One of the many new challenges for statistical inference in the information age is the increasing concern of data privacy protection. A particularly fruitful approach, that offers strong protection against privacy breaches, is the concept of ‘differential privacy’ (Dwork et al. (2006)). The idea is that instead of the original database, only a randomly perturbed version is released for further analysis. Such a randomization mechanism is said to provide differential privacy if the conditional distribution of the released database given the original data does not depend too much on any individual entry of the true database.

This talk attempts to provide a general introduction to the notion of differential privacy from the point of few of mathematical statistics, but is directed at a broad audience. After discussing the main ideas of differential privacy, we will briefly recall the concept of minimax optimal estimation and survey some of the few existing results from the statistics literature on estimation under differential privacy. In this setup, the objective is not only to come up with an optimal estimation procedure that efficiently recovers information from the randomized observations, but also to devise a randomization mechanism that best facilitates subsequent estimation while respecting the required privacy provisions.

In the second part of the talk, we will present some of our own results on minimax optimal locally private estimation of linear functionals. Our analysis allows for a quantification of the price, in terms of statistical accuracy, that has to be paid for achieving differential privacy. This price appears to be highly problem dependent.