<http://repository.cmu.edu/cgi/viewcontent.cgi?article=1008&context=jpc>

Journal of Privacy and Confidentiality (2009) 1, Number 2, pp. 135–154

Brief History of Differential Privacy

Before the term “Differential Privacy” was coined, there were different statistical analyses being done on the population without any consideration of the individuals that were part of the studies. Cynthia Dwork and Adam Smith then decided that this had to change and began trying to answer this question: How does one release statistical analysis results to the public without compromising the privacy of the individual respondents? [16] (match number to Bib)

Dwork and Smith started with the idea that data was *noninteractive*, that once some statistical result was published by the curator, the dataset could be destroyed because it wasn’t going to be used again. This was troublesome because the statistics released could be biased based on how the curator determined the privacy needed for the data. If significant privacy was needed, then important data would be withheld, and the statistical inference would be altered.

Then they considered how the statistical analysis would be affected if the data were *interactive*. In this scenario, the curator would act as a gate between the database and the users. The queries would pass through the curator and be cleaned until they didn’t cause privacy issues. This notion would ideally be better because then only the questions of interest would be answered. A user wouldn’t be able to scrub the dataset of every last bit of meaning and analyses.

Dwork ad Adams felt that this was still not an idyllic way to bring about privacy so they tried to prove the statement, “access to the statistical database does not help the adversary to compromise the privacy of any individual.” This almost immediately fell flat as this example demonstrates: Say john is two times heavier than the average American. Then as long as someone has access to a database such as a medical database that contains the weights of Americans, if a mean statistic can be drawn from the data, then John’s weight is calculated. The point being that regardless of if John was in that database or not, he would be vulnerable to attack. This is where “Differential Privacy” was born. Instead of avoiding the risk by holding back certain information, the goal was to minimize the risk so that the risk of being inside the database was the same as being outside.