Journal of Privacy and Confidentiality (2013) **5**, Number 1

Introduction to Special Section

Dan Kifer*
Guest Editor

This issue contains three invited papers from the 2011 IEEE International Workshop on Privacy Aspects of Data Mining (PADM) that was held on December 11, 2011 in Vancouver, Canada in conjunction with the International Conference on Data Mining (ICDM).

The workshop was organized by Raghav Bhaskar, Aris Gkoulalas-Divanis, Dan Kifer, and Srivatsan Laxman with the aim of bringing together researchers with a variety of different perspectives on privacy technology. Authors of selected papers from this workshop were invited to submit extended versions of their papers to the special PADM section of this issue. The following three papers were accepted.

In "Differential Privacy Applications to Bayesian and Linear Mixed Model Estimation," John Abowd, Matthew Schneider, and Lars Vilhuber investigate questions about the quality of differentially private statistical models. Using two frameworks for constructing differentially private algorithms (sample-and-aggregate and objective perturbation), they fit linear mixed models and Bayesian linear mixed models using data from the U.S. Census Bureau's Quarterly Workforce Indicators. While pointing out practical model-building issues for which differentially private solutions are still needed, they evaluate the quality that one can expect from privacy-preserving versions of these models.

In "On Regression-Tree-Based Synthetic Data Methods for Business Data," Joo Ho Lee, In Yong Kim, and Christine M. O'Keefe study the utility of statistical disclosure control techniques when applied to business data. Business data have have different characteristics than data about individuals—there are many outliers and many of the measured variables have highly skewed distributions. This paper evaluates the CART synthesizer (which samples synthetic data from decision trees built on the original data) using data from the sugar cane industry in Queensland, Australia and identifies challenges that still need to be overcome.

In "Privacy-Preserving Data Sharing for Genome-Wide Association Studies," Caroline Uhler, Aleksandra Slavkovic, and Stephen Fienberg apply differential privacy to genome-wide association studies (GWAS). The release of GWAS data is a sensitive issue because of the possibility of re-identification of individuals. This paper develops privacy-preserving methods for releasing minor allele frequencies, chi-square statistics, p-values, and genome-wide associations.

We are grateful to the authors and to the referees of these papers as well as the referees of PADM 2011 for their hard work and valuable suggestions.

*Department of Computer Science & Engineering, Penn State University, University Park, PA, USA, mailto:dkifer@cse.psu.edu

