

John M. Abowd

TITLE : An Overview of the NSF-ITR-Census Bureau Synthetic Data Project

ABSTRACT: This talk describes five different synthetic data development efforts produced by a collaboration of university-based and US Census Bureau-based researchers. These projects include (1) a partially synthetic version of the 1990-1993 and 1996 Surveys of Income and Program Participation linked to lifetime earnings and retirement benefit histories; (2) a partially synthetic version of the Longitudinal Business Database covering establishment microdata from 1976 to 2001; (3) a partially synthetic version of the Longitudinal Employer-Household Dynamics infrastructure file system with integrated employer-employee data; (4) a fully synthetic Origin-Destination matrix used to produce OnTheMap; and (5) partially synthetic data used as part of the confidentiality protection in the American Community Survey public-use micro samples. The talk covers an overview of synthesizer methods, analytical validity, and disclosure avoidance techniques.