

Privacy Protection When Data From Multiple Sources Are Linked

Stephen E. Fienberg

Department of Statistics, Machine Learning Department, and Cylab

Carnegie Mellon University

Pittsburgh PA 15213

Considerable effort has gone into understanding issues of privacy protection of individual information in single databases, and various solutions have been proposed depending on the nature of the data, the ways in which the data base will be used and the precise nature of the privacy protection being offered. Once data are merged across sources, however, the nature of the problem become far more complex and a number of privacy issues arise for the linked individual files that go well beyond those that are considered with regard to the data within individual sources. In this presentation we will discuss (a) the problem of record linkage, (b) the privacy issues arising from databases constructed from multiple sources, and (c) a strategy for combining information for analytical purposes without combining databases.