# Towards Privacy in Public Databases

Adam Smith

Massachusetts Institute of Technology

<adsmith@MIT.EDU>

## Abstract

This talk describes recent, ongoing work.

We begin a theoretical study of the *census problem*. Informally, in a census individual respondents give private information to a trusted party (the census bureau), who publishes a sanitized version of the data. There are two fundamentally conflicting requirements: *privacy* for the respondents and *utility* of the sanitized data. Unlike in the usual approach to secure function evaluation, in which privacy is preserved to the extent possible given a specific functionality goal, in the census problem *privacy* is paramount; intuitively, things that cannot be learned "safely" should not be learned at all.

An important contribution of this work is a definition of privacy (and privacy compromise) for statistical databases inspired by definitions of security for secure function evaluation, together with a method for describing and comparing the privacy offered by specific sanitization techniques. We obtain several privacy results using two different sanitization techniques, and then show how to combine them via cross training. We also obtain two utility results involving clustering.

Joint work with Shuchi Chawla, Cynthia Dwork, Frank McSherry and Hoeteck Wee.