

## Privacy and Utility in Data Sets

Felix T. Wu

Cardozo School of Law

Is it possible to publicly release useful data, while preserving the privacy of the individuals whose information is in the database? This question has been the subject of considerable controversy, particularly in the wake of well-publicized instances in which researchers showed how to re-identify individuals in supposedly anonymous data. Some have argued that privacy and utility are fundamentally incompatible, while others have suggested that simple steps can be taken to achieve both simultaneously. Both sides have looked to the computer science literature for support.

What the existing debate has overlooked, however, is that the relationship between privacy and utility depends crucially on what we mean by “privacy” and what we mean by “utility.” Apparently contradictory results in the computer science literature can be explained by the use of different definitions to formalize these concepts. Without sufficient attention to these definitional issues, it is all too easy to over-generalize the technical results. More importantly, formal definitions may never capture some of the nuances in common understandings of “privacy” and “utility,” nuances that are highly contextual and that depend on social factors, not just numbers. By analyzing some of those nuances, we can begin to understand the policy choices inherent in deciding whether and how to regulate data privacy across varying social contexts.