
OWASP
Open Web Application Security Project

# Secure Software Design For Data Privacy

Narudom Roongsiriwong, CISSP

MiSSConf(SP5), July 6, 2019

# WhoAmI




- Lazy Blogger
  - Japan, Security, FOSS, Politics, Christian
  - http://narudomr.blogspot.com
- Information Security since 1995
- Web Application Development since 1998
- SVP, Head of IT Security, Kiatnakin Bank PLC (KKP)
- Committee Member, Thailand Banking Sector CERT (TB-CERT)
- Consultant, OWASP Thailand Chapter
- Committee Member, Cloud Security Alliance (CSA), Thailand Chapter
- Committee Member, National Digital ID Project, Technical Team
- Contact: narudom@owasp.org

# Privacy By Design

## The 7 Foundational Principles

- Proactive not Reactive; Preventative not Remedial
- Privacy as the Default
- Privacy Embedded into Design
- Full Functionality – Positive-Sum, not Zero-Sum
- End-to-End Security – Lifecycle Protection
- Visibility and Transparency
- Respect for User Privacy

*Source: Privacy By Design – The 7 Foundational Principles, Ann Cavoukian, Ph.D. , Information & Privacy Commissioner, Ontario, Canada*

# Data Privacy Ground Rules

- If you don't need it, don't collect it.
- If you need to collect it for processing only, collect it only after you have informed the user that you are collecting their information and they have consented, but don't store it
- If you have the need to collect it for processing and storage, then collect it, with user consent, and store it only for an explicit retention period that is compliant with organizational policy and/or regulatory requirements
- If you have the need to collect it and store it, then don't archive it, if the data has outlived its usefulness and there is no retention requirement.

# Fundamental Security Concepts

| | | | |
|---|---|---|---|
| **Core** | Confidentiality | Integrity | Availability |
| | Authentication | Authorization | Accountability |
| **Design** | Need to Know | Least Privilege | Separation of Duties |
| | Defense in Depth | Fail Safe / Fail Secure | Economy of Mechanisms |
| | Complete Mediation | Open Design | Least Common Mechanisms |
| | Psychological Acceptability | Weakest Link | Leveraging Existing Components |

# Security in Privacy Design

| | | | |
|---|---|---|---|
| **Core** | Confidentiality | ?Integrity? | Availability |
| | Authentication | Authorization | Accountability |
| **Design** | Need to Know | Least Privilege | Separation of Duties |
| | Defense in Depth | Fail Safe / Fail Secure | Economy of Mechanisms |
| | Complete Mediation | Open Design | Least Common Mechanisms |
| | Psychological Acceptability | Weakest Link | Leveraging Existing Components |

# Privacy vs Integrity

- In most of data protection acts (such as GDPR) said that "*organizations must take necessary and reasonable steps to ensure the accuracy of personal data collected from data subjects*"
- Some privacy design approaches using referential integrity across datasets
- But some privacy design approaches using data distortion techniques
- Conclusion
  - Data as "Source of Truth" → Integrity is a must
  - Data in use → Integrity depends on utility

# PRIVACY Design

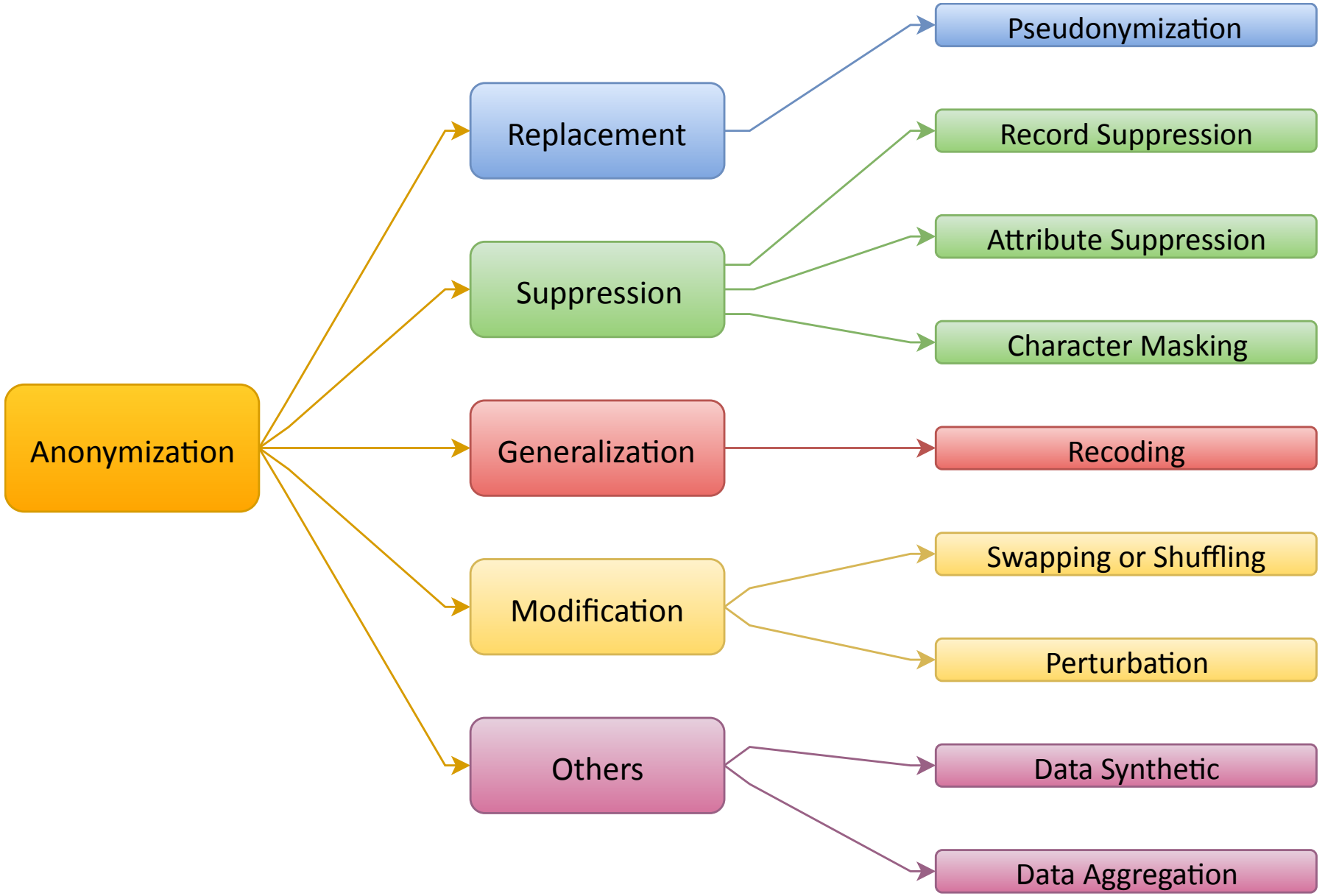
# Privacy with Data Anonymization

- Anonymization is the process of removing private information from the data
- Anonymized data cannot be linked to any one individual account

# What You Need to Aware of Anonymization

- Purpose of anonymization and its utility
- Characteristics of each anonymization techniques
- Inferred information after implementation
- Expertise with the subject matter
- Competency in anonymization process and techniques
- Recipients

# Anonymization Techniques



# Terminology

- Data Attribute:
  - Data field, data column or variable, an information that can be found across the data records in a data set
- Dataset:
  - A set of data records, conceptually similar to a table in a conventional database or spreadsheet, having records (rows) and attributes (columns)
- Direct Identifier:
  - A data attribute that on its own identifies an individual (e.g. fingerprint) or has been assigned to an individual (e.g. Citizen ID)
- Indirect identifier or Quasi-Identifiers:
  - A data attribute that, by itself/on its own, does not identify an individual, but may identify an individual when combined with other information
- Re-identification:
  - Identifying a person from an anonymized dataset

# Pseudonymization

- Decoupling identifiable data from the dataset, usually by means of identifier key references
- Pseudonym (aka Token) may represent one or more attributes
- Pseudonyms can be
  - Reversible (by the owner(s) of the original data), where the original values are securely kept but can be retrieved and linked back to the pseudonyms
  - Irreversible, where the original values are properly disposed and the pseudonymization was done in a non-repeatable fashion
- Pseudonyms persistence
  - Persistent – Same pseudonym values represent the same individual across different datasets
  - Non-persistent – Different pseudonyms represent the same individual in different datasets to prevent linking of the different datasets
- Pseudonyms generation
  - Random (Ex. UUID, GUID)
  - Deterministic (Ex. Hashing, Encryption, PCI DSS Tokenization)

# Pseudonymization – Example#1 (1/2)

## Before Anomymization:

| Name | Address | Phone |
|---|---|---|
| Jim Demetriou | 4290 Cheval Circle, Stow, OH 44224 | 330-805-4211 |
| Gary Furlong | 24 Steeple Drive, Hillsborough, NJ 08844 | 908-359-1754 |
| Maria Herring | 8096 Wild Lemon Lane, Manlius, NY 13104 | 315-682-4453 |
| John Sacksteder | 2480 Pendower Lane, Keswick, VA 22947 | 240-994-6728 |
| John Mantel | 23 College Street, South Hadley, MA 01075 | 413-532-5562 |
| Dan Okray | W1748 Circle Drive, Sullivan, WI 53178 | 262-593-5004 |

## After Pseudonymizing the Name Attribute:

| Name | Address | Phone |
|---|---|---|
| LAU5B90A | 4290 Cheval Circle, Stow, OH 44224 | 330-805-4211 |
| 1YXHL5K0 | 24 Steeple Drive, Hillsborough, NJ 08844 | 908-359-1754 |
| KOTACI4U | 8096 Wild Lemon Lane, Manlius, NY 13104 | 315-682-4453 |
| SDM1VHX3 | 2480 Pendower Lane, Keswick, VA 22947 | 240-994-6728 |
| UJQXYU27 | 23 College Street, South Hadley, MA 01075 | 413-532-5562 |
| 9NG6Y5VF | W1748 Circle Drive, Sullivan, WI 53178 | 262-593-5004 |

# Pseudonymization – Example#1 (2/2)

## Identity Database

| Pseudonym | Name |
|---|---|
| LAU5B90A | Jim Demetriou |
| 1YXHL5K0 | Gary Furlong |
| KOTACI4U | Maria Herring |
| SDM1VHX3 | John Sacksteder |
| UJQXYU27 | John Mantel |
| 9NG6Y5VF | Dan Okray |

# Pseudonymization – Example#2

**Identity**

First Name: Narudom
Last Name: Roongsiriwong
Age: 18

+

**Non-Identifiable Data**

Gender: Male
Nationality: Thai
Blood Type: O
Occupation: Engineer

=

**Full Data**

First Name: Narudom
Last Name: Roongsiriwong
Age: 18
Gender: Male
Nationality: Thai
Blood Type: O
Occupation: Engineer

# Pseudonymization Guideline

- When to use
  - Data values need to be unique and no need to keep original attribute
- How to use:
  - Replace the respective attribute values with made up values
  - The made up values should be unique, and should have no relationship to the original values
- Tips
  - GDPR separates Pseudonymization from Anonymization
  - This should be a key part of your Privacy by Design strategy
  - Ensure not to re-use pseudonyms that have already been utilized
  - Persistent pseudonyms are usually better for maintaining referential integrity across data sets
  - For reversible pseudonyms, the mapping tables or functions or secret encryption keys should be securely kept and can only be used by the organization

# Attribute Suppression

The removal of an entire part of data ("column" in database) in a data set.

## Before Anomymization:

| Name | Address | Phone |
|---|---|---|
| Jim Demetriou | 4290 Cheval Circle, Stow, OH 44224 | 330-805-4211 |
| Gary Furlong | 24 Steeple Drive, Hillsborough, NJ 08844 | 908-359-1754 |
| Maria Herring | 8096 Wild Lemon Lane, Manlius, NY 13104 | 315-682-4453 |
| John Sacksteder | 2480 Pendower Lane, Keswick, VA 22947 | 240-994-6728 |
| John Mantel | 23 College Street, South Hadley, MA 01075 | 413-532-5562 |
| Dan Okray | W1748 Circle Drive, Sullivan, WI 53178 | 262-593-5004 |

## After Suppressing the "Address" Attribute:

| Name | Phone |
|---|---|
| Jim Demetriou | 330-805-4211 |
| Gary Furlong | 908-359-1754 |
| Maria Herring | 315-682-4453 |
| John Sacksteder | 240-994-6728 |
| John Mantel | 413-532-5562 |
| Dan Okray | 262-593-5004 |

# Attribute Suppression Guideline

- When to use
  - That attribute is not required in the anonymized dataset, or when the attribute cannot otherwise be suitably anonymized with another technique
- How to use:
  - Delete (e.g. remove) the attribute(s), not hiding
  - If the structure of the data set needs to be maintained, clear the data (and possibly the header)
- Tips
  - This is the strongest type of anonymization technique, because there is no way of recovering any information from such an attribute
  - Less sensitive derived attribute may be create to suppress the original attribute(s). E.g. “Usage Duration” attribute base on “Check-In” and ‘Check-Out” date and time attributes

# Record Suppression

- The removal of an entire record in a data set

Can anyone guess
who should this person be?

| Name | Address | Phone |
|---|---|---|
| 3BRYAYN8 | Highlands Farm Woodchurch, Ashf[illegible], TN26 3RJ | 2087726222 |
| 3O7T78EZ | St Elizabeths, Much Hadham, SG10 6EW | 2083435600 |
| 3WVYDLCN | 10 Downing St, Westminster, London SW1A 2AA | 1322341162 |
| 6SSC98FX | Hermitage Court, Hermitage, Kent, ME16 9NT | 2086887666 |
| 9CSYE673 | Grimsby Road, Cleethorpes, North East Lincolnshire, DN35 7LB | 1908262860 |
| 9DIHFAQ9 | 14 High Street, Brompton, Gillingham, ME7 5AE | 2089440110 |

Google

10 downing street

All | Images | Maps | News | Videos | More | Settings | Tools | SafeSearch on

About 50,200,000 results (0.69 seconds)

## Prime Minister's Office, 10 Downing Street - GOV.UK

https://www.gov.uk/government/.../prime-ministers-office-10-downing-street

Prime Minister Theresa May delivered the following speech at a Downing Street reception for diabetes charities and the NHS. ... Prime Minister Theresa May hosts mental health and community leaders in Downing Street following independent review of Mental Health Act. ... PM Theresa May's ...

History of 10 Downing Street · Prime Minister's Office, 10 ... · Past Prime Ministers

## 10 Downing Street - Wikipedia

https://en.wikipedia.org/wiki/10_Downing_Street

Downing Street, also known colloquially in the United Kingdom simply as Number 10, is the headquarters of the Government of the United Kingdom and the ...

Architect: Kenton Couse | Designated: 14 January 1970

Town or city: City of Westminster, London, SW1 | Country: United Kingdom

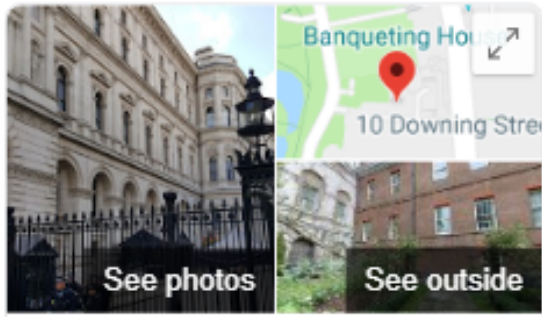
List of residents of 10 Downing ... · 10 Downing Street Guard Chairs · Gavin Barwell

People also search for

- 10 downing street cat
- downing street apartments
- prime minister house india
- where does theresa may live
- first lord of the treasury uk
- who received first bharat ratna award

People also ask

- Can you walk up to 10 Downing Street?
- Who lives at Number 9 Downing Street?
- What is written on the letterbox of 10 Downing Street?
- Does the prime minister have to live at 10 Downing Street?




## 10 Downing Street

Website | Directions | Save

3.4 ★★★½☆ 867 Google reviews

Government office in London, England

10 Downing Street, also known colloquially in the United Kingdom simply as Number 10, is the headquarters of the Government of the United Kingdom and the official residence and office of the First Lord ... Wikipedia

**Address:** 10 Downing St, Westminster, London SW1A 2AA, UK

**Opened:** 1684

**Phone:** +44 20 7925 0918

**Reference no:** 1210759

**Architects:** Christopher Wren, William Kent, Quinlan Terry, Raymond Erith, Kenton Couse

# Record Suppression Guideline

- When to use
  - The records are so unique and outliers can lead to easy re-identification
- How to use:
  - Delete the entire record, not just row hiding
- Tips
  - The removal of a record can impact the data set such as for statistical analysis

# Character Masking

- The change of the characters of a data value, e.g. by using a constant symbol (e.g. “*” or “x”)
- Masking is typically partial, i.e. applied only to some characters in the attribute

# Character Masking Guideline

- When to use
  - The data value is a string of characters and hiding some part is sufficient to provide anonymity
- How to use:
  - Replace the appropriate characters with a chosen symbol
    - Fixed number of characters (e.g. for credit card numbers)
    - Variable number of characters (e.g. for email address)
- Tips
  - Subject matter knowledge of each data type to be mask is needed to ensure the right characters are masked
  - The data owners are meant to recognize their own data

# Recoding

- A deliberate reduction in the precision of data
- Example:
  - Converting a person's age into an age range
  - Converting a precise location into a less precise location

# Recoding – Example

## Before Anomymization:

| Name | Address | Phone |
|---|---|---|
| LAU5B90A | 4290 Cheval Circle, Stow, OH 44224 | 330-805-4211 |
| 1YXHL5K0 | 24 Steeple Drive, Hillsborough, NJ 08844 | 908-359-1754 |
| KOTACI4U | 8096 Wild Lemon Lane, Manlius, NY 13104 | 315-682-4453 |
| SDM1VHX3 | 2480 Pendower Lane, Keswick, VA 22947 | 240-994-6728 |
| UJQXYU27 | 23 College Street, South Hadley, MA 01075 | 413-532-5562 |
| 9NG6Y5VF | W1748 Circle Drive, Sullivan, WI 53178 | 262-593-5004 |

## After Recoding the Address Attribute:

| Name | Address | Phone |
|---|---|---|
| LAU5B90A | Stow, OH | 330-805-4211 |
| 1YXHL5K0 | Hillsborough, NJ | 908-359-1754 |
| KOTACI4U | Manlius, NY | 315-682-4453 |
| SDM1VHX3 | Keswick, VA | 240-994-6728 |
| UJQXYU27 | South Hadley, MA | 413-532-5562 |
| 9NG6Y5VF | Sullivan, WI | 262-593-5004 |

# Recoding Guideline

- When to use
  - The data values that can be recoded and still be useful for the intended purpose
- How to use:
  - Design appropriate data categories and rules for translating data.
  - Consider suppressing any records that still stand out after the translation (see record suppression)
- Tips
  - Design the data ranges with appropriate sizes
    - Too large data range may cause the data too much modification
    - Too small data range may be easy to re-identify

# Shuffling

- Rearranging data in the data set where the individual attribute values are still represented in the data set, but generally, do not correspond to the original records

# Shuffling – Example

## Before Anomymization:

| Name | Address | Phone |
|---|---|---|
| Jim Demetriou | 4290 Cheval Circle, Stow, OH 44224 | 330-805-4211 |
| Gary Furlong | 24 Steeple Drive, Hillsborough, NJ 08844 | 908-359-1754 |
| Maria Herring | 8096 Wild Lemon Lane, Manlius, NY 13104 | 315-682-4453 |
| John Sacksteder | 2480 Pendower Lane, Keswick, VA 22947 | 240-994-6728 |
| John Mantel | 23 College Street, South Hadley, MA 01075 | 413-532-5562 |
| Dan Okray | W1748 Circle Drive, Sullivan, WI 53178 | 262-593-5004 |

## After Shuffling:

| Name | Address | Phone |
|---|---|---|
| Jim Demetriou | 23 College Street, South Hadley, MA 01075 | 262-593-5004 |
| Gary Furlong | 2480 Pendower Lane, Keswick, VA 22947 | 315-682-4453 |
| Maria Herring | 24 Steeple Drive, Hillsborough, NJ 08844 | 413-532-5562 |
| John Sacksteder | 8096 Wild Lemon Lane, Manlius, NY 13104 | 908-359-1754 |
| John Mantel | W1748 Circle Drive, Sullivan, WI 53178 | 330-805-4211 |
| Dan Okray | 4290 Cheval Circle, Stow, OH 44224 | 240-994-6728 |

# Shuffling Guideline

- When to use
  - Subsequent analysis only needs to look at aggregated data and there is no need for analysis of relationships between attributes at the record level
- How to use:
  - Identify which attributes to shuffle then shuffle or reassign the attribute values to any record in the data set
- Tips
  - Assess and decide which attributes need to be shuffled

# Perturbation

- The value modification from the original data set in order to be slightly different
- Two main techniques
  - Probability distribution: data replacement from the same distribution sample or from the distribution itself
  - Value distortion: modification by multiplicative or additive noise, or other randomized processes (more effective)

# Perturbation – Example

## Before Anomymization:

| Person | Height (cm) | Weight (kg) | Age (years) | Smokes? | Disease A? | Disease B? |
|---|---|---|---|---|---|---|
| 198740 | 160 | 50 | 30 | No | No | No |
| 287402 | 177 | 70 | 36 | No | No | Yes |
| 398747 | 158 | 46 | 20 | Yes | Yes | No |
| 498732 | 173 | 75 | 22 | No | No | No |
| 598772 | 169 | 82 | 44 | Yes | Yes | Yes |

## Perturbation Rules Using Base-X Rounding:

| Attribute | Anonymization Technique |
|---|---|
| Height (in cm) | Base-5 rounding (5 is chosen to be somewhat proportionate to the typical height value of, e.g. 120 to 190 cm) |
| Weight (in kg) | Base-3 rounding (3 is chosen to be somewhat proportionate to the typical weight value of, e.g. 40 to 100 kg) |
| Age (in years) | Base-3 rounding (3 is chosen to be somewhat proportionate to the typical age va lue of, e.g. 10 to 100 years) |
| (the remaining attributes) | Nil, due to being non-numerica l and difficult to modify without substantial change in value |

# Perturbation – Example

## Before Anomymization:

| Person | Height (cm) | Weight (kg) | Age (years) | Smokes? | Disease A? | Disease B? |
|---|---|---|---|---|---|---|
| 198740 | 160 | 50 | 30 | No | No | No |
| 287402 | 177 | 70 | 36 | No | No | Yes |
| 398747 | 158 | 46 | 20 | Yes | Yes | No |
| 498732 | 173 | 75 | 22 | No | No | No |
| 598772 | 169 | 82 | 44 | Yes | Yes | Yes |

## After Anomymization:

| Person | Height (cm) | Weight (kg) | Age (years) | Smokes? | Disease A? | Disease B? |
|---|---|---|---|---|---|---|
| 198740 | 160 | 51 | 30 | No | No | No |
| 287402 | 175 | 69 | 36 | No | No | Yes |
| 398747 | 160 | 45 | 18 | Yes | Yes | No |
| 498732 | 175 | 75 | 21 | No | No | No |
| 598772 | 170 | 81 | 42 | Yes | Yes | Yes |

# Perturbation Guideline

- When to use
  - Quasi-identifiers (typically numbers and dates) which may potentially be identifying when combined with other data sources, and slight changes in value are acceptable.
  - Should not be used where data accuracy is important
- How to use:
  - Depends on the exact data perturbation technique used

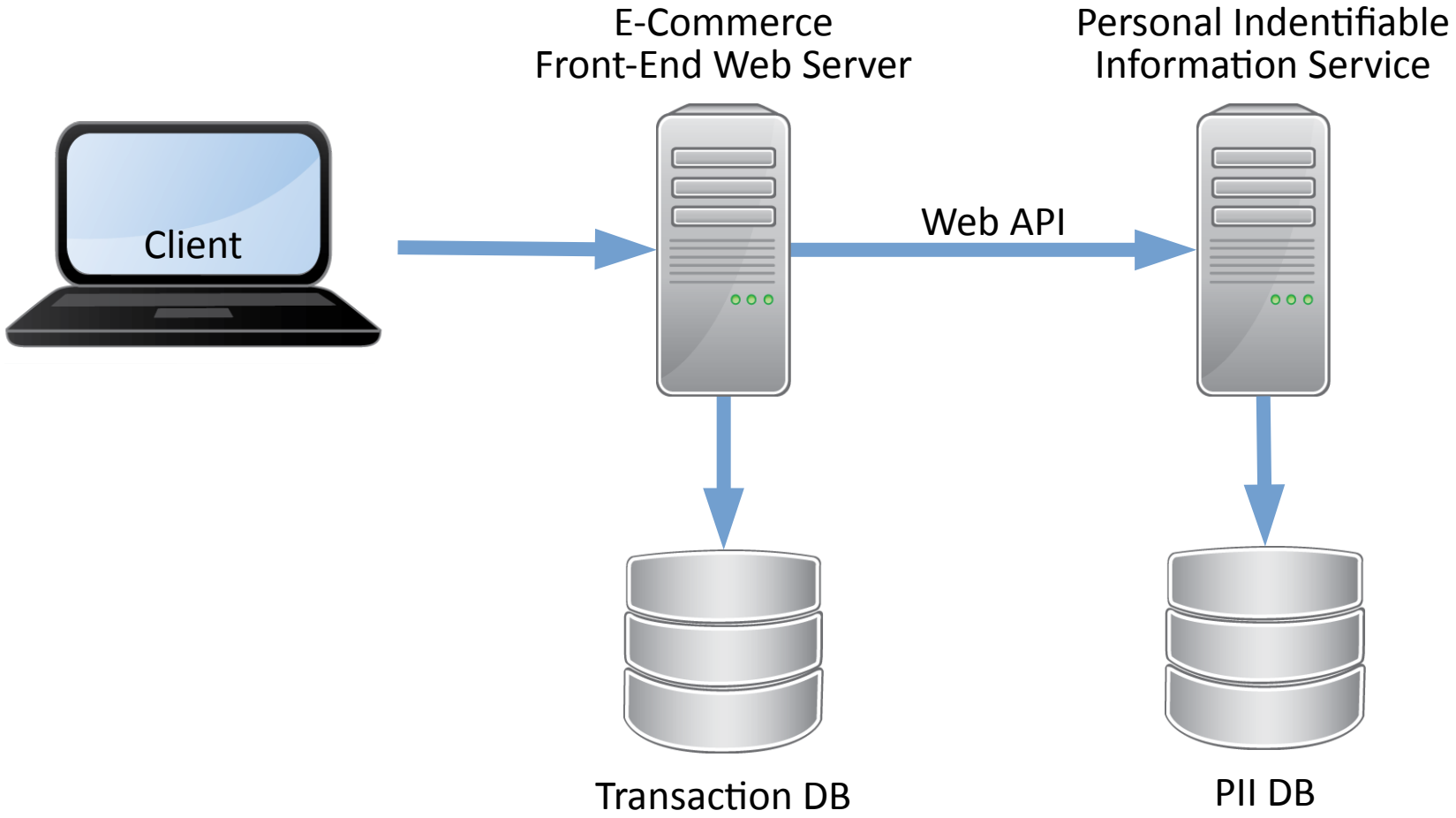
# Other Techniques

- Data Synthetic
- Data Aggregation

# Conclusion: Select the Right Anonymization

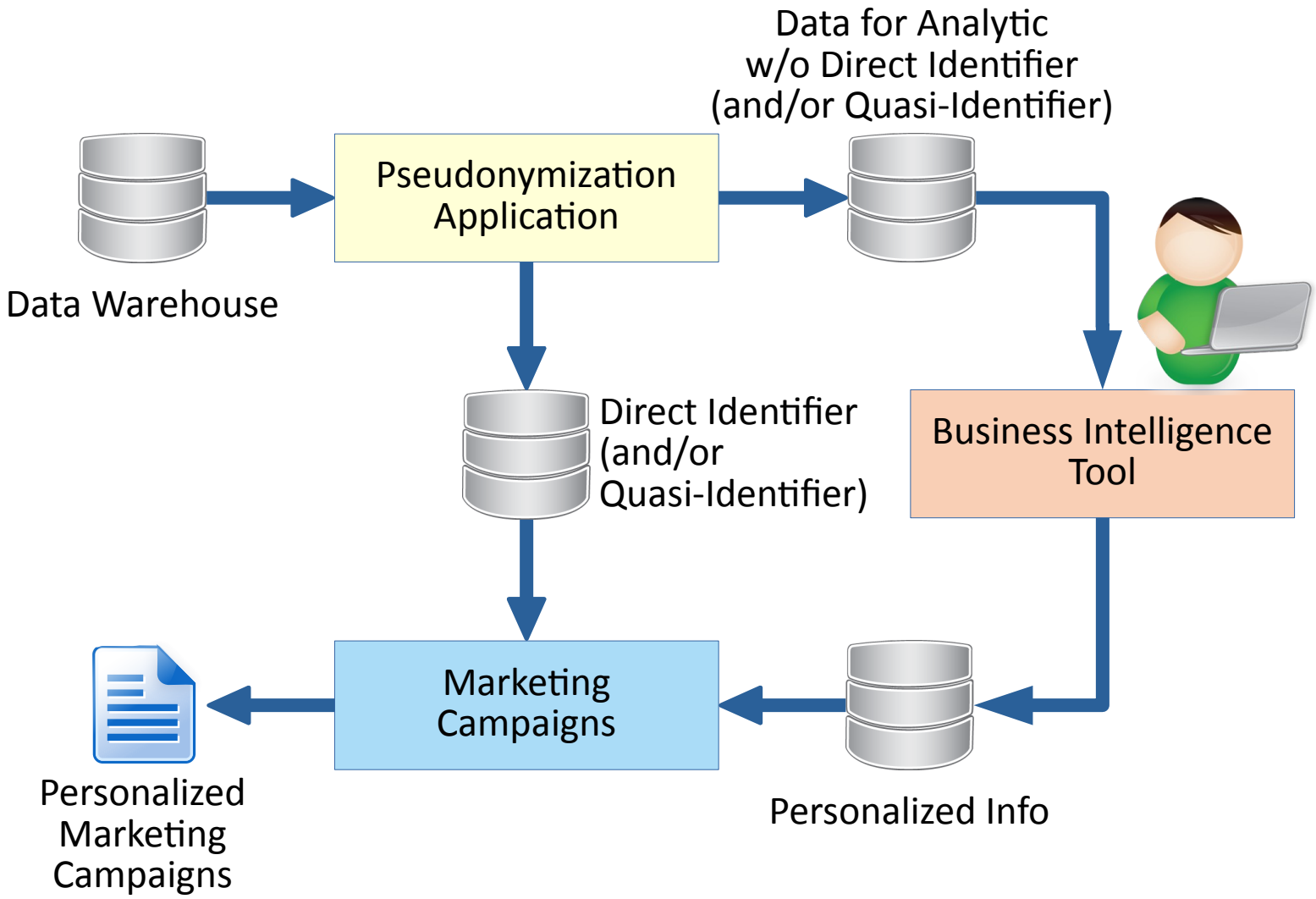
- Purpose of anonymization and its utility
- Characteristics of each anonymization techniques
- Inferred information after implementation
- Expertise with the subject matter
- Competency in anonymization process and techniques
- Recipients

# Example System Design:
## E-Commerce on the Cloud




**Pseudonymization**

# Example System Design:
## Personalized Marketing



# Protect Cardholder Data

## *Requirement 3: Protect stored cardholder data*

Protection methods such as encryption, truncation, masking, and hashing are critical components of cardholder data protection. If an intruder circumvents other security controls and gains access to encrypted data, without the proper cryptographic keys, the data is unreadable and unusable to that person. Other effective methods of protecting stored data should also be considered as potential risk mitigation opportunities. For example, methods for minimizing risk include not storing cardholder data unless absolutely necessary, truncating cardholder data if full PAN is not needed, and not sending unprotected PANs using end-user messaging technologies, such as e-mail and instant messaging.

Please refer to the *PCI DSS and PA-DSS Glossary of Terms, Abbreviations, and Acronyms* for definitions of "strong cryptography" and other PCI DSS terms.

| PCI DSS Requirements | Testing Procedures | Guidance |
|---|---|---|
| **3.1** Keep cardholder data storage to a minimum by implementing data retention and disposal policies, procedures and processes that include at least the following for all cardholder data (CHD) storage:<br>• Limiting data storage amount and retention time to that which is required for legal, regulatory, and/or business requirements<br>• Specific retention requirements for cardholder data<br>• Processes for secure deletion of data when no longer needed<br>• A quarterly process for identifying and securely deleting stored cardholder data that exceeds defined retention. | **3.1.a** Examine the data retention and disposal policies, procedures and processes to verify they include the following for all cardholder data (CHD) storage:<br>• Limiting data storage amount and retention time to that which is required for legal, regulatory, and/or business requirements.<br>• Specific requirements for retention of cardholder data (for example, cardholder data needs to be held for X period for Y business reasons).<br>• Processes for secure deletion of cardholder data when no longer needed for legal, regulatory, or business reasons.<br>• A quarterly process for identifying and securely deleting stored cardholder data that exceeds defined retention requirements. | A formal data retention policy identifies what data needs to be retained, and where that data resides so it can be securely destroyed or deleted as soon as it is no longer needed.<br>The only cardholder data that may be stored after authorization is the primary account number or PAN (rendered unreadable), expiration date, cardholder name, and service code.<br>Understanding where cardholder data is located is necessary so it can be properly retained or disposed of when no longer needed. In order to define appropriate retention requirements, an entity first needs to understand their own business needs as well as any legal or regulatory obligations that apply to their industry, and/or that apply to the type of data being retained. |
| | **3.1.b** Interview personnel to verify that:<br>• All locations of stored cardholder data are included in the data retention and disposal processes.<br>• Either a quarterly automatic or manual process is in place to identify and securely delete stored cardholder data.<br>• The quarterly automatic or manual process is performed for all locations of cardholder data. | |

*(Continued on next page)*

# Example System Design:
## PCI-DSS 3.2

Requirement 3: Protect stored cardholder data

Protection methods such as

- encryption, — Pseudonymization
- truncation, — Recoding
- masking, — Character Masking
- and hashing — Pseudonymization

are critical components of cardholder data protection. If an intruder circumvents other security controls and gains access to encrypted data, without the proper cryptographic keys, the data is unreadable and unusable to that person.


Q&A