保护个人隐私远远不止隐藏数据记录中的敏感属性( 如姓名住址年薪等) 那么简单，还要阻止敏感属性值与特定的实体或个人关联起来，以防止由非敏感属性信息推测出个人的真实身份

实施数据隐私保护主要是考虑以下两个方面：（１）如何保证数据应用过程中不泄露隐私；

（２）如何更有利于数据的应用．

隐私分为两类

链接攻击 Sweeney L．K-anonymity：a model for protecting privacy[J]．Inter-

national Journal of Uncertainty，Fuzziness and Knowledge—based

Systems，2002，10(5)：557—570．

e1,

这是一个基于数据失真（distorting）和限制发布的技术，随机化，需要注意的是，随意对数据

进行随机化并不能保证数据和隐私的安全

A Condensation Approach to Privacy Preserving Data Mining：以前的方法用perturbation approach去重建数据的分布以达到保护的目的，但是这种方法对每个dimension单独地进行处理而忽略了不同dimension之间的联系，并且只能对某种data mining有效。此算法不要求问题明确的目标，把原始数据集映射为一个新的匿名化数据集，并且紧紧保持了原始数据的特性。

On the Privacy Preserving Properties of Random Data Perturbation Techniques：随机加噪音，基于random matrix-based filter去加噪音。 定义了Signal-to-Noise Ratio

Privacy-Preserving Data Mining。最早的基于随机扰动技术的保护，扰动后数据不对，但是趋势还是一致的，对于分类分析来说依旧好用。 实验部分 ，借鉴

随机化应答

匿名化2个优化：1，更好的匿名化原则，即利用率高。2，高效的匿名化算法。 抑制和泛化。

l-diversity：K匿名的2个问题，改进：让同组的class也至少出现l次

t-closeness：改进l-diversity，考虑class的分布