IN5540 - Privacy by Design - Module

Tanusan Rajmohan - tanusanr@ifi.uio.no

UNIVERSITY OF OSLO

Autumn 2019

Contents

Learning outcome

After completing this course, you will:

- have knowledge of basic legal privacy concepts and data protection regulations
- have knowledge of security and privacy enhancing technologies
- have knowledge of concepts of privacy by design and privacy impact assessment
- have knowledge of principles of architectural tactics for privacy and privacy patterns
- be able to map legal privacy principles to technical privacy concepts
- be able to relate security and privacy goals to mechanisms and technologies
- be able to apply privacy by design and perform privacy impact assessments
- be able to apply appropriate architectural tactics for privacy and privacy patterns

Part I

Privacy- Philosophy and legislation, GDPR

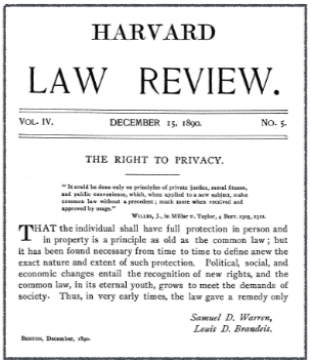
1 The Concept of Privacy - History and Definitions

1.1 Early History of Privacy

- **Aristotle:** Public sphere of politics (polis) vs. private/domestic sphere of family (oikos)
- **Long practiced concept**, e.g.:Hippocratic Oath, Seal of confessions, Secrecy of letter correspondence (e.g., in Prussia since 1712)

First Definition by Lawyers – Warren & Brandeis 1890

"The right to be let alone"



HARVARD

LAW REVIEW.

VOL. IV. DECEMBER 15, 1890. NO. 5.

THE RIGHT TO PRIVACY.

"It would be deep only as principles of private justice, moral fitness, and public convenience, which, when applied to a new subject, make common law without a precedent; much more when received and approved by usage."

WILLES, J., in Millar v. Taylor, 4 Burr. 2303, 2312.

THAT the individual shall have full protection in person and in property is a principle as old as the common law; but it has been found necessary from time to time to define anew the exact nature and extent of such protection. Political, social, and economic changes entail the recognition of new rights, and the common law, in its eternal youth, grows to meet the demands of society. Thus, in very early times, the law gave a remedy only

Samuel D. Warren,
Louis D. Brandeis.

Boston, December, 1890.

1.2 Construction of privacy by the EU

European Convention on Human Rights (1950): Article 8 - Right to respect for private and family life (and right to information). Everyone has the right to respect for his private and family life, his home and his correspondence

1.3 Definition - Alan Westin

("Privacy and Freedom", 1967)
"Privacy is the claim of individuals, groups and institutions to determine for themselves, when, how and to what extent information about them is communicated to others"

1.4 Definition - Census Decision

(German Constitutional Court, 1983)

- **Right to Informational Self-Determination:** *Right of individuals to make their own decisions as regards the disclosure and use of their personal data.*
- Derived from the basic rights of **Human dignity** & to **Self-Determination** (Art. 1 (1) & 2 (1) GG)
- Fundamental right & **elementary prerequisite for the functioning of a free democratic society**

1.5 Construction of privacy by the EU (II)

Charter of the Fundamental Rights of the European Union (2000):

Article 7 - Respect for private and family life

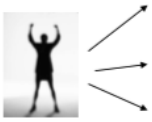
Everyone has the right to respect for his or her private and family life, home and communications.

Article 8 - Protection of personal data

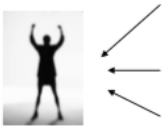
1. Everyone has the right to respect of personal data concerning him or her.
2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis [laid down by law. 5/35].

1.6 Privacy Dimensions

- Informational Privacy
- Privacy of communications



- Spatial privacy
- Territorial privacy
- Bodily privacy



2 Basic Privacy Principles

2.1 Basic Privacy Principles

(part of the OECD Privacy Guidelines & most Privacy/Data Protection Laws)

- **Lawfulness** of processing, e.g. by **Informed Consent** (c.f. OECD Collection Limitation Principle)
- **Data Minimization & Avoidance** (c.f. OCED Data Quality Principle)
  - Data should be adequate, relevant and not excessive
  - Minimization of data collection, use, sharing, linkability, retention

- **Purpose Specification & Purpose Binding** (c.f. OECD Purpose Specification Principle & Use Limitation Principle)
  - "Non-sensitive" data do not exist !

  Examples of Purpose Misuse ("function creep"):
  - Lidl Video Monitoring Scandal (2006)
  - Loyality Card Data use against customer interests
- **Transparency** and **Intervenability** (c.f. OECD Openness Principle & Individual Participation Principle)
- Appropriate **Security** (c.f. OECD Security Safeguards Principle)
- **Accountability** (c.f. OECD Accountability Principle)

3 Overview to Privacy & Data Protection Laws

3.1 Privacy Legislation - History

1950	Art. 8 ECHR: *"Everyone has the right to respect for his private and family life, his home and his correspondence."*
1970	First data protection act: "Hessisches Datenschutzgesetz"
1973	Swedish Data Act
1980 (updated in 2013)	OECD Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data
1981	The Council of Europe's Convention for the Protection of Individuals
1995	EU Data Protection Directive 95/46/EC
1998	Revised Swedish Data Act (PUL - Personuppgiftslagen)
2000	Safe Harbor Principles **(Sept. 2015: declared as invalid by European Court of Justice)** – replaced by the Privacy Shield in June 2016

2000	EU Charter of the Fundamental Rights recognises Privacy (Art. 7) & Data Protection (Art. 8)
2002	EU ePrivacy Directive 2002/58/EC (amended in 2009 – "Cookie Directive" 2009/136)
2006	EU Data Retention Directive 2006/24/EC **(April 2014: declared as invalid by European Court of Justice)**
since 2012	Reform of EU Data Protection Legislation – Draft EU General Data Protection Regulation (GDPR) & EU Police Data Protection Directive
April 2016	GDPR & Police Data Protection Directive 2016/680 were adopted. - GDPR will apply from 25 May 2018 - Directive 2016/680 enters into force 5 May 2016
January 2017	Draft ePrivacy Regulation adopted

3.2 Swedish Sectoral Data Protection laws (lex specialis) - Examples

- **Patient Data Act (2008:355)** (Patientdatalagen) **and the Pharmacy Data Act (2009:367)** (Apoteksdatalag)
- **Credit Information Act (1974:182)** (Kreditupplysningslagen) and the **Debt Recovery Act (1974:182)** (Inkassolagen)
- **Camera Surveillance Act (2013:460)** (Kameraovervakningslagen) - A Revision will enter into force in May 2018.

Note: All sectoral data protection laws are currently under revisions for achieving compliance with the GDPR.

3.3 General Data Protection Regulation/Freedom of the Press Act

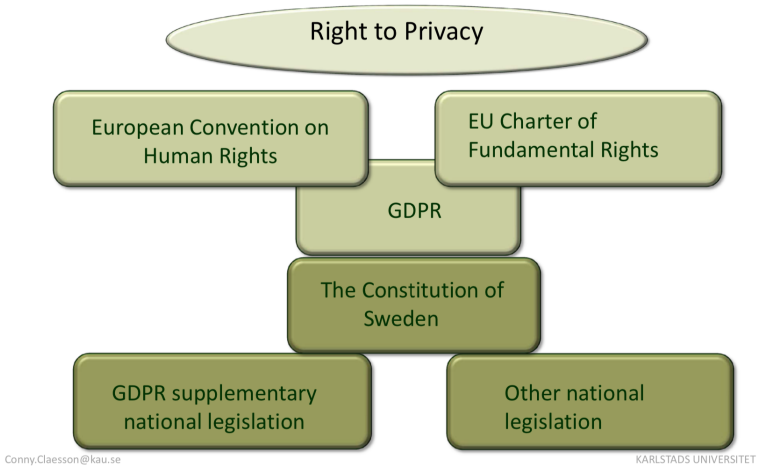
Protection of personal data and Public Access to Information

3.3.1 Public Access to Information / Right to Privacy

- Freedom to the Press Act: Every Swedish citizen are entitled to have free access to official documents held by public authorities
  - Except: Some official documents containing sensitive information are secret (assessment of potential damage or harm must be considered in each individual case)

  vs.

  - General Data Protection Regulation (GDPR): Contains general provisions about the processing and protection of all kinds of personal data



3.3.2 GDPR and Public Access to Information

- GDPR replaces PUL (The Personal Data Act, the implementation of the 1995 EU Data Protection Directive)
- EU regulation - Takes precedence over national legislation (a new game board)
- Does not replace Freedom of the Press Act
  - Personal data in official documents may be disclosed (art 86 GDPR) *But:*
  - A public authority is not obliged to provide official documents in electronic form.
  - The Public Access to Information and Secrecy Act states that secrecy shall apply to personal data, if it can be assumed that disclosure would cause the data to be processed in violation of GDPR

3.3.3 Proposed complementary legislation (Sweden):

- Data Protection Act
- Research Data Act
- Adjustment of other legislation

3.3.4 Swedish Research Data Act proposes:

- Pseudonymisation or equally strong protection necessary to process personal data for research purposes
- Ethical review necessary to process sensitive personal data for research purposes (Ethical Review Act)

4 Privacy Issues

4.1 Privacy Issues - Overview

4.2 Location tracking

Privacy Risks:

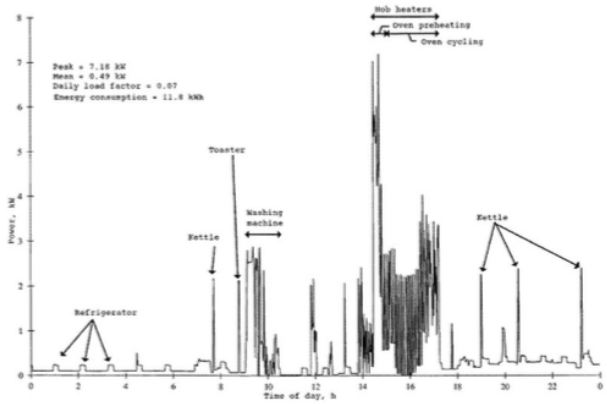
- Movement profiles
- Disclosure of the user's current context
- Activity recognition
- Disclosure of social networks

4.3 (Secret) Mobile Tracking via MAC addresses

4.4 Smart Grids/Smart Meters

- "The EU aims to [introduce] smart meters by 2020 [to] reduce emissions in the EU by up to 9% [...]"
- Optimised energy use in reaction the the households' energy use patterns

4.5 Smart meters and privacy risks

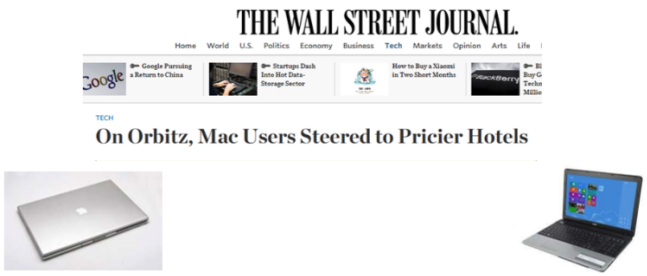


- Smart meters **"leak" information"** about individual's **lifestyles**
- 2009: Dutch government revokes plans for a mandatory smart meter deployment under **consumer pressure**
- New smart meter projects must balance privacy and energy saving by **PbD & PIA**

4.6 Privacy Risks & challenges of Big Data

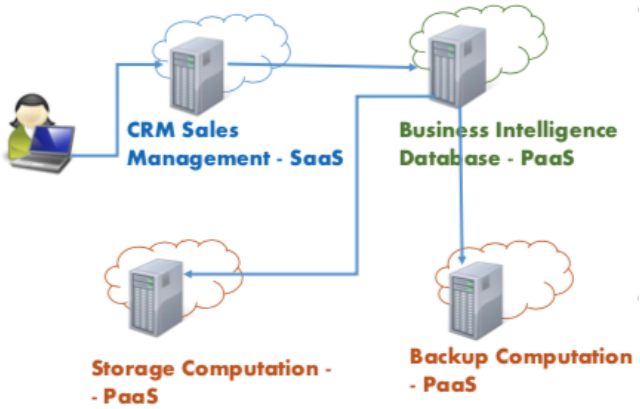
- The sheer scale of data collection, tracking, profiling & detail of the data, from many different sources
- Security of data
- Transparency
- Inaccuracy, discrimination, exclusion & economic imbalance
- Increased possibilities of government surveillance

4.7 Price Discrimination



4.8 Predictive Policing

4.9 Lack of control related to Cloud Computing



- Lack of Transparency in regard to Cloud Service's operations:
  - Chain with multiple processors & subcontractors
  - Different geographic locations within the EEA
  - Disclosure requests by law enforcement
- Lack of Intervenability:
  - Lack of tools provided for exercising data subjects' rights

5 GDPR

5.1 EU General Data Protection Regulation (GDPR) - Background

- Enters into affect on May 25th 2018 & replaces EU Directive 95/46/EC

Objectives:

- To **harmonize** data protection laws across Europe
- **Modernisation** of Data Protection Rules of the Digital Age
- **Strengthening** the existing **rights** and empowering individuals with more control
- Improved **level of compliance**

5.1.1 Scope

Lex generalis
Material Scope (Art. 2):

- Applies to the processing of **personal data** wholly or partly by automatic means, ...
- **Some main exceptions:** GDPR does not apply for data processing for:
  - **Public security, criminal law enforcement**
    - EU Police Data Protection Directive 2016/680 may apply
  - **Purely private or household activity ("household exemption")**
    - Note: Publishing data on a website or to a broad audience on Social Networks is not falling under the household exemption
      - *European Court of Justice decision in the Bodil Lindqvist case, 2003*
      - *Art. 29 WP Opinion 5/2009 on online social networking*

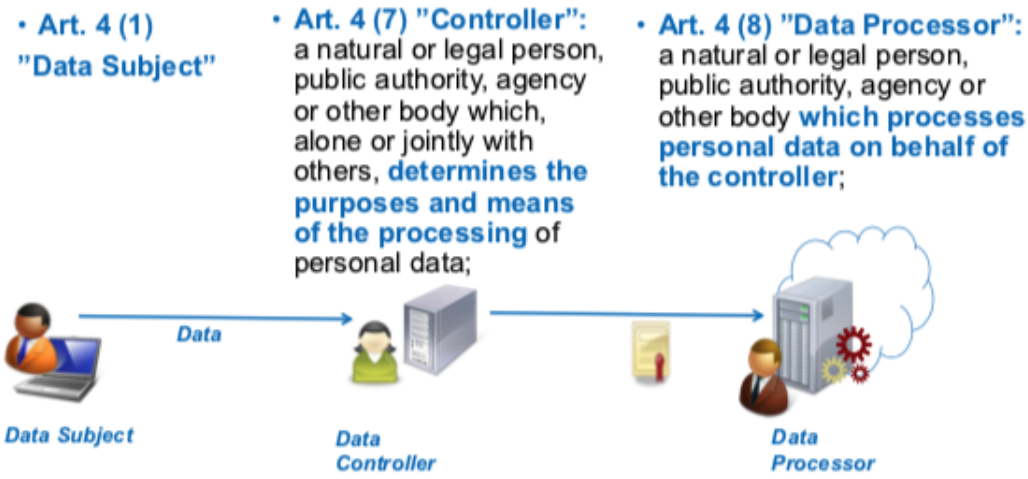
5.1.2 Territorial Scope

(Art. 3): The GDPR applies to personal data processing by controller/processor:

1. established in the EU
2. outside the EU that offer goods and services to, or that monitor, individuals in the EU
3. in places where EU Member State law applies by virtue of public international law.

5.1.3 Definitions

- Art.4 (I): **'Personal data'** means any information relating **to an identified or identifiable natural person ('data subject');**
- **Types of personal data:**
  - Explicitly disclosed data (e.g., name, delivery address)
  - Implicitly disclosed data incl. meta data (e.g., IP address, MAC address, cookies, location data, traffic data)
  - Derived data (e.g., user behavioral profiles)
  - Third party provided data (e.g., reputation scores)



5.1.4 Principles relating to processing of personal data

(Art. 5):

- lawfulness, fairness and transparency
- purpose limitation
- data minimisation
- data accuracy
- storage limitation
- integrity and confidentiality
- accountability

5.2 Lawfulness and Consent

5.2.1 Lawfulness of processing conditions

(Art. 6):

- Consent of the data subject

  or processing is necessary:

- for the performance of a contract with the data subject
- for compliance with a legal obligation
- to protect the vital interests of a data subject or another person
- for the performance of a task carried out in the public interest
- for the purposes of legitimate interests pursued by the controller or a third party

5.2.2 Lawfulness of processing "special categories of data"

(Art. 9(1)): processing of **"special Categories of Data" about**

- racial or ethnic origin
- political opinions
- religious or philosophical beliefs, or trade union memberships,
- genetic data, biometric data for uniquely identifying a person
- health, sex life or sexual orientation

shall be **prohibited.**

Exceptions (Art. 9 (2)):

- **Explicit** Consent
- Compliance with legal obligations related to employment and social security
- Vital interests
- Data manifestly made public by the data subject
- Legal claims
- Medicine, public health, research

5.2.3 Consent

(Art. 4 (11)): Consent has to be

- **Freely given** ($\rightarrow$ free choice, no negative consequences if no consent is given, unbundled)
- **Specific** ($\rightarrow$ for specific purpose(s), separate opt-in for each purpose)
- Unambigius indication of an agreement, by a statement or clear **affirmative action**
  - deliberate action, no pre-ticked opt-in boxes or opt-out constructions
- **Informed** at least about:
  - controller's identity,
  - purposes,
  - type of data
  - right to withdraw consent,
  - any use for decisions based solely on automated processings, risks of data transfers to third countries

5.2.4 Conditions of Consent

(Art. 7):

- Controller needs to keep evidence that the data subject consented
- Data Subject has the right to withdraw consent at any time
- Withdrawal shall be as easy as to give consent

5.3 Data Subjects Rights

5.3.1 Overview to Data Subject Rights

Transparency Rights:

- Right to Information (ex ante)
- Right to Access (ex post)
- (Data Breach Notification)

Intervenability Rights:

- Right to rectification
- Right to erasure ("Right to be forgotten")
- Right to restriction of processing
- Right to data portability
- Right to object to marketing & profiling
- (Right to withdraw consent)
- (Right to lodge complaint with supervisory authority)

5.3.2 Transparency

(Art. 12): General modalities:

- Information to be provided needs to be *concise, transparent, intelligible, in easily accessible form, using a clear and plain language* (→ Art. 29 WP recommends **multi-layered privacy policies**)
- Should be *provided by electronic means* for electronic requests
- May be provided in combination with **standardized (machine readable) icons**

5.3.3 Transparency Information

(Art. 13) ex ante Transparency & (Art. 15) ex post Transparency / Right of Access:

- Identity of controller & contact details (incl. DPO)
- Purposes & legal basis for processing
- Data recipients
- International data transfers
  and in addition:
- Storage retention periods
- Data subjects rights
- Existence of automated decision making, logic involved & significance and envisioned consequences

(preferably to be presented in multi-layered privacy policies
Right of Access: Controller shall provide an (electronic) copy of the data (Art. 15(2))
Art. 15 (4) & Recital 63 - Rights & freedoms of others, incl. trade secrets, intellectual property & copyright protection should not adversely be affected

5.3.4 Right to Rectification & Erasure

(Art. 16): Right to rectification of inaccurate personal data without undue delay

(Art. 17): Right to erasure ("Right to be forgotten") without undue delay, if:

- data are no longer necessary
- withdrawal of consent & no legal basis
- objection
- unlawful processing
- required for compliance with legal obligation

If made public, reasonable steps need to be taken to inform other controllers that process such data

(Art. 19): Notification obligation to each data recipient

5.3.5 Right to Data Portability

(Art. 20): Right to Data Portability:

- **Processing based on consent or contract - Art. 6 (1) (a), (b), 9 (2)**
- Right to receive disclosed personal data in a structured, commonly used and machine-readable format &
- Right to transmit those data to another controller, or
- Right to have data transmitted directly between controllers

Note: This does not include derived data or data received by third parties (e.g. reputation scores).

5.3.6 Right to Object

(Art. 21): Right to object to

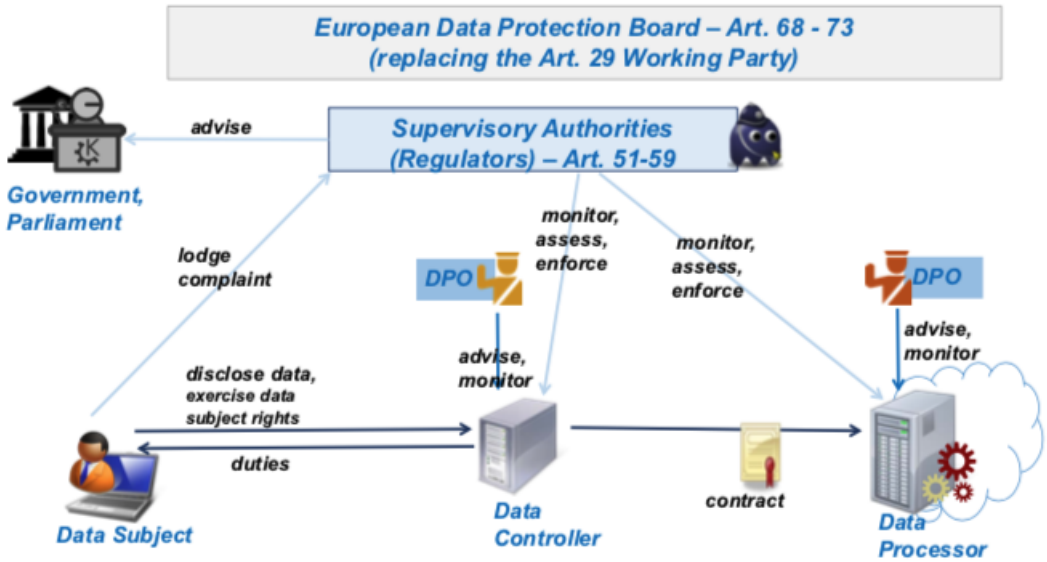
- Data processing (incl. profiling) - based on legitimate interests or the performance of a task of public interest
- Direct marketing (incl. profiling); and
- Data processing for scientific/historical research purposes and statistics

(Art. 22): Right not to be subject to a decision based solely on automated processing, which:

- produces legal effects concerning him/her
- significantly affects him/her

5.4 Obligations & Rules

5.4.1 Advising, Monitoring, Enforcing



5.4.2 Clear Rules for Business

- **One single set of rules** - Which will make it simpler / cheaper for companies to do business in the EU
- **One-stop-shop** - businesses will only have to deal with one single (lead) supervisory authority.
- **European rules on European soil** - companies based outside of Europe will have to apply the same rules when offering services in the EU.
- **Risk-based approach** - measures tailored to the respective risks.

5.4.3 Obligations - Controller

- Implement **appropriate technical & organizational data protection measures** (Art. 24, 25)
- built into products and services from the earliest stage of development (**Data Protection by Design** - Art. 25 (1))
- to ensure that only the data **necessary** should be processed, short stage period, limited accessibility (**Data Protection by Default** - Art. 25 (2))
- Select only processors with sufficient guarantees to implement appropriate technical & organizational measures (Art. 28)

5.4.4 Obligations - Controller (II)

- **Data breach notification** to
  - the supervisory authority (Art. 33) - without undue delay & within 72 hours if feasible (Art. 33)
  - the data subject - in case of high risk to their rights and freedom (Art. 34)
- **Data Protection Impact Assessment** (Art. 35) - for high risk data processing
- **Prior Consultation** (Art. 36) - with supervisory authority

5.4.5 Obligations - Processor & Controller

- Processing by processor governed by **contract** or **legal act** (Art. 28)
- **Security of Processing** (Art. 32)
  - Appropriate measures, such as pseudonymisation and/or encryption for protecting Confidentiality, Integrity and Availability
- Main **records of processing activities** (Art. 30)
- Designate a **data protection officer - DPO** (Art. 38)
  - Unless data processing is not their core business activity.

5.4.6 Data Transfers to Third Countries

(Art. 45): Adequacy:
Personal data can only be transferred to third country, where the Comission has decided an "**addequate level of data protection**".

- **Special adequacy decisions: Privacy Shield**
  - Privacy shield replaced Safe Harbor after CJEU 2014 Decision on Screms vs. Facebook
  - However: Concerns by EDPS & Art. 29 Working Party
- Examples of exceptions:
  - Standard contractual clauses (Art. 46)
  - Binding corporate rules (BCRs - Art. 47)
  - Explicit consent (Art. 49)

5.4.7 Administrative Fines

(Art. 83): Supervisory Authority shall impose administrative fines for infringements of the GDPR, which shall be **effective, proportionate and dissuasive**.

Two tier structure:

- Greater of 10 Million € or 2% of global turnover
- Greater of 20 Million € or 4% of global turnover (for serious breaches)

6 ePrivacy Regulation

6.1 Draft ePrivacy Regulation - Background

- **Proposed in January 2017 by EU Comission**
- **Lex specialis to GDPR** - specific privacy rules for electronic communications;
- **Directly applicable** Regulation, replacing the ePrivacy ("Cookie" Directive);
- Applies now also to services, such as Skype, Whatsapp, Facebook, Messenger, Gmail...

6.2 Proposed main rules

- Privacy rules for communication content & metadata
- Simpler rules on cookies:
  - No consent needed for non-privacy intrusive cookies
  - Privacy-friendly browser settings rule
- Protection against SPAM
- Traditional telecom services may process communication data for providing additional services or business development;
- Enforcement by DPAs in charge of the rules under the GDPR;

6.3 "Grave" Concerns by the Art. 29 Working Party with regard to

- Tracking of the location of terminal equipment;
- Conditions under which the analysis of content & metadata is allowed;
- Tracking walls;
- Default settings of terminal equipment/software.

7 Mapping privacy principles to technology

7.1 Principles relating to processing of personal data

Privacy Principles (Art. 5 GDPR):

- lawfulness, fairness and transparency
- purpose limitation
- data minimisation
- data accuracy
- storage limitation
- integrity and confidentiality
- accountability

Algorithmic Transparency & Ethical Data Management, Fairness by Design, Ex ante and ex post Transparency Enhancing Tools,…

Privacy Principles (Art. 5 GDPR):

- lawfulness, fairness and transparency
- purpose limitation
- data minimisation
- data accuracy
- storage limitation
- integrity and confidentiality
- accountability

Privacy policy languages, sticky policies, Functional separation,…

Privacy Principles (Art. 5 GDPR):

- lawfulness, fairness and transparency
- purpose limitation
- data minimisation
- data accuracy
- storage limitation
- integrity and confidentiality
- accountability

PETs for achieving Anonymity / Pseudonymity on:

Communication level: Mixes, DC-nets, Tor, Steganographic Tools …

Application level: Anonymous Credentials, Anonymous Payment schemes, Private Information Retrieval,…

Data level: k-Anoymity & Differential Privacy tools/applications, Secret Sharing,….

Privacy Principles (Art. 5 GDPR):

- lawfulness, fairness and transparency
- purpose limitation
- data minimisation
- data accuracy
- storage limitation
- integrity and confidentiality
- accountability

Integrity controls, Intervenability tools allowing online corrections/deletions,…

Privacy Principles (Art. 5 GDPR):

- lawfulness, fairness and transparency
- purpose limitation
- data minimisation
- data accuracy
- storage limitation
- integrity and confidentiality
- accountability

Anonymisation tools (e.g., k-anonymity based), Obligations for data deletions in privacy policy/access control languages,…

Privacy Principles (Art. 5 GDPR):

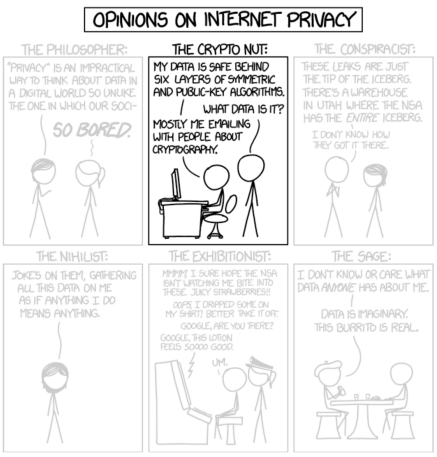
- lawfulness, fairness and transparency
- purpose limitation
- data minimisation
- data accuracy
- storage limitation
- integrity and confidentiality
- accountability

Security controls, incl. access control, authentication, encryption (e.g., TLS), logging,…

Part II

Privacy Enhancing Technology (PET)

8 Introduction



Have Objectives:

- Control over Personal Data
- Transparency
- Lawful Processing of Data
- Data Minimization / Avoidance
- Data Security and Integrity
- Identity Management

8.1 What are PETs?

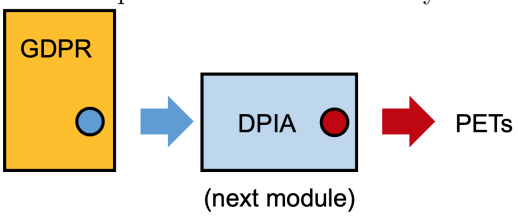
8.1.1 Technical Means for Protecting Personal Data

Standards & Protocols → Tools & Mechanisms

- Transport Layer Security (RFC 5246)
- XACML (eXtensible Access Control Markup Language)
- Tor, Mixes, etc.
- Lets Encrypt
- Secure Messaging apps
- Browser extensions, etc.

8.1.2 PETs and the GDPR

to be compliant with the GDPR → you need to know PETs



8.2 Security Technologies

Have Objectives:

- Confidentiality
- Integrity
- Availability
- Authentication
- Authorization
- Accounting

8.2.1 Technical Means for Securing Data

Data Security

- Confidentiality, Integrity, Availability
- Authentication, Authorization, Account

requirement

Implementing Data Protection

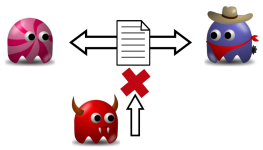
- Control over Personal Data
- Data Minimization / Avoidance
- Identity Management
- Lawful Processing of Data

8.2.2 Confidentiality

Information Not available or disclosed to unauthorized parties

- Stored Data

- Data in Transit



8.2.3 Integrity

Information Not modified by unauthorized parties or in an unauthorized manner

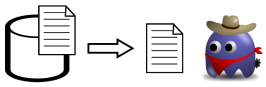
- Unauthorized Parties

- Unauthorized Manner

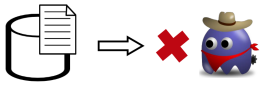
8.2.4 Availability

Information available when needed

- Available



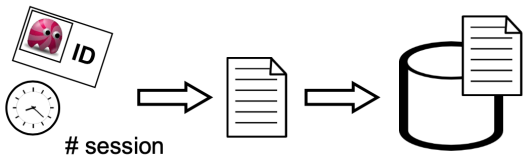
- NOT Available



8.2.5 Accounting

Keeping track of information (users and data)

- Building and storing log data

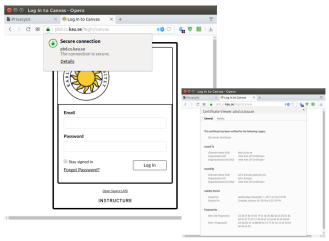


8.2.6 Authentication

- Assurance of an identity claim

Are you really who you claim to be?

- Digital certificates



- ID cards

8.2.7 Authorization

- Grant or deny access to resources

operations over resources

(once authenticated)

- Authorized

- NOT Authorized

8.3 Why Do We Need Technologies?

8.3.1 Security and Privacy: Multiple Facets

Legislation and Regulation

- EU GDPR, US Computer Fraud and Abuse act

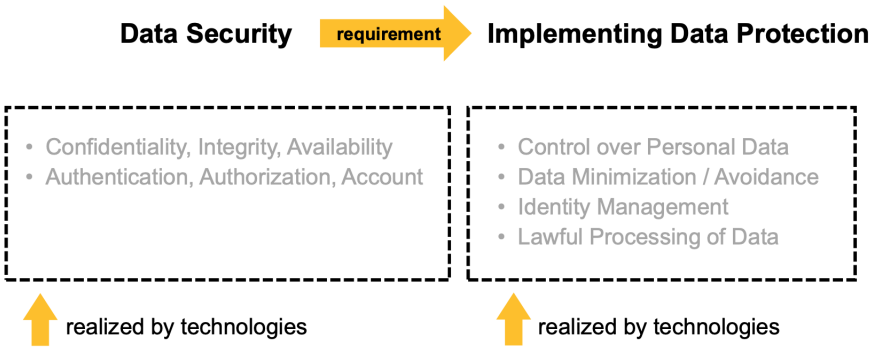
Procedures and Standards

- ISO 27000 (Security Management)

Security and Privacy Technologies

implement + realize data protection

8.3.2 Technical Means for Protecting Data



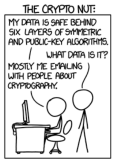
8.4 What do we need to know?

What are the available tools? **What problem do they solve?**

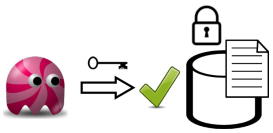
8.4.1 Example

• Problem:
How to provide data confidentiality?

• Hint:



• Solution:
Encryption



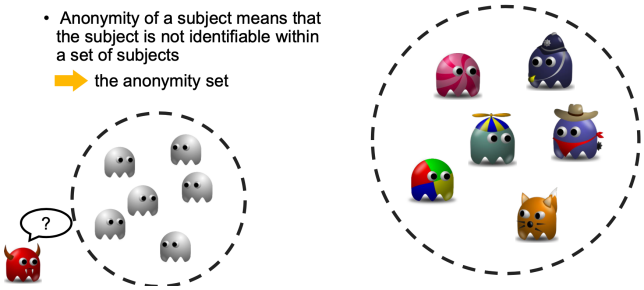
8.5 Pfitzmann and Hansen's Terminology

8.5.1 A terminology for talking about privacy by data minimization

- Anonymity
- Pseudonymity
- Identity Management
- Unlinkability
- Undetectability
- Unobservability

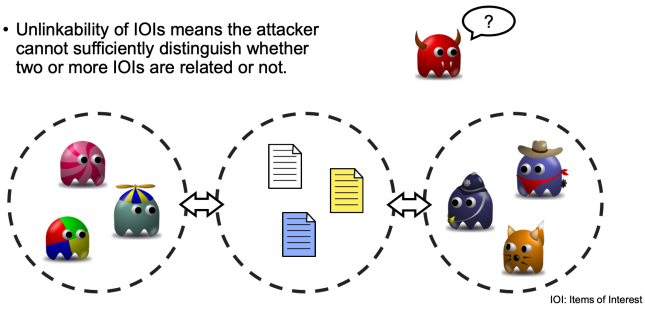
8.5.2 Anonymity

- Anonymity of a subject means that the subject is not identifiable within a set of subjects
  - ➡ the anonymity set



8.5.3 Unlinkability

- Unlinkability of IOIs means the attacker cannot sufficiently distinguish whether two or more IOIs are related or not.

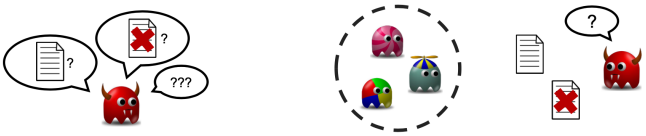


8.5.4 Undetectability

- Undetectability of an IOI means that the attacker cannot sufficiently distinguish whether it exists or

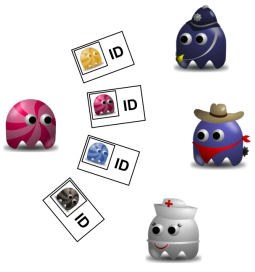
8.5.5 Unobservability

- Unobservability of an IOI means:
  - undetectability of the IOI against all subjects uninvolved in it and
  - anonymity of the subject(s) involved in the IOI even against the other subject(s) involved in that IOI



8.5.6 Pseudonymity

- A pseudonym is an identifier of a subject other than one of the subject's real names.
- multiple types of pseudonyms
  - ➡ check the literature!



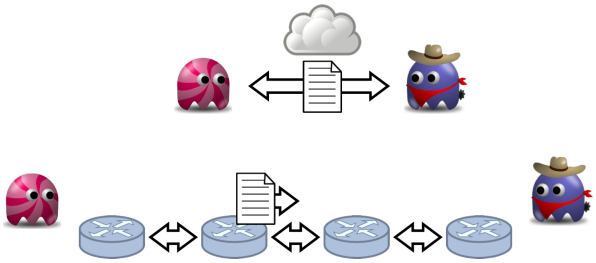
8.5.7 Identity Management

- Managing various partial identities
  - ➡ pseudonyms
- Privacy-enhancing IDM
  unlinkability between partial IDs

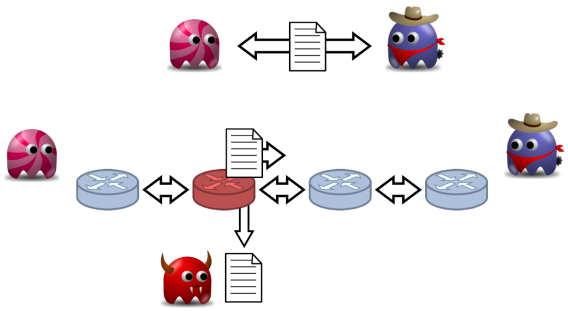
9 Secure Communication

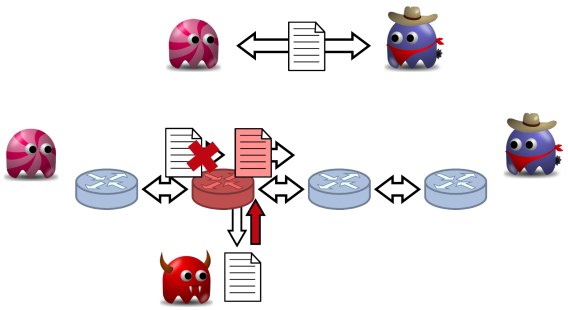
9.1 Why Do We Need Secure Communications?

9.1.1 Information Over the Internet



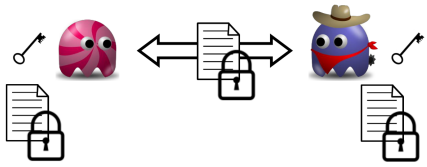
9.1.2 The Internet is Shared Medium





9.1.3 Secure Communications

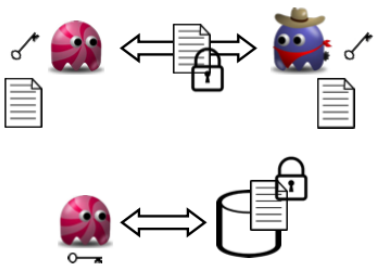
- Enforce: confidentiality, integrity and authentication → Encryption



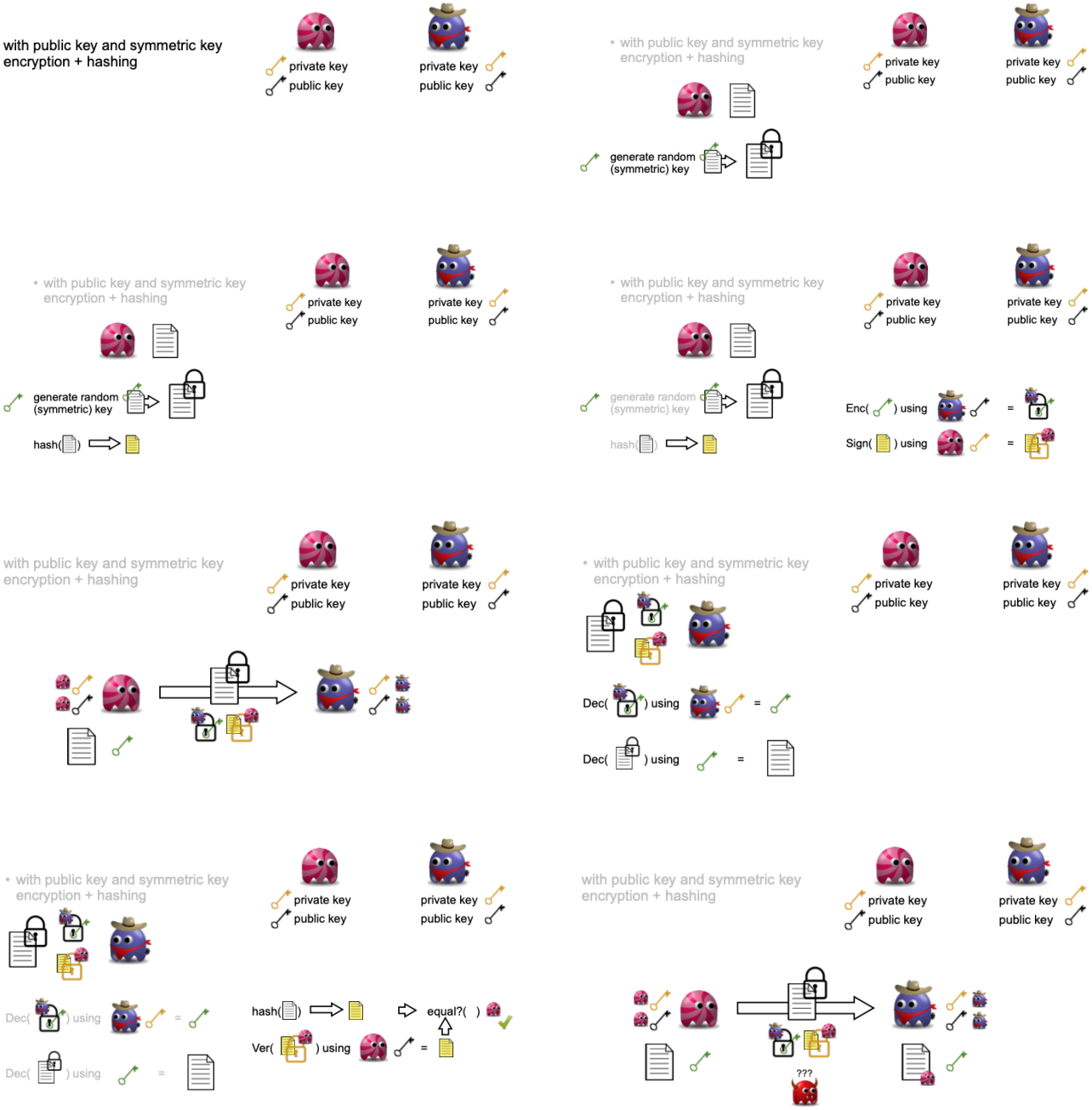
9.2 Pretty Good Privacy (PGP)

Security tool for

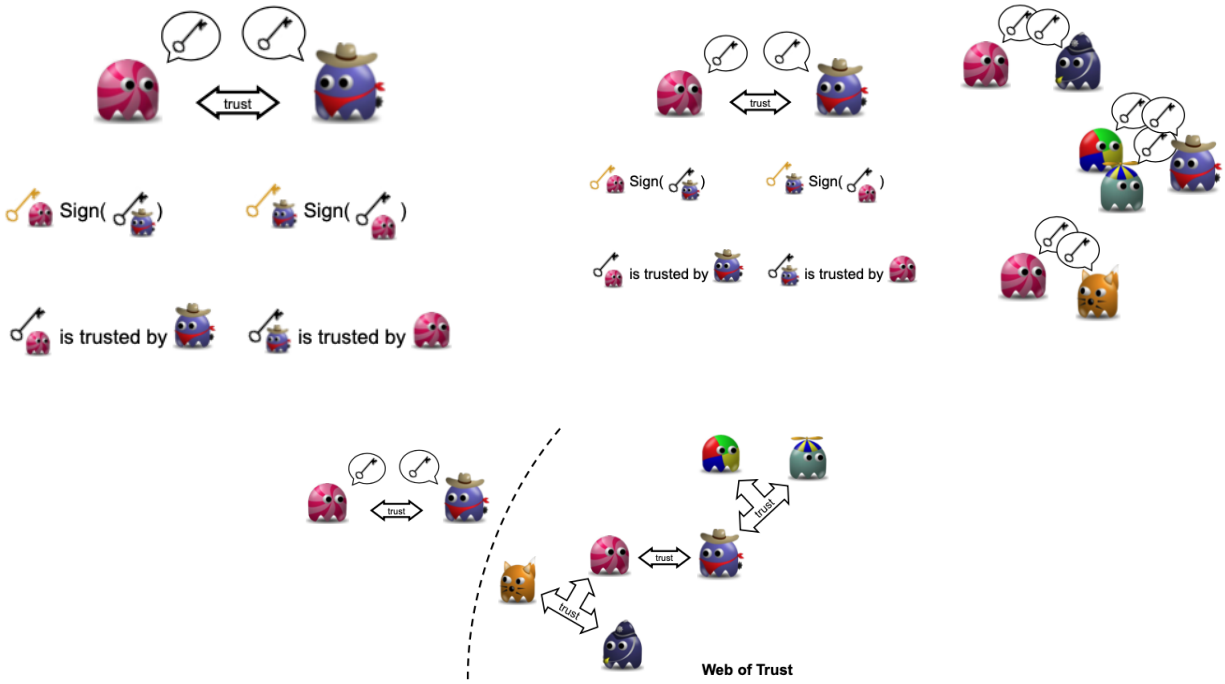
- confidentiality
- integrity
- authentication



9.2.1 How does PGP work?



9.2.2 Public Key Distribution: Web of Trust



9.2.3 PGP in the Real World

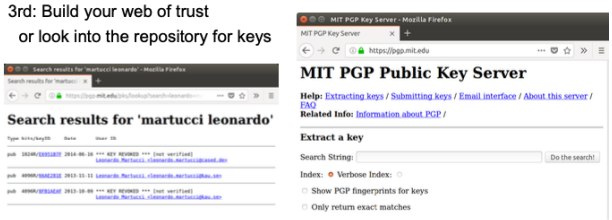
Exchanging encrypted email ⇨ with OpenPGP (RFC 4880)

1st: generate a key pair — with ssh-keygen
e.g. ssh-keygen -t rsa -b 4096

2nd: upload to a public repository — e.g. the MIT PGP Public Key Server

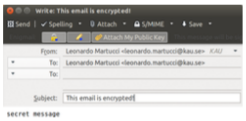
Exchanging encrypted email ⇨ with OpenPGP (RFC 4880)

3rd: Build your web of trust
or look into the repository for keys



Exchanging encrypted email ⇨ with OpenPGP (RFC 4880)

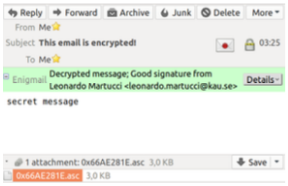
4th: start encrypting and signing your emails! ⇨ suggestion: with the support of an email client extension e.g. Enigmail

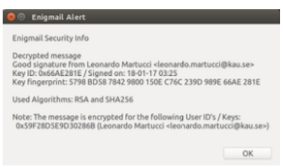


⇦ enigmail toolbar

Exchanging encrypted email ⇨ with OpenPGP (RFC 4880)

5th: receiving emails ⇨ suggestion: with the support of an email client extension e.g. Enigmail





9.2.4 The Caveats

PGP is

- old (from the 90's)
- not really usable
- keys are really long
- no key management
- no forward secrecy

9.3 Transport Layer Security (TLS)

9.3.1 TLS in a Nutshell

Properties of the secure channel

- Confidentiality (forward secrecy)
- Integrity
- Server authentication
- Optional client authentication

9.3.2 History

Secure Sockets Layer (SSL) in 1̃996
TLS versions

- 1.0 in 1999
- 1.1 in 2006
- 1.2 in 2008
- 1.3 as draft in January 2018

A *subset* of TLS 1.1 and 1.2 secure today

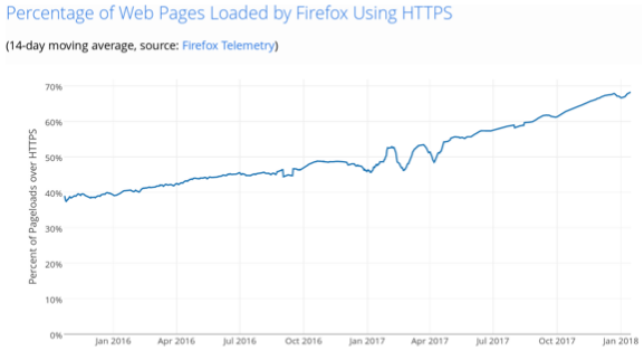
- Depends on selected algorithms
- ...and implementation!

9.3.3 TLS is hard

Product	CVE ID	Issue source
OpenSSL	2013-4353, 2015-0206, 2014-[3567, 3512, 3569, 3508, 3470, 0198, 0160]	Memory management
	2015-0205, 2015-0204, 2014-3572, 2014-0224, 2014-3568, 2014-3511	State machine
	2014-8275	Certificate parsing
	2014-2234	Certificate validation
	2014-3509, 2010-5298	Shared mutable state
	2014-0076	Timing side-channel
	2014-3570	Wrong sqrt
GnuTLS	2014-8564, 2014-3465, 2014-3466	Memory management
	2014-1959, 2014-0092, 2009-5138	Certificate validation
NSS	2014-1544	Memory management
	2013-1740	State machine
	2014-1490	Shared mutable state
	2014-1569, 2014-1568	Certificate parsing
	2014-1492	Certificate validation
	2014-1491	DH param validation
SChannel	2014-6321	Memory management
Secure Transport	2014-1266	State machine
JSSE	2014-6593, 2014-0626	State machine
	2014-0625	Memory exhaustion
	2014-0411	Timing side-channel
Applications	2014-2734	Memory management
	2014-3694, 2014-0139, 2014-2522, 2014-8151, 2014-1263	Certificate validation
	2013-7373, 2014-0016, 2014-0017, 2013-7295	RNG seeding
Protocol-level	2014-1771, 2014-1295, 2014-6457	Triple handshake
	2014-3566	POODLE

Table 1: Vulnerabilities in TLS implementations in 2014.

9.3.4 Adoption and Use as HTTPS



9.3.5 Improvements to TLS

TLS 1.3 makes the protocol *simpler*

- No legacy algorithms, compression, renegotiation
- Encrypt as much as possible of handshake
- Split between authentication, key exchange, cipher

Significant efforts around implementation

- Remove old code, support for old versions, architectures, refactoring
- BoringSSL
- OpenSSL cleanup (Levchin prize 2018)
- Formal methods to verify implementations

9.3.6 Conclusions

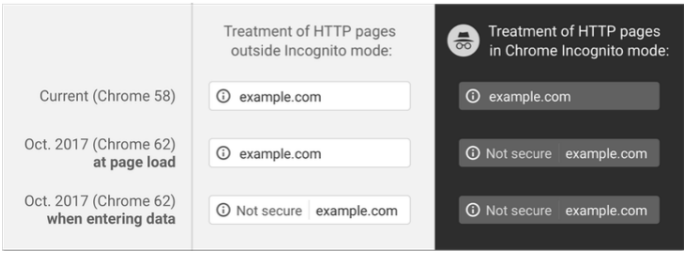
TLS = secure channel between client and server

- Confidentiality (forward secrecy)
- Integrity
- Authentication (client optional)

Long history

- TLS 1.1 and 1.2. widely adopted
- TLS 1.3 soon here
- Soon (already?) a must default

The S in HTTPS, but many other applications as well



9.3.7 Certificate Authorities in TLS

"Who is example.com?"

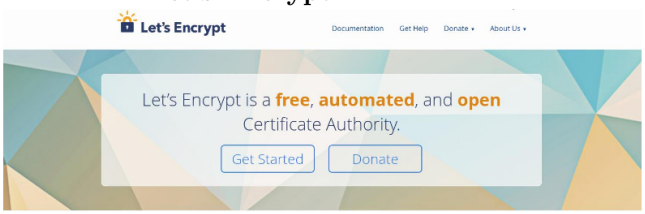
Certificate Authorities (CAs) says who is who

- Different levels of verifying who is who
- Domain, organization, extended

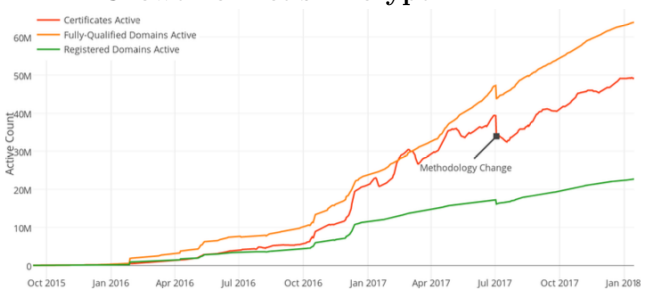
Who says who is a CA?

- Browsers do
- 100+ CAs trusted by default
- Any CA can sign for any domain
- → Lot's of problems

Let's Encrypt



Growth of Let's Encrypt



9.3.8 Certificate Transparency

Make certificates transparent

- Public append-only log
- Monitors and auditors

Name and Shame
Browsers require proof of log inclusion
Right now

- Required for extended validation certificates
- Plan: all newly issued certificates after April 2018

9.3.9 Wrap Up

TLS creates a secure channel

- Authenticated by CAs
- Browsers specify CAs they trust

Let's Encrypt: free, automated and open CA
Certificate Transparency

- All certificates publicly logged
- Soon required by Chrome, others to follow?

→ HTTPS a reasonable measure today

9.4 Secure Messaging

9.4.1 Secure Messaging in a Nutshell

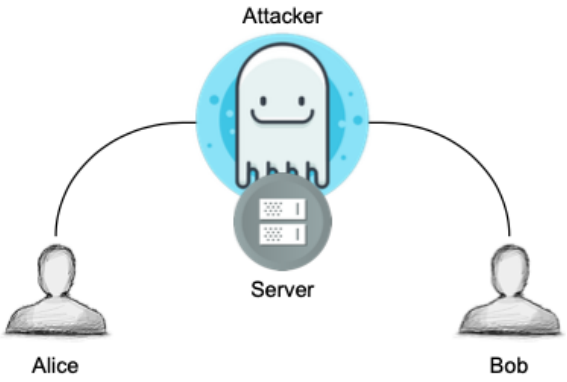
High-level properties

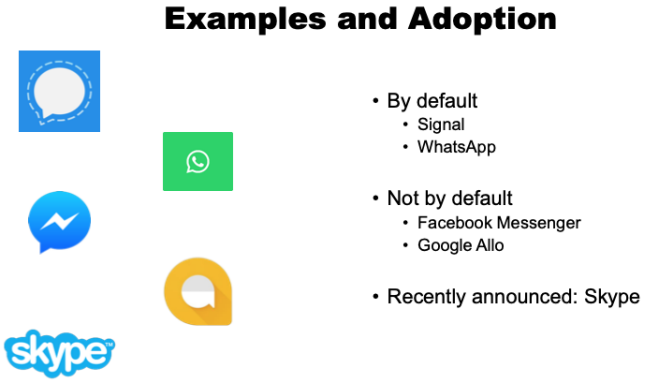
- Goal: end-to-end security
- Trust establishment
- Conversation security
- Transport privacy

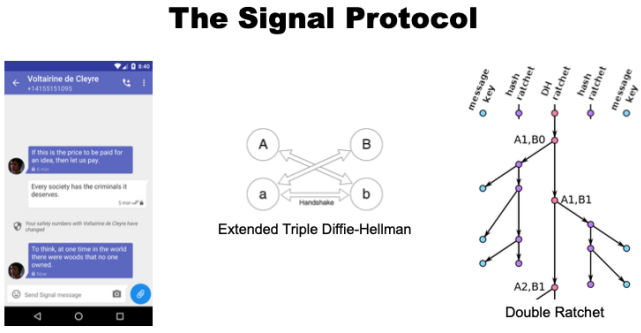
Highly usable

- No tasking key management
- Mostly as smartphone apps

One-to-one OK
Group chats a work in progress







9.4.2 Conclusions

- Secure messaging brings end-to-end security to the masses
- Strong security and widespread use brings new use-cases
- Community working on
  - Group chats
  - Transport privacy

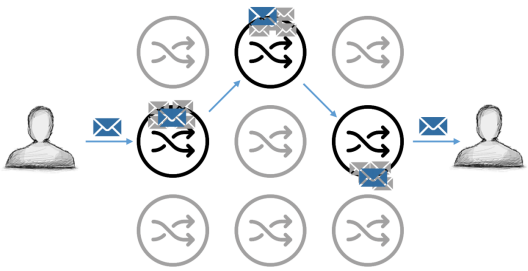
10 Anonymous communication

10.1 Mixnets

10.1.1 Introduction

- 1981 by David Chaum "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms"[1]
- anon.penet.fi
- Strong anonymity even against strong adversaries

10.1.2 Mixnets in a Nutshell



Two key design decisions

- Mix format
- Mixing strategy

Properties

- Sender anonymity
- Recipient anonymity

[1] https://www.cs.umd.edu/class/spring2015/cmsc414/papers/chaum-mix.pdf

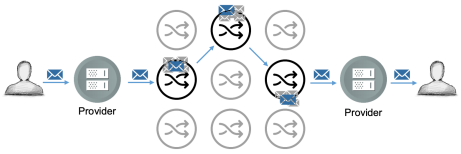
10.1.3 The Anonymity Trilemma

TABLE I
Latency vs. bandwidth vs. strong anonymity of AC protocols, with the number of protocol-nodes K, number of clients N, and message-threshold T, expected latency ℓ' per node, dummy-message rate β.

Protocol	Latency	Bandwidth	Strong Anonymity
Tor [10]	$\theta(1)$	$\theta(1/\mathsf{N})$	impossible
Hornet [47]	$\theta(1)$	$\theta(1/\mathsf{N})$	impossible
Herd [48]	$\theta(1)$	$\theta(\mathsf{N}/\mathsf{N})$	possible
Riposte [49]	$\theta(N)$	$\theta(\mathsf{N}/\mathsf{N})$	possible
Vuvuzula [20]	$\theta(\mathsf{K})$	$\theta(\mathsf{N}/\mathsf{N})$	possible
Riffle [21]	$\theta(\mathsf{K})$	$\theta(\mathsf{N}/\mathsf{N})$	possible
Threshold mixes [14]	$\theta(T \cdot \mathsf{K})$	$\theta(1/\mathsf{N})$	impossible*
Loopix [24]	$\theta(\sqrt{K} \cdot \ell')$	$\theta(\beta)$	possible
DC-Net [15], [46]	$\theta(1)$	$\theta(\mathsf{N}/\mathsf{N})$	possible
Dissent-AT [22]	$\theta(1)$	$\theta(\mathsf{N}/\mathsf{N})$	possible
DiceMix [16]	$\theta(1)$	$\theta(\mathsf{N}/\mathsf{N})$	possible

* if T in $o(poly(\eta))$

10.1.4 Loopix



	GPA	Corrupt mixes	Corrupt provider	Insider
Sender-Recipient Third-Party Unobservability	✓	✓	✓	✓
Sender online unobservability	✓	✓	✓	•
Sender anonymity	✓	✓	✓	✓
Receiver unobservability	✓	✓	✗	•
Receiver anonymity	✓	✓	✗	•

Table 1: The summary of security properties of the Loopix system in face of different threats. For the insider column we write • to denote that this concept doesn't apply to the respective notion.

10.1.5 Wrapping up

- Strong anoymity, at the cost of latency and bandwidth
- All security from the mixing
  - Mix format and mixing strategy
- No wide deployments yet, but
  - Loopix and Sphinx
  - Panoramix and Katzenpost
- Applications beyond messaging: evoting, surveys...

10.2 Tor

10.2.1 Tor in a Nutshell

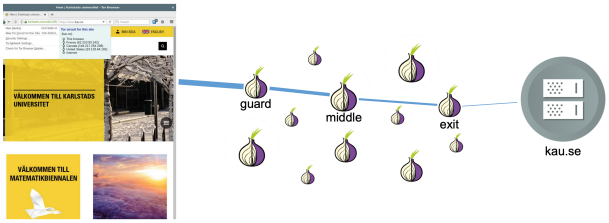
The Tor project, US non-profit 2006

- Many projects: Tor Browser, Orbot, Tails, OONI...
- Tor network of 6000 relays and 2000 bridges (> 48 Gbps)
- Low-latency anonymity network

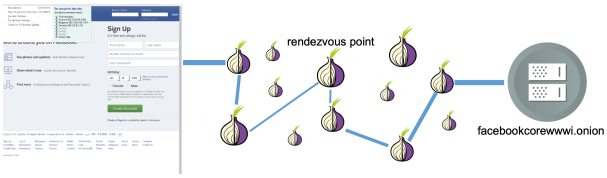
Use cases

- Anonymous browsing
- Onion services
- Single onion services
- Censorship circumvention

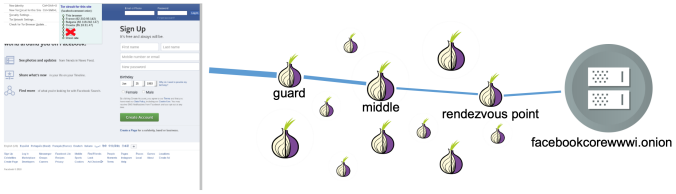
10.2.2 Anonymous Browsing



10.2.3 Onion Services

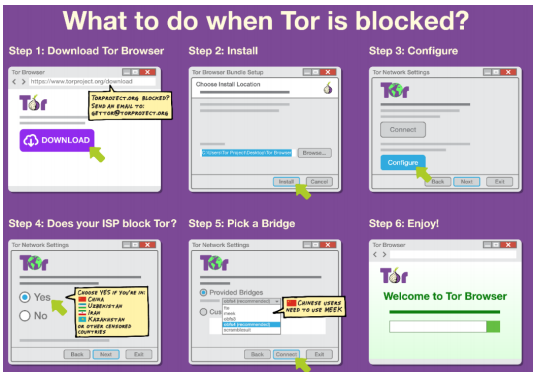


10.2.4 Single Onion Services



10.2.5 Censorship Circumvention

- Traffic exits at exit relays, bypasses national censors or regional restrictions
- → Censors block Tor
- Pluggable Transports
  - Obfs4
  - Meek
  - Snowflake



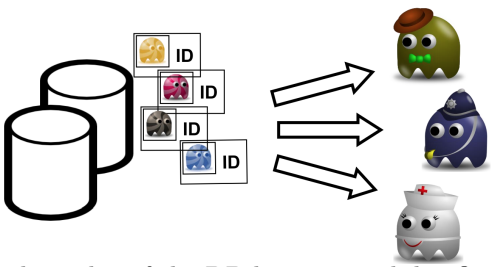
10.2.6 Wrapping up

- Tor is a *low latency network*, 6000 relays and 2000 bridges
- Anonymous browsning
  - Sender anonymity ("who is sending requests to a website?")
- Onion services & single onion services
  - Recipient anonymity ("who is receiving requests?")
  - Self authenticated, end-to-end encrypted, NAT punching, limit surface area
- Censorship circumvention

11 Databases

11.1 Why Do We Need Privacy in Databases?

11.1.1 Storing Personal Data



but what if the DB has personal data?

11.1.2 Not Easy to Release Data

Data anonymization is a difficult problem

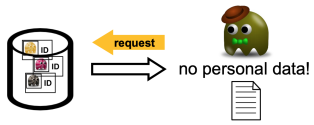
AOL released a dataset of search queries from ca. 650K users → replaced usernames with numbers
2006

The New York Times

NETFLIX released 100M ratings from ca. 480k users → claimed that all personal data was removed from the set

11.1.3 DB and Data Protection

How to use a DB that has personal data stored? and NOT disclose personal data?



11.2 k-anonymity

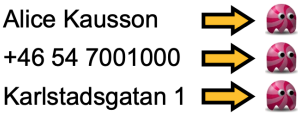
Goal: to prevent re-identification of individuals when releasing data

K-anonymity property: on data release, information about a subject cannot be distinguished from at least k-1 individuals

11.2.1 Types of Identifiers

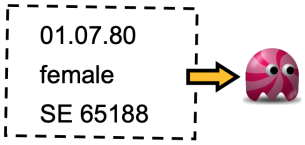
Explicit Identifiers
Uniquely attributes

- name
- phone number
- address



Quasi-Identifiers
In combination, can uniquely identify

- birth date
- gender
- ZIP code



11.2.2 K

• Measure for the anonymity set
where min(k) = 2 (k = 1 means NO anonymity)

11.2.3 Example: building a k=2 release

Name	Birth date	Gender	ZIP	Civil Status	Duration	Diagnosis
	11.03.79	male	1072	married	1	A
	17.03.79	male	1276	married	7	B
	01.07.80	female	1073	single	2	B
	07.09.84	female	1077	single	0	C
	02.07.89	male	1016	single	2	D
	21.09.91	female	1267	it's complicated	4	E
	24.12.98	female	1268	it's complicated	4	A

Remove Name Field

Name	Birth date	Gender	ZIP	Civil Status	Duration	Diagnosis
	11.03.79	male	1072	married	1	A
	17.03.79	male	1276	married	7	B
	01.07.80	female	1073	single	2	B
	07.09.84	female	1077	single	0	C
	02.07.89	male	1016	single	2	D
	21.09.91	female	1267	it's complicated	4	E
	24.12.98	female	1268	it's complicated	4	A

Generalize Birth date to Range

Name	Birth date	Gender	ZIP	Civil Status	Duration	Diagnosis
	1970's	male	1072	married	1	A
	1970's	male	1276	married	7	B
	1980's	female	1073	single	2	B
	1980's	female	1077	single	0	C
	1980's	male	1016	single	2	D
	1990's	female	1267	it's complicated	4	E
	1990's	female	1268	it's complicated	4	A

The Gender Field

Name	Birth date	Gender	ZIP	Civil Status	Duration	Diagnosis
	1970's	male	1072	married	1	A
	1970's	male	1276	married	7	B
	1980's	female	1073	single	2	B
	1980's	female	1077	single	0	C
	1980's	male	1016	single	2	D
	1990's	female	1267	it's complicated	4	E
	1990's	female	1268	it's complicated	4	A

NOT k=2 here

Generalize Gender Field

Name	Birth date	Gender	ZIP	Civil Status	Duration	Diagnosis
	1970's	male	1072	married	1	A
	1970's	male	1276	married	7	B
	1980's	ghost	1073	single	2	B
	1980's	ghost	1077	single	0	C
	1980's	ghost	1016	single	2	D
	1990's	female	1267	it's complicated	4	E
	1990's	female	1268	it's complicated	4	A

OR Suppress Information

Name	Birth date	Gender	ZIP	Civil Status	Duration	Diagnosis
	1970's	male	1072	married	1	A
	1970's	male	1276	married	7	B
	1980's	female	1073	single	2	B
	1980's	female	1077	single	0	C
*	*	*	*	*	*	*
	1990's	female	1267	it's complicated	4	E
	1990's	female	1268	it's complicated	4	A

Generalize ZIP data

Name	Birth date	Gender	ZIP	Civil Status	Duration	Diagnosis
	1970's	male	1***	married	1	A
	1970's	male	1***	married	7	B
	1980's	ghost	10**	single	2	B
	1980's	ghost	10**	single	0	C
	1980's	ghost	10**	single	2	D
	1990's	female	12**	it's complicated	4	E
	1990's	female	12**	it's complicated	4	A

Civil Status Field is k=2!

Name	Birth date	Gender	ZIP	Civil Status	Duration	Diagnosis
	1970's	male	1***	married	1	A
	1970's	male	1***	married	7	B
	1980's	ghost	10**	single	2	B
	1980's	ghost	10**	single	0	C
	1980's	ghost	10**	single	2	D
	1990's	female	12**	it's complicated	4	E
	1990's	female	12**	it's complicated	4	A

11.2.4 l-diversity and t-closeness

l-diversity

- Addresses two attacks on k-anonymity
  - Homogeneity attack
  - Background knowledge attack

BUT

- Difficult, sometimes unnecessary
- Insufficient to prevent attribute disclosure
- it does not consider overall data distribution
- it does not consider semantics

t-closeness

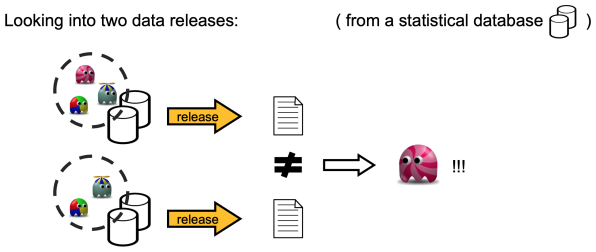
- Addresses l-diversity limitations
- Metric is the attacker's information gain

BUT

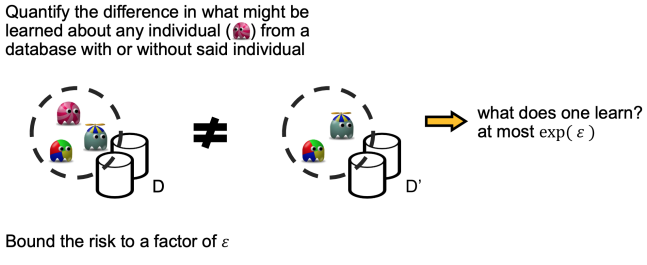
- No computational procedure
- Limitations on the utility of data releases

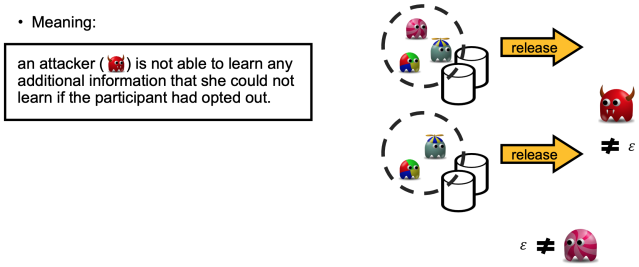
11.3 Differential Privacy

11.3.1 Releasing Personal Data



11.3.2 Differential Privacy





11.3.3 How to do it?

Add noise to the query result

- how? it depends on the mechanism design and the type of data
- exponential mechanism → categorical data
- Laplace mechanism → numerical data

11.3.4 Limitations

Differential Privacy does not mean that the attacker learns nothing about the individual from the results → mind the background information!

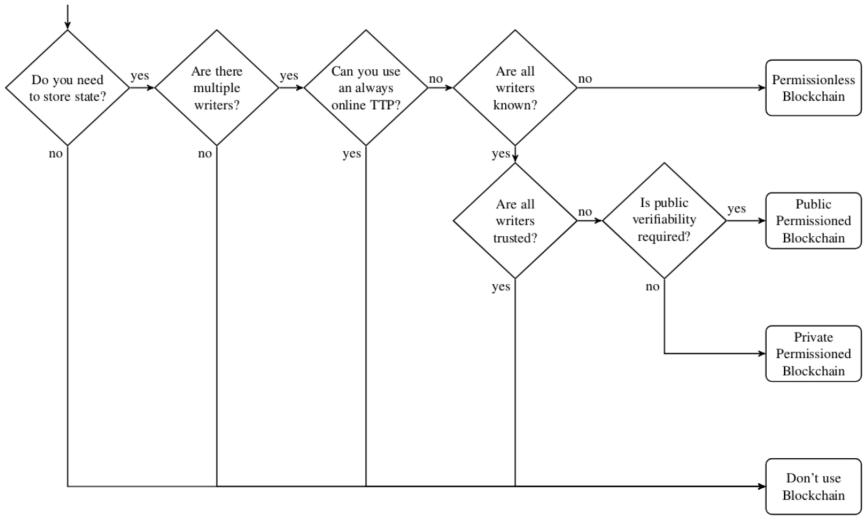
12 Other PETs

12.1 Blockchains

12.1.1 Three Parts of Any Blockchain

1. **Data Structure:** blockchain, DAGchain...
2. **Network:** permissionless or permissioned
3. **Consenus:** PoW, PoS, PoA, PBFT...

12.1.2 An Appropriate Technical Solution?



12.2 Idemix

How to verify an identity without revealing the identity?

12.2.1 Identity Mixer

- Identity Mixer → idemix
- anonymous credential system
- developed at IBM Research
- strong authentication and privacy → at the same time.

Watch Maria Dubovitskay talking about Idemix: `https://www.youtube.com/watch?v=_UHTKjsedQY`

12.3 Transparency Enhancing Tools (TETs)

12.3.1 Motivation: Transparency & Intervenability

Legal privacy principles

- GDPR:
  - General Art. 5 I (a) – lawfulness, fairness and transparency
  - Data subject rights to Transparency & Intervenability (GDPR – Chapter III)
- Swedish Data Patient Act:
  - Rights to access health records and log information

12.3.2 Transparency vs. Confidentiality - Examples

Log files in eHealth - privacy issues:

- Information about who (e.g., psychiatrist) accessed EHR is sensitive for patients
- Monitoring of performance/quality of work of medical personnel

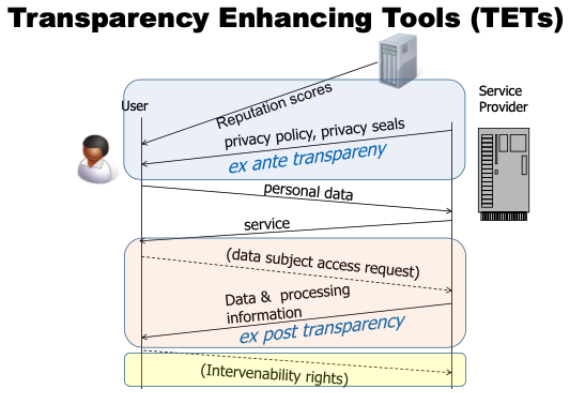
Requirements:

- Privacy-preserving
- Considering Tradeoffs with rights of others

Business secrets in relation to profiling

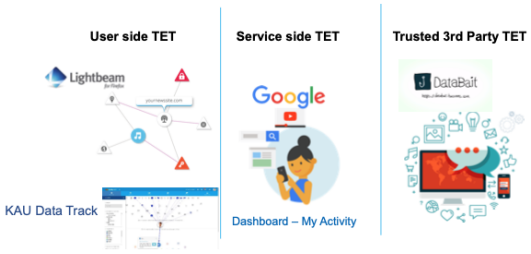
- (cf. Recital 63 GDPR)

12.4 TETs - Examples

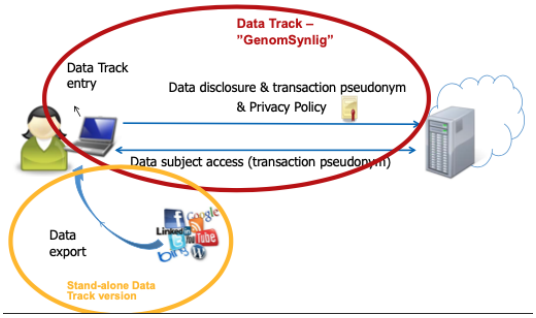


12.4.1 Ex-ante TETs - Examples

- Privacy Policy Languages: e.g., P3P, PPL, A-PPL
- Multi-Layered Structured Policies (Art. 29 WP), complemented by Policy Icons,
- Examples: Examples of suggested Cloud-specific policy icons (A4Cloud): SE, OUTSIDE EEA, EEA



12.4.2 User controlled ex post TET: Data Track



12.4.3 Data Track – Trace View: Viewing attributes in common

Part III

Designing for Privacy / Data Protection impact assessment

13 Privacy and Data Protection

What Does it Mean to Design for "Privacy"?

13.0.1 Privacy is a Human Right

From the United Nations Universal Declaration of Human Rights:

Article 12

- No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks.

From the European Convention on Human Rights:

Article 8 - Right to respect for private and family life

1. Everyone has the right to respect for his private and family life, his home and his correspondence.
2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

13.0.2 Privacy $\neq$ Data Protection

- Privacy is fuzzy, contextual, social construct, depends...
- Data protection, by necessity, has to be more concrete
- Proportionality a key consideration
- $\rightarrow$ Data protection necessary but not sufficient for privacy

13.0.3 Rights from the EU charter

From the EU Charter of Fundamental Rights:

Article 7

Respect for private and family life

- Everyone has the right to respect for his or her private and family life, home and communications.

Article 8
Protection of personal data

1. Everyone has the right to the protection of personal data concerning him or her.
2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
3. Compliance with these rules shall be subject to control by an independent authority.

Article 52
Scope of guaranteed rights

- Any limitation on the exercise of the rights and freedoms recognized by this Chapter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognized by the Union or the need to protect the rights and freedoms of others.

13.1 Ann Cavoukian's Privacy by Design

13.1.1 Overview

- Seven principl es, by Ann Cavoukian
- End of 1990's beginning of 2000's
- Information and Privacy Commissioner of Ontario
- Data protection centric (control, see paradigms later in the course)

1. **Proactive** not Reactive; **Preventative** not Remedial
2. Privacy as the **Default**
3. Privacy **Embedded** into Design
4. **Full** Functionality – Positive-Sum, **not Zero-Sum**
5. End-to-End Security – Full **Lifecycle** Protection
6. Visibility and Transparency – Keep it **Open**
7. Respect for User Privacy – **Keep it User-Centric**

13.2 Privacy Paradigms

13.2.1 Three Privacy Paradigms

Privacy as Confidentiality

Privacy as Control

Privacy as Practice

13.2.2 Privacy ad Confidentiality

- Data disclosed → privacy lost
- Data minimisation
- Centralised → bad
- Cryptography community
- Open source, reproducability

13.2.3 Privacy as Control

- Ability to exercise control over personal data → privacy
- May be in your interest to disclose personal data (e.g., healthcare)
- Data protection
  - Purpose
  - Intervenability
  - Transparency
  - Accountability

13.2.4 Privacy as Practice

- Freedom to understand and control privacy decisions
- Industry: "do not scare users"
  - Over time, get people to share more and more about themselves, but not perceive it as invasive
- Like a mirror
  - Understand how you are perceived
  - Control how you are perceived
  - Feedback and nudges

13.2.5 Designing for "Privacy"?

- Neither paradigm is wrong
- Industry likes privacy as practice
  - For the wrong reasons? (more data)
  - But does also good?
- GDPR
  - Privacy as control
  - Data minimization important principle
  - High fines → personal data is a risk → push for privacy as confidentiality?

13.3 Technology in Hostile States

1. Do not rely on the law to protect systems or users.
2. Prepare policy commentary from quick response to crisis.
3. Only keep the user data that you currently need.
4. Give users full control over their data.
5. Allow pseudonymity and anonymity.
6. Encrypt data in transit and at rest.
7. Invest in cryptographic R&D to replace noncryptographic systems.
8. Eliminate single points of security failure, even against coercion.
9. Favor open source and enable user freedom.
10. Practice transparency: share best practices, stand for ethics, and report abuse.

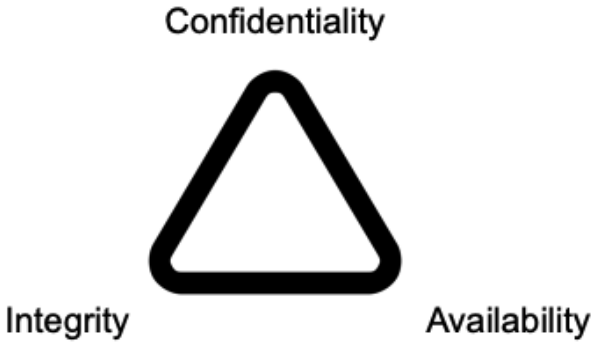
13.3.1 Comparison to the GDPR

1. Do not rely on the law to protect systems or users.
2. Prepare policy commentary from quick response to crisis.
3. Only keep the user data that you currently need.
4. Give users full control over their data.
5. Allow pseudonymity and anonymity.
6. Encrypt data in transit and at rest.
7. Invest in cryptographic R&D to replace noncryptographic systems.
8. Eliminate single points of security failure, even against coercion.
9. Favor open source and enable user freedom.
10. Practice transparency: share best practices, stand for ethics, and report abuse.

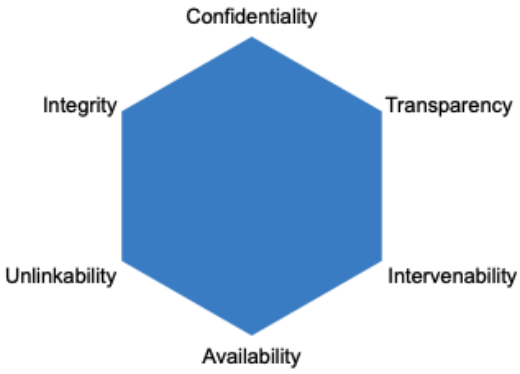
- Privacy as confidentiality
- GDPR similarities
  - Control
  - Data minimisation
  - Transparency
  - Pseudonymity
- GDPR a reason to invest in cryptographic R&D?
- User-centric, freedoms, and rights
- GDPR prescribes a lot of process (1), but also ”safeguards, security measures and mechanisms”, i.e., also technical protections

13.4 Privacy Protection Goals

13.4.1 Security: The CIA Triad



13.4.2 Privacy Protection Goals



13.4.3 Complementing CIA

- Add privacy to the security triad
  - CIA already considered in procedures, processes etc
  - → privacy protection goals help with including privacy
- Three important privacy goals
- Stems from the German data protection community

13.5 Data Protection by Design and by Default

Data Protection by Design, GDPR Article 25 §1

Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organizational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.

The following sentences are referenced out from the paragraph to indicate the following statements

"Who should do something?"
the controller shall

"What is the objective?"
to meet the requirements of this Regulation and protect the rights of data subjects

"What has to be done?"
implement appropriate technical and organizational measures & **safeguards into the processing**

"When should it be done, and how?"
both at the time of the determination of the means for processing and at the time of the processing itself & **in an effective manner**

"Under what conditions?"
Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing

Data Protection by Default, GDPR Article 25 §2

The controller shall implement appropriate technical and organizational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.

"Who should do something?"
The controller

"What has to be done?"
implement appropriate technical and organizational measures & **by default**

"To what?"
to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility

"Under what conditions?"
appropriate

"What should be required for more privacy invasive features?"
the individual's intervention

14 Privacy and Data Protection Impact Assessments

14.1 From DPbD and by Default to DPIAs

14.1.1 Data Protection by Design and by Default

- Data protection by design
  - Consider rights of people in the full life-cycle of systems
  - Reasonable measures should be taken
- Data protection by default
  - Strong protections should not be opt-in, but opt-out
  - Data minimization and purpose binding are key principles
- The controller's responsibility
  - High fines in the GDPR
  - Responsibility will likely "trickle down" to processors

14.1.2 The Impact of the GDPR

- The GDPR is a legal document, nobody really knows for sure yet
- Important principles:
  - Data minimization
  - Purpose binding
  - Transparency
  - Intervenability (data subject rights)
- Fines and forced transparency (organizational threats)

14.1.3 Implementing Data Protection

"This is a fundamental tension between the requirements of Data Protection regimes, and the goals of most common anonymizers as well as the desires of users, that is not resolved and hardly discussed: in order to implement an effective and comprehensive data protection regime, we fist have to implement the most extensive surveillance and tracking infrastructure."
- George Danezis & Seda Gürses, A critical review of 10 years of Privacy Technology

14.1.4 What is clear: PIAs are Important

- Understanding privacy risks fundamental
- DPIAs strongly encouraged, at times mandatory
- Beyond DPbD value: identified threats and risks
  - Part of getting ready for the GDPR (inventory)
  - Incident response (personal data breaches)
  - Understanding privacy risks → understanding risks related to the GDPR to the organization → informs organizational risk management)

14.2 PIA is a Process

14.2.1 What is Data Protection Impact Assesment?

"Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, **is likely to result in a high risk** to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data."
– Art. 35 GDPR.

14.2.2 What is Privacy Impact Assesment?

"A **privacy** impact assessment (PIA) is an instrument for assessing the potential impacts on **privacy** of a process, information system, programme, software module, device or other initiative which processes personally identifiable information (PII) and, **in consultation with stakeholders**, for taking actions as necessarily in order to treat **privacy** risk."
– ISO/IEC 29134:2017.

14.2.3 Who benefits from PIAs?

They do:

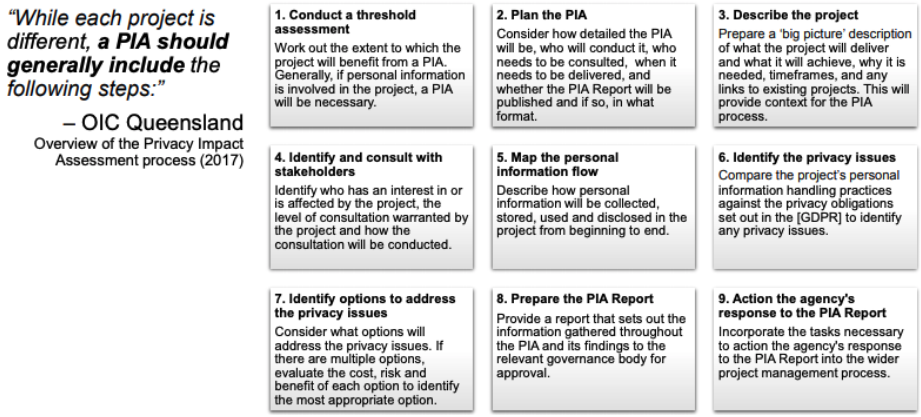
- Your **customers and general public** – because you are looking out for their privacy interests
- Your **organisation** – because you demonstrate to your employees and contractors that you take privacy seriously and expect them to the same
- The **regulators** – because when you carry out a proper PIA you clarify your project information dealings, making their work easier

Not sure yet?

- A PIA helps to **reduce costs** in management time, legal expenses and potential negative media (i.e., PR also likes it)
- A PIA helps to **demonstrate compliance** as an element of accountability
- A PIA enhances **informed decision-making** and exposes internal communication gaps or hidden assumptions
- A PIA helps to **avoid privacy pitfalls** of a project
- And, well... it might be mandatory...

14.2.4 How do you do PIA?

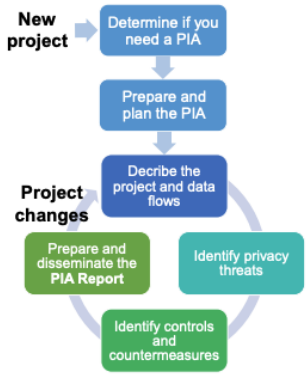
"[PIA] is a **process** which should begin at the earliest possible stages, when there are still opportunities to **influence the outcome** of a project. It is a process that should **continue** until and even after the project has been deployed."
– David Wright The state of art in PIA (2012)



14.3 Overview of frameworks

14.3.1 PIA Articulation and Systematisation

- PIAs require **multiple** technical and organizational **methods**
  - Project planning
  - System documentation
  - Privacy risk analysis
  - Reporting and action plan
- Methods have to be studied, selected and systematized to create methodology, i.e., a **PIA framework**



14.3.2 PIA Frameworks: A Few Examples

- **Privacy and Data Protection Impact Assessment Framework for RFID Applications** (PIA RFID), 2011.
- UK – **Conducting privacy impact assessments code of practice**, Information Commissioner's Office (ICO), 2014.
- FR – **Privacy Impact Assessment (PIA)**, Commision nationale de l'informatique et des libertés (CNIL), 2015.
- ISO/IEC 29134 Information technology – Security techniques – **Guidelines for privacy impact assessment**, 2017.

14.3.3 PIA Frameworks: PIA RFID

- **Privacy and Data Protection Impact Assessment Framework for RFID Applications** (PIA RFID), 2011.
- Oetzel, M.C. and Spiekermann, S., 2014. **A systematic methodology for privacy impact assessments: a design science approach**. European Journal of Information Systems, 23(2), pp.126-150.

14.3.4 Ok, but which one should I use?

- They are all relatively similar
- They all aim for Privacy, i.e., PIA (not merely 'Data Protection')
- Choose and adapt them to your organisation's needs
- Engage with your DPA
- ISO 29134 definitely shows that we're reaching some level of maturity regarding PIAs

14.4 DPIAs according to WP29

Clarifications on DPIA's obligatoriness

- A DPIA is only required when the processing "**is likely to result in a high risk to the rights and freedoms of natural persons**" (Art. 35(1))
- To ensure **consistent interpretation** of the circumstances in which a DPIA is mandatory
- To provide a **criteria** on whether a DPIA is required

The following criteria should be considered:

1. Evaluation or scoring
2. Automated decision-making with legal effect
3. Systematic monitoring
4. Sensitive data
5. Data processed on a large scale
6. Datasets matched or combined
7. Data concerning vulnerable data subjects
8. Innovative use (e.g., new technology)
9. Data transfer across borders outside the EU
10. When the processing in itself "prevents [...] from exercising a right or using a service [...]"

Examples

Examples of processing	Possible Relevant criteria	DPIA required?
A hospital processing its patients' genetic and health data (hospital information system).	• Sensitive data • Data concerning vulnerable data subjects	Yes
The use of a camera system to monitor driving behavior on highways. The controller envisages to use an intelligent video analysis system to single out cars and automatically recognize license plates.	• Systematic monitoring • Innovative use or applying technological or organisational solutions	
A company monitoring its employees' activities, including the monitoring of the employees' work station, internet activity, etc.	• Systematic monitoring • Data concerning vulnerable data subjects	
The gathering of public social media profiles data to be used by private companies generating profiles for contact directories.	• Evaluation or scoring • Data processed on a large scale	
An online magazine using a mailing list to send a generic daily digest to its subscribers.	• (none)	Not necessarily
An e-commerce website displaying adverts for vintage car parts involving limited profiling based on past purchases behaviour on certain parts of its website.	• Evaluation or scoring, but not systematic or extensive	

14.4.1 Examples of EU DPIA frameworks

EU generic frameworks

- DE: Standard Data Protection Model (2016)
- ES: Guía para una Evaluación de Impacto en la Protección de Datos Personales (EIPD), AGPD (2014)
- FR: Privacy Impact Assessment (PIA), CNIL (2015)
- UK: Conducting privacy impact assessments code of practice, ICO (2014)

EU sector-specific frameworks

- PIA Framework for RFID Applications
- DPIA Template for Smart Grid and Smart Metering systems

Standards

- ISO/INEC 2913430

14.4.2 Criteria for an acceptable DPIA

Criteria for DPIA (30+ items, Annex 2)

- a systematic description of the processing is provided
- necessity and proportionality are assessed
- risks to the rights and freedoms of data subjects are managed
- interested parties are involved

All in all...

- WP29 uses terms PIA and DPIA interchangeably
- Previous work on PIA frameworks are the main references
- How do you do a DPIA? Just do a PIA

14.5 PIA RFID framework in detail

PIA RFID – Main references

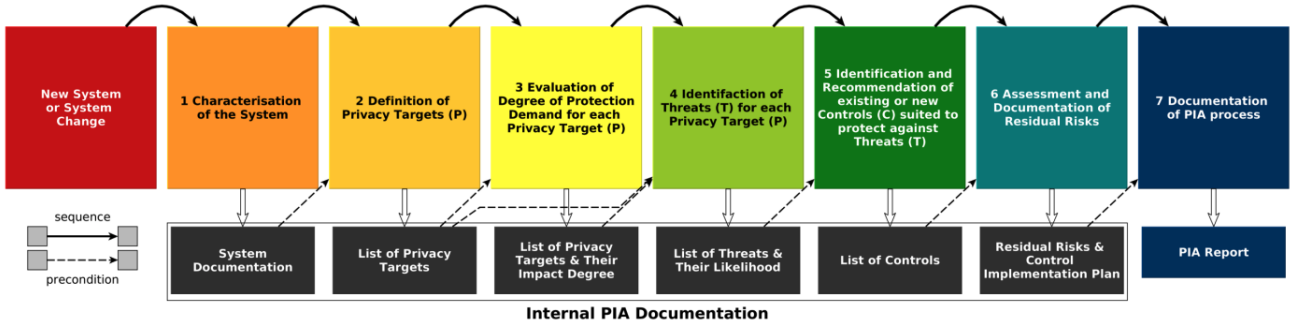
Technical Report

Oetzel, C., Spiekermann, S., Grüning, I., Kelter, H. and Mull, S., 2011. Privacy Impact Assessment Guideline for RFID Applications. Bundesamt für Sicherheit in der Informationstechnik (BSI).

Research Paper

Oetzel, M.C. and Spiekermann, S., 2014. A systematic methodology for privacy impact assessments: a design science approach. European Journal of Information Systems, 23(2), pp.126-150.

14.5.1 PIA methodology and constructs



14.5.2 Step 1 - Characterisation of the System

"**To describe** the system in such a **comprehensive** and detailed way that potential **privacy problems** can be **detected**."

- **System View**
  - Applications and system components
  - Hardware and software
  - Internal and external interfaces
  - Network topology
- **Functional View**
  - Generic business processes
  - Detailed use cases
  - Roles and users
  - Technical controls
- **Data View**
  - Categories of processed data
  - Data flow diagrams
  - Actors and data types
- **Physical View**
  - Environment's physical security
  - Operational controls

14.5.3 Step 2 - Definition of Privacy Targets (P)

- **Privacy targets** are directly **derived** from the established **privacy principles** formulated in the **GDPR**

P1 – Data quality
- P1.1 Ensuring fair and lawful processing through transparency
- P1.2 Ensuring processing only for legitimate purposes
- P1.3 Providing purpose specification
- P1.4 Ensuring limited processing for specified purpose
- P1.5 Ensuring data avoidance
- P1.6 Ensuring data minimisation
- P1.7 Ensuring data quality, accuracy and integrity
- P1.8 Ensuring limited storage

P2 – Processing legitimacy
- P2.1 Ensuring legitimacy of personal data processing
- P2.2 Ensuring legitimacy of sensitive personal data processing

P3 – Information right of data subject
- P3.1 Providing adequate information in cases of direct collection of data from the data subject
- P3.2 Providing adequate information where data has not been obtained directly from the data subject (e.g., from third parties)

P4 – Access right of data subject
- P4.1Facilitating the provision of information about processed data and purpose
- P4.2 Facilitating the rectification, erasure or blocking of data
- P4.3 Facilitating the portability of data
- P4.4 Facilitating the notification to third parties about rectification, erasure and blocking of data

P5 – Data subject's right to object
- P5.1 Facilitating the objection to the processing of personal data
- P5.2 Facilitating the objection to direct marketing activities
- P5.3 Facilitating the objection to disclosure of data to third parties
- P5.4 Facilitating the objection to decisions that are solely based on automated processing of data
- P5.5 Facilitating the data subject's right to dispute the correctness of machine conclusions

P6 – Security of data
- P6.1 Ensuring the confidentiality, integrity and availability of personal data storage, processing and transmission
- P6.2 Ensuring the detection of personal data breaches and their communication to data subjects

P7 – Accountability
- P7.1 Ensuring the accountability of personal data storage, processing and transmission

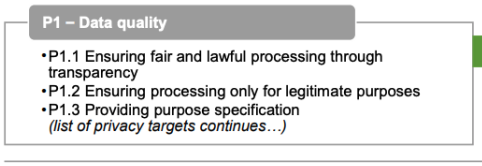
14.5.4 Step 3 - Degree of Protection Demand

To determine the right **level of protection demand** you should ask: "What would happen if . . . ?"

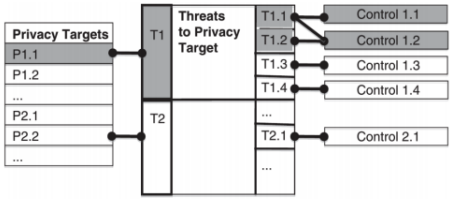
What could be impacted if the privacy target was not met…?				
System operator perspective			Data subject perspective	
Reputation or brand value	Financial situation	Social standing, reputation	Financial situation	Personal freedom
Low – 1: The impact of any loss or damage is *limited and calculable*				
Medium – 2: The impact of any loss or damage is *considerable*				
High – 3: The impact of any loss or damage is *devastating*				

14.5.5 Step 4 - Identification of Threats (T)

"For each privacy target, we systematically **identify the threats** that could **prevent us from reaching** them."



Threat code and name	Sub-threat code	Description of threat	Likely (y/n)	Associated privacy target
T1 Lack of transparency – Missing or insufficient service information	T1.1	Incomplete or insufficient information describing the service. The operation details (data flows, data locations, ways of transmission, etc.) and the impacts of the RFID application are not sufficiently explained to the data subject. An RFID emblem is not displayed on the website of the RFID operator.	y	P1.1
	T1.2	Existing information describing the service is not easily accessible for the data subject. The information is not well-indexed and/or searchable.	y	P1.1
	T1.3	The basic concept as well as the purpose underlying the service is not clearly explained.	y	P1.1 P1.2



Controlled matching of privacy targets with threats and controls

14.5.6 Step 5 - Identification and Recommendation of Controls (C)

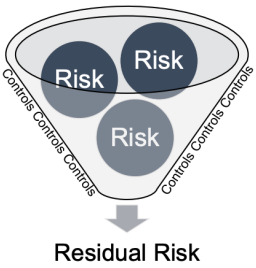
- "The crucial step in a PIA is to identify **controls** that can **minimize**, **mitigate** or **eliminate** the identified **threats**."

- Three levels of rigor: satisfactory (1), strong (2) and very strong (3). (E.g., High protection (3) x Likely (y) = Very Strong (3))

Sub-threat code	Control code(s) and name(s)		Assigned overall category (from step 3)	Description
T1.1	C1.1	SERVICE DESCRIPTION	2 (P1.1)	Extensive informational material (e.g. flyers, RFID emblem, websites) is made available that is easily understandable and accessible.
	C1.4	INFORMATION TIMELINESS		The information describing the service is checked at regular intervals for its timeliness, especially when questions from data subjects could not be answered.
	C6.2	HANDLING OBJECTIONS TO AUTOMATED DECISIONS		The logic involved in any automatic processing of data and automated decisions is described and made available to the data subjects. They are informed of their right to object to this automated decision making. A contact address is given. Objections are individually processed and automated decisions are disabled on request.

14.5.7 Step 6 – Assessment and Residual Risks

- To produce a **control implementation plan** that clearly show how threats are mitigated or remain misaddressed.
- Residual risks should be **well documented** in a PIA report.
- Upper management, corporate risk management and IT staff will be **held accountable** if privacy breaches occur.



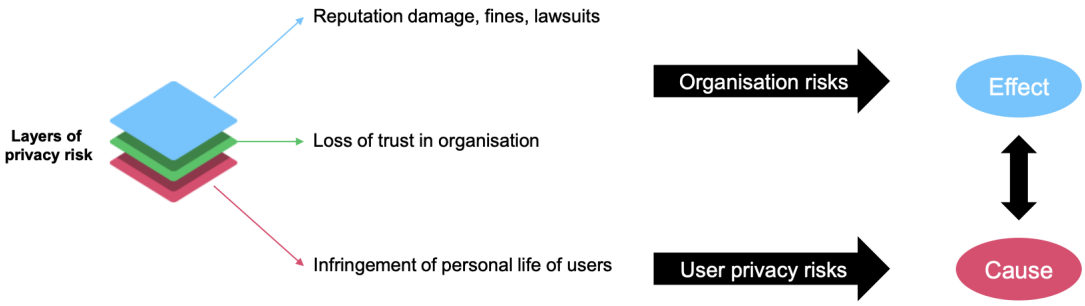
14.5.8 Step 7 – Documentation & PIA Report

- Companies should have a strong interest in **comprehensively documenting** the entire PIA process.
- It is **advisable** to produce two versions of a **PIA Report**:
  - **Internal** and **auditing** purposes (DPAs & Staff)
  - **Public** and less detailed (Customers & Media)

	Data protection authorities	Company staff	Customers	Media
General system information				
System overview	x	x		
System boundaries	x	x		
System purposes	x	x	x	x
Assessment information				
Relevant privacy targets identified	x	x	x	x
Relevant privacy threats addressed	x	x		
Chosen privacy controls	x	x	x	x
Residual risks encountered	x	x	(x)	(x)
PIA quality signals				
Stakeholders involved	x	x		
Legal compliance checked	x	x		
PIA start date/System's start date	x	x		
Accountability				
Person(s) involved in the PIA	x	x		
Organisation who conducted the PIA	x	x	x	x
Person who approved the PIA	x	x		
Privacy responsible in company	x	x	x	x
Date of PIA completion	x	x	x	x
Time frame of PIA validity	x	x		

14.6 PIAs in Practice

1. Perceiving PIAs as Mandatory
   - PIAs are not mandatory, DPIAs only in particular cases
   - May lead to "PIA fatigue"
2. Not Adapting Questions
   - Same questionnaire for assessing data processing
   - Different needs for different activities
   - Should first perform a light-weight PIA to determine if full PIA is necessary
3. Focus on the Wrong Stakeholder
   - Organization-centric, to avoid fines
   - Should be user-centric, and consult users as part of PIA
4. PIA as a Task
   - Treating PIA as a one-time task early in development
   - Revised years after first creation
   - PIA is a process, not a task
5. Mixing Cause and Effect



14.6.1 Conclusions

- Five common mistakes
  1. Perceiving PIAs as mandatory
  2. Not adapting questions
  3. Focus on the wrong stakeholder
  4. PIA as a task
  5. Mixing cause and effect
- Being organization-centric instead of user-centric
  - PIAs (at best) → data protection compliance
  - → no privacy-friendly systems
- Focus on avoiding risks, not only mitigating

14.7 Wrapping Up

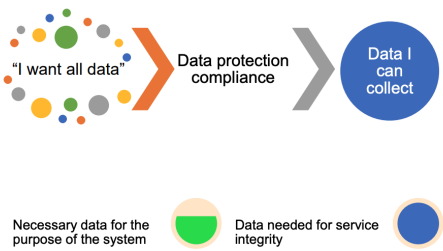
14.7.1 Designing for Privacy

- Privacy is multifaceted
  - An essential human right, data protection closely related
  - Paradigms as confidentiality, control, practice
  - CIA+Unlinkability+Transparency+Intervenability
- Data protection by design and by default
  - Reasonable measures, protect rights, full life-cycle
  - Strong protections by default
  - Depends on how the GDPR is enforced and interpreted
- DPIAs/PIAs are essential to design for privacy
  - A process, understanding privacy risks
  - Added value for organization: incident response, risks related to GDPR

14.7.2 Privacy Engineering

- Newly formed field of research and practice
- From tradecraft and know-how to engineering
- We don't really have good and solid methods, but we have starting points that show promise
  - PIAs we already covered
  - Module 4 on Privacy Management touches on high-level analysis methods, like LINDDUN
  - Module 5 on Privacy Patterns for Software Design covers software engineering perspective

14.7.3 Change of Mindset



Part IV

Privacy risk assessment

15 Privacy Impact Assessment and Privacy Risk Analysis as integral part of privacy management

15.1 Privacy Impact Assessment (PIA)

Privacy Impact Assessment (PIA) is the name for a process that analyzes and documents the future impact that processing their personal data will have on data subjects. "Privacy impact" are the (possibly unwanted and invisible) consequences that data processing may impose on individuals or society. Privacy impact is the result of personal data processing, data breaches or data collection.

The idea of privacy impact assessments has been around since the 1970ies. Roger Clarke summarizes the complete history of the concept and provides a global perspective on PIA in his article (Clarke, 2009).

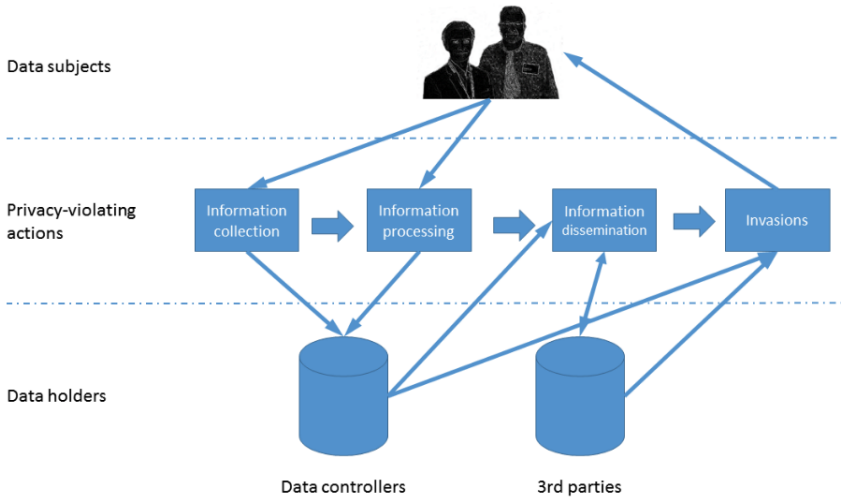
Certain regulation requires privacy impact analysis before a new application can start with the collection and processing of personal data. In many countries, a mandatory PIA is to be performed when sensitive personal data is being processed (e.g. data about health conditions, sexual orientation, religion).

Although the term "privacy impact analysis" is very popular, there is no systematic method to perform a PIA to this date. National authorities and expert organizations have published guidance and checklists. Sometimes, a PIA report is a required attachment for notifications of data processing to a supervisory government authority. A detailed analysis for various regulatory environments and their demands on PIA was published in (Tanrock, Pearson and Charlesworth, 2010).

In the following sections, we will recall Solove's taxonomy of privacy and privacy invasions, then we will examine the PIAF project's very extensive PIA process. Next, we compare carious PIA standards and guidelines to each other. Concluding, you will have a reading assignment about how to perform a PIA for RFID-enabled wireless technology.

15.1.1 Privacy impact on data subjects

The figure below shows the privacy-violating actions **information collection, information processing, information dissemination and invasions.**



All four actions impact data subjects.

1. **Information collection**: Increases the information pool, thereby shifting power balances between data subjects and data processors. Collecting incorrect or outdated information may cause harm later.
2. **Information processing**: Processing personal data, electronic decision making and accumulation of data may result in price discrimination, exclusion, denial of service, incomplete or wrong service, wrong conclusions, unjustified data combination, classification errors.
3. **Information dissemination**: Information spreads to other processors, private or professional circles, the public, criminals, or personal enemies. Causes may be a part of the legitimate application, or as data breaches (hacking, data theft or system failures). Dissemination includes freedom-of-information inquiries and police, government and homeland security access.
4. **Invasions**: Invasions are all kinds of direct or indirect interactions back to the data subject that are based on personal data. From uninvited advertising via stalking to blackmail, any invasion and the effort to mitigate and prevent them should get considered.

We observe that data subjects are being put at risk already from the data collection step, and that risk is increasing with each step towards invasion. A privacy impact analysis must therefore consider the risks for data subjects that occur in the early phases before invasion occurs!

15.1.2 Conducting a PIA

In this part of the course, we will take a very brief look at the recommendation of the PIAF project about PIA to the European Union.

Core PIA elements

According to (PIAF D3, 2012), PIA is an overall process composed of the following steps:

1. **An on-going process**: A PIA should be regarded and carried out as a process and not only as a single task of completion of a report. A PIA process starts early and continues throughout the life cycle of the project. The PIA process is summarized as:
   (a) Early start in development project
   (b) Project description
   (c) Stakeholders' consultation
   (d) Risks management
   (e) Legal compliance check
   (f) Recommendations and report
   (g) Decision and implementation of recommendations
   (h) Audit and review

   Please note that steps b. to h. are similar to the Plan-Do-Check-Act cycle and the OASIS privacy management steps shown in section 1!

2. **Scalability**: A PIA policy should allow organizations to carry out a PIA appropriate to their own circumstances. A PIA policy should allow scalability of the PIA process.

3. **All privacy types**: As the rights to privacy and to protection of personal data are fundamental rights in the European legal order, for ensuring the highest level of protection thereof, a PIA should address all types of privacy issues and not only the protection of personal data.

4. **Accountability**: An organization should be able to demonstrate that a PIA has been carried out adequately.

5. **Transparency**: A PIA process should enjoy at least a minimum level of transparency. Both the assessor and the stakeholders must have all relevant information to assess the privacy and data protection implications of the proposed project. This does not preclude due respect for sensitive information.

6. **Stakeholders' involvement**: Stakeholders, as representative as possible, including the public, if applicable, should be identified and informed about the planned project and of the PIA process. Their views should be sought and taken into consideration. The PIA policy should provide explicit mechanisms for stakeholders' consultation.

7. **Publication of the PIA report**: The PIA report should be made public and should be easily accessible.

8. **Central public registry**: There should be a public register of PIAs actually carried out and it should be easily accessible.

9. **Sensitive information**: State secrets and commercially sensitive information should not be made public.

10. **Risks management and legal compliance check**: Risks management and legal compliance check are core elements of PIA. To that end, effective procedures for risk management should be identified and/or developed.

11. **Audit and review**: A PIA process should be subjected to external review and/or audit.

Impact based on understanding of risk

When assessing privacy impact, a thorough understanding of privacy risk is needed. PIA needs to base its conclusions on privacy risks to data subjects and society, and gives these indications about how privacy risk is connected to impact assessment:

- A PIA process requires a relative quantification of these risks. The Assessor should consider the likelihood and consequences of privacy risks occurring. Finally, the risk assessment requires evaluating the applicable risks. Thus the assessor should consider: (1) the significance of a risk and the likelihood of its occurrence, and (2) the magnitude of the impact should the risk occur. The resulting risk level can then be classified as low, medium or high.

- The assessor should identify, assess and mitigate all possible risks and other negative privacy impacts. Residual risks should be justified.
  - Any risk management is only as good as the methodology underlying it. This means if the methodology is flawed, then so is the assessment.
  - The risk assessment should take into account the impacts on both the individual and on society.
- Based on risk assessment: The assessor must define controls for privacy risk! Controls are Preventive controls (prevent violation) or detective controls (detect violation)
  - Technical controls: go into the technical implementation of the project (e.g. security and PET mechanisms,anonymity, data minimization).
  - Non-technical controls get implemented in processes, procedures, policies and operations.

However, as we have seen in the previous section on limitations of PETs in practice, the empirical foundations of the above risk assessment and control selection processes are scarcely(almost not) available, and hence need to get built up with own resources.

Practical assessment of impact

The European Union Agency For Network and Information Security (ENISA) has published guidelines for privacy risk assessment for Small and Medium Enterprises. (SME) (ENISA, 2016) that contain guidance on privacy impact assessment focused on individual data subjects in chapter 3 on page 19. There, four levels of privacy impact are defined:

Level of Impact	**Description**
Low	Individuals may encounter a few minor inconveniences, which they will overcome without any problem (time spent re-entering information, annoyances, irritations, etc.).
Medium	Individuals may encounter significant inconveniences, which they will be able to overcome despite a few difficulties (extra costs, denial of access to business services, fear, lack of understanding, stress, minor physical ailments, etc.).
High	Individuals may encounter significant consequences, which they should be able to overcome albeit with serious difficulties (misappropriation of funds, blacklisting by financial institutions, property damage, loss of employment, subpoena, worsening of health, etc.).
Very high	Individuals which may encounter significant, or even irreversible consequences, which they may not overcome (inability to work, long-term psychological or physical ailments, death, etc.).

15.1.3 Overview over PIA standards

(Marnau, 2013) surveyed a number of PIA standards and guidelines published in various countries. Below in figure 2, you see the classification of the PIA guidelines. Full references can be found in (marnau, 2013).

All guidelines provide checklists or question lists for PIA. Only one of them, the BSI RFID PIA, has protection targets stated. The same guidelines are the only one with risk and attacker information available, and in addition provides risk treatment suggestions. The other guidelines are more or less lists of open questions with the intention to provide a PIA report. Most PIA frameworks must therefore get considered as an invitation to extensive philosophical work.

PIA relates to data protection principles

The PIA assessments should for each transaction on personal data that gets assessed relate to the data protection principles. The PIAF project recommends consideration of these principles as part of the PIA assessment checklist:

- Fair and lawful processing of data
- Purpose specification
- Adequate, relevant and not excessive for the purpose (data minimization)
- Accurate and kept up to date
- No longer than necessary for the purpose
- In accordance to the rights of the data subjects
- Appropriate measures for misconduct (including loss of data)
- Transfer of personal data outside the EU/EEA

Overview	IPC Federated PIA	PbD PIA F Framework	ICO Handbook V2	ISO	IPC PHIPA	PIA Health and Social Care	BSI RFID PIA
Issuer/Year	IPC (Canada) 2009	IPC (Ontario, Canada) 2011	ICO(UK) 2009	ISO 2012	IPC (Ontario, Canada) 2005	Health Information and Quality Authority (Ireland) 2010	BSI (Germany) 2012
Character	Framework	Framework	Handbook	Framework/Draft	Guideline	Guideline	Guideline
Target Audience	Federated Identity Management Services	Organizations processing Personal Information (PI)	Organizations handling personal data	Organizations processing personal identifiable information (PII)	Health Information Custodians (1)	Health and Social Care	RFID operators(EU)
Questionnaire/ Checklist	☑	☑	☑	N/A	☑	☑	N/A
Number of Questions	24	115	104 (2)	N/A	30	11(3)	N/A
Questions Publicly Available	☑	☑	☑	N/A	☑	☑	N/A
Intended Type of Answers	Yes/No+notes	Yes/No+notes	*Yes/No+notes*	N/A	*Yes/No/In Progress/N/A + Notes*	Yes/NO	N/A
Answers Verifiable	N/A	N/A	☑	N/A	☑	☑	N/A

Key Issues Addressed	IPC Federated PIA	PbD PIA F Framework	ICO Handbook V2	ISO	IPC PHIPA	PIA Health and Social Care	BSI RFID PIA
Policy Information (3)	N/A	N/A	☑	N/A	☒	☒	N/A
Methodology	3 Phases PIA, guided by the Global Privacy Standard (5)	Guided by the PbD Principles (6)	5 phases PIA (7)	4 Steps PIA (8)	Answering the questionnaire as a self-assessment tool	4 stages PIA (9)	6 Steps PIA (10)
Legal Compliance Check (11)	N/A	N/A	☑(12)	☒	N/A	☒	☒
Data Protection Targets	☒	☒	☒	☒	☒	☒	☑(13)
PIA Report available for the DPA	N/A	N/A	☑	N/A	N/A	☑	☑(14)
PIA Report Public	N/A	N/A	N/A	N/A	N/A	Recommended, but not mandatory	☒
Stakeholders involvement	N/A	N/A	☑	N/A	N/A	☒	☑
Privacy risk scale	☒	☒	☒	☑(15)	☒	☒	☑(16)
Privacy risk treatment	N/A(17)	N/A(17)	☒	☑	☒	☒	☑
Mandatory PIA	N/A	N/A	☒	N/A	N/A	☒	☑(18)
Possible attackers	☒	☒	☒	☒	☒	☒	☒

Legend: ☑ yes; ☒ no; ☑ probably, but could not be verified; ◈ could not be determined, N/A Not applicable

Figure 2 Survey of PIA methods from TCLOUDS Deliverable 1.2.4 (Marnau, 2013)- Note that ISO/IEC 29134:2017 was not available at the time of analysis.

15.2 Privacy Risk Analysis

15.2.1 Definitions

According to ISO 13335-1 and ISO 27001, risk management terminology is defined in these terms:

- **Risk acceptance** - decision to accept a risk;
- **Risk analysis** - systematic use of information to identify sources and to estimate the risk;
- **Risk assessment** - overall process of risk analysis and risk evaluation;
- **Risk evaluation** - process of comparing the estimated risk against given risk criteria to determine the significance of the risk;
- **Risk management** - coordinated activities to direct and control an organization with regard to risk;
- **Risk treatment** - process of selection and implementation of measures to modify risk;
- **Statement of applicability** - documented statement describing the control objectives and controls that are relevant and applicable to the organizations information security management system (ISMS).

15.2.2 Introduction to risk assessment

Risk Assessment is the systematic study of assets at risk; threats to the assets; vulnerabilities of the system; and impacts (consequences). The goal of risk assessment is the analysis of the probability and consequences of risk when realized.

In a corporate context, potential damages in monetary units are assessed and set in relation to the probability of the damage occurrence. Then a decision is made whether to ignore the risk, buy insurance, invest in technology, or abandon the particular product or service. Much data has been collected by consulting and insurance companies about types of risks and the resulting damages to the owning business. The usual method to guess monetary damages is an analysis of past occurrences of similar problems, the damages caused by them, and the financial loss that has occurred. Additional factors like the value of transactions or the number of customers involved can be used to increase precision of the calculation.

Qualitative risk assessment
Qualitative risk assessment investigates the risks for assets through a principal analysis of threats, threat agents, weaknesses and risk vectors. Qualitative analysis is focused on all possible risks that could get realized. This form of risk analysis is an extensive analysis of system properties. It may require extensive resources, detailed knowledge about the functionality and inner workings of system components, their context and their limitations. This form of analysis is often considered impractical for applications that operate on tight budgets. Practitioners usually skip this principal step by selecting applicable risks for their assessment from risk and threat catalogs.

Quantitative risk assessment
Quantitative risk assessment aims at ranking risks and their impact (or losses realized) by putting likelihood and impact values on risks and their effects: Specifies likelihood and impact of riskful events on assets, Based on historic data for both likelihood and impact, In new contexts often guesswork.

Most risk assessment methods multiply a likelihood parameter with an impact factor to generate the overall risk level or risk impact: *Loss(A) = impact(T(A))*likelihood(T(A)) where T(A) is threat T* effective on asset A. The resulting loss value is then prioritized according to threshold values for low-, medium- and high-priority risks. Some methods assign numerical values to the parameters, others map them scales of 3, 5 or 7 values ranging from "no risk" to "very high risk".

Figure 1 below shows a risk analysis chart. It is a graphical visualization of the risk calculation formula. It shows a 6x7 matrix for impact and likelihood. Assessors can check of their judgment of impact and likelihood for an identified risk on an information asset. The color of the field the mark is placed in gives an indication of the priority of the risk: green indicates low risk, yellow marks medium risk, and red highlights the high-priority risks.

Likel / Imp	Negli	V low	Low	Med	High	V High	Extr
None							
Minor							
Med							
High							
V High							
Extr							

Figure 1: Chart for risk assessment. See text for explanation.

15.2.3 Specification of privacy risks

Unlike the "perimeter security" paradigm that was central to information security for many years, privacy risks occur inside and outside an information system. Where the perimeter security paradigm took care that all critical information stays inside the secured systems, many open systems on the Internet trade personal information and process it as the very purpose of the system. The effects of a breach of private information security could affect the owner of the information system – but also the person the data is about. This invokes a duality of privacy risks effective on both the processor and the data subject, as shown in figure 2. However, in the areas of risk management and investment decisions, the duality of privacy risks has until recently not been the subject of major concern.

When focusing on privacy breaches, little history of damages is known. Two observations make it hard to implement privacy risk management. First, unlike the security risk calculation, in the privacy domain the question of risk is not focusing exclusively on the owner of an IS and the respective damages caused to his business operation. Privacy management also involves the data subject's private data and potential damages caused to the users and their personal business following privacy breaches. The two entities involved complicate the generation of a simple database with cost and probability of privacy breaches, as each type of user - depending on the application - has different personal value at stake. Fundamental questions in privacy risk assessment are: How much damage is a particular privacy breach going to cause? How long will the personal information that got out be a risk? Is the risk constant over time, does it degrade, or will it increase? How does the risk change when personal information is combined with other information? How does the entity using the personal information influence the risk?

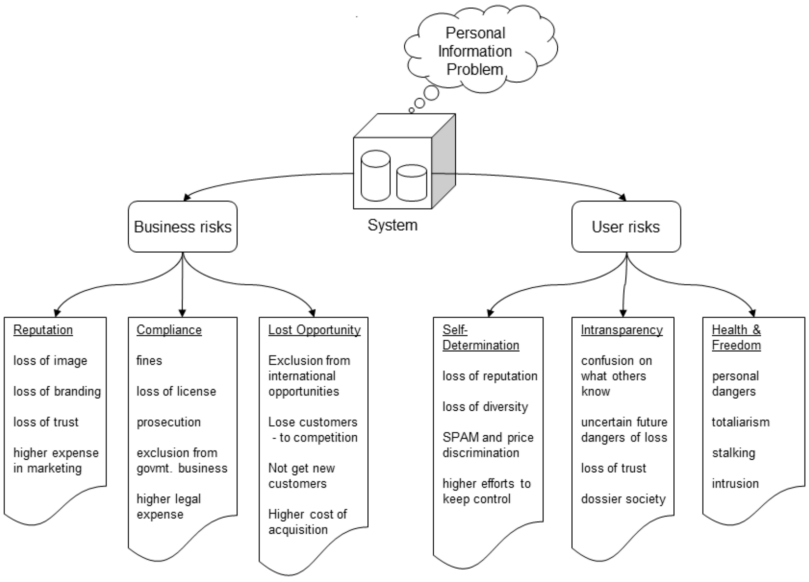


Figure 2: Duality of privacy risks and consequences for organizations and data subjects (Fritsch and Abie 2008).

How should we then, in practice, proceed with risk assessment of privacy risks? In the remainder of this section, we'll take a look at the ENISA Guidelines for SMEs on the security of personal data processing (ENISA, 2016). This booklet contains an introduction to risk assessment, including templates for privacy impact assessment (PIA) and privacy risk assessment. The approach is somewhat restricted in contrast to the analysis in the sections above. It focuses on protection system components, not personal data. However the guidelines were made with small organizations in mind, and therefore apply the idea that good information security will protect personal data processing infrastructure.

16 Privacy controls

16.1 The concept of privacy controls

Privacy controls are lists of measures that will reduce privacy risk contained in an information system. They respond to risks identified in a risk analysis process. They correspond to the impact levels identified in a privacy impact analysis (PIA) . The risk manager chooses matching controls that fulfill a number of requirements.

A **privacy control** is chosen from one of two categories: a technical control; an administrative control.

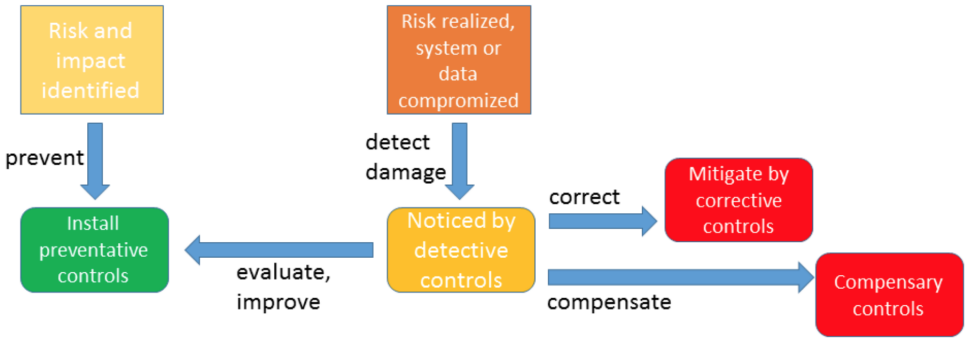
Technical controls are controls that are part of the information technology used to process personal data. technical controls are often functions of information security such as access control, encryption, integrity protection, availability insurance. Safety functions such as fire protection, redundant power supplies and reserve hardware are part of the technical controls, too. Finally, PETs and other privacy support technology such as data hiding, stenography, mathematical data obfuscation and TETs are technical controls. Technical controls are implemented into the technical infrastructure at project development and on the occasion major changes of the technological base.

Administrative controls are all non-technical controls. Administrative controls are, for example, staff qualification management, staff security clearance, the proper administration of data subject consent and privacy policies in harmony with the data transactions performed. The administration of physical access to computing hardware and storage devices, the management of roles and privileges that lead to authorization in access control systems, and the authorization and monitoring of subcontractors are other examples of administrative controls. Sometimes, physical controls are listed separately, implying the securing of physical access to systems. Non-technical controls get implemented in processes, procedures, policies and operations.

16.1.1 Types of privacy controls

Controls are divided into four types of controls. For each type, there may exist technical and administrative (non-technical) controls.

- **Preventative controls** are used to prevent a threat from being realized in a damage.
- **Detective controls** help recognizing that a threat has affected a system.
- **Corrective controls** reduce the effect of a threat by reducing the damage realized.
- **Compensatory controls** are used to mitigate or lessen damage by providing alternative resources.



Please observe the central role of detective controls! Often, only preventive controls are being discussed, in particular when we discuss the use of cryptography or of PETs. However, detective controls are an important trigger for these important reactions on a data breach or compromise that actually happened:

1. They trigger treatment of the compromise with corrective controls;
2. They trigger the use of compensatory controls in case of insufficient correction result or in parallel to correction;
3. They feed back new knowledge about risks and about the effectiveness of the existing preventive controls.

Step 3 is of particular importance as part of a privacy management process! Its feedback is fed into the "act" phase of the Plan-Do-Check-Act (PDCA) process. There, the information is used to evaluate and improve the risk analysis and the corrective actions.

16.2 Risk treatment with privacy controls

Treatment of identified risks is part of a larger process of privacy risk management. Many established standards focus solely on technical aspects, aspects of information security or aspects of managerial responsibilities and processes. Many aspects need consideration. While IT management standards often focus on aspects of technology and reliability, privacy management reaches out into domains such as law, business, optimization of investments and compliance.

In (Fritsch and Abie, 2008), a structured process for privacy investment decisions was introduced. It is composed of the following five steps:

1. System environment analysis
2. Privacy impact analysis
3. Countermeasure selection
4. Total cost of ownership analysis
5. Design & deployment

As shown in figure 1, in step 1 the context of the system is established by collecting the legal and technical framework, the user requirements and business model constraints. The 2nd step assesses privacy threats, models personal data assets, and performs a risk analysis and a privacy impact assessment (PIA). During step 3, candidate countermeasures against the threats are being selected. These include privacy controls and other countermeasures. In step 4, the Return-on-privacy-investment (ROPI) perspective will help to understand the limitations of chosen controls. Finally, in step 5, the chosen controls and countermeasures get deployed, documented and evaluated as part of the application lifecycle.

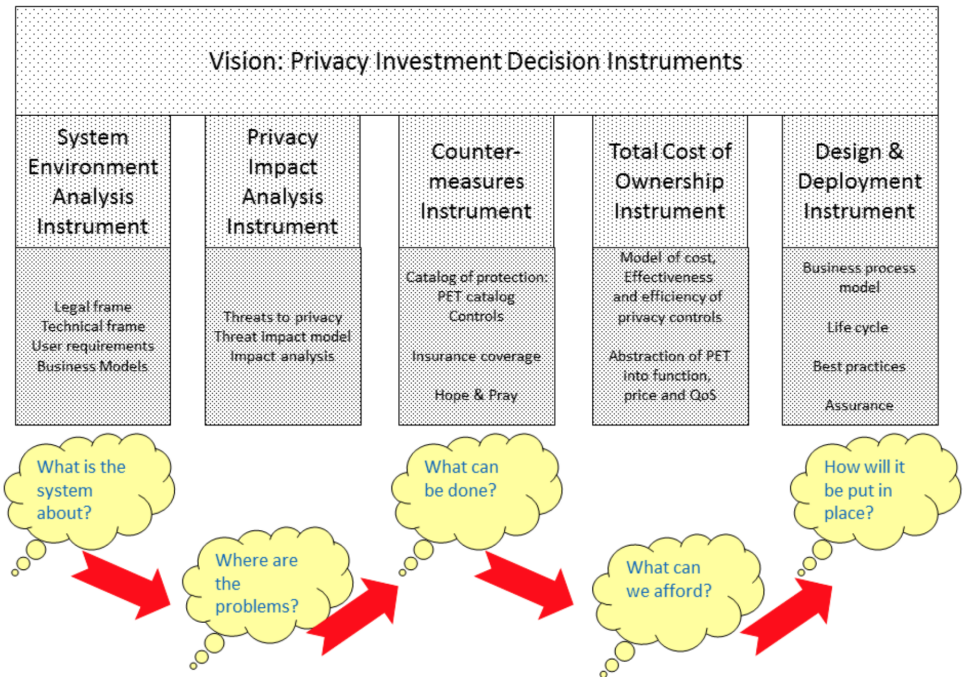


Figure 1: Overall privacy management process supporting control selection (Fritsch and Abie, 2008)

16.2.1 From privacy impact levels to controls in NIST 800-53

How are identified risks treated with controls? Most methods provide a procedure that maps identified risks and privacy impacts through a classification of their magnitude to the respective groups of controls. As controls provide various levels of security, complexity, cost and effectiveness, the mapping aims at matching the level of risk and the level of impact with the appropriate level of protection. Please recall the section "Impact based on understanding of risk" (PIA section of the lecture) and the colored risk analysis chart in figure 1 in the lecture section on Privacy Risk Analysis. Did you read the reading assignment on privacy impact in the privacy overlays document? If not, not, then:

- Please familiarize yourself with section 3.2 "Understanding and evaluating impact" in (ENISA, 2016) on pages 19-23!
- Please read the privacy impact levels shown in table 1 and table 2 on pp.11-13 in CNSSI 1253F Attachment 6-Privacy Overlays

16.3 Properties of privacy controls

16.3.1 Privacy controls: Lists and specifications

This section shows examples of privacy control collections from various sources. The collection is incomplete, as much standardization activity is ongoing. In particular the ISO standards for privacy controls are not yet complete. However the chapters below show a selection of different approaches and standards that illustrate the ideas.

Privacy Strategies

In (Colesky, Hoepman, and Hillen, 2016), the authors discuss how privacy requirements can get modeled into particular actions that support privacy. They identify eight different strategies that get deployed to improve privacy: **ENFORCE, DEMONSTRATE, INFORM, CONTROL, MINIMIZE, ABSTRACT, SEPARATE, HIDE**. Each of the privacy strategies is defined in (Colesky, Hoepman, and Hillen, 2016) . They provide a model on how those strategies can get deployed against privacy-reducing actions. These actions are shown in table 1 as: **Operate, Store, Retain, Collect, Share, Change, Breach**. For each action, one can follow the table horizontally to the eight strategies. The strategies that apply are highlighted with a gray background.

Actions	Taxonomy Group	Applies to	ENFORCE	DEMONSTRATE	INFORM	CONTROL	MINIMIZE	ABSTRACT	SEPARATE	HIDE
Operate	Processing	Processing	•	•	•	•	•	•	•	•
Store										
Retain										
Collect	Collection	Collection								
Share	Dissemination	Dissemination								
Change	Invasion	Invasion								
Breach			•	•	•					

Table 1: Controlling strategies mapped to privacy-consuming actions in privacy strategies from (Colesky, Hoepman, and Hillen, 2016)

You see in table 1 for example that:

- **Store** is an action that gets its privacy improved by all eight strategies **ENFORCE, DEMONSTRATE, INFORM, CONTROL, MINIMIZE, ABSTRACT, SEPARATE, HIDE;**
- **Breach** is an action that related to the strategies **ENFORCE, DEMONSTRATE, INFORM**.

Once the actions on personal data have been mapped out, the relevant strategies for improving privacy are chosen. The implementation of the strategies is then done by applying privacy patterns that implement the respective strategies. The web page https://privacypatterns.eu is a collection of privacy patterns intended to be used as part of the privacy strategy implementation.

NIST 800-53

NIST Special Publication 800-53 (NIST, 2017) on Security and Privacy Controls for Information Systems and Organizations is the specification of privacy and security controls for public offices in the United States. It contains an extensive collection of specified controls including appendices that show how to select controls that respond to various risk and impact levels. NIST has defined several families of controls that are shown in table 2 below.

Privacy Control Family	PRIVACY CONTROLS
AP	**Authority and Purpose**
AP-1	Authority to Collect
AP-2	Purpose Specification
AR	**Accountability, Audit, and Risk Management**
AR-1	Governance and Privacy Program
AR-2	Privacy Impact and Risk Assessment
AR-3	Privacy Requirements for Contractors and Service Providers
AR-4	Privacy Monitoring and Auditing
AR-5	Privacy Awareness and Training
AR-6	Privacy Reporting
AR-7	Privacy-Enhanced System Design and Development
AR-8	Accounting of Disclosures
DI	**Data Quality and Integrity**
DI-1	Data Quality
DI-2	Data Integrity and Data Integrity Board
DM	**Data Minimization and Retention**
DM-1	Minimization of Personally Identifiable Information
DM-2	Data Retention and Disposal

Table 2: NIST Special Publication 800-53 Revision 4: Privacy control families. (NIST, 2017) contains all controls with extensive descriptions for reference.

CNSS Privacy Overlays to NIST

On April 23, 2015, the Committee on National Security Systems (CNSS) published the Privacy Overlay to CNSS Instruction (CNSSI) 1253, "Security Categorization and Control Selection for National Security Systems." The Privacy Overlay is Appendix F, Attachment 6 to CNSSI 1253.

The Privacy Overlay is comprised of four Privacy Overlays that identify security and privacy control specifications from NIST Special Publication (SP) 800-53, rev. 4, "Security and Privacy Controls for Federal Information Systems and Organizations" to protect personally identifiable information (PII), including protected health information (PHI), in National Security Systems (NSS) and reduce privacy risks to individuals throughout the information life cycle. The Privacy Overlay is required for all member agencies of the CNSS (https://www.cnss.gov/CNSS/about/structure.cfm#members). In addition, the Department of Defense (DoD) requires use of the Privacy Overlay for all DoD information systems (NSS and non-NSS) including cloud based systems containing PII (see, DoD's Cloud Computing Security Requirements Guide). The Privacy Overlay selects controls and control extensions from both the Security Control Catalog and the Privacy Control Catalog of NIST SP 800-53, rev. 4 and provides additional supplemental guidance and legal references. The Privacy Overlay contains tables that help translate the result of a privacy impact assessment into a list of appropriate controls. Please note that a public draft of revision 5 of SP 800-53 has been published (NIST, 2017). However, the document is not finalized. Privacy control families are summarized in appendix E, while guidance for privacy control selection is provided in appendix F. There, the aforementioned appendix J has been integrated.

LINDDUN

LINDUN maps it threats via mitigation strategies directly to PETs from research literature. Where Colesky, Hoepman, and Hillen map their strategies into privacy patterns, the LINDDUN process finishes with references to research literature. Table 3 below shows the LINDDUN mapping (literature references omitted - refer to LINDDUN web page for references!).. It is not overly useful for practitioners, however it gives good insights into the research horizon on privacy controls.

Mitigation Strategy				Privacy Enhancing Techniques (PETs)
Protect ID	Pseudonyms			Privacy enhancing identity management system [HBC+04], User-controlled identity management system [CPHH02]
	Attributes			Privacy preserving biometrics [STP09], Private authentication [AF04, ABB+04]
	Properties			Anonymous credentials (single show [BC93], multishow [CL04])
Protect data	Transactional data	Remove		(see awareness to minimize information sharing)
		Hide	Data-flow specific	Multi-party computation (Secure function evaluation) [Yao82, NN01], Anonymous buyerseller watermarking protocol [DBPP09]
			General	see guard exposure - Confidentiality - encryption
		Replace		/
		Generalize		see guard exposure - minimization - generalize
	Contextual data	Remove		Mix-networks (1981) [Cha81], , ISDN-mixes [PPW91], Onion Routing (1996) [GRS96],Tor (2004) [DMS04])
		Hide	General	Crowds (1998) [RR98], Low-latency communication (Freedom Network (1999-2001) [BGS01], Java Anon Proxy (JAP) (2000) [BFK00]
			Undetectability	Steganography [AP98] , Covert communication [MNCM03], Spread spectrum [KM01]
			Non-repudiation	Deniable authentication [Nao02], Off-the-record messaging [BGB04]
		Replace		Single proxy (90s) (Penet pseudonymous remailer (1993-1996), Anonymizer, SafeWeb), anonymous Remailer (Cipherpunk Type 0, Type 1 [Bac], Mixmaster Type 2 (1994) [Mixa],Mixminion Type 3 (2003) [Mixb])
		Generalize	Undetectability	dummy traffic, DC-networks (1985) [Cha85, Cha88]
	Awareness	Feedback and awareness tools		Feedback tools for user privacy awareness [LHDL04, PK09, LBW08]
		User-friendly privacy support		Data removal tools (spyware removal, browser cleaning tools, activity traces eraser,harddisk data eraser)

Guard exposure	Compliance	Policies and Consents		Policy communication (P3P [W3C]), Policy enforcement (XACML [oo], EPAL [IBM])
		Notice and Transparency		/
	Confidentiality	Encryption		Symmetric key & public key encryption [MOV97], Deniable encryption [Nao02],Homomorphic encryption [FG07] , Verifiable encryption [CD98]
		Access control		Context-based access control [GMPT01], Privacy-aware access control [CF08, ACK+09]
	Minimization	Remove		/
		Hide	Receiver privacy	Private information retrieval [CGKS98], Oblivious transfer [Rab81, Cac98])
			Database privacy	Privacy preserving data mining [VBF+04, Pin02], Searchable encryption [ABC+05],Private search [OS05]
			General	see guard exposure - confidentiality - encryption
		Replace		/
		Generalize		K-anonymity model [Swe02b, Swe02a], l-Diversity [MGKV06]
Maximize accuracy	Review data			/
	Update/ request deletion			/

Table 3: Mapping of Mitigation Strategies to Privacy Enhancing Solutions in LINDDUN

As we can see in table 3, the LINDDUN method aims at mapping countermeasures to known cryptographic solutions or PETs. At the LINDDUN web page, you will find a fully referenced version of table 3.

16.3.2 ISO standards

As of today, the ISO has not published a standard on privacy controls. Controls are mentioned in some of ISO's standards, e.g. in the privacy management framework, however the ongoing work has not led to a published standard yet. For the reason of the ISO's standards being protected by a paywall, they are not used in this course.

16.4 Selection criteria for privacy controls

The selection of appropriate privacy controls is an unstructured process that has to balance several factors. Controls are different in their risk treatment, in the cost they impose, in their compatibility with administrative procedures and in many other parameters. The privacy officer, the risk manager, the business process owner and the technical experts should choose, evaluate and device upon the appropriate controls for risks. Figure 1 below shows the seven factors that the selection of controls has to take into consideration.

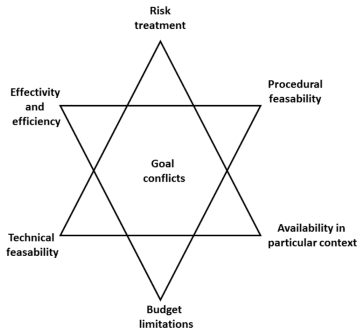


Figure 1: Selection criteria for controls.

16.4.1 Risk treatment

Privacy controls will vary in their particular approach to risk treatment. Some may deploy PETs, others may minimize the collection of risky personal data, yet others may install safety procedures. A clear understanding of how the risk is mitigated by the control, and how the overall impact of a data breach will be reduced by the control should get assessed under control evaluation.

16.4.2 Availability in particular context

Privacy controls may not be available in certain contexts. Technological solutions may be restricted from being used in certain countries due to export control regimes. Location may restrict remote access to data. Other situations may require certified and security-cleared systems and staff - which may or may not be available for a particular control. When evaluating candidate controls, a thorough understanding of the application contexts of the control is necessary. You should, in addition, consider the consequences of changes to the business model that may change the context and use environment of the control in the near and medium future.

16.4.3 Budget limitations

Controls come at different cost. Besides direct licensing, controls have a cost of operation, a cost of maintenance and possible additional cost for more complex overall system handling. The total cost of ownership of deploying and maintaining particular privacy controls should be assessed.

16.4.4 Goal conflicts

Controls may cause goal conflicts with other important features or requirements of the system or the processes they get applied to. Controls for data minimization, anonymization or encryption may, for example, violate archiving requirements, public transparency laws or other internal requirements. Therefore, the conflicts that a control may cause with respect to other system goals than data protection should get analyzed before deployment.

16.4.5 Effectivity and efficiency

Controls will be of different effectiveness in mitigating risks, while at the same time mitigating the risk with varying degrees of efficiency. Proven controls should provide information about both the effect and the efficiency. However, as shown in this lecture, those parameters are widely unknown as of today.

16.4.6 Technical feasibility

Technical privacy controls may or may not be available for particular technical contexts. Operating system variations, programming languages, availability of portable source code in suitable programming languages and system demands may make particular controls unusable for a particular component. The technical managers, operators and developers should inspect such controls before decisions are made about these.

16.4.7 Procedural feasibility

Some controls are by their structure not compatible with the existing administrative processes of the application, of the IT department or of other units involved in processes with identified risks. Risk managers will therefore, in collaboration with the business process owner and possibly other managers, discuss the alignment of privacy controls with the existing procedures and processes the control will get deployed into.

Part V

Privacy Management

17 Introduction

17.1 Context of privacy management

Digital privacy, today is regarded as the right and the ability of persons to interact digitally while in control over information dissemination and informed about who is participating in the interaction, what information about the person is exchanged, what the information is being used for. Digital privacy can be seen as two concepts side-by-side – the right to be left alone, not get observed while on-line; and the concept of fair-play and lawfulness expressed in data protection philosophy and regulation. Recent regulation like GDPR shows signs of matching both concepts by including person-centric mandatory risk and impact analysis into regulation of personal data processing. Please proceed to the following sections that will explore the origins and the context of the concepts of information privacy and data protection! You will first read about the development of the concept of privacy as an individual's right. Then you will learn about the development of data protection regulation that governs the use of personal data in today's information systems. Next, we will have a look at the stakeholders involved in information privacy. Finally, you will learn the background of information privacy and for privacy enhancing technology. This is an important section, since it lays the foundation for your understanding of the thinking that is put into technical privacy.

17.2 What is "Privacy"?

17.2.1 Definitions

Oxford dictionary:
A state in which one is not observed or disturbed by other people. The state of being free from public attention.
Merriam-Webster dictionary:

- the quality or state of being apart from company or observation : seclusion
- freedom from unauthorized intrusion - one's right to privacy
- archaic : a place of seclusion
- secrecy
- a private matter : secret

plural: privacies - first known use: 15th century

17.2.2 The right to be let alone

"Recent inventions and business methods call attention to the next step which must be taken for the protection of the person, and for securing to the individual what Judge Cooley calls the right "to be let alone" Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life; and numerous mechanical devices threaten to make good the prediction that "what is whispered in the closet shall be proclaimed from the house-tops." (Warren/Brandeis, 1890).

"It is our purpose to consider whether the existing law affords a principle which can properly be invoked to protect the privacy of the individual; and, if it does, what the nature and extent of such protection is." (Warren/Brandeis, 1890). Warren/Brandeis defined privacy as a "right to be let alone" (1890) inspired by social invasion caused by new media, especially invasive press photography. They conceptualized "privacy invasion" inspired by property invasion and bodily invasion (e.g. physical punishment or assault), and thereby let the concept of privacy enter the legal discussion. In their article, they argue how privacy regulation would not infringe on the freedom of the press - and how sanctions for violations could be implemented.

Privacy and privacy violations

"Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others. Viewed in terms of the relation of the individual to social participation, privacy is the voluntary and temporary withdrawal of a person from the general society through physical or psychological means, either in a state of solitude or small-group intimacy or, when among larger groups, in a condition of anonymity or reserve." (Alan Westin, 1967) With the appearance of computers and more and more automated state surveillance, the right to privacy was transformed into a perspective on the control of data dissemination. Unauthorized collection and use of personal data was considered a risk to the freedom of the individual. The article "Taxonomy of Privacy" by Dan Solove (2006) extends the concept to a multi-step concept applied to privacy into the digital age. Solove shows how personal data can cause harm to persons. Figure 1 below shows the four areas of actions that violate privacy from Solove's taxonomy:

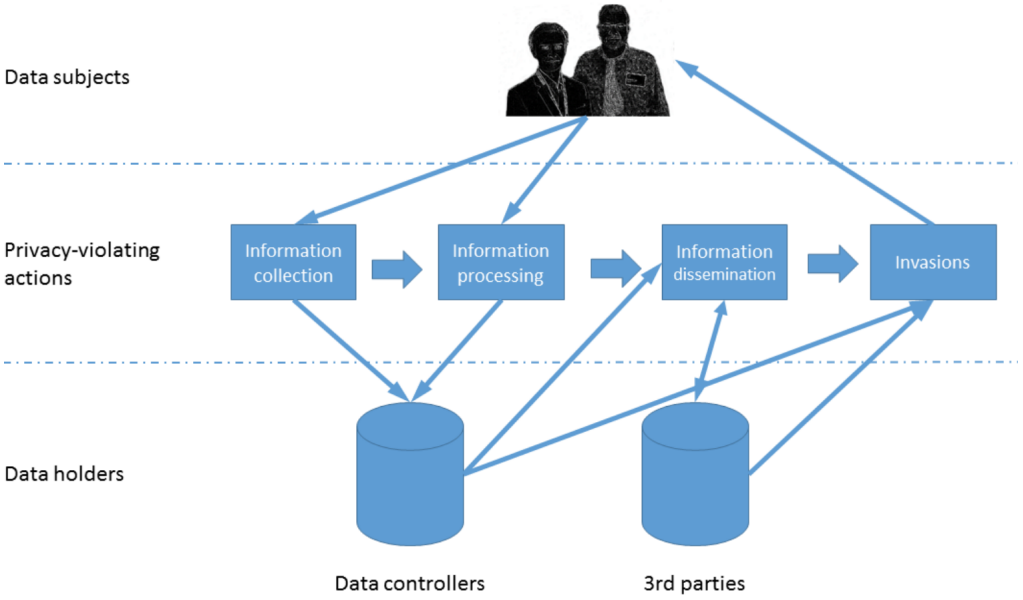


Figure 1: Privacy-volating actions after (Solove, 2006).

The four privacy-violating actions are:

1. **Information collection**: Solove groups information collection about data subjects in two groups: surveillance and interrogation. Surveillance refers to observation and collection without the data subject's active support. Interrogation refers to data collection as a result of asking the data subject in one way or another.

2. **Information processing**: Information processing is the activity that uses or aggregates personal data, binds it to an identified person or to other identified data. Secondary use outside of a collection purpose as well as concerns of data subject discrimination and other issues of power are part of these activities.

3. **Information dissemination**: Any disclosure of personal information to other parties, e.g a confidentiality breach, data lost to hackers, or intentional transfer of data to 3rd parties are part of the dissemination activities. Consequences for data subjects are disclosure, exposure, increased accessibility, and possible distortions of life such as blackmail, appropriation and others.

4. **Invasions**: These activities invade directly into the lives of data subjects. Solove mentions intrusions (uninvited contact, e.g. for advertising or by stalkers) and decisional interference against the data subject (e.g. when systems price-discriminate based on personal data collections).

17.3 What is "Data Protection"?

"No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks." (Article 12, The Universal Declaration of Human Rights, 1948)

17.3.1 From "being let alone" to data protection regulation

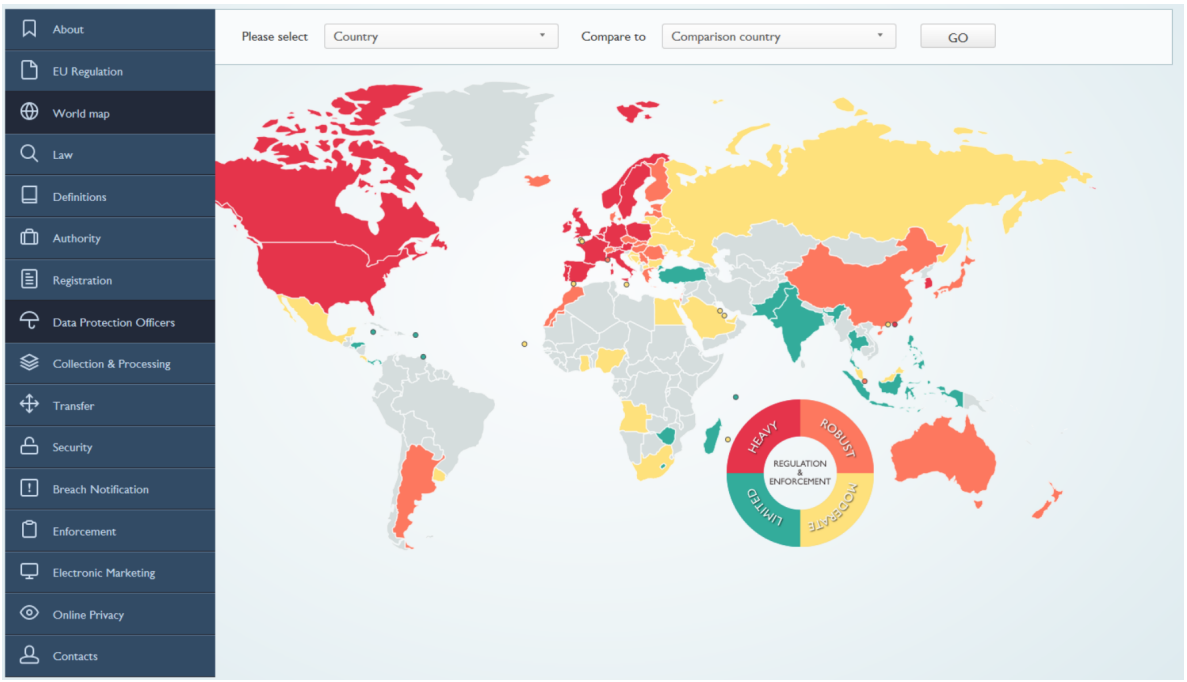
With computerization, governments and large corporations began to collect and process citizen data. Starting in the United States of America in the 1960s, a legal and philosophic discourse about freedom, data collection and government power led to the idea of the regulation of the processing of personal data. In 1974, the first U.S. Privacy Act formalized regulation of government data processing. The privacy debate was taken up in Europe at the same time. In 1970 in Hessen (Federal state of Germany) and in 1977 for all of Germany, data protection legislation was introduced. It was mainly intended to protect the individual from state power. According to Deutsche Welle, "France was one of the first countries in Europe to enact a privacy law. The French parliament in 1978 decreed that any person company or government agency receiving or processing personal information without authorization could be punishable by up to six months in prison and a maximum fine of 20,000 francs (3,000 euros, $4,115)".

A milestone was the Council of Europe's Convention 108 "Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data" in 1981. It lay the grounds for converging regulation of data protection in what was to become the European Union. National refinements and EU directives further defined the concept of data protection in Europe. Finally, in 2012, a standardized European regulation for data protection was agreed upon: the General Data Protection Regulation (GDPR), to be applied latest from May 25, 2018.

Current regulation defines the rights of citizens (information, accession and transparency) as well as the duties of data collectors and data processors. It, in addition, defines the roles of national supervisory authorities and of data protection officers. In data protection regulation, often "sensitive personal data" is especially protected: Health data, sexual orientation, membership in unions, religion, among others. Processing of such data requires either special permissions, or is regulated by specific laws.

17.3.2 International regulation

Internationally, regulation is very varied. Beyond EU borders, different privacy legislation or data protection regimes apply. Some countries may have very little to no regulation. The EU legislation is therefore very concerned with cross-border transfer of personal data, which is only allowed to organizations and countries that are considered equivalent to GDPR standards.



Compare data protection laws around the world. DLA Piper map of data collection legislation.

17.4 Stakeholders and forces

17.4.1 Stakeholders in information privacy and data protection

When referring to responsibilities and roles in personal information processing, many names are used for the involved stakeholders. This page introduces the important stakeholder names and concepts.

17.4.2 Data protection terminology

In data protection regulation, three major stakeholder roles are commonly used:

- **Data subject**: A natural person governed by data protection legislation whose personal data is being collected, stored and/or processed.
- **Data controller**: An entity (often organization) that is responsible for data processing of a data subject's personal data. A data controller is in control of the processing application, and in addition has the legal relationship with the data subjects that governs privacy policies and informed data subject consent.
- **Data processor**: A processing entity engaged by the data controller. The processor activities are governed by and held liable against the controller.

The roles and responsibilities of controllers and processors are extensively defined in the European General Data Protection Regulation (GDPR) in chapter 1, article 4, which defines the rules for personal data processing in the European Union.

17.4.3 Stakeholders in business contexts

Stakeholders in privacy and data protection are those human beings or organizations involved in the processing of personal information. They represent business interests, society as a whole (societal values as well as government interests), individuals being processed, government regulators, and technology vendors. Other stakeholders may be consumer groups and activist groups. For the sake of this course module, we focus on the stakeholders in those organizations that are responsible for digital products and services that process personal data.

OASIS defines the organizational stakeholders as:

- **Chief privacy officer (CPO / DPO)**: Manager responsible of all matters concerning privacy and data protection. Also called "DPO" - Data protection officer.
- **IT Architect**: IT specialist designing software and systems.
- **Chief Information Officer**: Overall responsibility for IT operations.

- **Business Analyst**: Business specialist developing the business aspects of an application.
- **Team Privacy Champion**: Team specialist responsible for privacy issues.
- **Senior Developer**: Often team leader or project leader, responsible for strategic development decisions.
- **Line of Business Owner**: Manager responsible for operations or overall business based on application. Could be product manager, operations manager or other functions.
- **Legal Department**: Processes legal and policy related matters, compliance documents and complaints.

OASIS suggests that the organizational stakeholder responsibilities should be defined with the use of a table:

Organizational Stakeholder	Data Subjects	Applications	Personal Information	Policies	Domains	Data Flows & Touch points	Privacy Controls	Services and technical functionality
CPO								
IT Architect								
Business Analyst								
Team Privacy Champion								
Senior Developer								
Line of Business Owner								
Legal Departemt								
Chief Information officer								

17.4.4 Other stakeholders

Governments are stakeholders in data protection. They enact legislation that sanctions violations. They appoint, in addition, **supervisory authorities** such as **data protection offices** and **consumer protection authorities** that regulate societal and market forces. The near future may see the appearance of **certification services** and **audit consultancies** for privacy compliance. **Non-government organizations (NGO)** and **activists** are stakeholders that lobby governments or other stakeholders.

17.5 Privacy & Technology

Information privacy must get considered by and implemented into technological systems that process personal information. Information technology is therefore in the core focus of privacy technology. However, additional important aspects provide to the issue: Information security, IT management (organization of business processes and responsibilities, and technology perspectives on the data subject's involvement. How will IT systems provide "good privacy"?

17.5.1 The information security perspective

One understanding of technical privacy focuses on securing the personal data that is being processed with the means of information security techniques. By implementing the C-I-A (confidentiality, integrity, availability) properties with the help of cryptography, access control technology and through other means, privacy is supported. However, C-I-A does not support many data protection requirements, such as consensual data collection, or the alignment of processing with privacy policies. Therefore, we shall remember: "privacy" is not equal to "personal data encryption plus access control". Cryptography is an essential tool in providing good information privacy, however the cryptographic protocols that enable and maintain information privacy are far more varied and complex than encryption algorithms.

17.5.2 Privacy as information flows and policy

Privacy in IT should be seen as the management of consensual processes and information flows with personal data, and as interactions based on data subject interventions. Many applications of IT implement business processes that consume personal data. Starting from a business model that defines product, target market, assets and processes, systems are build to implement business process models. As part of such business processes, personal data flows from one human or digital agent to the next until the business transaction is delivered. When designing such business process models, privacy is an important consideration. Processes are often modeled with graphical tools related to business software. Modeling methods such as "Business Process Modeling" (BPM) are used for their specification.

How will privacy and data protection enter business process design? IBM Research invented the "Enterprise Privacy Authorization Language" (EPAL) that "enables an enterprise to formalize the exact privacy policy that shall be enforced within an enterprise. It formalizes the privacy promises into policies and associates a consented policy to each piece of collected data. This consented policy can then be used in access control decisions to enforce the privacy promises made". When used to model privacy policies, researchers found that privacy policies still offer ambiguities that are open to interpretation by human beings (see Stufflebeam et. al.). Due to its complexity, EPAL was not widely taken up. As of today, there are no general methods that transform privacy requirements into system design.

17.5.3 Privacy as user-centric regime

Data protection regulation is assuming a mature and free individual citizen in the role of the data subject. Therefore, data subjects have extensive rights in knowing what is being collected and processed. They should have the right to determine which information they share with whom. In addition, data subjects can withdraw themselves from data processing - and require their identifiable personal data deleted or anonymized.

These transparency and intervention rights are described with the term "user-centric privacy". User/centric privacy is concept opposed to intransparency, to "being managed". Its basic idea is that data subjects make conscious decisions about releasing personal data. Since the release of personal data can change power distribution in relationships as well as create personal harms and risks, data subjects shall be able to assess the status of their personal data as well as the risks. User-centric privacy therefore has two major components:

1. Transparency of data collection and processing to and at controllers through transparency-enhancing technology (TET)

2. Control over release, use and dissemination of personal data through privacy-enhancing technology (PET)

To implement user-centricity, the idea of **intervenability** has come into data protection. Intervenability is the concept of data subject intervention into data processing. Current legislation such as the GDPR define mandatory data subject interventions (accession, deletion, correction, withdrawal of consent). Together with being informed about processing (privacy policy) and being notified about data breaches, intervenability covers much of user-centricity at regulation level.

17.5.4 Privacy as a set of properties of systems and data

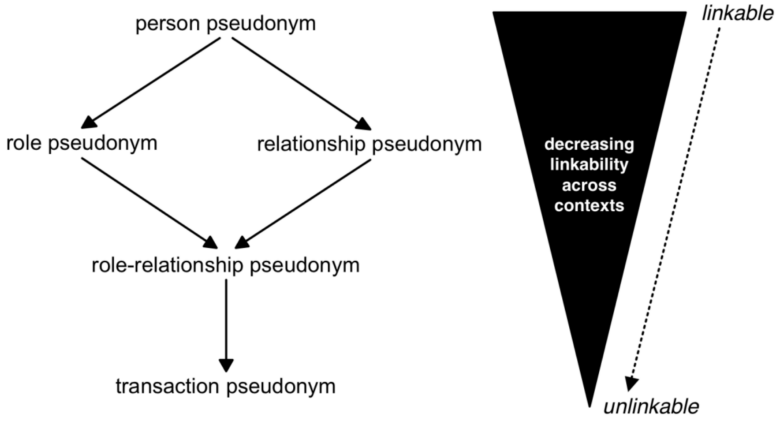
In information privacy research, privacy is described as a set of system properties. Such properties are formalized as technical properties. Pfitzmann and Hansene provided an in-depth definition of terms and technical concepts in Version 0.34 of the "*Terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management*". It contains definitions and concepts for:

- Anonymity & Identifiability
- Unlinkability
- Undetectability & Unobservability
- Pseudonymity
- Identity and Identity Management
- Partial Identities

A large part of the scientific study of PET (Privacy enhancing technology) uses the concepts and definitions from this document. You should study the document well! The document contains an appendix with translations of the terminology to Czech, Dutch, French, German, Greek, Italian, Japanese, Russian, Slovak and Turkish.

Example: Pseudonym quality lattice

The figure below shows the degrees of unlinkable transactions provided by different types of pseudonyms. On the top, personal pseudonyms provide the least degree of unlinkability, since the person will be recognized with every transaction. Role or relationship pseudonyms specialize for a particular role or business relationship. Finally, transaction pseudonyms that will not get re-used after "business is done" provide the highest degree of unlinkability.



18 Privacy management approaches

18.1 IT Management: Plan-Do-Check-Act cycle

The Plan-Do-Check-Act (PDCA) model is a process-oriented approach for continuous improvement of business processes. It is credited to W. Edwards Deming. Its intention is the improvement of processes or the removal of problems as part of an iterative process that includes planning, assessment, implementation and evaluation. Figure 1 below shows the four stages Plan, Do, Check, Act as a cyclus that starts with the planning phase.

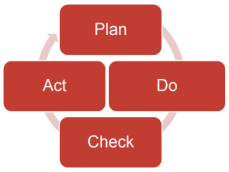


Figure 1: The PDCA cycle

PDCA is the framework of many IT management standards. The ISO 2700x family of standards for information security management is an implementation of PDCA that is very relevant for privacy management. Figure 2 below shows the activities of security management mapped to the phases of PDCA.

1. In the Plan phase, the business objectives are identified. Management support is obtained, the scope of the ISMS is defined. Risk analysis methods are chosen, and an appropriate inventory of assets at risk with ranked risk assessments is produced.
2. The Do phase manages the risks by generating a treatment plan for the risks, by allocating budgets, training staff and by the creation of policies.
3. The Check phase monitors the implementation of the security management activities, and possibly prepares for the certification of its results.
4. The Act phase carries our re-assessment audits that evaluate the overall outcome of the corrective actions and the initiates a new round of the cycle with corrective input, if necessary.

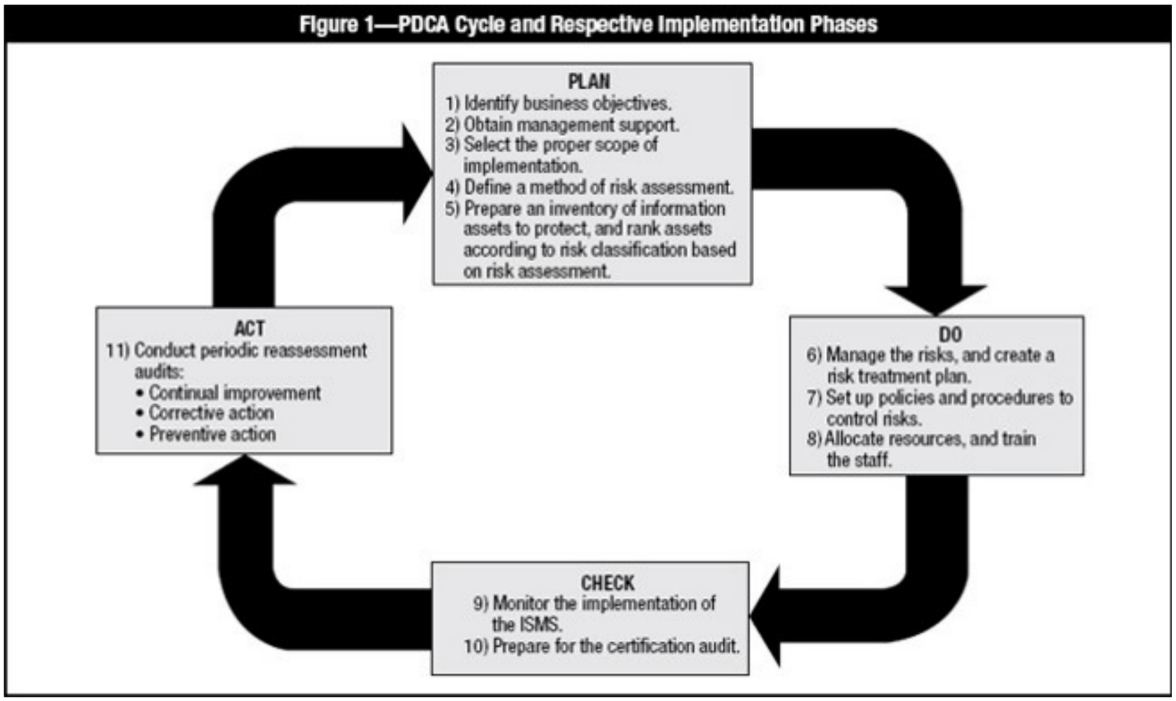


Figure 2: PDCA cycle for ISO 27000 Information Security Management System (ISMS). From "Planning for and Implementing ISO 27001" by Charu Pelnakar, ISACA Journal, Vol. 4, 2011.

If the first iteration of a PDCA cycle does not remove the risks in satisfactory ways - or if new assessments should introduce new risks - then the cycle will be re-run again and again, until the major security issues are resolved in satisfactory ways. Figure 3 shows three iterations of PDCA that finally in the 3rd iteration resolved the problem:

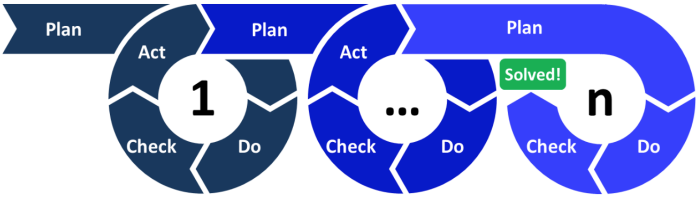


Figure 3: Multiple iterations of the PDCA cycle are repeated until the problem is solved.

A ISMS in compliance with ISO27001 will require very detailed preparation, and will need in addition dedicated staff resources. In Figure 4 below, you can see a detailed view of an ISMS. A full ISMS is therefore often feasible for large organizations with both corporate IT policies, training and professional information security managers. An existing ISMS can be used as a framework for the integration of privacy management by introducing privacy risk assessment elements, personal data assets and corrective actions for privacy and data protection into the existing ISMS. However there will arise a need for more staff trained for privacy management.

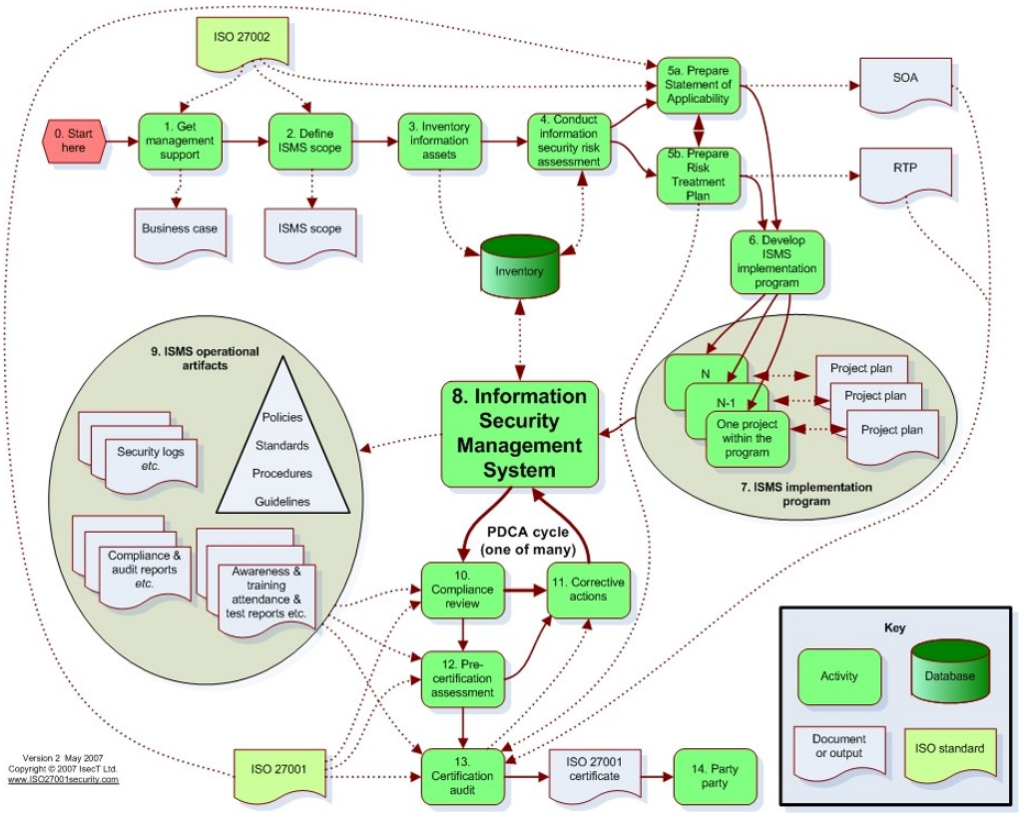


Figure 4: Detailed sketch of an ISO27001 ISMS.

If you consider the integration of privacy management into your existing ISMS, the article "Mapping between GDPR (the EU General Data Protection Regulation) and ISO27k" published by the ISO27k forum will be helpful to you. It lists overlaps and similarities between GDPR requirements and ISO2700x security management. The document even suggests relevant ISO 27001 controls for managing privacy.

18.2 LINDDUN method for threat-based privacy design

18.2.1 Overview

LINDDUN (short for **L**inkability, **I**dentifiability, **N**on-repudiation, **D**etectability, **D**isclosure of information, **U**nawareness, **N**on-compliance) is a privacy threat analysis method that assesses the privacy threat situation. It supports the privacy analyst with threat trees that help to select corresponding privacy controls. LINDDUN was developed by the DistriNet Research Group at KU LEUVEN in Belgium. LINDDUN is a two-phase process that first defines the problem space, and then charts the solution space for privacy threats. Step 1 to 3 are part of the problem space definition, while steps 4 to 6 are part of the solution space mapping. To define the problem space, LINDDUN works based on data flow diagrams (DFD). They map the flow of personal data between actors and components of an information system. This is represented in step 1 of the LINDDUN process diagram in figure 1.

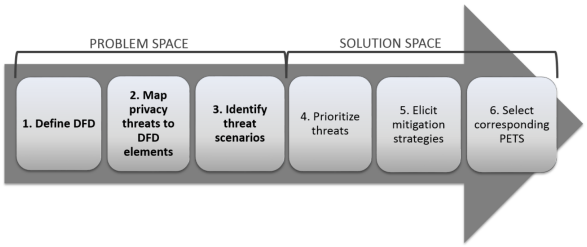


Figure 1: LINDDUN process, from www.linddun.org

Step 2 maps privacy threats to the identified data flows. In step 3, the overall threat scenarios on the system are specified. With step 4, we enter the solution space. Step 4 prioritizes the identified threats. In step 5, the threat mitigation strategies are chosen. Finally, step 6 selects the effective countermeasures by mapping identified mitigation strategies to privacy technologies. To facilitate analysis, LINDDUN provides threat lists, lists of PETs (Privacy enhancing technologies) and strategy guidance that is to be used under analysis.

18.3 Privacy by Design

18.3.1 Term and Context

"Privacy by Design" is a term that refers to the idea of designing privacy functionality into information systems from the start. It is a term that evolved in the 1990s and matured in the 2000s. The term was used often in political and activist contexts that were promoting the idea of user-centric information privacy. To understand the term, one has to imagine the context of that time. Information systems were designed with the mindset of the web client connected to the large server. The purpose of a system was the implementation of a business purpose. Only the business purpose and the cost of the system were relevant. Regulation such as data protection was perceived as a cost factor that had to get dealt with at minimum expense.

This lead to a number of strategies and tactics deployed for data protection compliance, such as:

- Creative legalese writing of complex and unreadable privacy policies that were made to hide the actual vast data processing consent (see Annie Anton's article).
- Power play: Users who disliked data collection were denied service and participation.
- Ignorance: requests for insight into data processing were ignored or denied based on claims of business secrecy.
- Users were asked to protect their privacy with external actions and tools such as early user-centric PETs.
- Software that had options for privacy-friendly configuration came with default configurations that consumed personal data (a strategy still deployed by web browser and operating system vendors today).

In this context, privacy researchers recognized that end users (data subjects) will never possess enough power, resources or time to actually execute their rights guaranteed in data protection acts. The researchers concluded that the burden of delivering compliant IT systems should be put on the system vendors, not on their users. The vendors were seen responsible for designing privacy-friendly systems. This was the birth of the "Privacy by Design" concept.

18.3.2 Concept

The seven principles of Privacy by Design, as postulated by Anne Cavoukian, are described below.

Principle 1: Proactive not Reactive; Preventative not Remedial

The Privacy by Design approach is characterized by proactive rather than reactive measures. It anticipates and prevents privacy invasive events before they happen. PbD does not wait for privacy risks to materialize, nor does it offer remedies for resolving privacy infractions once they have occurred – it aims to prevent them from occurring. In short, Privacy by Design comes before-the-fact, not after.	Whether applied to information technologies, organizational practices, physical design, or networked information ecosystems, PbD begins with an explicit recognition of the value and benefits of proactively adopting strong privacy practices, early and consistently (for example, preventing (internal) data breaches from happening in the first place). This implies: • A clear commitment, at the highest levels, to set and enforce high standards of privacy – generally higher than the standards set out by global laws and regulation. • A privacy commitment that is demonstrably shared throughout by user communities and stakeholders, in a culture of continuous improvement. • Established methods to recognize poor privacy designs, anticipate poor privacy practices and outcomes, and correct any negative impacts, well before they occur in proactive, systematic, and innovative ways.

Principle 2: Privacy as the Default

We can all be certain of one thing – the default rules! Privacy by Design seeks to deliver the maximum degree of privacy by ensuring that personal data are automatically protected in any given IT system or business practice. If an individual does nothing, their privacy still remains intact. No action is required on the part of the individual to protect their privacy – it is built into the system, by default.	This PbD principle, which could be viewed as Privacy by Default, is particularly informed by the following FIPs: • Purpose Specification – the purposes for which personal information is collected, used, retained and disclosed shall be communicated to the individual (data subject) at or before the time the information is collected. Specified purposes should be clear, limited and relevant to the circumstances. • Collection Limitation – the collection of personal information must be fair, lawful and limited to that which is necessary for the specified purposes. • Data Minimization – the collection of personally identifiable information should be kept to a strict minimum. The design of programs, information and communications technologies, and systems should begin with non-identifiable interactions and transactions, as the default. Wherever possible, identifiability, observability, and linkability of personal information should be minimized. • Use, Retention, and Disclosure Limitation – the use, retention, and disclosure of personal information shall be limited to the relevant purposes identified to the individual, for which he or she has consented, except where otherwise required by law. Personal information shall be retained only as long as necessary to fulfill the stated purposes, and then securely destroyed. Where the need or use of personal information is not clear, there shall be a presumption of privacy

Principle 3: Privacy Embedded into Design

Privacy by Design is embedded into the design and architecture of IT systems and business practices. It is not bolted on as an add-on, after the fact. The result is that privacy becomes an essential component of the core functionality being delivered. Privacy is integral to the system, without diminishing functionality.	Privacy must be embedded into technologies, operations, and information architectures in a holistic, integrative and creative way. Holistic, because additional, broader contexts must always be considered. Integrative, because all stakeholders and interests should be consulted. Creative, because embedding privacy sometimes means re-inventing existing choices because the alternatives are unacceptable. • A systemic, principled approach to embedding privacy should be adopted – one that relies upon accepted standards and frameworks, which are amenable to external reviews and audits. All fair information practices should be applied with equal rigour, at every step in the design and operation. • Wherever possible, detailed privacy impact and risk assessments should be carried out and published, clearly documenting the privacy risks and all measures taken to mitigate those risks, including consideration of alternatives and the selection of metrics. • The privacy impacts of the resulting technology, operation or information architecture, and their uses, should be demonstrably minimized, and not easily degraded through use, misconfiguration or error.

Principle 4: Full Functionality – Positive-Sum, not Zero-Sum

Privacy by Design seeks to accommodate all legitimate interests and objectives in a positive-sum "winwin" manner, not through a dated, zero-sum approach, where unnecessary trade-offs are made. Privacy by Design avoids the pretence of false dichotomies, such as privacy vs. security, demonstrating that it is possible, and far more desirable, to have both.	Privacy by Design does not simply involve the making of declarations and commitments – it relates to satisfying all legitimate objectives – not only the privacy goals. Privacy by Design is doubly-enabling in nature, permitting full functionality – real, practical results and beneficial outcomes to be achieved for multiple parties. • When embedding privacy into a given technology, process, or system, it should be done in such a way that full functionality is not impaired, and to the greatest extent possible, that all requirements are optimized. • Privacy is often positioned in a zero-sum manner as having to compete with other legitimate interests, design objectives, and technical capabilities, in a given domain. Privacy by Design rejects taking such an approach – it embraces legitimate non-privacy objectives and accommodates them, in an innovative positive-sum manner. • All interests and objectives must be clearly documented, desired functions articulated, metrics agreed upon and applied, and trade-offs rejected as often being unnecessary, in favour of finding a solution that enables multi-functionality. Additional recognition is garnered for creativity and innovation in achieving all objectives and functionalities in an integrative, positive-sum manner. Entities that succeed in overcoming outmoded zero-sum choices are demonstrating first-class global privacy leadership, having achieved the Gold Standard.

Principle 5: End-to-End Security – Full Lifecycle Protection

Privacy by Design, having been embedded into the system prior to the first element of information being collected, extends securely throughout the entire lifecycle of the data involved — strong security measures are essential to privacy, from start to finish. This ensures that all data are securely retained, and then securely destroyed at the end of the process, in a timely fashion. Thus, Privacy by Design ensures cradle to grave, secure lifecycle management of information, end-to-end.	Privacy must be continuously protected across the entire domain and throughout the life-cycle of the data in question. There should be no gaps in either protection or accountability. The "Security" principle has special relevance here because, at its essence, without strong security, there can be no privacy. • Security – Entities must assume responsibility for the security of personal information (generally commensurate with the degree of sensitivity) throughout its entire lifecycle, consistent with standards that have been developed by recognized standards development bodies. • Applied security standards must assure the confidentiality, integrity and availability of personal data throughout its lifecycle including, inter alia, methods of secure destruction, appropriate encryption, and strong access control and logging methods.

Principle 6: Visibility and Transparency – Keep it Open

Privacy by Design seeks to assure all stakeholders that whatever the business practice or technology involved, it is in fact, operating according to the stated promises and objectives, subject to independent verification. Its component parts and operations remain visible and transparent, to both users and providers alike. Remember, trust but verify!	Visibility and transparency are essential to establishing accountability and trust. This PbD principle tracks well to Fair Information Practices in their entirety, but for auditing purposes, special emphasis may be placed upon the following FIPs: • Accountabilty – The collection of personal information entails a duty of care for its protection. Responsibility for all privacy-related policies and procedures shall be documented and communicated as appropriate, and assigned to a specified individual. When transferring personal information to third parties, equivalent privacy protection through contractual or other means shall be secured. • Openness – Openness and transparency are key to accountability. Information about the policies and practices relating to the management of personal information shall be made readily available to individuals. • Compliance – Complaint and redress mechanisms should be established, and information communicated about them to individuals, including how to access the next level of appeal. Necessary steps to monitor, evaluate, and verify compliance with privacy policies and procedures should be taken.

Principle 7: Respect for User Privacy – Keep it User-Centric

Above all, Privacy by Design requires architects and operators to keep the interests of the individual uppermost by offering such measures as strong privacy defaults, appropriate notice, and empowering user-friendly options. Keep it user-centric!	The best Privacy by Design results are usually those that are consciously designed around the interests and needs of individual users, who have the greatest vested interest in the management of their own personal data. Empowering data subjects to play an active role in the management of their own data may be the single most effective check against abuses and misuses of privacy and personal data. Respect for User Privacy is supported by the following FIPs: • Consent – The individual's free and specific consent is required for the collection, use or disclosure of personal information, except where otherwise permitted by law. The greater the sensitivity of the data, the clearer and more specific the quality of the consent required. Consent may be withdrawn at a later date. • Accuracy – personal information shall be as accurate, complete, and up-to-date as is necessary to fulfill the specified purposes. • Access – Individuals shall be provided access to their personal information and informed of its uses and disclosures. Individuals shall be able to challenge the accuracy and completeness of the information and have it amended as appropriate. • Compliance – Organizations must establish complaint and redress mechanisms, and communicate information about them to the public, including how to access the next level of appeal. Respect for User Privacy goes beyond these FIPs, and extends to the need for human-machine interfaces to be human-centered, user-centric and user-friendly so that informed privacy decisions may be reliably exercised. Similarly, business operations and physical architectures should also demonstrate the same degree of consideration for the individual, who should feature prominently at the centre of operations involving collections of personal data.

18.3.3 Critique

"Privacy by Design" was perceived by regulators, activist and politicians as a methodology for embedding privacy in newly designed systems. However, while the seven principles are expressing principal requirements for built-in privacy, they still are closer to the legal philosophy of privacy than they are to the software engineering process. Therefore, two different types of criticism against "privacy by Design" have evolved:

1. The seven principles for privacy by design do not relate to the software engineering process. They do not provide any functional requirements that are meaningful to a software architect or to a requirements engineer. The principles still have to get interpreted by privacy experts into a particular specification process. Technicians and software engineers are often left confused concerning what they, in practice, should do with the seven principles when they go about with designing a new information system that consumes personal data. Watch Claudia Diaz' elaborating the weaknesses and show privacy engineering with examples in this lecture video: Claudia Diaz' presentation "Engineering Privacy by Design" ($https://www.youtube.com/watch?v=AEjWsxZkjD0$)

2. The Privacy by Design principles are focused on designing new systems. However the world is full of decade-old legacy systems and databases that are already there. Only a fraction of the information systems in use will be rebuilt from scratch at any time in digital history. What the principles do not support is the process or re-engineering existing systems.

The term "Privacy by Design" has since been included in new regulation, and is well-present in the privacy debate around the world. However, it is yet far away from a functional or technical specification that will be instructive to software architects.

18.4 OASIS privacy management: Privacy by Design for Software Engineering

OASIS provides us with a template for the systematic analysis of systems that process personal data. OASIS (Organization for the Advancement of Structured Information Standards) hosts a technical committee that works on the **OASIS Privacy by Design Documentation for Software Engineers** (PbD-SE). Although the documentation is not yet finished, its draft version is available at the OASIS web page. In this document, the authors elaborate how a use case or business case shall get analyzed in a structured 12-step process supported by the "**Privacy by Design Use Case Template for Privacy Requirements**". The result of the analysis is a privacy requirements specification. Step by step, analysts will assess a use case, the involved stakeholders, the personal data and data flows, and then finally specify the necessary policies and functions.

12-step process for "Privacy by Design"

Below in the table, you find the Privacy by Design Use Case Template for Privacy Requirements including comments on the use of the template.The template is used by privacy engineers and system owners to discuss the use case, the data subjects and the processed personal data in connection to to systems used to process the personal data. The "description" column is used to document the findings.

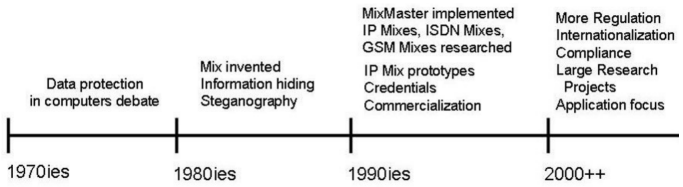
#	**Step**	**Description**
1	Use Case Title	
2	Use Case Category	
3	Use Case Description	
4	Applications associated with Use Case • Relevant applications and products requiring software development where personal data is communicated, created, processed, stored or deleted	
5	Data subjects associated with Use Case • Includes any data subjects associated with any of the applications in the use case	
6	PI and PII and the legal, regulatory and /or business policies governing PI and PII in the Use Case • The PI and PII collected, created, communicated, processed, stored or deleted within privacy domains or systems, applications or products • The policies and regulatory requirements governing privacy conformance within use case domains or systems and links to their sources	
7	Domains, Domain Owners, and Roles associated with the Use Case – Definitions: • Domains - both physical areas (such as a customer location or data center location) and logical areas (such as a wide-area network or cloud computing environment) that are subject to the control of a particular domain owner • Domains - both physical areas (such as a customer location or data center location) and logical areas (such as a wide-area network or cloud computing environment) that are subject to the control of a particular domain owner • Roles - the roles and responsibilities assigned to specific participants and systems within a specific privacy domain	
8	Data Flows and Touch Points Linking Domains or Systems • Touch points - the points of intersection of data flows with privacy domains or systems within privacy domains • Data flows – data exchanges carrying PI and privacy policies among domains in the use case	
9	Systems supporting the Use Case applications • System - a collection of components organized to accomplish a specific function or set of functions having a relationship to operational privacy management	
10	Privacy controls required for developer implementation • Control - a process designed to provide reasonable assurance regarding the achievement of stated objectives [Note: to be developed against specific domain, system, or applications as required by internal governance policies, business requirements and regulations	
11	Services • Service - a collection of related functions and mechanisms that operate for a specified purpose	
12	Underlying functionality	

19 Why there is a need for privacy management

19.1 Will PETS solve the problem? An excursion into privacy enhancing technology

19.1.1 History of PETs

PET as a research topic has been opened by David Chaum in 1981. In his 1981 MIX paper, he describes a method for anonymous and unobservable delivery of electronic messages called "Mix". Chaum uses security protocols and subsequent layers of encryption to provide privacy protection by "mixing" several people's email traffic in encrypted form. The concept later was implemented in the MixMaster email anonymization system which is the first practically available PET system.



Historic development of privacy enhancing technology, (Fritsch, 2007).

The appearance of technological measures for privacy protection coincides with strengthening legal regulation of the use of personal data on information systems. Starting in the 1970s, regulatory regimes were put on computers and networks. Starting with government data processing, along the lines of computerization of communication and workflows, explicit rules like the European Data Protection Directive of 2002, and later the GDPR, have been put in place. Several companies turned privacy protection into a business model for anonymous internet access, for one-time e-mail addresses, Spyware detection and other purposes (Anonymizer.com, Zeroknowledgesystems.com, dossier services, XeroBank, Anti-Spyware, Virus tools).

With the globalization of the economy and the IT infrastructure supporting it, in the years staring the 3rd millennium, privacy management turned into a matter of corporate governance and compliance, with legislation targeting this issue. Standardization bodies and interest groups such as ISO, W3C and IETF initiated privacy technology standardization work. Global players such as IBM and HP targeted corporations with their privacy compliance services. In this context, recent efforts on using Trusted Computing to implement privacy-compliant data handling, show the path to the future of information privacy as a matter of compliance.

19.1.2 Typology of PETS

PETS are divided into two categories: transparency tools and opacity tools. Transparency tools are intended to create insight into data processing. Their effect is a better understanding of procedures, practices and consequences of personal data processing at a data processor. Because they enhance understanding and visibility, they are called transparency tools. Opacity tools are intended to hide a user's identity or his connection to personal data that occurs at a data processor. As they hide identities, reduce visibility, or camouflage connections, they are called opacity tools.

	Transparency tools	Opacity tools
Definiton	*Tools that show clearly to a person what personal data is being processed, how it is processed, and by whom it is processed.*	*Tools that hide a person's identity or his relationship to data as it is processed by someone else.*
Non-technical example	• Legal rights to be informed about data processing; • Privacy audits.	• Pseudonymous access to online services; • Election secrecy.
Technical example	• Database audit interfaces; • Audit Agents, • Log files.	• MixMaster anonymous e-mail; • TOR anonymizing web surfing; • Pseudonyms, such as PrivacyABCs.

Examples for transparency and opacity tools (Fritsch, 2007).

Opacity tools

Opacity tools are technical mechanisms and tools that hide personal data, hide connections between personal data, and between personal data and identity information. Techniques used in opacity tools are:

- obfuscation;
- encryption;
- data separation;
- anonymization and pseudonymization;
- access restrictions;
- use restrictions.

Historically, the term "privacy enhancing technologies" (PETs) refered to tools that hide, restrict or control personal data in the interest of data subjects. Therefore, many PETs are opacity tools. Many opacity tools are summarized in (Fritsch, 2007).

Transparency tools

Hedbom defines transparency tools in the following way:

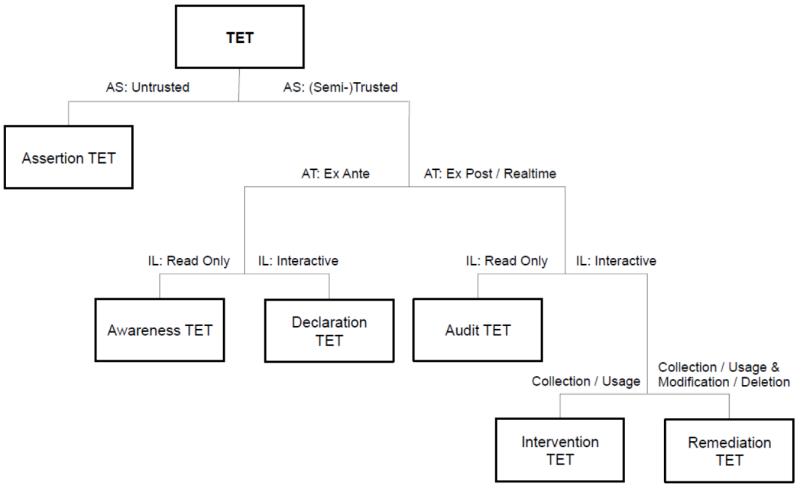
A transparency tool for privacy purposes is a technological tool that has one or more of the following characteristics:

- gives information on intended collection, storage and/or data processing to the data subject, or a proxy acting on the behalf of the data subject, in order to enhance the data subject's privacy;.
- provides the data subject, or a proxy acting on the behalf of the data subject, with access to stored data and/or to logic of data processing in order to enhance the data subject's privacy;.
- provides counter profiling capabilities for a data subject, or a proxy acting on behalf of the data subject, in order to 'guess' how her data match relevant group profiles that may affect her risks and opportunities, implying that the observable and machine readable behavior of her environment provides enough information to anticipate the implications of her behavior.

Transparency tools (or transparency enhancing tools, TETs) are generally seen as a combination of technological solutions and legal or procedural frameworks. Transparency tools can, however, create new privacy problems when they log data processing activity or communication with other data subjects, as Pulls points out:

In general, transparency can be in conflict with the privacy principle of data minimization. For example, ensuring the privacy (in particular, the confidentiality) of a private conversation is natural, while making the conversation transparent to a third party is a violation of the expectation of privacy of the conversing parties.

An important distinction for transparency tools is the division into **ex-ante** (inform before processing) and **ex-post** (inspect after processing) TETs. Zimmermann provided a classification of TETs in TetCat that is based on the degree of interactivity and intervention provided to data subjects. TetCat is shown in the figure below:



TetCat classification of transparency enhancing tools (Zimmermann, 2015).

Zimmermann divides TETs in untrusted TETs - based on **assertions** or declarations by data processors - and trusted TETS that actually inform about data processing. The trusted TETs are divided in before-processing (**ex ante**) and after-processing (**ex-post**) TETs.

Ex-ante TETs are further divided into **awareness** TETs and **declaration** TETs, where the first category is informative, while latter category involves data subject interaction in inspecting and accepting data processing.
Ex-post TETs are divided into **audit** TETs for inspection by data subjects, **intervention** TETs that enable limitation of data collection and use, and **remediation** TETS that in addition accommodate the manipulation and removal of personal data.

19.2 Limitations of PETs in practice

This section of the course will look at PETs' limitations. PETs have technological limitations, limitations in their ergonomy and user friendliness as well as in their maturity as part of a decision-making process in IT investment.

19.2.1 Technological performance

Privacy-enhancing technology deploy massive amounts of cryptography to hide information. MIX-based anonymizers or anonymous credential technology such as IDEMIX comes with computational cost. The use of asymmetric cryptography and the layered application of cryptography both impose considerable computational load both on servers and on end user devices. Both the relative slow web surfing experience of using the TOR anonymizer and the computational turnover times times of "anonymous credential" systems (IDEMIX, U-Prove, Privacy ABCs) have been widely discussed. Policy-based work-flow enforcement such as Sticky Policies come with an enormous increase in system complexity, both for software engineers and for the system users.

Usability Privacy policies and their underlying security models restrict data subjects and data processors in how they can use their systems. Privacy handling procedures and technology normally degrades the user experience of data processing in two ways. First, the system imposes procedural restrictions on information handling (e.g. approval of data processing, restriction of data collection). Second, the application of controls such as PETs and TETs change system performance, and may require users to understand and master complex rules such as access control policies (f.ex. social media person-based discriminatory access restrictions). For each application, there is a limit on how much degradation users will accept until the individual value of using the application disappears (See Fritsch and Fuglerud, 2010). The research field "Usable Privacy" works in this area. Its goal is the provision of usable user interface to privacy technology.

19.2.2 Privacy management beyond PETs

Only few studies exist on the cost of privacy management on the business side. In 2004, the Ponemon Institute conducted a study for IBM (Ponemon, 2004). It provides a cost factor model (see table 1) and provides some insight into corporate spending patterns for privacy management in large corporations. The authors define a "total privacy cost framework". The approach is to compare the cost of non-compliance to privacy requirements to the cost of investing in privacy management with respect to its effect.

The assumption is that the optimum in privacy spending is where the expenditure equals the non-compliance cost. This results in the calculation of privacy protection cost not with the goal of maximum privacy, but cheapest compliance.

Business side cost factors
***Privacy Office*:** Costs associated with dedicated staff, office overhead, travel and business equipment.
***Policy & Procedures*:** Costs associated with the creation, review, publication and dissemination of the privacy policy (and privacy notice when applicable).
***Downstream Communications*:** Costs associated with the communication and outreach activities for the privacy program both within the company and to outside stakeholders.
***Training & Awareness*:** Costs associated with the education of employees and other key company stakeholders about the privacy policy, program and related concepts.
***Enabling Technologies*:** Costs associated with technologies that help mitigate privacy risk, enhance responsible information management, or protect the critical data infrastructure.
***Employee Privacy*:** Costs associated with the protection of sensitive employee records, including heath care and OSHA claims.
***Legal Activities*:** Costs associated with legal review and counsel concerning the privacy program as well as legal defence costs in the event of a privacy violation.
***Audit & Control*:** Costs associated with the monitoring, verification and independent audit of the privacy program, including use of controlled self-assessment tools.
***Redress & Enforcement*:** Costs incurred to provide upstream communication of a privacy or data protection breach to appropriate parties within the organization, including the cost of investigation and collaboration with law enforcement. In addition to the above cost center activities, the current research captured additional information

Table 1: Cost of privacy from (Ponemon, 2004).

From (Ponemon, 2004), some significant insight can be gained. The survey lists the privacy costs ranked by direct cost. IT systems (e.g. PET or IDM), are on the third position of the most expensive cost factors, amounting about one-third of the cost of privacy office, and less than 50% of the cost for training. Beyond PET, there eight other cost factors exist that are policy-intense or involve specialized employees, e.g. lawyers. Privacy technology by itself is not a main cost driver – policies, enforcement, legal counsel and many other factors outnumber the cost of PET used. When deciding on the deployment of privacy-enhancing technologies into a business infrastructure, return-on-investment (ROI) considerations will play an important role in any investment decision – both on the business and the user sides.

19.2.3 Return on investment

When choosing components, software architects are interested in their properties beyond the technical properties. Components come with a cost, so a "return on investment" perspective on privacy controls is important. Every solution comes with its individual strengths and weaknesses. Some may impose a heavier computational load, others may reduce usability of the overall system. Other solutions will require more system administration time. Therefore, it is important to parametrize privacy controls according to the expected return on investment.

As a complement to the **"Return on security investment"** (ROSI) instrument, a **"Return on privacy investment"** (ROPI) instrument can be used to assess the return of privacy investments (Fritsch and Abie, 2008).

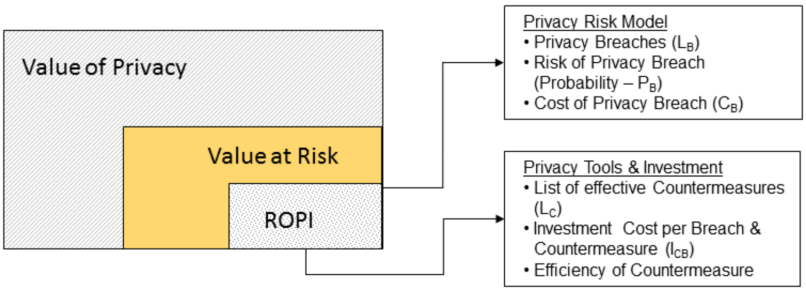


Figure 1: Return on privacy investment instrument.

It enhances the ROSI-like approach Likelihood*Damage – Cost. Figure 1 shows that from the total value of privacy protection, ROPI reduces financial risks through investments that avoid the risks. ROPI states the effect that a particular investment has on the privacy-relevant value of an information system. With the parameters in figure 1, ROPI is defined as:

$$Value\_after\_investment = Value\_of\_Privacy - (Value\_at\_Risk - ROPI)$$

where for any privacy breach $L_B : ROPI = P_B * C_B - I_{CB}$

ROPI is based on concepts and empirical data that have yet to be scientifically explored and defined. ROPI is the reduction of a privacy risk through an investment. It takes into account the likelihood of a privacy problem expressed in P_B, the cost of a privacy problem expressed in C_B, and the investment cost for a countermeasure, expressed as I_{CB} (where C is the countermeasure, B the breach). As you can see the ROPI method requires defined lists of breaches, cost of breach, and cost of investment as an input.

From a risk management perspective, ROPI will reduce privacy risks that have been assessed as "value at risk". An investment in PET – or insurance – then is sought after by analyzing countermeasures, their cost, and their effectiveness. Unfortunately, the components of the ROPI model are almost exclusively the white spots on our privacy risk management road-map. The section below will elaborate this.

19.2.4 Poor data on privacy risks and PETs make PET usage difficult

On the practical side, a methodology for privacy-risk reduction in information systems is needed. It should select the right amount of privacy protection – for a tolerable investment - to reduce the risks. For quick and efficient construction of privacy-respecting infrastructures, tools for process modeling, lifecycle management of personal data are necessary. Within the personal information treatment process, some form of "black box" abstraction for the PET basic functions is needed. This abstraction reduces a PET component to its functionality, cost of acquisition and cost of operation. Of particular interest are:

- The value of privacy in IT systems;
- The cost or damage that occurs upon privacy breaches;
- The cost-benefit distribution between companies and users (called the "dual nature" or privacy risks below);
- An abstraction of PET components into building blocks with functions, effectiveness measure and cost;
- A model of privacy risks and their magnitude of impact;
- A model of cost versus risk versus investment;

From the above sections on the cost factors of privacy as well as the ROPI model, we see the need for good information about the technological, economic and risk-reducing properties of privacy controls. However,very little concrete data is available. Data was so insufficient that many countries, including the EU members through their 2018 GDPR implementation, regulated privacy incident reporting with laws. We can at least expect figures about breaches and possible causes of breaches to accumulate in the near future. A recent report from the Ponemon Institue (Ponemon, 2017) shows figures about data breach handling cost in a global scale. Where it comes to other factors, such as the cost of ownership of PETs, the effectivness and efficiency of particular solutions, or the product quality of PETs, privacy managers are mostly left to their own assessments. Many companies maintain their own internal figures with breach data, countermeasures, efforts spent, maintenance efforts and accumulate thereby experiential knowledge about the quality and the other properties of PETs and other privacy solutions.

For any deployment of PET into information systems, the effectiveness of the PET measure against threats is important. Since risk and associated cost are not easy to quantify, the verification of effectiveness of a PET system relative to its cost is one more unknown parameter. It is a base to economic and technical decision-making that is – so far - hard to express in numbers. While PET cost of installation and operation, although non-existent, could be assessed with experiments, the efficiency of their deployment remains unknown. Summarizing this section, we note that PETs are:

- often poorly aligned with business models or regulation;
- costly in integration and operation;
- poorly usable due to complexity and resource use;
- unclear about their actual performance;
- complicating development;
- of varying product quality, often simply research prototypes.

Summarizing, we see that there is a need for empirical data on PETs and their properties. This was summarized in (Fritsch and Abie, 2008) as a road-map to the management of privacy risks in figure 2.

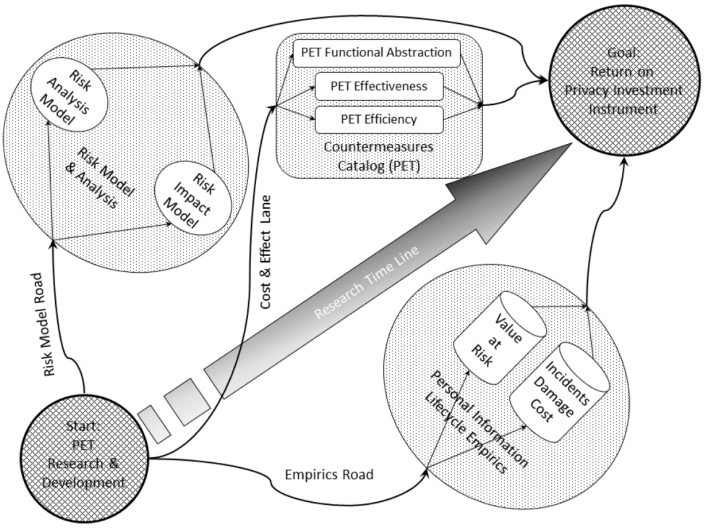


Figure 2: Road-map to privacy risk management from (Fritsch and Abie, 2008).

The road-map offers three "roads" that are build along the timeline: the **risk model road**, the **cost & effect lane**, and the **empirics road**. Each of the three roads provides a path to collect and increase knowledge about privacy risk mitigation.

- The **risk model road** is used to improve privacy risk analysis and privacy impact analysis models.
- The **cost & effect lane** is a track that collects data on PET effectiveness and efficiency, and in addition provides functional abstractions of PETs that relate to privacy risks (e.g. privacy patterns).
- The **empirics road** is traversed to collect empirically sound base data on the value of personal data, the cost and frequency of data breaches and other information that will help quantify actual risk and damage as part of a privacy management model.

However, as of today, mandatory privacy breach reporting to government authorities is the only solid source of empirical data. Any organization interested in long-term, empirically funded privacy management will need to set up their own task force to collect data on all three paths of the road-map - e.g. with researchers at Karlstad University!

19.3 Privacy management: data, processes and administration

19.3.1 Working with personal data

The processing of personal data should be aligned with a business purpose and a business process. Often, personal data is used as part of a business work-flow. This work-flow is aligned with the privacy policy that describes personal data processing as part of the data subject consent collection. When handling personal data, two aspects need attention: The control of personal data flows, and the management of identities of data subjects and data processors. Figure 1 below shows on the left side how data flows are composed of collection, storage, processing, transfer and deletion of personal data. On the right side, figure 1 shows how identity management is an important aspect of personal data processing through identification of authorized access, authentication of involved parties, linkability of data sets (or the absence thereof through anonymization), expression of ownership of data, and delegation of rights to other parties.

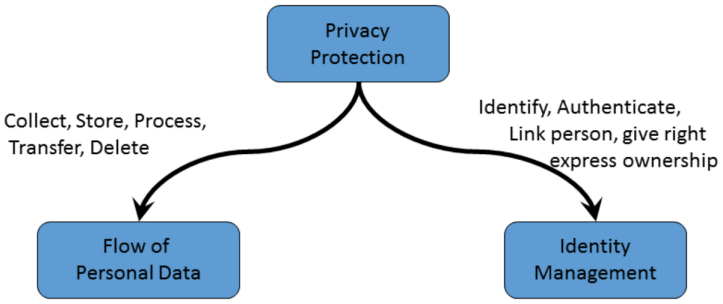


Figure 1: Aspects of privacy protection: Data flow and identity management.

The handling of personal data as "business process" therefore includes a number of mechanisms known from business process management systems. Data flow models aligned with the privacy policy and the business models are a precondition for successful privacy compliance. They, in turn, demand role and access control models for staff. Additionally, identity management of the data subjects over the complete data processing and storage period is needed.

19.3.2 Formalizing data flow and privacy constraints

A formalized description of access restrictions is called a "privacy policy", an analog to "security policies". However, the term is ambiguous, since privacy policies for end users that are written in prose (a form or technique of language that exhibits a natural flow of speech and grammatical structure) are referred to by the same name. For this section, we presume that there are formalized, machine-readable privacy policies that have been specified in harmony with the prose versions written by legal experts.

The binding of data flows to privacy policies and to security and privacy controls is a challenge for software engineers. Two XML-based languages for the description of privacy constraints have been specified:
- EPAL: the Enterprise Privacy Authorization Language
- XACML: the eXtensible Access Control Markup Language
In (Anderson, 2005), you can read how these two languages compare.

In practice, information systems need to consider the rules laid forth in privacy work-flow specifications through security functionality. Hewlett-Packard's research group on privacy specified a hardware-supported system called Sticky Policies for this purpose. It consists of an architecture of trusted hardware that executes data processing only according to policies that are attached to the encrypted personal data. The concept included the management of user consent statements as part of the data flow and access control model. The Sticky Policy concept is a very complex cryptographic system that needs to span over the overall infrastructure of data processing, including all subcontractors. An introduction to the overall architecture with an example application is provided in (Pearson and Casassa Mont, 2011).

19.3.3 Managing personal data

For the sake of completeness and with reference to the data subject intervention rights defined in the EU GDPR, I mention a number of issues of importance when working with a data model for storing and handling personal data. To satisfy data subject rights, data processors will need to be able to provide detailed information about what data they store, how they processed it, for what purpose (under what privacy policy/data subject consent version), and which other processors they shared it with.

The ability to answer such inquiries requires a stringent householding approach to personal data. It may include:

- stringent data labeling with metadata (provenance, data subject identity assurance, references to privacy policies and consent statements);
- stringent control over modifications in business models, business practice, change of organizational structure or change of hardware/software that may cause update of privacy policies and renewal of data subject consent;
- privacy-friendly logging of processing transactions and their purpose;
- mechanisms for the archival and reference of past versions of privacy policies personal data was collected under, including links to data subject consent versions for the respective occasion;
- identity management for data subject identification and long-term recognition of legitimate data subjects.

The correct and law-abiding management of personal data collections will therefore impose serious housekeeping efforts on the responsible organizations.

20 Summary

20.1 Privacy in the life cycle of IT management

It is time to summarize the course! You have made it to the last portion of course content! This section will present you a connected summary of the course content compiled as a reference framework. It will show you when you use the learned approaches in which phase of system design and maintenance. This work is based on the article "Privacy in the life-cycle of IT Services" (Rother, Schiering 2013). In their article, authors create an overlay over contemporary information technology (IT) development and IT management that shows how privacy design and management activities are related to those activities. The IT life cycle models COBIT, TOGAF and CMMI are applicable to the V model of software development, while ITIL is mostly concerned with operations of IT systems. Figure 1 below illustrates how the IT management models overlay with the V model.

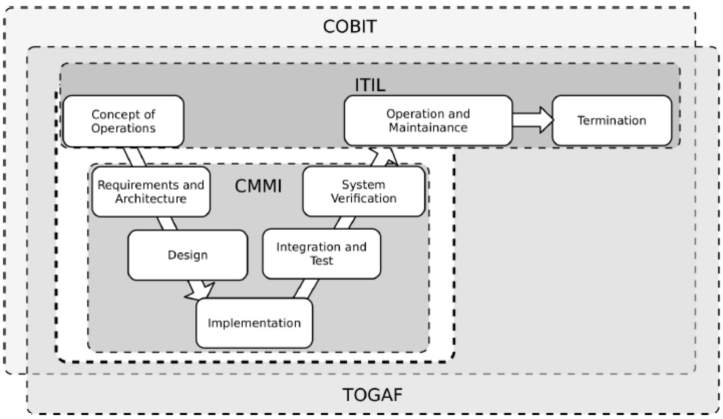


Figure 1: IT management frameworks and their relationship to the V-model of software development (Rother, Schiering 2013).

Unlike the previous parts of this course which focused on plan-do-check-act cycles, the V model spans from early design to implementation of a system. The PDCA activities from earlier course sections will mainly be carried our in the "operations and maintenance" phase of the IT life cycle (although feeding back information to re-engineering activities). Roter and Schiering place LINDDUN into the requirements engineering and architecture specification phase of the V model. Some may argue that LINDDUN reaches forth into the design phase, however as of now LINDDUN is very little helpful in determining actual design decisions. As shown in Figure 2, privacy patterns are mainly relevant in the design phase, while data protection and privacy impact assessments are carried out during conceptualization, requirements engineering, design and implementation as well as during the testing and certification phase.

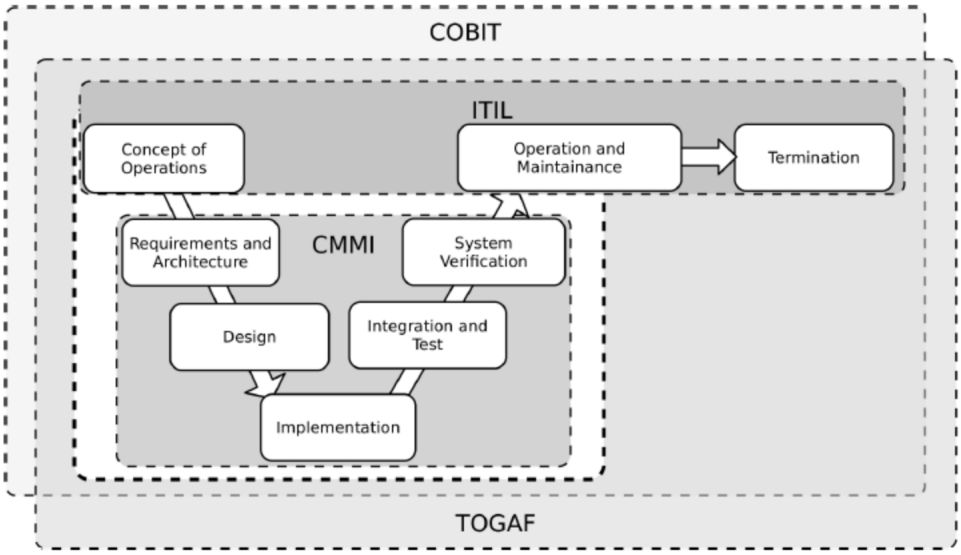


Figure 2: Privacy design and management activities as part of the IT life cycle. From (Rother, Schiering 2013).

Risk-driven operational privacy management is then applied to the operations and maintenance phase as well as to the termination of the IT system. For a closer look at the privacy design and management support in ITIL, COBIT, TOGAF and CMMI, please refer to the article (Rother, Schiering 2013).

20.2 Management summary

Privacy management will not be possible without proper support and budget from management. Managers will not use their precious time on this course. Instead of providing a 1-page-summary of this course to convince management, I provide you with two links that should get the privacy management project going.

On LinkedIn, Constantine Karbaliotis, Director, Leader Managed Privacy Services, at PwC Canada, has posted what he called the "The Nightmare Letter: A Subject Access Request under GDPR". Later, he added the equivalent letter for inquiries from the data protection authority, and a data subject rectification request letter to his blog.

I conclude the course on privacy management with my recommendation to read and save the "nightmare letter". It should have its effect when used in communication with the CEO and the legal department.

Part VI

Privacy engineering & privacy patterns

21 Software Architecture and Design Primer

Goals of this module

- To give a basic introduction to software architecture and design

- To introduce some basic terminology, such as quality attributes and architectural tactics

Let's start with a cliffhanger...

A piece of architecture

- Burj Dubai, Dubai, UAE. World's tallest skyscraper at 828m
- Embodies elements from the style of Islamic architecture, e.g. spiral minarets.
- Y-shaped footprint: Y-shaped plan optimized for hotel and residential usage
- Newly developed bearing structure: hexagon-shaped core + buttresses avoid twisting
- Coated anti-glare glass to shield from extreme heat
- Complex air-conditioning system, using cooler and cleaner air from higher levels

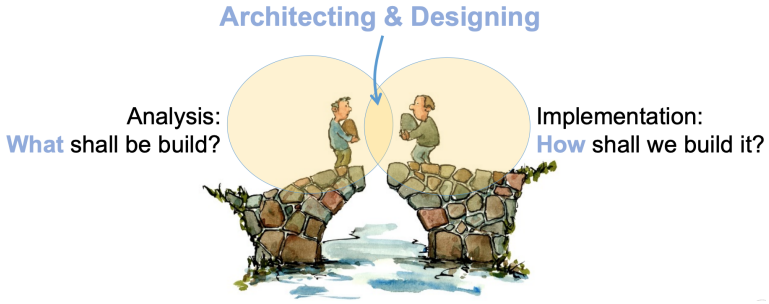
21.1 So, what do architects do?

- Architects,
  - Design, plan, and develop residential and commercial structures
  - Based on their customer's wishes, objectives and budget,
  - Considering the building's style, safety, sustainability and other concerns
- Or more abstractly speaking
  - They take as input their customers' requirements
  - Make principal design decisions regarding the structures of the building
  - Considering the desired quality properties

21.1.1 Do we have something similar for software(-intensive) systems?

- Software is a fundamental piece in all sectors of society
- Software(-intensive) systems become increasingly complex
  - Flight software in a Boeing 787: 14 MLOC
  - Windows 7: 40 MLOC
  - Facebook: 60 MLOC
  - Avg. high-end car: 100 MLOC
- Do you think this works without spending some thoughts on designing and planning?
- Software architects
  - as **input requirements and general constraints**,
  - make **principle design decisions** from a set of alternatives that determine **macro-structures** aiming to adhere to these requirements
  - and **strives to embed, evaluate and evolve** them in the developed system as it is implemented and evolved
- Software designers
  - Refine those designs decisions and the corresponding structures

21.2 Mind the gap



21.3 Attempts to define software architecture

- Many definitions state similar to IEEE standard 1471: "the **fundamental organization** of a system embodied in its **components**, their **relationships** to each other and to the environment and the **principles guiding its design and evolution**."
- You can find over 100 definitions at www.sei.cmu.edu/architecture/start/community.cfm
- Many definitions state: "The software architecture of a system is the **set of principal design decisions**."
- "The decisions that are the **most difficult to change**."
- **Shift** from the **result** of architecting (components + relationships) to the rationale – the **decisions**.

21.4 Things influencing this decision making

- Functional requirements
  - Inventory functionality, reservation functionality
  - Multi-user support and support of different user types (student, staff)
  - Accessible via the internet
- Quality attribute requirements
  - Security: How to deal with which kinds of security threats?
  - Performance: How many simultaneously active users?
  - Usability: is the GUI convenient on all supported platforms?

Constraints: Standards to be used, legal issues, resources, organizational constraints

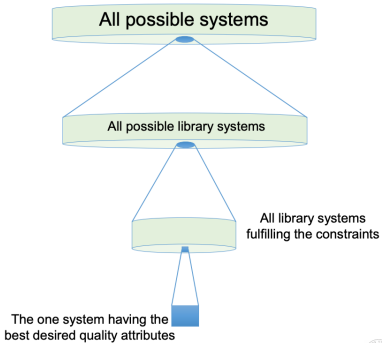
21.5 Quality attributes and quality models

- What is **quality**?
  - Latin "qualitas": the nature/distinguishing characteristic of something
  - Quality: "**the degree of excellence of something**"
- Quality **attributes** reflect the multiple dimensions of quality:
  - A software can be great w.r.t. performance...
  - ...and pretty bad w.r.t. maintainability
- Quality attributes are categorized and refined in **quality models**
- **ISO 25010** defines a quality model with eight top level attributes
  - Functional Suitability
  - Performance Efficiency
  - Compatibility
  - Usability
  - Reliability
  - Security
  - Maintainability
  - Portability

21.6 What is driving the software architecture the most?

- Functional requirements?
  - Naah, any structure will do.
- Constraints?
  - Often imply quite easy design decisions
- Quality attribute requirements?
  - Most important drivers
  - Often competing and requiring trade-offs

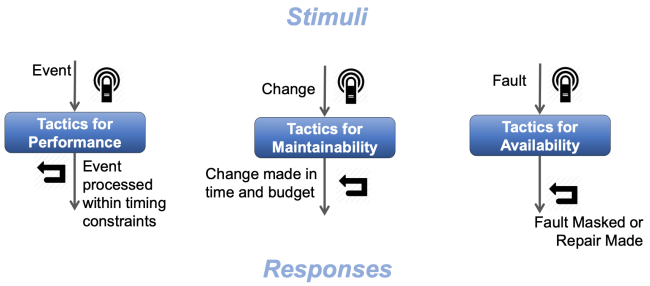
A piece of architecture



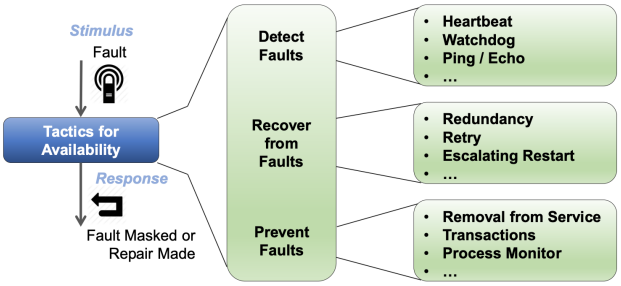
21.7 Architectural tactics

- **Creating** the (software) **architecture** of a system means to a large degree to make **decisions** towards **fulfilling** and **balancing** the desired **quality attributes**.
- We call these decisions **architectural tactics**.

- Architectural tactics are "reusable" design decisions that can be applied to influence how a (part of a) system addresses a single quality attribute.
- They can be categorized and organized hierarchically.



Example: Architectural tactics for availability



21.8 What got privacy to do with this?

- **Privacy protection** can be considered one of a system's quality attributes
- In most quality models, privacy is neglected.
- The GDPR makes **neglecting privacy** an **extensive mistake**!
- So, are there **architectural tactics (and patterns)** for **privacy privacy protection**?

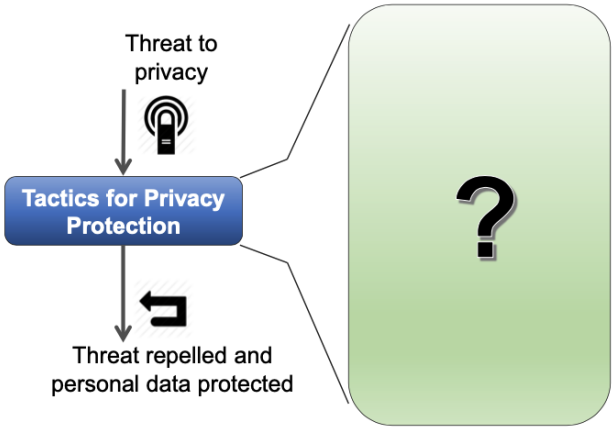
22 Privacy Design Strategies

Goals of this module

- To introduce and explain architectural tactics for protecting privacy

- Summarize the main points of the reading material for this module

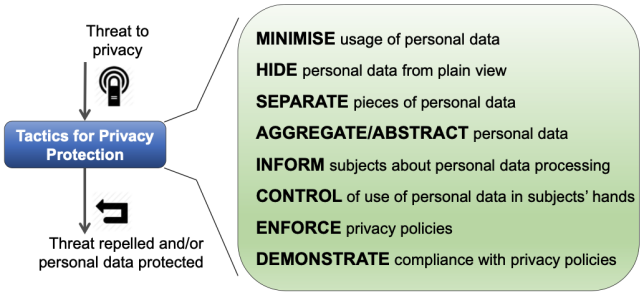
What we want to find out...



22.1 Privacy design strategies

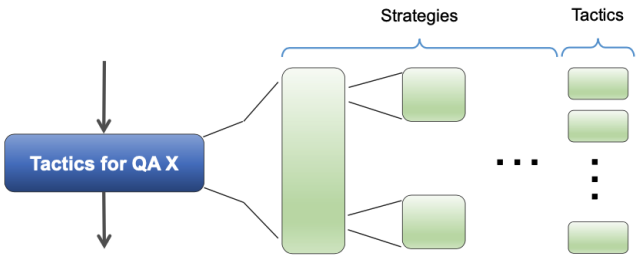
- Privacy design strategies are defined by Hoepman
  - They describe fundamental approaches to achieve privacy protection
  - For that purpose, they favor certain structures in software over others
- Architectural tactics for privacy protection are
  - design decisions that influence how a system (or a part of it) addresses privacy protection
- Privacy design strategies seem to be appropriate

22.2 Overview of privacy design strategies



22.2.1 Strategies = Tactics?(!)

- Confusing terminology from two different communities:
  - Privacy design strategies from the privacy and security research community
  - Architecture tactics for privacy protection from the software architecture community
- We use them synonomously



22.2.2 Minimize

"The amount of personal data that is processed should be restricted to the minimal amount possible."

- Is the amount of personal data collected justified by the purpose?
- Is there another way of fulfilling the same purpose with less personal data?
- Examples of implementation
  - Use of pseudonyms in a system because there is no need for persons' real names

22.2.3 Hide

"Any personal data, and their relationships, should be hidden from plain view."

- Is personal data stored/transported/etc "as it is" or is it, in some way, transformed such that it cannot easily be used by others
- Data in plain view is easier to abuse
- Who the "others" are, depends on the usage context
- Examples of implementation
  - Anonymization or encryption of data

22.2.4 Separate

"Personal data should be processed in a distributed fashion, in separate compartments whenever possible."

- Makes it harder to create full profiles of persons based on their personal data
- Prefer distributed processing over centralized processing
- Prefer local processing over remote processing
- Examples of implementation
  - Storing customer contact information and purchase information in separate databases

22.2.5 Aggregate/Abstract

"Personal data should be processed at the highest level of aggregation and with the least possible detail in which it is still useful."

- Process personal data at the level of detail that is absolutely necessary (and not in more detail)
- Aggregate data over groups of individuals, over groups of attributes, over time, ...
- Examples of implementation
  - Age ranges or regional categories instead of birthday and address in surveys

22.2.6 Inform

"Data subjects should be adequately informed whenever personal data is processed."

- Inform data subjects about
  - Which of their personal data is processed by which means for which purpose
  - The security mechanisms used to protect their personal data
  - Third parties with which data is shared
  - Their data access rights
- Examples of implementation
  - Provide a clear, understandable privacy policy

22.2.7 Control

"Data subjects should be provided agency over the processing of their personal data."

- Provide appropriate means to data subjects to exert their data protection rights
- Provide appropriate means to data subjects for deciding whether or not to use a system and for controlling the processing of personal data
- Examples of implementation
  - Notifications of desired access rights of apps
  - Customizable privacy settings in, e.g., social network systems
  - Means to execute subjects' right to be forgotten

22.2.8 Enforce

"A privacy policy compatible with legal requirements should be in place and should be enforced."

- Create, maintain, and update a privacy policy
- A privacy policy accounts for technical controls and organizational controls to privacy protection
- It should cover the full life-cycle of a system
- Examples of implementation
  - Access control systems

22.2.9 Demonstrate

"The data controller must be able to demonstrate compliance with the privacy policy and any legal requirement."

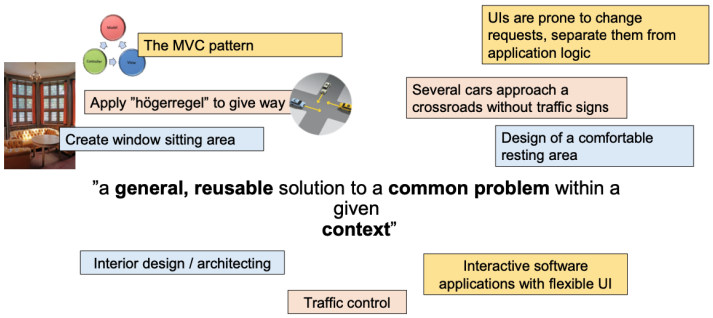
- Always be able to show how the privacy policy in place is implemented
- Explicitly required by the GDPR!
- Examples of implementation
  - Publish a recent audit certificate confirming compliance

23 Privacy Design Patterns

Goals of this module

- To explain what patterns are & To show examples of privacy design patterns

23.1 What are patterns?

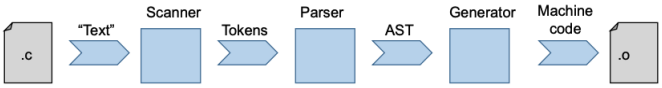


23.1.1 Description of patterns

- Name
- Context: The situation/class of system in which the pattern can be applied
- Problem: Description of what the pattern tries to solve, often express as the forces that it tries to balance
- Solution: description of the structure, i.e. configuration of elements that solves the problem and how they interact
- Often added categories
  - Summary of pattern
  - Goals: what is achieved by applying the patterns
  - Constraints and consequences: which benefits and potential disadvantages has the patterns
  - Motivating example
  - Known uses.

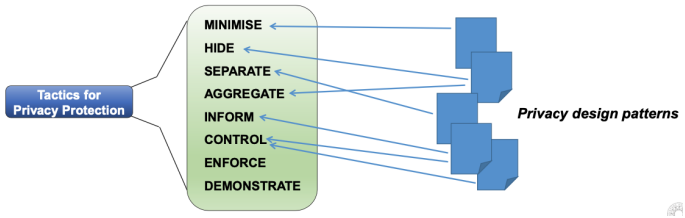
23.1.2 How do patterns relate to tactics?

- A single tactic addresses a single quality attribute
- A pattern "bundles" several tactics, potentially for different quality attributes, and describes how to apply them



23.2 What are privacy design patterns?

According to the previous definition: a general, **reusable software design solution** to a **common privacy protection problem** within a given context.

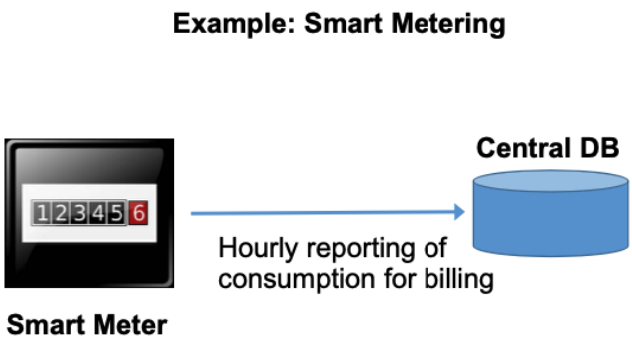


23.2.1 Example of Privacy Design Patterns

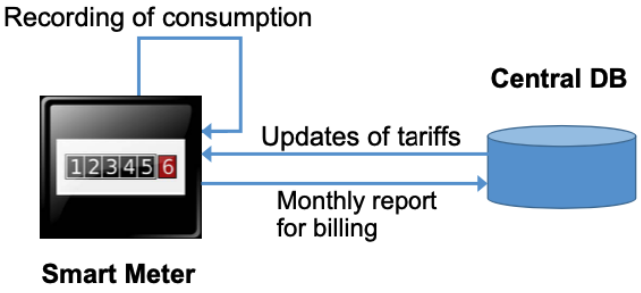
- User data confinement
- Asynchronous notice
- Location Granularity

23.3 User data confinement

- Context: Any service collecting personal data from the user of the service.
- Problem:
  - Many system architecture tend to collect and store data in central entities, e.g. databases
  - The user is forced to trust these entities and share sensitive data with them



- Solution: "turn around the trust relationship"
  - Leave storage and processing of sensitive data to the data subject / customer.
  - Instead of customers having to trust the service provider that their privacy is protected...
  - ...the service provider has to trust the customer's data storage and processing
- Applied strategies: Aggregate (Minimize), Separate



23.4 Asynchronous notice

- Context: Any service that processes personal information over a longer period of time
- Problem:
  - A user might have forgotten that he initially gave consent for his data being tracked
  - The initial consent might have been forged by an attacker
  - User may get surprised / upset about the processing of his data
- Examples
  - A webshop showing ads related to and based on previously visited content, consent provided at first visit.
  - A mobile app tracking the device's geographic position, consent provided at installation.
- Solution:
  - Proactive notification of the user about data processing after the time of consent.
  - Message should include information on kind of processing and options for accessing and controlling (see also "Control" strategy)
  - Notice may also include summary of data collected most recently.
  - Make sure notification can reach user (e.g., message to verified e-mail address)
- Applied Strategies: Inform

23.5 Location Granularity

- Context: A service collecting location data about a user.
- Problem
  - Many services require location-based data
  - Too much location information may harm the user's privacy
- Solution
  - Introduce different levels of precision of geographical data
  - Choose most coarse-grained level still useful for service
- Example
  - Weather app might still be precise enough based on ZIP-based location
- Applied strategy
  - Abstract

23.6 Pattern Catalogues

- Two publicly available catalogues of patterns
  - *https://privacypatterns.eu/*
  - *https://privacypatterns.org/*
- Both catalogues categorize patterns according to design strategies
- Beware!
  - Not an established canon of patterns as for design patterns.
  - Description quality varies a lot
  - Categorizations differ between the catalogues

24 The Dark Side – Privacy Dark Patterns

Goals of this module

- To introduce the concept of privacy dark patterns

- To present examples of privacy dark patterns

24.1 What are privacy dark patterns?

- **Recap**: privacy patterns are general, **reusable software design solutions** to **common privacy protection problems** within a given context.
- Privacy dark patterns are general, recurring software design solutions that constitute **common privacy "infringements"** within a given context.
- Not to be confused with anti-patterns

Examples

- Privacy Zuckering
- Bad Defaults
- Forced Registration

24.1.1 Privacy Zuckering?

- **Context:** Systems allowing user- specific, modifiable privacy settings
- **Description:**
  - Service provider allows users to modify privacy settings.
  - Settings are overly complex or fine- grained, or difficult to understand.
  - Users are likely to give up on modifying settings or to make unintended changes
- **Effect:**
  - Service provider can claim that users have control over their privacy settings.
  - Users are discouraged from making changes.
  - Changes might be unintentional and have undesired effects on the user's privacy
- **Example:**
  - In the past, Facebook was accused to apply Privacy Zuckering

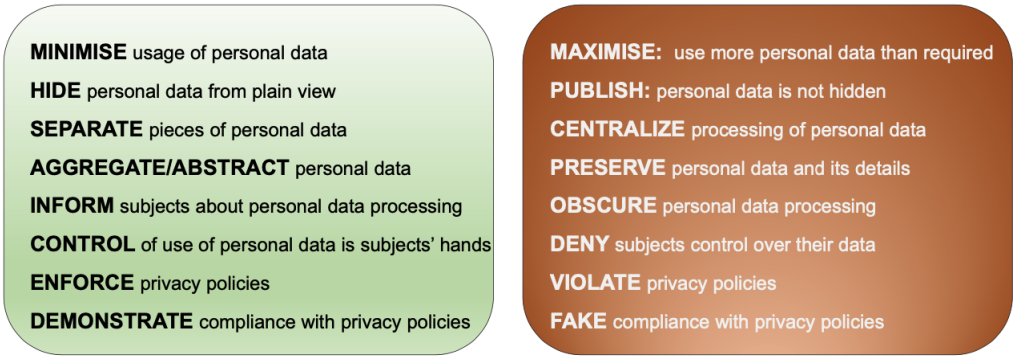
24.1.2 Bad defaults

- **Context:** Systems allowing user- specific, modifiable privacy settings
- **Description:**
  - After creating an account with a service provider, privacy settings are set to bad default values.
  - With these values, sharing of personal data is easy or encouraged.
  - Most users will not go through the options and change the settings to better values.
- **Effect:**
  - Users share more information than they might have intended to.
- **Example:**
  - Almost every social network service (at least in the past).
- **Remark:**
  - Very powerful if combined with Privacy Zuckering

24.1.3 Forced registration

- **Context:** Any service technically *not* requiring personal accounts
- **Description:**
  - User wants to use some functionality that is only accessible after registration.
  - The registration is technically unnecessary but gives the service provider access to the user's personal data.
- **Effect:**
  - Users register with service provider.
  - Allows provider to track user.
  - Sloppy configuration of privacy settings is likely.
- **Example:**
  - Numerous webshops

24.2 Dark strategies



24.2.1 Why do dark privacy patterns work?

(Because we are humans!)

- Prompting System 1 thinking
  - Automatic, fast, unconscious as opposed to effortful, slow, controlled
- Fundamental human needs
  - The need to belong – "All your friends will miss you..."
- Other psychological aspects

24.3 Summary

- Privacy dark patterns describes ways of infringing privacy.
- Similarly to privacy patterns, they implement dark strategies.
- They have a strong psychological component.