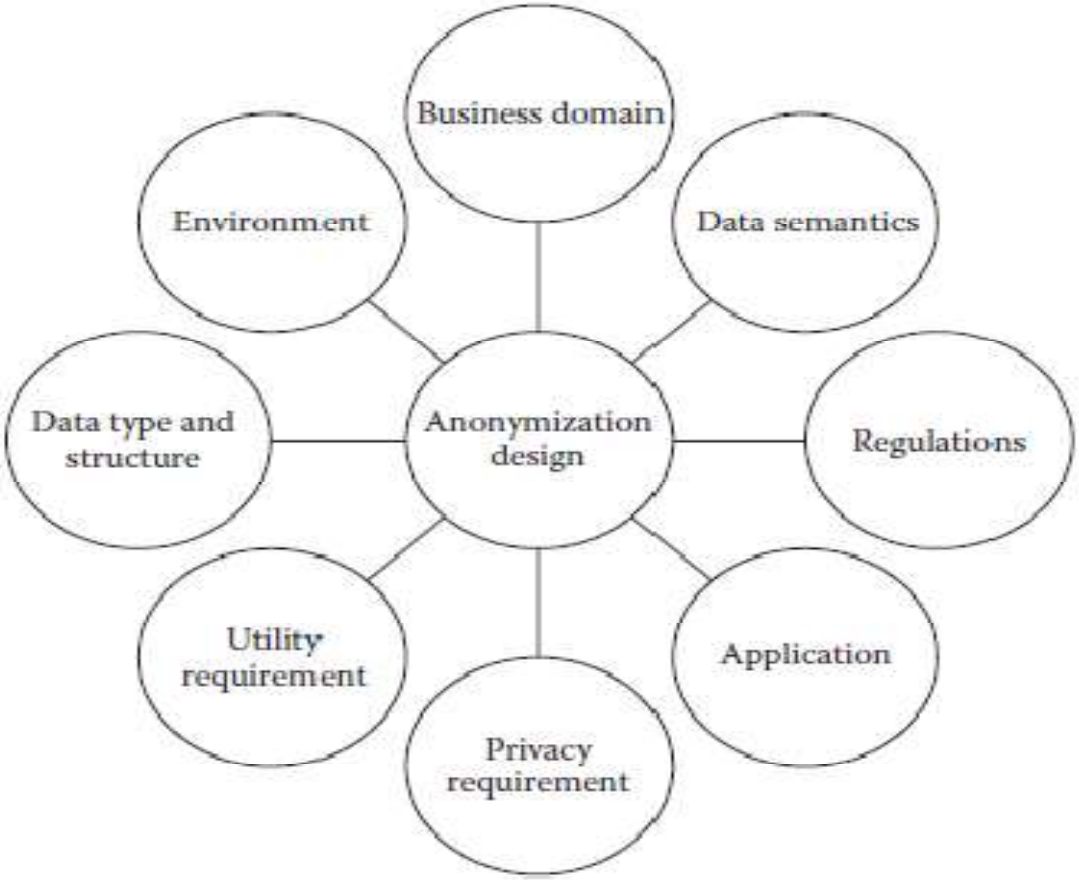
Anonymization design principles

Anonymization Design Principles

- Anonymization design is not straightforward
- achieving a balance between privacy and utility has many dependencies
- for a given requirement, many Factors that drive anonymization design

Anonymization Design Principles



Anonymization Design Principles

- When there is a need for data privacy, organizations generally use either a commercial or a home-grown product for anonymizing data.
- It is critical to ensure that an organization's data anonymization program is not limited by the features of the product.
- Many organizations fail to maintain a balance between privacy and utility.
- It is generally difficult to determine how much anonymization is required, which results in either loss of information or the anonymized data set becoming unusable.

Anonymization Design Principles

- Even with adoption of the best of breed data anonymization products, an organization's anonymization program may not be successful.
- In addition to this, the pressures of regulatory compliance force many organizations to be very defensive and adopt very high privacy standards that will render the data unusable for any research.
- If enough care is not taken, then the anonymized data could have very little utility

Anonymization Design Principles

- In this context, irrespective of which tool an organization uses, there is a need for a mechanism to monitor privacy versus utility for various privacy requirements.
- Unfortunately, quantifying privacy and utility is nontrivial. Therefore, it is critical to provide assurance of high quality of data anonymization during the initial phase of the anonymization life cycle.
- To support this, we felt it is necessary to define a set of design principles.
- These principles will provide the required guidelines for the data anonymizer to adopt the correct design for a given anonymization requirement.

Anonymization Design Principles

- Classifying principles into two broad types—scientific and normative.
- Scientific principles are laws of nature and form the fundamental truths that one can build upon. Normative principles act as a guide and need to be enforced.
- Similarly, a data anonymizer needs guidance, and the anonymization design principles should be enforced to ensure proper anonymization design.
- These principles are fundamental in nature and are applicable to all aspects of anonymization.
- They connect the high-level privacy and utility requirements to low-level implementation.

Data

Anonymization in multidimensional data

- **Multidimensional Data**
  - Multidimensional data also referred to as relational data are the most common format of data available today in many enterprises.
  - In a relational table, each row is a vector that represents an entity
  - The columns represent the attributes of the entity.
  - As relational data are the most common data format, a lot of attention has been paid to privacy preservation of relational data

Data

Anonymization in multidimensional data

- A row of data in a relational table is classified into explicit identifiers, quasi-identifiers, sensitive data, and nonsensitive data. Both perturbative and nonperturbative techniques could be used to protect the data.
- As a rule, EI are completely masked out, QI are anonymized, and SD are left in their original form.
- Depending on the sensitivity of data, appropriate data protection techniques can be applied.

Data

Anonymization in multidimensional data

- The fundamental differences between anonymizing multidimensional data and other data structures are as follows:
- In a multidimensional data table, each record or row is independent of others; therefore, anonymizing a few of the records will not affect other records.
- Anonymizing a tuple in a record will not affect other tuples in the record.

Data
Anonymization in multidimensional data

- Depending on the sensitivity of data, appropriate data protection techniques can be applied.
- The fundamental differences between anonymizing multidimensional data and other data structures are as follows:
- In a multidimensional data table, each record or row is independent of others; therefore, anonymizing a few of the records will not affect other records.
- Anonymizing a tuple in a record will not affect other tuples in the record.

Data

Anonymization in multidimensional data

- Depending on the sensitivity of data, appropriate data protection techniques can be applied.
- The fundamental differences between anonymizing multidimensional data and other data structures are as follows:
- In a multidimensional data table, each record or row is independent of others; therefore, anonymizing a few of the records will not affect other records.
- Anonymizing a tuple in a record will not affect other tuples in the record..

Challenges in Privacy Preservation of Multidimensional Data

Challenges in Privacy Preservation of Multidimensional Data

The challenges in this kind of data preservation are as follows:

- Difficulty in identifying the boundary between QI and SD in the presence of background knowledge of the adversary
- High dimensionality of data poses a big challenge to privacy preservation Clusters in sensitive data set
- Difficulty in achieving realistic balance between privacy and utility