



Harmonization of the protection of social statistics at Statistics Finland

Miika Honkala

Work Session on Statistical Data Confidentiality, Skopje

21 September 2017

STIINA project

- STIINA (Social Statistics Integrated Information Architecture) started at Statistics Finland in 2016
- Some objectives
 - To move all population data to a same data store
 - To facilitate the production of social statistics
- Consist of many subprojects
 - STIINA3 is a subproject that works with statistical disclosure control
- The project will continue at least until 2020

STIINA3 (Methods and new data contents)

- The project started at Statistics Finland in January 2017
- The tasks of the project in the spring 2017
 - Current protection practices and tools in social statistics were studied
 - Other potential protection methods and tools were studied
- The project will continue in the beginning of 2018
 - Aim is to to standardize protection practices in different social statistics

Current protection practices

- Statistics Finland has protection guidelines for tabular social statistics
- Protection methods that are in use
 - Suppression
 - Changing classification of the variables
 - Rounding
- Protection tools that are in use
 - Tau-Argus
 - Some SAS macros
 - Protection using own SAS codes

Problems with the current practices

- The current protection guidelines offer many possibilities
 - Suppression: a lot of different threshold rules
 - Many solutions for changing classifications of the variables
 - Different rounding practices
- Protection practices can be a bit different in each social statistics
 - Each protection task can have its own expert
 - If some expert changes a job, there is a problem
 - Confusion: no one knows what practices are used with other statistics

Objectives of the STIINA3

- To create new protection practices for social statistics
 - More accurate rules
 - Fewer opportunities for protection
 - More similar practices in different statistics
- To offer a common software (or softwares) for protection
 - Easy to use, sufficient properties
 - Some SAS macro would be good
 - More automatic protection procedure
 - Harmonizing of the protection is easier

The challenges of harmonization

- Statistics Finland releases social statistics related to about 70 topics (population structure, families, migration, education, wages, ...)
 - Frequency tables and magnitude tables
 - Different statistics may need different protection practices
- The number of methods can not be very high
 - Which methods are allowed, which are rejected?
 - Several methods may be needed, however

The challenges of harmonization

- Limitation and perturbation methods
 - Limitation methods (suppression etc.) are currently used
 - Perturbation methods would also be interesting
- The released tables have to provide as much information as possible
 - Harmonizing the protection practices must not prevent it

Similar ESS project

- Harmonised protection of Census data in the ESS started in 2016
- ESS countries will carry out Census in 2021
- Objective: harmonize the protection of tables for the Census
 - STIINA3 project has quite similar objective
- Some methods were tested in the project
 - Record swapping
 - Cell key method
- Statistics Finland can possibly utilize the results of the project

Conclusion

- Harmonization of the protection practices is in the beginning
 - Current protection methods and tools have been cleared
 - Decisions on new practices have not been made
- The desired end result of the STIINA3 project
 - All social statistics are protected using common, clear practices
 - There is SAS software for protection
 - Easy to use and contain enough properties
 - Protection procedure is as automatic as possible

- Thank you!