Exploring Confidentiality Issues in a Data-Driven World: Challenges for Public Statistics

Dominik Rozkrut, Statistics Poland and University of Szczecin, Poland
Monika Rozkrut, University of Szczecin, Poland
Wlodzimierz Okrasa, Cardinal Stefan Wyszynski University in Warsaw and Statistics Poland,

Abstract: There is growing concern among statistical methodologists and experts in confidentiality protection procedures about the risk of a widening gap between the operational capabilities of the tools (techniques and approaches) available to confidentiality researchers, official statistics managers and policy makers, and the challenges posed by new types of data and storage (including Big Data, data from multiple sources, administrative data, spatially explicit data files, etc.), and recent developments in computing power, algorithms and software for record linkage, data sharing and data analytics. A catalogue of issues involved changes along the dynamics of risk factors occurring both on the part of the data provider and the producer of statistics, and on the part of the data/statistics user. Therefore, structuring the questions that arise in this context, focusing on the factors related to the overflow of new data requiring verification (for their credibility and usability) and adjustment for statistical purposes, and to the new tools used by the parties involved in the statistical process, forms the first part of this presentation. A comprehensive outline of the context, including the issues of basic protection of confidentiality, privacy and access, is the starting point for these considerations. The second part attempts to formulate a methodological framework for providing strategies to deal with the above-mentioned challenges and other problems related to the disclosure limitations, from an institutional point of view.   The conclusions will outline a proposal for a new type of research avenue, one that would ultimately include stakeholders / institutions that are trying to solve similar problems due to the emergence of new data types and related challenges. This could start in the form of an exchange of research experiences on the problems of maintaining confidentiality in public statistics in the face of the threat of alienation from the data-driven world.

Keywords: confidentiality and data access; disclosure limitation and risk factors; data protection procedures; new data and analytics.

References

Duncan, G. T., Elliot, M., Salazar-González, J-J., 2011. Statistical Confidentiality Principles and Practices. Springer.

McKinseyGlobal Institute, 2016. The Age Of Analytics: Competing In A Data-Driven World (December). https://www.mckinsey.com/~/media/McKinsey/Industries

Okrasa, W., 1994. Private Lives, Public Policies. Items. Social Science Research Council. Vol. 48, No. 1. N.Y.

Rozkrut, D., Rozkrut, M., Okrasa, W., 2020. Confidentiality Issues in Integrated Data Access from an Institutional Perspective. Presented at Joint Statistical Meeting/JSM2020. Philadelphia, August 1-6, 2020.

Vanwey, L. K., et al., 2005. Confidentiality and spatially explicit data: Concerns and challenges. Proceedings of the National Academy of Sciences 102(43):15337-42

National Academies of Sciences, Engineering, and Medicine 2017. Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy. Washington, DC: The National Academies Press. https://doi.org/10.17226/24652