted p%-rule takes into account the fact that the largest respondent with waiver has agreed to potentially risk indirect disclosure. This clearly provides a strong argument against the use of the dominance rule in the context of waivers.

These observations also suggest that including the adapted dominance rule in τ-Argus will probably not be of much added value. It would be much more useful if it were possible for τ-Argus to temporarily switch to an equivalent p%-rule for the cells where there are waivers and to apply the corresponding decision criterion described in the handbook. The return result of such a temporary check with the adapted p%-rule would be :

- either to keep the confidentiality status of the cell produced by the initial check conducted with the dominance rule, provided that the disclosure of the aggregate risks to expose the largest respondent without waiver ;
- or to change the cell status to “safe”, provided that the disclosure of the aggregate doesn’t pose any threat to the largest respondent without waiver and that any minimum frequency rule is satisfied.

This may be too complicated to implement in τ-Argus though. Alternatively, this can be done manually. In that case, the protection flags can be used to set the cell status from “unsafe” to “safe” and the minimum frequency rule should be verified manually for these cells. Waivers do not allow overriding a minimum frequency rule unless waivers for all respondents in a cell are available.

While all these problems may be evident, we would like the handbook to be more explicit about them.

Update in June 2009: This section of the SDC handbook has unfortunately not seen any changes since version 1.01 yet. As waivers are sometimes the only way to address confidentiality issues in certain areas of statistics, it could be useful to make this section more complete. Therefore, the following reformulation is suggested:

“Sometimes respondents authorize the publication of an aggregate even if this publication might cause a risk of disclosure for their contributions. Such authorizations are also referred to as ‘*waivers*’. Sensitivity rules are directly impacted by the use of waivers. Therefore, they need to be reformulated so that we will be able to deal with such a situation properly.

The objective of the reformulation is to avoid that the largest respondent who has provided a waiver would be able to reveal the contribution of the largest respondent for whom no such waiver is available. With s being the largest respondent from which no waiver has been obtained and with r being the largest respondent except for s, for any pair (i,j), $i \neq j$ of respondents, where no waiver has been obtained from j, it holds $x_i + x_j \leq x_r + x_s$. Therefore, the sensitivity rules can be reformulated as follows:

p%-rule: $X - x_s - x_r < p/100 \cdot x_s$
(2,k)-dominance rule: $x_r + x_s > k/100 \cdot X$

However, while this reformulation works seamlessly for the p%-rule, the reformulated dominance rule is not practical. This is inherently due to the way the dominance rule works: as it aims at protecting the contributions of the top n respondents simultaneously and not separately (as is the case with the p%-rule), the use of the dominance rule would require the NSI to obtain the waivers for all of the top n respondents. This could obviously lead to a significant overprotection compared to the use of p%-rule and would also seriously question the utility of waivers. There are two alternative solutions to deal with this problem: either switch completely to the p%-rule (recommended solution) or only apply the p%-rule for those cells where there are respondents with waivers.

Finally, waivers should never override the minimum frequency rule, unless waivers have been obtained from all the respondents in a cell.”

HB Item 11. Chapter 4 – 'How to process linked tables' not understandable anymore

Issue reference(s)	HB version 1.10, Chapter 4 Magnitude tabular data, 4.2.1 Evaluation of Secondary Cell Suppression algorithms offer by τ-Argus, How to process linked tables, pages 152-153
Reported	June 2009
Updates	/

The linked tables problem is a key issue when it comes to tabular data protection, e.g. in the area of Structural Business Statistics. Therefore, the proper understandability of this section of the handbook should be considered as crucial. We have identified the following issues in this section of the handbook:

- In the version 1.10 of the handbook, the link between the first and the second paragraph has disappeared. In the handbook version 1.01, the passage "We begin by protecting T1 […] This means we have to set the status of those cells to 'unsafe'." was linked perfectly to the 2nd paragraph "The question is then […]". This passage has probably become the victim of a copy&paste procedure.

- We understand that it is difficult to describe this process with non-jargon language. However, the use of mathematical language in order to make this topic more accessible also fails because of the complexity of the subject. After the link between the first and the second paragraph has been re-established (see previous comment), there is probably no other way than providing a detailed accompanying example. Regarding the degree of detail, it should be mentioned that the introductive example provided on pages 155-157 does unfortunately not fill this gap because it only refers back to the theoretical section for the details (see *Processing of linked tables* on page 156). The example does not necessarily need to be as complex as the introductive example. A few very simple tables with fictive numbers as well as a description of the problem should be sufficient. Any remaining complexities or generalizations can then be completed by additional comments and/or mathematical language.

HB Item 12. Chapter 4 – Handling periodical data

Issue reference(s)	τ-Argus manual versions 3.3 and 3.2 HB versions 1.01 (pp. 151-152) and 1.10 (pp. 153-154), Chapter 4 Magnitude tabular data, 4.3.3. Processing table protection efficiently, paragraphs "a-priori file" and "periodical data"
Reported	June 2008
Updates	July 2009, November 2008

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 within τ-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 handbook also mentions the a priori file, which may be helpful :

[...] Specifying all this information manually during a τ-ARGUS-run for a large table can be rather cumbersome and error prone. In the batch-version of τ-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. SDC handbook [1.01] 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. τ-Argus manual version 3.2 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.

An alternative workaround to the above problem may be to perform a first confidentiality pattern calculation with τ-Argus and to compare the output with the confidentiality flags applied for the previous year. Then, on a case by case basis, the a priori 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 being less optimal than the method suggested in the SDC handbook.

Furthermore, the modification of the cost function in the a priori file is not documented in an easily understandable manner, even though it is a powerful feature. 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. Update in November 2008 : Such guidance could be completed by a comparison of the various features[3] and cases[4]. Given the fact that this feature is strongly related to the software, the additional documentation should be added into the τ-Argus manual rather than into the SDC handbook.

Update in July 2009: The relevant sections in both the SDC handbook and the τ-Argus manual haven't been adapted.

[3] e.g. when to adjust the cost rather than to use the status flags, what is the relation with the actual cell safety status, etc.
[4] e.g. implications of a high/low/zero value, how to choose the values, relation with each of the suppression methods, etc.

HB Item 13. Chapter 4 – minor issues

Issue reference(s)	HB version 1.10, Chapter 4 Magnitude tabular data
Reported	June 2009
Updates	/

In an attempt to increase readability, the following points could be considered:

- The subchapter *4.1 Introduction* should be provided with section titles. Here below are some suggestions:

    - *Magnitude table characteristics*: this title would enclose the paragraphs 1 to 4 on page 116: “Statistical magnitude tables display the sums of observation of a quantitative table […] are the contributions to the cell value.“

    - *Disclosure risks in magnitude tables*: this title would enclose the 5th paragraph on page 116 to the 1st paragraph included on page 118: “At first sight, one might find it difficult […] the so-called ‘secondary’ tabular data protection methodologies.”

    - move the title *Example 1* from the 3rd paragraph on page 117 to the 2nd paragraph on page 118, which starts with “Assume that a table displays the sum of a variable ‘production’ […].”

- The paragraph on the top of the page 132 “Considering this as the basic idea of secondary protection … to be part of a table.” appears to be out of context with the previous paragraph on the same page “Obviously … will be sensitive”. “This” probably refers to an idea described in the subchapter *4.1 Introduction* whereas in its current context it seems to refer to the topic on sampling weights, which is confusing. We recommend moving this sentence to the end of subchapter *4.1 Introduction*. In addition, given the first sentence of *4.2.2 Secondary tabular data protection methods* on page 132, there is a visible redundancy.

- It could be useful to wrap up the section *4.2.1 Sensitive Cells in Magnitude Tables* with a table which describes the key aspects of the minimum frequency rule and the sensitivity rules. Here below is an example of how such a table could look like:

Rule	Notes on protection from disclosure
Minimum frequency rule	• n = 3 should be the normal choice because it prevents all the relevant risks of exact disclosure ; • n > 3 should only be chosen if there is a risk of coalitions within cells ; • best used together with a sensitivity rule because on its own it only prevents exact disclosure – in a combined approach the minimum frequency rule should always be applied last in order to allow for an optimal secondary protection ; • if the number of contributions in a cell is less than n, the resulting unsafe cell status should never be overridden by waivers unless the latter are available for all contributions in that cell.
(n,k)-dominance rule	• if n = 1, the rule ignores the case in which the contributor with the second largest contribution to a cell is able to derive a close upper estimate for the largest contribution by subtracting her own contribution ; • if n ≥ 2, it takes into account the additional knowledge of the second largest contributor (or coalition of contributors) but also has a tendency overprotection compared to the p%-rule ; • n ≥ 3 should only be chosen if there is a risk of coalitions within cells ; • cannot handle waivers adequately because it requires waivers to be available for each of the top n contributors.
(p,q)-rule	• if q = 100, the rule corresponds to the p%-rule and provides comparable protection to a (2,k)-dominance rule but without the tendency for overprotection ; • if q < 100, it can take into account the risk that the second largest respondent could estimate the smaller contributions within q% ; • recommended approach to handle waivers.

- Singleton and multi cell disclosure on page 135: “The most prominent case is that of two cells with only one contributor (a.k.a. ‘singletons’) within the same aggregation. No matter how large the cell values, because they both know their own contribution of course, both can use this additional knowledge to disclose the other’s contribution.” After a few re-readings due to a minor paradox, these two sentences tend to become clearer. Maybe the following reformulation would help to make them more understandable: “[…] two cells within the same aggregation with each cell having only one single contributor (a.k.a. ‘singletons’).” In case that the new formulation leads to similar confusions, one should consider providing a brief example.

- The section *Recommendation* on page 147 should be renamed *Modular or Hypercube ?*.