

William E. Winkler

TITLE: General Discrete-data Modeling Methods for Producing Synthetic Data with Reduced Re-identification Risk that Preserve Analytic Properties

ABSTRACT: General modeling methods for representing and improving the quality of discrete data (Winkler 2003) extend and connect the editing methods of Fellegi and Holt (1976) and the imputation ideas of Little and Rubin (2002). This paper describes a modeling framework to produce synthetic microdata that better corresponds to external benchmark constraints on certain aggregates (such as margins) and on which certain cell probabilities are bounded both below and above to reduce re-identification risk. Rather than use linear constraints (Meng and Rubin 1993), the modeling methods use convex constraints