



Exercise 07:

Information Systems & Data Warehouses, Data Attributes & Types, Anonymization

Privacy-Preservation Technologies in Information
Systems

Dr. Wiem Fekih Hassen / Dr. Armin Gerl / Felix Bölz

WS 21/22



Task 1: Information Systems & DataWarehouses

Privacy-Preservation Technologies in Information
Systems

Dr. Wiem Fekih Hassen / Dr. Armin Gerl / Felix Bölz

Information System (IS)

Definition

Information systems is the study of complementary networks of hardware and software that people and organizations use to **collect, filter, process, create, and distribute data.**



IS Components

1

Hardware

Digital devices that you can physically touch (e.g, laptop, mobile phones, tablet, e-readers, etc.).

2

Software

Set of instructions that tell the hardware what to do. It is divided into two categories: *operating systems* and *application software*.

3

Databases

A collection of facts (e.g, street address, the city, phone number, etc.)

4

Networks

The different components that allow the communication between hardware and software

5

Procedures

A series of steps undertaken to achieve a desired outcome or goal.

IS Classification

DSS: is used by senior management to make non-routine decisions.

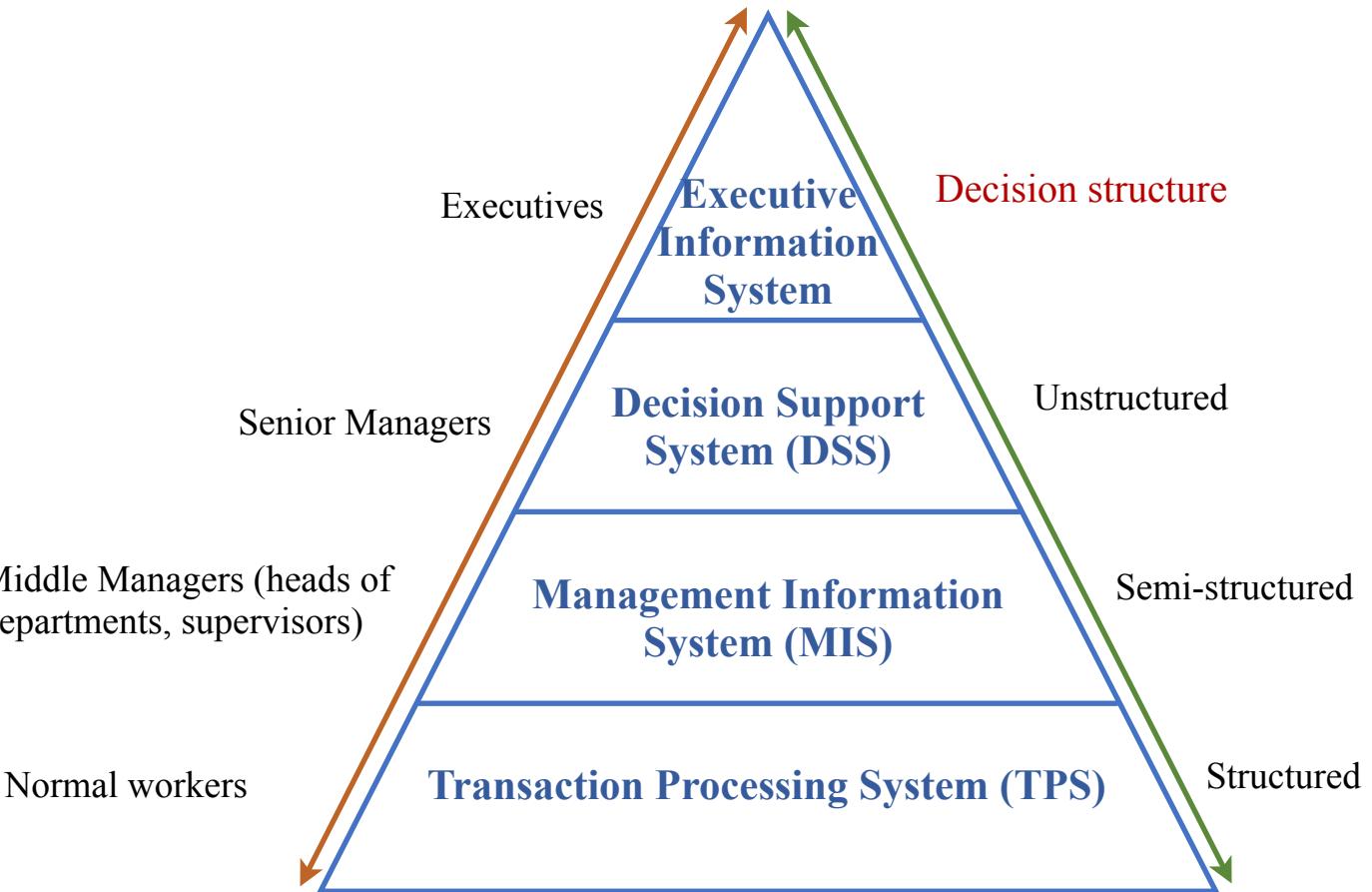
- Senior level managers are concerned with the long-term planning of the organization. They use information from middle managers and external data to guide them when making unstructured decisions.
- The main objective of DSS is to provide solutions to problems that are unique and change frequently.

MIS: Are used to monitor the organization's current performance status.

- Middle managers make semi-structured decisions. The decisions are partly based on set guidelines and judgmental calls.

TPS: Data base systems and normal software

- Users at this level use make structured decisions. they have defined rules that guides them while making decisions.

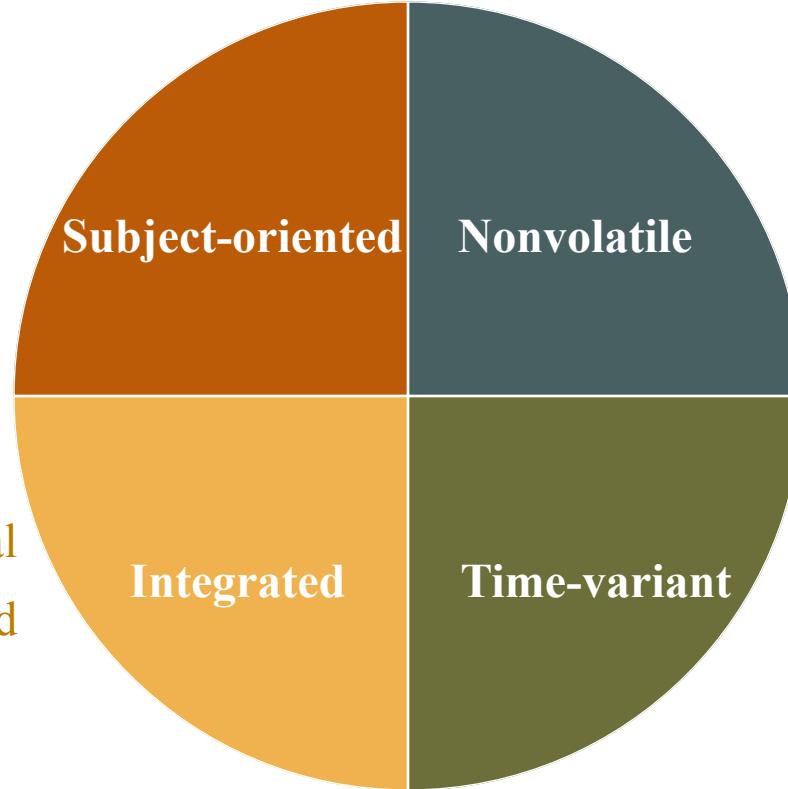


DataWarehouse (DW) – Properties



The data DW revolves around specific subjects/topics. No normalization of schema.

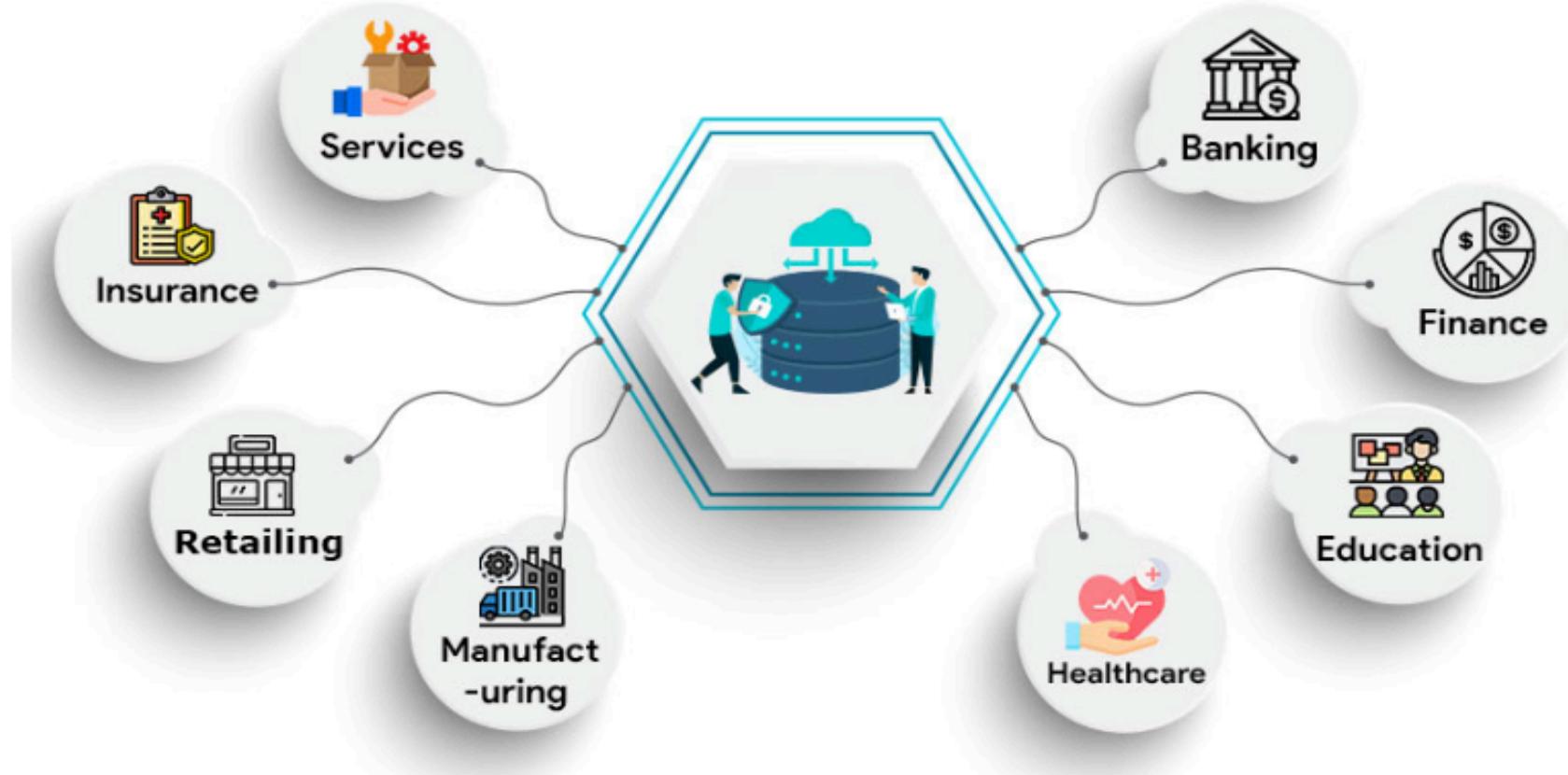
The data in DW comes from several operational systems. Data is processed to remove inconsistencies.



Data in DW is optimized read-only, which means it should not be updated, created, or deleted.

DW data represents current and historical data (a long time horizon up to 10 years).

DataWarehouse — Examples



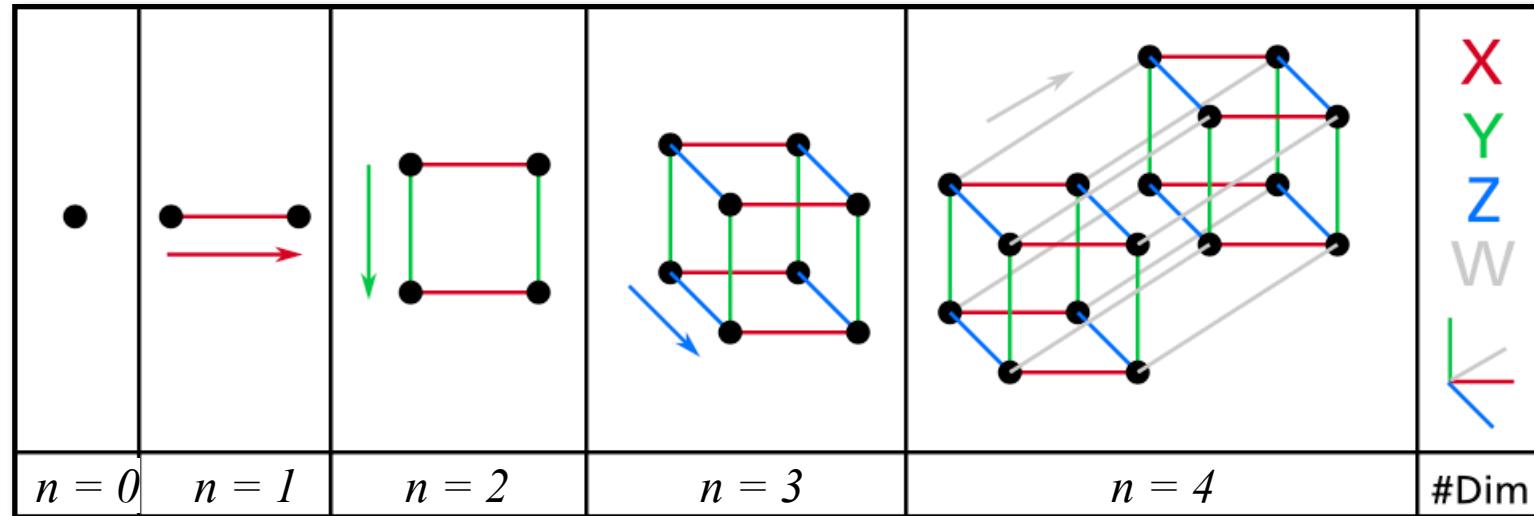
Source: <https://datachannel.co/blogs/best-applications-of-data-warehousing/>



Hypercube – Definition & Example

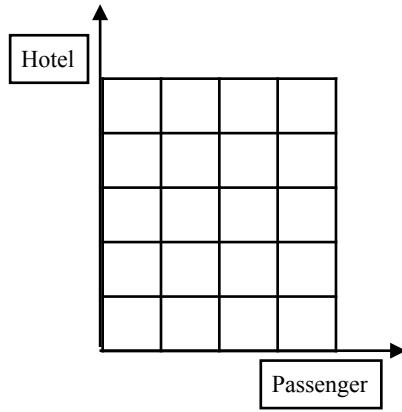
Definition

In geometry, a **hypercube** is an **n -dimensional** analogue of a square ($n = 2$) and a cube ($n = 3$).

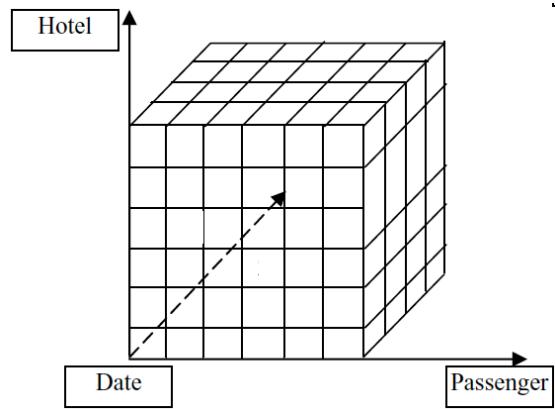


Source: <https://en.wikipedia.org/wiki/Hypercube>

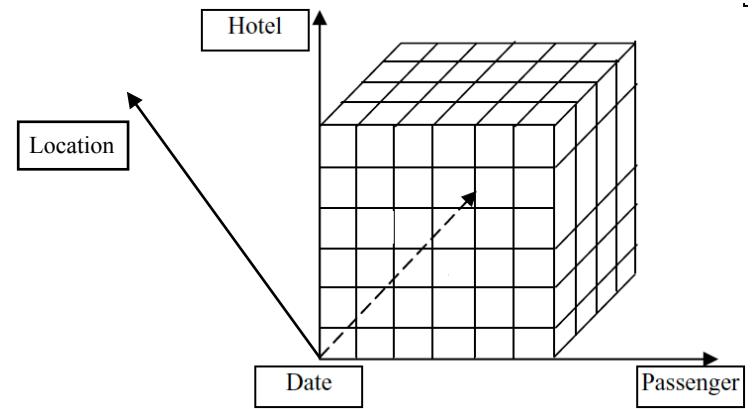
Hypercube — Data Models



$n = 2$



$n = 3$



$n = 4$



Task 2: Data Attributes & Types

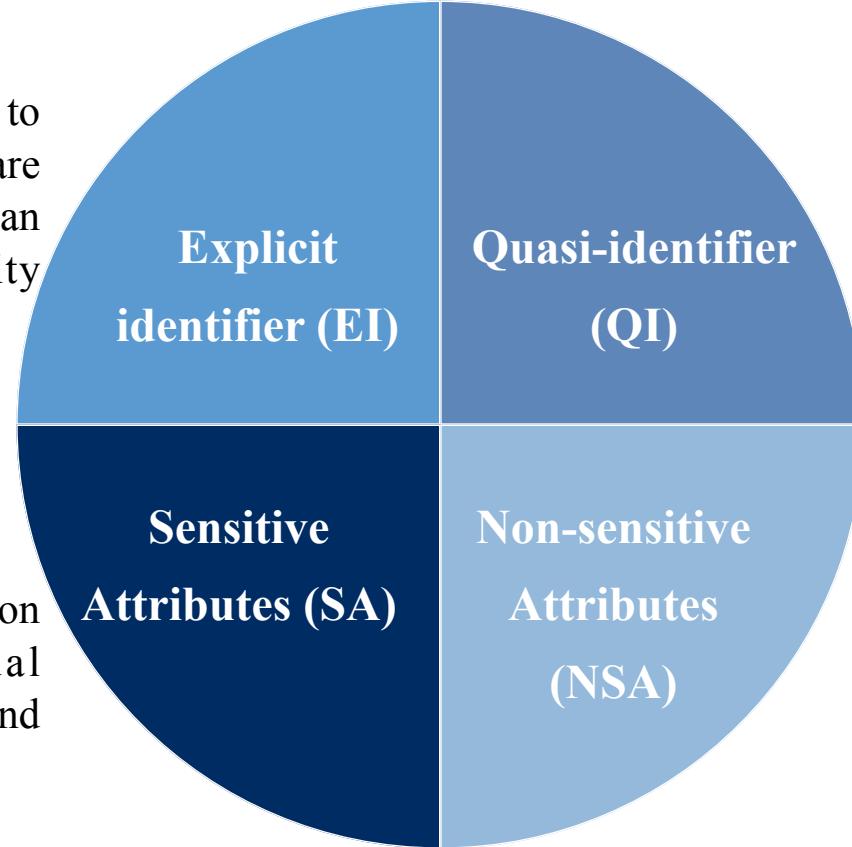
Privacy-Preservation Technologies in Information
Systems

Dr. Wiem Fekih Hassen / Dr. Armin Gerl / Felix Bölz

WS 21/22

Attributes Classifications

Attributes that clearly identify the entity to which the record refers to, because they are used as a primary key for the entity in an official context, e.g., Social security number, Tax identification number, etc.

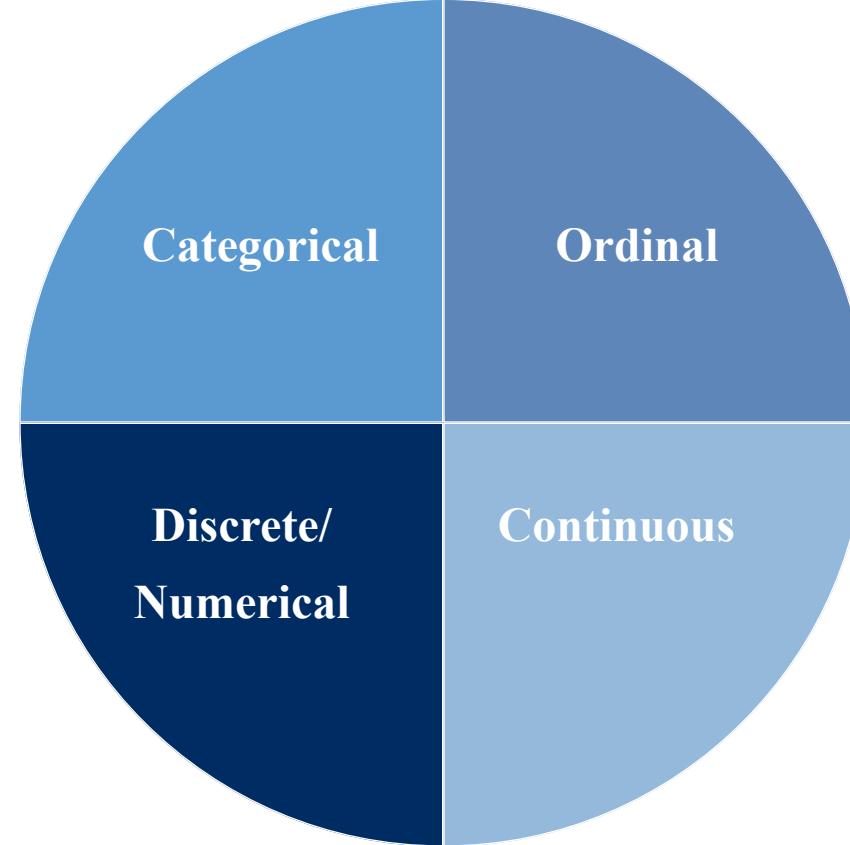


These attributes hold private information about the respondent (e.g. sexual orientation, salary, medical conditions and illnesses etc.)

When an attribute doesn't immediately lead to re-identification but needs to be combined with others to do so.

All information that doesn't directly or indirectly imply a connection to entity the record belongs to.

Attributes Types



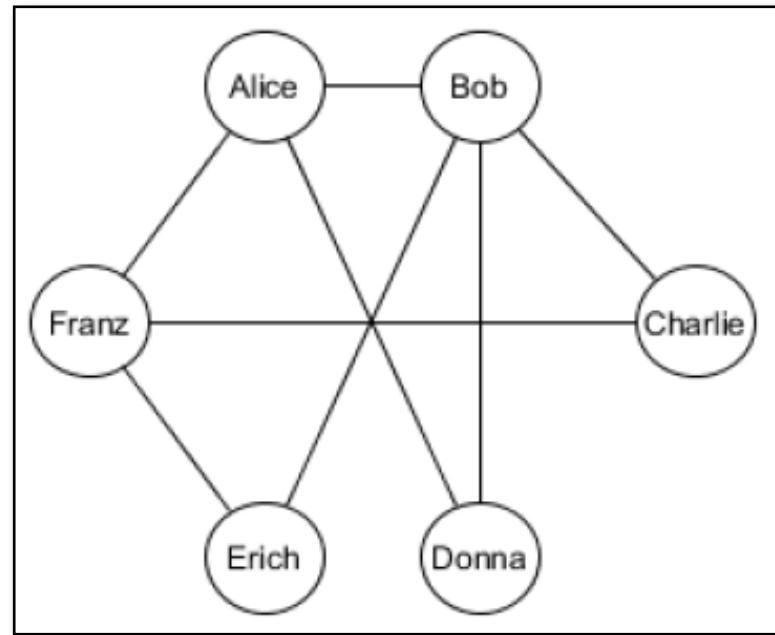
A limited domain of values belongs to a categorical (or discrete) attribute (e.g. Gender, State, Date,...)

If a sufficiently large or infinite range of values is possible, the attribute's domain is continuous (e.g. Temperature, height, speed,...)

Microdata vs. Macrodata

- **Microdata** refers to information about a *specific* entity, besides a private citizen, this can also be a company
- **Macrodata** in contrast refers to *aggregated* or grouped data.

Storing a Graph Data into Data Base



Transform the graph into Table

	Alice	Bob	Franz	Erich
Alice	1	1	1	0
Bob	1	1	0	1
Franz	1	0	1	1
Erich	0	1	1	1



Task 3: Anonymization

Privacy-Preservation Technologies in Information
Systems

Dr. Wiem Fekih Hassen / Dr. Armin Gerl / Felix Bölz

Anonymization – Basic Concepts

Definition

The goal of anonymization is the prevention of the leakage of the identity of a user based upon personal data

1

Suppression

Values are being replaced by an replacement character, e.g., asterix (*).

2

Generalization

Using Hierarchies to cluster attribute values in similar groups.

3

Deletion

Delete singular entries, that carry high re-identification risk or would require inappropriate effort to anonymize.

Attribute Classifications & Types

Name	Number-Plate	Non-payed penalties
Wiem	PA X 293	0 €
Armin	DEG L 32	100 €
Felix	M XX 002	50 €
Ashish	NU LL OOOO	9,90 €

- **Name:** Explicit identifier (EI)
 - ➡ Type: Categorical/ordinal
- **Number-Plate:** Explicit identifier (EI)
 - ➡ Type: Categorical/ordinal
- **Non-payed penalties:** Sensitive Attributes (SA)
 - ➡ Type: Continuous

Anonymization Concept

- **Name:** Explicit identifier (EI)
 - ➡ Anonymization concept: Suppression & Deletion
- **Number-Plate:** Explicit identifier (EI)
 - ➡ Anonymization concept: Suppression & Deletion
- **Non-payed penalties:** Sensitive Attributes (SA)
 - ➡ Anonymization concept: Suppression & Generalisation

Anonymization of First Row

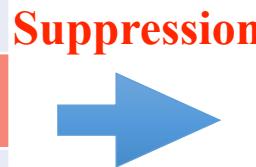
Name	Number-Plate	Non-payed penalties
Wiem	PA X 293	0 €
Armin	DEG L 32	100 €
Felix	M XX 002	50 €
Ashish	NU LL OOOO	9,90 €



Name	Number-Plate	Non-payed penalties
	PA X ***	0 €
Armin	DEG L 32	100 €
Felix	M XX 002	50 €
Ashish	NU LL OOOO	9,90 €

Anonymization of Second Row

Name	Number-Plate	Non-payed penalties
Wiem	PA X 293	0 €
Armin	DEG L 32	100 €
Felix	M XX 002	50 €
Ashish	NU LL OOOO	9,90 €



Name	Number-Plate	Non-payed penalties
Wiem	PA X 293	0 €
Armin	DEG L **	100 €
Felix	M XX 002	50 €
Ashish	NU LL OOOO	9,90 €

Anonymization of Third Row

Name	Number-Plate	Non-payed penalties
Wiem	PA X 293	0 €
Armin	DEG L 32	100 €
Felix	M XX 002	50 €
Ashish	NULL OOOO	9,90 €

Generalisation



Name	Number-Plate	Non-payed penalties
Wiem	PA X 293	0 €
Armin	DEG L 32	100 €
Felix	M XX 002	0 - 100 €
Ashish	NULL OOOO	9,90 €

Anonymization of Fourth Row

Name	Number-Plate	Non-payed penalties
Wiem	PA X 293	0 €
Armin	DEG L 32	100 €
Felix	M XX 002	50 €
Ashish	NULL OOOO	9,90 €



Name	Number-Plate	Non-payed penalties
Wiem	PA X 293	0 €
Armin	DEG L 32	100 €
Felix	M XX 002	50 €
Ashish	NULL OOOO	0 - 10 €

See you next year 😊