Privacy Enhancing Technologies

A software engineering approach to design PETs

Mike de Roode (m.deroode@student.fontys.nl)

April 18, 2016

Abstract

Privacy is an increasingly important subject for organizations. Nowadays, organizations (unknowingly) process vast amounts of personal data of their customers in numerous different information systems. There are strict legislations regarding the processing of personal data, and from mid-2016, these legislations will only get more strict with the introduction of the General Data Protection Regulation in the European Union.

In an approach to ensure compliance with these legislations, different techniques such as *Privacy Enhancing Technologies, Privacy-by-Design and Privacy Design Strategies* were introduced in the past decades. However, these techniques tend to be defined in such a high-level of abstraction that they are hard to use in practice.

This paper discusses and explains various software techniques which can help to design information systems that can better protect the privacy of their users. Next, these techniques are combined as a solution named *Privacy Management System*. This system is able to ensure and enforce full data processing transparency of an organization and should close the gap between the privacy legislations and software development.

Introduction

Privacy is an increasingly important subject for organizations. There are many strict laws and regulations which regulate how organizations should process data of their customers, and these laws will become even more strict in 2016[1].

In previous years, organizations profited from major improvements in both software development techniques and in computing power by increasingly automating business processes using large, advance business information systems. These information system can not only contain vast amounts of customer data, they are also able to process it in numerous ways. This rapid evolution of information systems made them valuable business assets with evermore functionality.

Ensuring that these information systems are compliant with privacy rules and regulations might not be top priority because the strict rules and regulations might limit the systems' functionality and make it more challenging to design such systems. However, aiming for compliance with the privacy laws is not a nice-to-have feature. Since January 2016, there are large consequences when an organization is not compliant with the privacy laws. The local Dutch Data Protection Authority (Dutch DPA) can fine companies up to 820.000 euros when organizations are not compliant with privacy laws[1]. The consequences will be even become larger in July 2016, when the European Parliament will enforce the new General Data Protection Regulation (GDPR)[2]. The GDPR enables DPAs to fine organizations up to 100.000.000 euros or two percent of the worldwide turn-over and is even able to shutdown organizations[2] when they are not compliant with the GDPR.

To help organizations comply with privacy laws, the term *Privacy Enhancing Technologies (PET)* was introduced. These PETs can be described as a wide range of measures to better protect the privacy of customers. Some PET subjects such as encryption, digital message exchange and physical protection are already extensively researched in the past few decades. However, there is little research conducted on designing information systems in such a way that they are able to better protect customers' privacy. Because information systems are the basis of all information processing within organizations, it is important to design these systems in such a way that they are more *privacy-friendly*. This paper will therefore focus on designing information systems which are able to better protect customers' privacy, by combining various PET measures into a *Privacy Management System*.

This paper first looks into the current privacy legislation and the new upcoming privacy legislation. Next, existing PET techniques are investigated and explained from a software engineering perspective. Finally, these techniques are combined and merged into a *Privacy Management System*-architecture. Note that because there exist large differences in privacy legislation between countries, the scope of this paper is limited to European member states.

1 Privacy

Privacy is described as *"the claim of individuals [...] to determine for themselves when, how and to what extend information about them is communicated to others"* [3]. Privacy is a basic human right which is regulated in numerous laws and regulations. The basis of privacy, for the Netherlands, is defined in the Constitution of the Netherlands (Article 9 to 13)[4]. In 1995, the European Parliament created the Data Protection Directive 95/46/EC[5] based on the European Convention on Human Rights (Article 8)[6]. Each European member state had to translate this Directive 95/46/EC into national Data Protection Acts (DPA). In the Netherlands, this resulted into the *Wet Bescherming Persoonsgegevens (wbp)* managed by *Authoriteit Persoonsgegevens* (Dutch Data Protection Authority)[7].

Because each member state had to create own laws in order to comply with Directive 95/46/EC, there are small differences in each country resulting in the situation that multinational organization had to comply with different sets of rules in each European country. To tackle this problem, the European Parliament created the General Data Protection Regulation (GDPR)[2]. When this regulation is completed and enforced, it instantly replaces all national Data Protection Acts of the European member states, and therefore eliminating previous mentioned problem. Each time when new privacy regulations were defined, they tend to be more stricter and complying with these regulation becomes a challenge.

An important aspect of privacy, especially when taking into account the massive increase in digital processing of data in the previous years, is the protection of personal data. This is regulated in Directive 95/46/EC and elaborated on in the Dutch Data Protection Act (Wet Bescherming Persoonsgegevens) and in the future General Data Protection Regulation.

Data protection is a complex issue nowadays in the era of digital data processing and IT automation, mainly because it is challenging to translate the *soft* requirements of laws and regulation into *hard* software and hardware requirements for information systems and business processes. Additionally, the privacy laws is not one set of rules, they are defined in many different places and contain room for interpretation.

Historically, data protection was mostly covered by the Information Security field. However, privacy protection requires a much broader and integral approach then only securing data. Besides digitally securing data, it is also important to take into account the physical protection of information (e.g. lost notebooks, secured data centres, leaving classified documents on a desk), whether there is a legal basis for processing data and measures to prevent the leakage of data by human errors.

The protection of personal data is a current topic in the Netherlands. Since January 2016, the Dutch Data Protection Authority (Authoriteit Persoonsgegevens) is able to fine companies up to 820.000 euros or two percent of the total yearly revenue when no sufficient actions are taken in order to prevent the leakage of personal data[7]. Additionally, all leakage of personal data have to be reported to the Dutch DPA. These measures are taken to motivate companies and organization to rethink the way personal data is processed, and to take preventive measures to ensure the protection of this data.

The following section defines important legal definitions of privacy. Next, the current legislation and the new General Data Protection Regulation are discussed.

1.1 Definitions

In order to continue the discussion about privacy and personal data, it is important to know how certain methodologies are defined in a legal perspective. The following descriptions help to understand how these terms are used in a legal context and are composed from definitions of the *Handbook of Privacy and Privacy-Enhancing Technologies*[8] and Directive 95/46/EC[5].

Data subject is defined as: a natural person, legal person, agency or organization who is the subject of the data.

Personal data is defined as: any information relating to an identified or identifiable natural person. An **identifiable person** is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more attributes that are specific to his or her physical, physiological, mental, economic, cultural or social identity.

A person is **directly identifiable** when, based on the available set of data, a natural person can be correlated to this data. A personal is **indirectly identifiable** when a certain amount of steps need to be undertaken in order to correlate a natural person to the set of information. Such steps might be the composing of new information based on the available set of information and the combination of other data sources with available set of information. Storing the birthdate and last name of a person might be sufficient to correlate a natural person with this data (when combining data with external sources such as Facebook).

Data **anonymization** can be described as: an irreversible process of altering data in such a way that the resulting data cannot be directly or indirectly related to an individual.

Processing of personal data is defined as: any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collecting, storing, organizing, recording, adapting, altering, retrieving, consulting, using, disclosing by transmission, disseminating, or otherwise making available, aligning or combining, blocking, deleting or destructing personal data.

Note that storing personal data is also defined as processing. The definition does not make differences between storing personal data in a database or storing personal data in a paper document.

A **business process** is defined as a structured collection of related actions which produce a certain output from a certain input. For this paper, it is assumed that a business process either requires personal data as input or uses personal data somewhere during the lifespan of the business process.

1.2 Current Privacy Legislation

As mentioned in the previous sections, currently, all EU member states have own privacy regulations based on the EU Directive 95/46/EC. The privacy regulations can therefore differ per country. It is very challenging to describe the

complete rules and regulations for all countries[1]. Therefore, it has been decided to only describe the Dutch privacy legislation. The next section describes the General Data Protection Regulation which will replace all national privacy legislations of EU member states.

EU Directive 95/46/EC applies to all (semi-)automated or manual data processing by an organization (be it a commercial, governmental or non-profit organization) outside the scope of public security, defence, state security. The Directive describes when data processing is lawful, namely when one of the following cases is true:

- The data subject has granted explicit permission to process its data.
- Data processing is required in order to execute a service for the data subject. The data subject should have explicitly agreed to use this service.
- Data processing is required for performing a task in the public interest or on request by an official authority.

The Directive also describes that data processing should be fair and lawful; meaning that there should be a legal basis to process the data, that the data should have a certain quality and it should be adequately used and destroyed. Additionally, there is described that the data subject has certain rights, namely the:

- *Right to obtain information.* The organization responsible for data processing should provide the data subject information about the processing such as: the origin of the collected data, who is processing the data, purpose of processing etc.).
- *Right of data access.* The data subject may request an organization to provide all information related to himself.
- *Right to object to data processing.* The data subject can deny permission to process its data when this is not conflicting with other legislations.
- *Right to make changes.* The data subject may request to change or update its data.

The Directive also describes that the organizations should take 'sufficient' measures to prevent unlawful access to data. Additionally, the data subject should be adequately informed when its data is processed and the data should only be used for the purpose that it has been originally intended for.

On top on the rules of the Directive 95/46/EC, the WBP also defines additional measures for data leakages (the event that a third party can unlawfully access data). The WBP states that such data leakages should be reported to the Dutch DPA. Additionally, the Dutch DPA might fine companies if no sufficient security measures were taken, or if the data processing is not conform the WBP.

1.3 General Data Protection Regulation

From July 2016, the General Data Protection Regulation (GDPR) from the European Parliament will replace the Directive 95/46/EC and the national Data Protection Acts. The GDPR is relevant to all companies, organizations and even

[1]This was one of the main reasons to introduce the new European privacy legislation

governments who process data about EU citizens. The GDPR does not have to be translated into national laws, it is effective immediately in all member states of the European Union. Companies and organizations have two years from July 2016 to change their privacy policies, information systems and business processes such that they are compliant with the GDPR.

Currently, the exact content of the GDPR is not yet final. There are not much resources which describe the current status of the GDPR. It is also still possible that small changes are made to the current draft. Therefore, there are still some uncertainties. In July 2016, the final version of the GDPR will be released and enforced.

Some major changes of the GDPR in comparison with current legislation in the Netherlands are described in the overview below[2].

- The GDPR is an EU law. Therefore all EU member states have the same legislation.
- Stakeholders (customers, employees) of an organization have the *Right to be Forgotten*, organizations should acknowledge this right and implement corresponding policies and processes to handle requests to be forgotten.
- Organizations with 250 or more employees must appoint a Data Protection Officer (DPO).
- Data leakages have to be reported within 24 hours to the national Data Protection Authority.
- Insufficient data protection measures can result in fines up to two percentage of the worldwide turnover or 100,000,000 euros. The Data Protection Authority can also shut down the company until the company is compliant with the GDPR.

As shown in the overview above, the new General Data Protection Regulation enforces stricter policies and rules than the current legislation. It also grants Data Protection Authorities the power to give much higher fines.

As previously mentioned, companies have until July 2018 to adapt their system so that they are compliant with the GDPR. Two years is a very short period to adapt all business processes, information systems and the organization so it is compliant with the new data protection regulation. Current systems should be assessed on whether they are already compliant with the new regulations or not. If not, a proposal should be written on how to change the systems, this includes both changes in the technical architecture and changes in the organization. Next, project should be started to implement those changes and finally these changes should be extensively tested, assessed on compliance and deployed into the organization. This process could easily span multiple years and it will be challenge to meet the deadline of two years. Therefore, it is important to start as soon as possible.

Next sections describe an approach on how to design an information system in such a way that it can processes information in a more *privacy friendly* way then in traditional system design.

[2]A full impact analysis of the GDPR is in progress. This is different per country because until the GDPR, each European member state has its own Data Protection Act. The Dutch impact analysis is currently under development.

2 Privacy Enhancing Technologies

Privacy is a complicated subject, as extensively mentioned in the previous section. Translating all laws, rules and regulations regarding privacy and the protection of personal data into technical specifications is a challenge, especially because some rules are defined rather vague and leave room for interpretation. Researchers made numerous approaches to tackle this problem. This section briefly describes a selection of these approaches. Additionally, next section describes a design which incorporates the previous approaches.

Note that this paper mainly focuses on the software perspective of PETs. There are also other imported aspects such as organizational and physical measures to ensure data protection. However, the latter two aspects are already extensively described in previous research and industry standards such as the ISO/IEC 27002:2013 and the Dutch *Baseline Informatiebeveiliging Rijksdienst* (BIR).

2.1 History

An early approach to close the gap between privacy laws and information technologies, named Privacy Enhancing Technologies (PET), was introduced by TNO, College Bescherming Persoonsgegevens (former Dutch DPA) and the Canadian DPA in 1995 [8]. Privacy Enhancing Technologies (PET) can be defined as a term which encapsulates all software, hardware and organizational measures which allow organizations to better protect the privacy of the customers[3]. PET has been extensively researched in the early 2000s. Researchers have described different approaches to integrate privacy enhancing technologies in software applications and organizations without changing their functionality[8]. The European Union and the Dutch government have financed many initiatives to promote PETs and PET research. PET is also mentioned in the Dutch Personal Data Protection Guidelines as an *"essential tool to guarantee the safety of personal data in information systems"* [9].

Because Privacy Enhancing Technologies is defined rather vague, it can consist of a broad range of measures, from introducing biometric scanners for granting employees access to a facility, to anonymizing personal data in Hadoop clusters. Previous research tends to only describe examples of how to apply these privacy enhancing technologies in specific cases, there is no readymade tool or defined approach on using and integrating PETs into information systems. Therefore, an approach to integrate privacy laws into information systems named *Privacy by Design* (PbD) was introduced [10]. This approach consisted of seven foundational design principles. This design approach received much attention over the years and was adopted by many Data Protection Authorities. However, Privacy by Design also received much criticism; the seven foundational design principles are defined vague, incomplete and not compatible with system engineering methodologies [10], [11].

In a need for concrete design guidelines for integrating privacy laws into information systems, Jaap-Henk Hoefman defined eight Privacy Design Strategies (PDS)[12]. These strategies are defined from an information technology point

[3]Note that customers can also help protecting their own privacy. Examples of this can be found in section 4: Privacy Management System.

of view instead of a legal perspective. The following overview summarizes the eight privacy design strategies.

- **Minimize** the processing of personal data.
- **Hide** any personal data from plain view (secure data).
- **Separate** the processing of personal information whenever possible.
- **Aggregate** personal data such that it contains the least amount of detail possible.
- **Inform** users adequately when personal information is processed.
- **Control** the processing of personal information in systems.
- **Enforce** a privacy policy which is compatible with the relevant legal requirements.
- **Demonstrate** compliance with privacy policies and legal requirements.

The Privacy Design Strategies can be used to categorize PET concepts and measures. Although these eight design strategies cover a large part of the privacy laws, they should not be used to check compliance with the privacy laws. If an organization is fully compliant with the eight privacy design strategies, it is not automatically compliant with the privacy law. However, they form a good starting point when developing new information systems and business processes.

3 PET concepts

This section describes concepts of privacy enhancing technologies which can be implemented into Information Systems and Information Processes and can be linked to the software engineering field (other measures such as physical access prevention are out of scope for this paper). Most PET patterns are simple patterns which do not require a lot of effort to implement when designing a new information system. However, implementing these patterns in existing systems might be a challenge [8], [13].

3.1 Database example

The following example will be used to describe the PET patterns of the next sections. In databases it is common to identify a person with an identifier (ID). This ID should be unique for each person in order to be able to identify everyone correctly. In (large) databases, such person identifier is often used in many places to link data to a specific person. The following example illustrates three tables in a database, each table contains information about a person. The information of all database tables can be linked using the person ID attribute.

Person ID	Name
1	Alan
2	Mike
3	Peter

Table 1: Contact Details

Person ID	Bill	Paid
1	550.00	Yes
1	850.00	No
2	499.99	No

Table 2: Bills per person

Person ID	Product
1	Soap
2	Beer
3	Basketballs

Table 3: Most bought products

3.2 Pseudo-Identifier

If someone acquires access to the example case database, or an employee (deliberately) leaks this database, one could easily combine all data about all persons. Even if the *Contact Details*-table is not leaked, databases often contain so much data about a person that this data can still be used to (indirectly) identify a person. If the latter is the case, the data is still considered personal data and the privacy laws, including fines, still apply.

This issue can, in most cases, not be fully prevented by securing the database. In an organization, there are always persons who have access to databases (e.g. Database Administrators), and security measures should work even if a person deliberately tries to leak data according to the Dutch DPA[13]. On the other hand, no database is fully secured, it cannot be assured that a database does not contain any security flaws. This is why data leakages cannot be fully prevented. However, it is possible to limit the impact of a data leakage. The pseudo-identifier pattern will help to limit the impact of a data leakage.

The pseudo-identifier pattern requires an additional database. This database should be designed and maintained by a Trusted Third Party (TTP). The TTP is an external, trusted, reliable party which has no strong affiliation with the organization. The TTP should configure a secure database with a *Table T* containing attributes *Person ID* and *Pseudo-ID*. The attribute *Person ID* in the Tables 2 and 3 should be replaced with *Pseudo-ID* (see figures below). The pseudo id is randomly generated and has no relation to any other attributes. When someone would like to request which contact details belongs to a certain bill, the TTP must be contacted.

Person ID	Name
1	Alan
2	Mike
3	Peter

Table 4: Contact Details

Pseudo-ID	Bill	Paid
1568	550.00	Yes
1568	850.00	No
8643	499.99	No

Table 5: Bills per person

Pseudo-ID	Product
1568	Soap
8643	Beer
9845	Basketballs

Table 6: Most bought products

Person ID	Pseudo ID
1	1568
2	8643
3	9845

Table 7: Table T of TTP

This solution ensures that when the organization's database is leaked, it is not possible to directly relate all data to a person because the *Person ID* is not used in all tables. The database is now divided into two parts: an **identity domain** and a **pseudo domain**. The identity domain contains all information which can be used to directly identify a person. The pseudo domain contains no information which can be used to directly identify a person, instead it uses an arbitrary pseudo identifier. In order to link this pseudo identifiers to a

person, the TTP should be contacted. The identity domain should be as small as possible; personal data should be used as less as possible. The pseudo id domain should be as large as possible, preferable all information systems and business processes should only use pseudo IDs (except one information system in the identity domain which contains all personal data).

Because the TTP is an external organization with their own infrastructure, security measures and employees; the change that both the organization's database and the TTP's database is leaked, is significant lower than that only one of the databases is leaked. If only the TTP's database is leaked, no important data is lost because this database only contains arbitrary identifiers which own their own have no value. When this leakage is detected, all pseudo identifiers could be re-generated.

However, as mentioned before, it might be the case that a database contains such an amount of data about a person that even all data in the pseudo domain might be enough to (indirectly) identify a person. In this case, additional measures should be taken. An enhanced version of the pseudo-identifier pattern might help in this situation. Instead of creating one *Pseudo ID* for identifying a person, it is also possible to generate for each table a different Pseudo ID. The TTP should store all these additional Pseudo IDs. An even more secure approach is to generate a new Pseudo ID each time data is processed. This way, a person can have several Pseudo IDs, even in the same table. This ensures that if someone acquires access to one table, it is even not possible to related data of a person from the same table to each other! Finally, an even more secure method is to periodically re-generate all pseudo identifiers. This final method is however more challenging to integrate compared to previous methods.

ID	Pseudo ID
1	1568
1	8661
1	7894
2	8643
2	1238
3	9845

Table 8: Table T

Pseudo-ID	Bill	Paid
1568	550.00	Yes
7894	850.00	No
8643	499.99	No

Table 9: Bills per person

Pseudo-ID	Product
8661	Soap
1238	Beer
9845	Basketballs

Table 10: Most bought products

As can be seen in Table 9, it is not possible to related the first two rows to each other. However, table 8 shows that these pseudo-ids are actually linked to the same person. This enables more security within the pseudo-domain. Without the TTP, there is no way to relate the data to each other. Even Database Administrators with full database permissions are now unable to find information about a certain person! The TTP still only has arbitrary identifiers. Only if both the TTP's and the organization's databases are leaked, the data can be related to each other. In the worst case scenario, the *Contact Details* table is leaked. Although this table contains a lot of personal data and leakage of this data might still result it fines, the impact of the leaked data is highly limited because no additional data about the persons is leaked.

This PET pattern can be related to the **separate** and **hide** privacy design strategy. A graphical representation of the pseudo identifier concept is shown in the Figure 1.

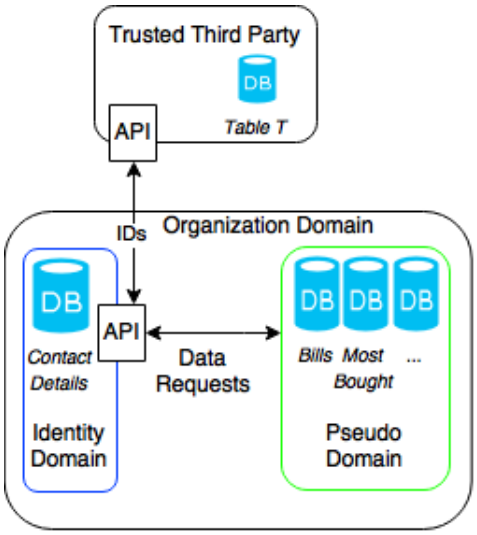

Figure 1: Privacy Management System including the pseudo identity concept

As can be seen in the figure above, the organization is split into two domains: the identification domain and the pseudo-identification domain. The domains can only communicate by a controlled interface. Table 11 summarizes the main advantages and disadvantages of the pseudo-identifier pattern.

Advantages	Disadvantages
Relative easy to implement	Performance is less because an external TTP has to be contacted each request
Separation of data protection responsibilities	TTP contains critical data, losing this data has great consequences
The Trusted Third Party has no confidential data	

Table 11: Advantages and disadvantages of Pseudo-Identifier pattern

3.3 Aggregating Data in Identification Domain

Data in databases is stored in the **highest granularity**, it consist of a single value and cannot be divided into multiple smaller or more precise attributes. For example, an address has a low granularity because it can consist of multiple smaller, more precise attributes (street, house-number, city, state, country etc.). Another example is a person's date of birth. The date of birth has a much higher

granularity then a person's age (in years). Additionally, a person's age has a higher granularity than age ranges (e.g. 0-18 year, 18-35, 35+).

As noted in the previous example of the identification and pseudo identification domain, a large set of data without personal data can still indirectly identify a person and is therefore threaten legally as personal data. Nowadays, taking into account social media such as FaceBook, only a person's home-town and birthday might be enough to identify a person.

The Aggregating Data pattern requires a change between the identification domain and the pseudo identification domain. Figure 1 shows a clear separation between the two domains, a good place to implement the aggregating data pattern is inside the *Application Programming Interface* (API) of the identity domain because all personal data requests are handled by this API. The main goal of this PET concept is to only provide the lowest granularity of data possible to the pseudo identification domain so it can execute it's tasks.

A simple example of this PET concept is: When a process has to know whether a person is older than eighteen, it does not request the exact date of birth from the identification domain. The process sends a request to the identification domain's API with the question if the person is older than eighteen years. The identification domain will reply with a simple *yes* or *no*[4]. This ensures that personal data is kept, as much as possible, in the identification domain. The pseudo domain only contains data with a low granularity and arbitrary pseudo IDs. This makes it harder to use the data in the pseudo domain to (indirectly) identify a person.

When designing processes, it is important to take into account the aggregating concept in an early stage because it can have a rather large architectural impact. The first step which should be taken when designing a new business process is inventorying which kinds of data the process requires in order to complete its goals. Together with the Data Protection Officer (DPO) should be assessed whether there is a legal basis to process the data. The next step is to decided, together with the DPO, which granularity level of the data should be used. It should always be the case that the lowest granularity possible should be selected without effecting the functionality of the business process. Finally, the owners of the data should be informed, and in some cases requested permission. More information about informing users and requesting permission can be found in *section 4: Privacy Management System*. All these steps and decisions should be documented in a structured way for auditing purposes.

This pattern can be related to the **aggregate** privacy design strategy. Table 13 summarizes the main advantages and disadvantages of the aggregating pattern.

[4]In this simple example, it is ignored that the Identity domain should first request which person id belongs to the pseudo id as described in the Pseudo-Identifier pattern.

Advantages	Disadvantages
Easy to implement	This pattern is contradicting with other patterns which are against implementing business logic across several systems.
Minimizes the use of data with high granularity	

Table 12: Advantages and disadvantages of aggregating pattern

3.4 Separating Data Processing

The separate pattern[5] is a relative easy to implement PET pattern which can be described with the following rules:

- Each business process should be secured in such a way that it is isolated from other business processes and business process instances
- Each database should be secured in such a way that it is isolated from other databases and that is only accessible via a controlled interface
- If a business process can be separated into smaller, isolated sub-processes, this should be done
- If a database can be separated into smaller, isolated databases, this should be done

The first rule implies that a business process instance can only access it's own data, it cannot access data from other business processes. The second rule implies that similar measures should be taken for databases. When one acquires (legitimately) access to a database, one should not have automatically access to other databases. All requests to a database should be via a controlled interface which automatically handles authentication, logging and monitoring of the requests.

The third rule implies that large business processes should be split up into smaller business processes. Large business processes require lots of information (personal data). If such a business process requires large sets of data about a person, this data can be (indirectly) related to a person. To prevent this scenario, the business process could be divided into smaller business processes which each are responsible a part of the computation of the large process. Each business process could output its result into a database and trigger the next sequential business process.

A similar rule counts for databases; instead of having one large enterprise database, smaller and secured databases should be constructed, each with a Database Owner and Database Administrators. The separation of ownership and responsibilities prevent that there are some *superusers* who are authorized to view, manage and maintain all data. Figure 2 illustrates the separation of databases into smaller databases. This example is based on the pseudo-identifier concept of subsection 3.2: Pseudo-Identifier.

[5]Please note that, as previous mentioned in subsection 1.1: Definitions, storing data is legally seen as *processing* data.

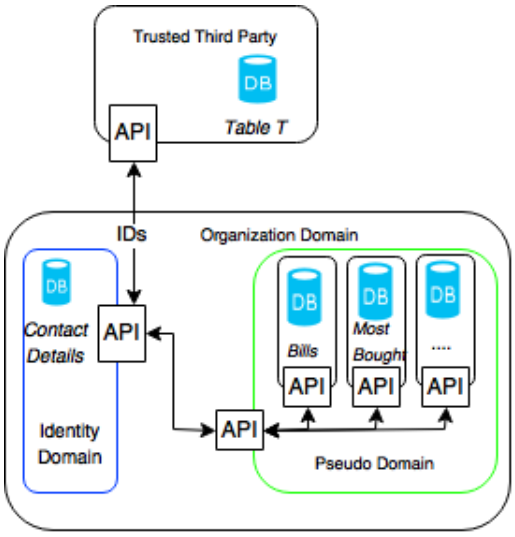

Figure 2: Separation of databases applied to the pseudo identifier concept

There are tools which can help calculating when a certain data is easily related to a natural person, namely K-Anonymity[14], L-diversity[15] and T-Closeness[16]. These can help defining when a certain database or process should be split into sub-processes/sub-databases.

Table 13 summarizes the main advantages and disadvantages of the separation pattern.

Advantages	Disadvantages
All data requests are handled by a controlled interface Processes and databases only use a small subset of the total data	Managing more databases might require more effort

Table 13: Advantages and disadvantages of separation pattern

3.5 Data Logging

Data Logging can be described as *the recoding and collection of data about actions, requests or events for analysis purposes.* Data logging is a well known concept in Computer Science field, but is mainly used to detect errors, faults or to analyse performance of a system. However, data logging can also help to identify which data is processed in which way and if this is compliant with privacy policies [13]. In this paper, the following types of data logging are recognized:

- **Application logging**, contains application- and process-specific data (input/output of a process-component).
- **Process logging**, contains information about the life-cycle of process instances.

- **Event logging**, contains information about an occurred event (server down, user created, order paid)
- **Performance logging**, contains data about the performance (duration of process instance, calls per second) of a process.

Performance logging should not contain any personal data (only meta-data about requests). Therefore, performance logging is not relevant for PET purposes. Event logging, process logging and application logging might contain personal data, this implies that corresponding measures should be taken to prevent that unauthorized people can acquire this information. Therefore, it is important to separate the different kinds of data logging.

As can be seen in figure 2, data requests are typically handled by APIs. These APIs are also a good place to implement data logging systems. Each request could be logged by the APIs, this can result in both performance logging as application logging. Business processes can generate application, process, event and performance logging.

The Dutch Data Protection Act and the General Data Protection Regulation state that users have the right to request an overview of all systems which did process their data, including a detailed description about this data and what the legal basis is for processing this information. Process and event logging might help to identify which systems and processes did use the user's personal data. Application logging might help to identify in what way the data is processed. An automated collection and processing of logging data could ultimately result in a system which automatically can handle such personal data requests. More about this can be read in section 4: Privacy Management System.

Table 14 summarizes the main advantages and disadvantages of the data logging pattern.

Advantages	Disadvantages
All requests are recorded	The logged data should also be secured
It is possible to find which systems use certain data	Processing all logged data might require many resources

Table 14: Advantages and disadvantages of data logging pattern

4 Privacy Management System

A Privacy Management System (PMS) can be described as *a shield around personal data which protects the personal data by controlling and logging which systems want to use the personal data based on user preferences.* In practice this means that a user can manage, via for example: a web interface, which personal data can be used and can therefore make its own custom privacy policy. A Privacy Management System can enforce custom privacy policies within an organization. Additionally, a PMS can inform users when their data is used, by which system and for what purpose. The goal of a Privacy Management System is to be transparent about data processing in order to increase users' confidence.

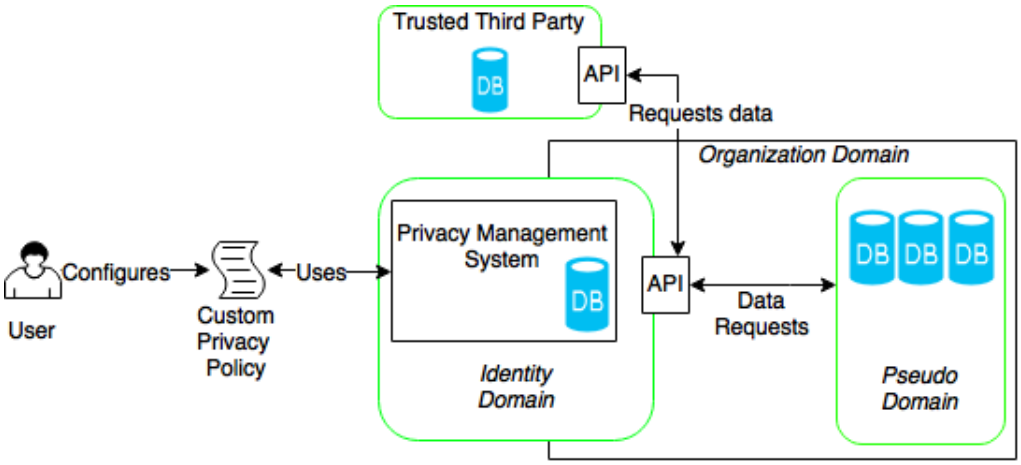

Figure 3: High level overview of a Privacy Management System

The Privacy Management System enables users to configure which data may be used in which way. Organizations can process personal data for various purposes; for example: a company may use an address as shipping address for new orders, at the same time the company can also use this address to send promotions such as vouchers. When using a PMS, the users can decide for which purposes their data can be used. The figure above illustrates a high-level schematic overview of a PMS.

A Privacy Management System is not a readymade tool which can be bought and easily integrated into a organization's internal environment, it is a custom made solution which can differ for each company due to different technical and organizational requirements. A PMS is however a very interesting PET concept because it can be built using all previous PET concepts. Most privacy management systems consist of the following parts:

1. (Web) interface which enables users to configure a custom privacy policy (CPP)
2. Storage facility which contains the users' personal data
3. Interfaces for information systems and business processes to request personal data
4. Data logging system which records meta-data about all personal data requests
5. (Web) interface, preferable the same as previous mentioned, which enables users to view when their data is used for which purposes

Item one and five can be implemented using a public website. The user can log-in into its account and configure its custom privacy policy. Note that the user might disallow the usage of data which is essential for an organization in order to execute its services (e.g. a shipping address for an online webshop). In this case, the user should be informed that this data has to be used by the organization and revoking permission might result in a situation where the organization is no longer able to execute its services. Although this is not convenient for the organization, users have the *Right to be Forgotten* so this

option should be present. However this right does not apply if an organization needs personal data for executing tasks which they are legally bonded to (for example: a company needs information about its employees in order to be able to pay salaries and report the salaries to the government).

Another use of the online web portal is that the user can view the history of data requests which are made by information systems and business processes. Additionally, the online web portal also enables the user to update its personal data. This *Right to change personal data* is also described in both the General Data Protection Regulation as the current Dutch Data Protection Act.

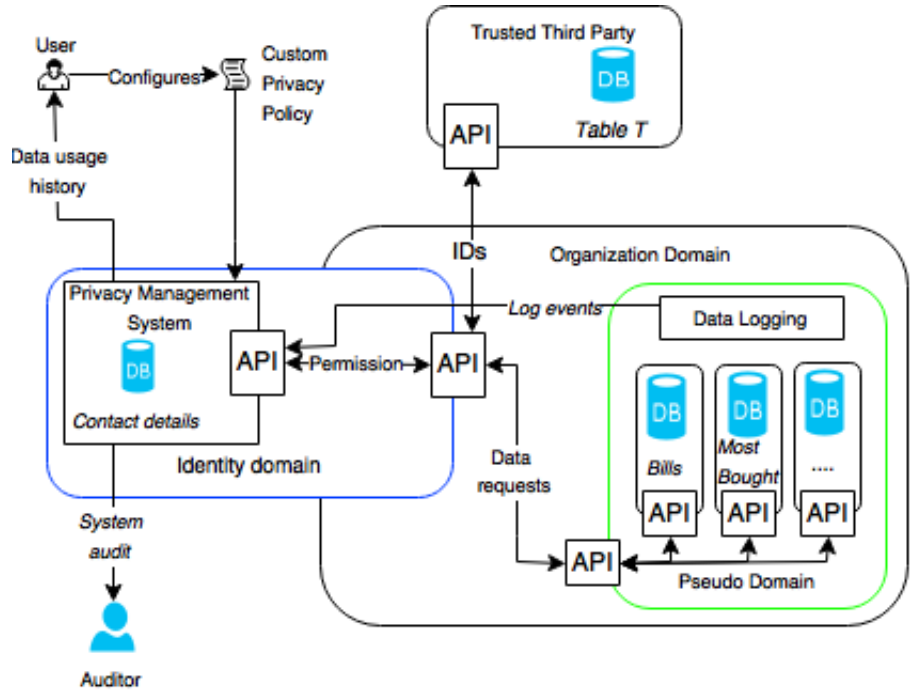

Figure 4: A Privacy Management System which integrates all previous PET concepts.

Item two can be implemented using an isolated and secured database using the rules mentioned in section 3.4. Item four is described in detail in section 3.5. The Privacy Management Systems uses the logging information to generate a complete history about which system or process used which data. First of all, this is interesting for users because they can see how their data is used and this automatically handles personal data requests. Secondly, this is interesting for auditors because they can use the collected data of the PMS to check whether the data requests have a legitimate basis.

Item three is an interesting subject because it can be combined with the pseudo-identity PET concept. The pseudo-identity concept creates a clear division between a pseudo-identity domain and an identity domain. The Privacy Management System should located into the identity domain because it contains directly identifiable personal data. All other information processing systems are located into the pseudo-identity domain. Figure 4 illustrates a Privacy Manage-

ment System which also integrates the pseudo identifier concept, the aggregating data concept, separating data processing concept and the data logging concept.

Something which is not visible in this figure is that all APIs export their logging information to the Data Logging system. For convenience reasons, the figure only implicit, and not graphically, contains the connections between the APIs and Data Logging system.

As mentioned in subsection 1.3: General Data Protection Regulation, organizations are required to test (audit) regularly if their organization is compliant with the current privacy laws. A PMS can help in executing such audits because all requests for personal data are documented into the privacy management system. It is not possible to request personal data without first contacting the PMS, because the PMS should first grant access before the corresponding pseudo identifiers are provided. Without these pseudo identifiers, it is technically not possible to acquire personal data (as explained in section 3.2). Additionally, the privacy management system also contains all user preferences regarding data processing. The auditor can use all this data to check whether the systems are compliant with the custom privacy policies and with the privacy laws.

This concludes that the Privacy Management System does not only facilitate a transparent method for processing data, it can also help to demonstrate compliance by supporting auditors with all the data they need.

5 Discussion

As previously noted, it is important to think about privacy protection in an early system development phase. Researchers advised to manage privacy requirements the same as functional system requirements[17]. Re-designing existing information systems and business processes in such a way that it integrates PET concepts is very challenging due to dependencies of the systems[13].

Researchers found that integrating PET measures in a new system increased the total development cost with one to ten percent, depending on how large the system is and how many PET measures should be integrated [13]. Implementing the same measures into an existing system significantly increases these costs[13].

Depending on which data and the quantity of processed data within an organization, some PET concepts might be more important than other. Due to differences in the (technical) environment of an organization, some PET concepts might be more challenging to integrate then others. Each organization should therefore discuss, preferable in corporation with the Data Protection Officer, which measures should be taken.

The Privacy Management System design incorporates some important PET concepts into one design. This example design can be used visualise what the impact of privacy laws has on information systems. Although implementing a Privacy Management System is a step in the right direction for complying with the privacy laws, it is still not sufficient. The organization's network should be sufficiently secured, the data should be physically protected (e.g. notebook theft) and employees should be trained. These topics are extensively covered in industry standards such as the ISO/IEC 27002:2013.

There are however still exception situations which require even more preventive measures, some example situation are when:

- The organization processes special personal data[6].
- The organization outsources data processing to another organization.
- Personal data of EU citizens is shared with non-European countries.
- The organization is a Telecom provider, healthcare organization or a government organization.

More research is required in order to extend the Privacy Management System so that it also can be used in these exception situations.

[6]There is a special category for data about a person's religion, race, criminal records, political opinion, health information, sexual behaviour/life and membership of labour unions.[7]

6 Conclusion

This paper first briefly discussed current and upcoming privacy legislations. It also discussed the major differences between the current and upcoming legislations, and addresses the importance and need to comply with the privacy legislation. Next, the history of Privacy Enhancing Technologies was discussed and some generic PET concepts were discussed, explained and summarized. These PET concepts were combined in the design of a Privacy Management System. This Privacy Management System enables customers to manage and control their personal data, and the processing of personal data.

This paper did only focusses on the (very) high level design of PET concepts. Further research will focus on how to translate these high level designs into functional requirements and ultimately into a working proof of concept. The results of this research will be published in the near future.

Further research should be conducted into the impact of the General Data Protection Regulation. Additionally, it is important to research the new legislation effects current PET concepts and to help organizations comply with the privacy legislation by designing ready-to-use PET concepts/strategies.

7 Further Reading

There are many other studies conducted in the Privacy Engineering field. One of the earliest research in this field is published as *Handbook of Privacy and Privacy-Enhancing Technologies: The case of Intelligent Software Agents*[8]. This study first focusses on the legal aspects of privacy and then works out some detailed cases. It should however be mentioned that the privacy laws did change in the past years and this can not be found in this study. Additionally, this study tends to only discuss high level software design of PET, detailed implementation guides and patterns are missing.

A more software engineering approach of integrating PET measures into information systems is described as *Privacy Design Strategies*[12]. These design strategies are also extensively used in this paper.

An interesting approach on how to ensure that organizations are compliant with the law, in participially the Dutch privacy law, is Ampersand. In a Dutch study published by H. Joosten of TNO[18], an approach is described to use a Rule-Based System (RBS) to ensure that an organization complies with a certain set of laws or regulation. Additional highly detailed information about this Rule-Based System can be found in the study *Rule Based Design*[19]. Note that Ampersand is a mathematical approach to integrate *rules* into an organization, some mathematical background knowledge is required.

It should be mentioned that there is many research conducted in the field of integration privacy measures into digital communication[20], [21] and that there are many industry standards, such as ISO/IEC 27002:2013, on the organizational aspect of privacy enhancing technologies.

Finally, it is important to look at studies of the local Data Protection Authorities and news regarding the new General Data Protection Regulation. The latter could have great impact on organizations.

References

[1] *De meldplicht datalekken in de wet bescherming persoonsgegevens (wbp)*, `https://zoek.officielebekendmakingen.nl/stcrt-2015-46128.html`, 2015.

[2] M. van Veiligheid en Justitie, "Regulation of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation)", 2012.

[3] A. Westin, *Privacy and freedom*. Londen: The Bodley Head, 2016.

[4] M. B. Z. en Koninkrijksrelaties, "Grondwet voor het koninkrijk der nederlanden", 2008.

[5] E. Parliament and of the Council, "Directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data", 1995.

[6] E. C. of Human Rights, "European convention on human rights", 1994.

[7] M. van Veiligheid en Justitie, "Wet bescherming persoonsgegevens", 2016.

[8] G. van Blarkom, J. Borking, and J. Olk, *Handbook of Privacy and Privacy-Enhancing Technologies: The case of Intelligent Software Agenst*. The Hague: College Bescherming Persoonsgegevens, 2003.

[9] B. van persoonsgegevens, *CBP Richtsnoeren: College Bescherming Persoonsgegevens*. 2013.

[10] J. van Rest, D. Boonstra, M. Everts, M. van Rijn, and R. van Paassen, *Designing Privacy-by-Design*. Springer Berlin Heidelberg, 2014.

[11] S. Gursus, C. Troncoso, and C. Diaz, *Engineering Privacy by Design*. 2012.

[12] J.-H. Hoepman, *Privacy Design Strategies*. 2012.

[13] M. van Binnenlandse Zaken en Koninkrijksrelaties, *Privacy Enhancing Technologies: Witboek voor beslissers*. 2014.

[14] L. Sweeney, "K-anonymity: A model for protecting privacy", 2002.

[15] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity", 2006.

[16] N. Li and T. Li, "T-closeness: Privacy beyond k-anonymity and -diversity", 2006.

[17] C. Kalloniatis, E. Kavakli, and S. Gritzalis, "Addressing privacy requirements in system design: The pris method", 2007.

[18] H. Joosten, "Aantoonbaar voldoen aan (komende) privacywet en regelgeving in een it-rijke context", 2013.

[19] S. Joosten, L. Wedemeijer, and G. Michels, *Rule Based Design*. 2010.

[20] M. Hafiz, "A pattern language for developing privacy enhancing technologies", 2011.

[21] M. Hafiz, "A collection of privacy design patterns", 2006.