

In this talk I will define key concepts of microdata protection and describe some relevant Statistical Disclosure Limitation (SDL) methods. I will start with basic definitions and describe a structure of a microdata file. Before releasing such data to the public, statistical agencies have an obligation by law to protect the confidentiality of the respondents/data providers and at the same time they strive to release a product that would satisfy the ever growing demands of potential data users. Thus, the goal of microdata protection is two-fold: minimize the risk of disclosure of respondents' confidential information and maximize the utility of the released data for the user. The key issue here is that these goals are conflicting goals. To decrease the disclosure risk, data protector typically has to perturb microdata in some way, which often leads to decreased utility of the resultant data to the user. On the other hand, when the data are modified with the goal to improve the utility, some protection maybe undone which will increase disclosure risk. Hence, a trade-off between data utility and disclosure risk is the main issue of SDL practice. This is why a decision about how to define and measure data utility and disclosure risk should be among the first steps in the process of microdata protection. It will help to better understand and compare the existing SDL methods, choose the most appropriate one, and develop the most appropriate protection strategy for a particular scenario of data release. I will give examples of such definitions and discuss their advantages and disadvantages. In what follows, I will present several SDL methods suitable for the protection of microdata and discuss their effectiveness based on the proposed metrics of utility and disclosure risk.