RZ 3391 (# 93437) 01/21/02
Computer Science 9 pages

# Research Report

## Privacy-enabled Services for Enterprises

Günther Karjoth, Matthias Schunter and Michael Waidner

IBM Research
Zurich Research Laboratory
8803 Rüschlikon
Switzerland
{gka,mts,wmi}@zurich.ibm.com
http://www.research.ibm.com/privacy

IBM Research
Almaden · Austin · Beijing · Delhi · Haifa · T.J. Watson · Tokyo · Zurich

# Privacy-enabled Services for Enterprises

Günther Karjoth, Matthias Schunter and Michael Waidner

*IBM Research, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland*

## Abstract

The IBM Enterprise Privacy Architecture (EPA) is a methodology for enterprises to provide an enhanced and well-defined level of privacy to their customers. EPA is structured in four building blocks. The privacy regulation analysis identifies and structures the applicable regulations. The management reference model enables an enterprise to define and enforce an enterprise privacy strategy and the resulting privacy practices. The privacy agreement framework is a methodology for privacy-enabling business process re-engineering. It outputs a detailed model of the privacy-relevant players and activities as well as the privacy policies that govern these activities. The technical reference architecture defines the technology needed for implementing the identified practices. The Platform for Enterprise Privacy Practices (E-P3P) is a refinement of EPAs technical reference architecture: Enterprises collect a certain amount of personal data while promising fair information practices to their customers. E-P3P enables an enterprise to keep the privacy promises made. It formalizes these 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.

# 1 Introduction

Consumer privacy is a growing concern in the marketplace. While the concerns are most prominent for e-commerce, the privacy concerns for traditional transactions are increasing as well. Some enterprises are aware of these problems and of the market share they might loose if they do not implement proper privacy practices. As a consequence enterprises publish privacy statements that promise fair information practices. In addition, laws increasingly impose baseline privacy regulations. Unfortunately, companies willing to implement fair privacy practices usually face the following problems:

- Business processes are designed without considering privacy requirements. Thus enterprises are forced to create stockpiles of personally identifiable information (PII) instead of collecting the PII when needed for the business at hand.

- Existing services often identify users even though their identity is not needed for the business at hand. Privacy-friendly security technology that provides security with less data is rarely used.

- Enterprises store a variety of PII. Larger enterprises may not know what types of PII are collected and where it is stored.

- Enterprises may neither know the consent a customer has given nor the legal regulations that apply to a specific customer record.

The goal of the IBM Enterprise Privacy Architecture (EPA) is to solve these problems while concentrating on the enterprise-related aspects. Essentially, EPA is a methodology for enterprises to provide a well-defined and enhanced level of privacy to its customers. It provides the foundation for the privacy part of IBM's Security and Privacy Services. EPA is presented in Section 2.

In Section 3 we present the Platform for Enterprise Privacy Practices (E-P3P). E-P3P is a refinement of the technical reference architecture of EPA. It enables enterprises to formalize and enforce privacy practices and to manage the consent of their customers.

# 2 The IBM Enterprise Privacy Architecture

The IBM Enterprise Privacy Architecture (EPA) is an architecture that allows enterprises to maximize the business use of personal information while respecting privacy concerns and regulations. It provides a sustainable privacy management system, which can be customized to the total set of privacy regulations and privacy choices facing an enterprise.

EPA introduces privacy-awareness and privacy services into enterprises in a systematic and complete way. As a prerequisite, the EPA privacy regulation analysis identifies and structures the applicable regulations. The topmost part of EPA is the Management Reference Model (see top 3 layers in Figure 1). The Management Reference Model defines the privacy strategy and practices of the enterprise. The Privacy

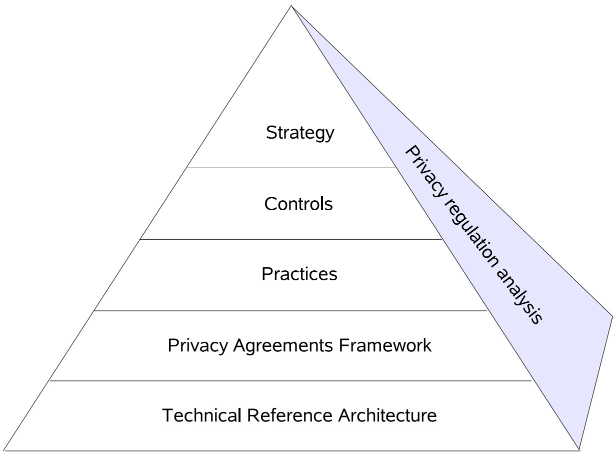
Strategy
Controls
Practices
Privacy Agreements Framework
Technical Reference Architecture
Privacy regulation analysis

Figure 1: Building Blocks of the IBM Enterprise Privacy Architecture (EPA)

Agreements Framework provides a privacy-enabled model for privacy-enhanced business process re-engineering. The lowest layer is the Technical Reference Architecture that defines the technology for implementing the required privacy services. The following advantages and requirements were considered when designing EPA:

*Enhance and preserve the value of data assets.* The data model of the Technical Reference Model provides for the identification and categorization of personal data within the organization and thereby allows the establishment of appropriate protection measures.

*Operate consistently with multiple privacy regulations and standards.* The privacy regulation analysis helps to identify compliance obligations across different jurisdictions and express these in common terms. The applicable regulations are formalized by an enterprise privacy policy that is associated with any collected data. This so-called “sticky policy paradigm” supports identifying the applicable regulations and privacy promises for all personal data in an enterprise.

*Build and promote trust in the marketplace.* EPA enables customers to retain control over their data. The Management Reference Model enables and promotes responsiveness and privacy awareness of the enterprise. Together with external auditing, these measures promote trust of the customers.

*Realize substantial privacy management choices.* The regulatory analysis reveals compliance choices. This analysis highlights choices for uses of less sensitive data types and shows high risk and redundant privacy relationships.

*Operate a sound platform for persistent privacy management.* The Requirements Process within the Management Reference Model ensures ongoing environmental input on privacy.

## 2.1 Privacy Regulation Analysis

Regulatory compliance is a primary driver of privacy-related activity in the marketplace. Thus, it is clear that a useful picture of the regulatory landscape is a pre-requisite to developing any kind of privacy architecture. The challenge is that regulations are typically written in dense legal style with formats and terminology that tend to differ depending on their origin and purpose.

EPA addresses this challenge by *regulatory summary tables* and *regulation rules tables*. Regulatory summary tables summarize the applicable regulations using a unified terminology. The regulation rules tables identify data that is are in the enterprise as well as the legal restrictions on using such data. The regulation rules tables are enterprise-specific and more formal than the regulation summary table. An entry describes which party can perform which action on which type of data, the resulting privacy obligations, and a reference to the legal regulation. In addition, the four business-use phases Collection, Retention, Processing and Use ("CRPU") are used to categorize the scope of privacy regulations.

## 2.2 Management Reference Model

The EPA management reference model addresses the enterprise-wide processes necessary for a comprehensive privacy management program. These processes are structured and linked to drive the program starting from a strategic view down through the implementation of privacy practices (see Figure 1).

**Strategy** defines the privacy philosophy, the high-level policies and identifies the applicable regulations.

This represents the highest level of an enterprise's privacy program and embodies its philosophy, its policies and the regulations it will adhere to. The outputs are a privacy strategy as well as a security strategy. Both define what an enterprise will do for protecting privacy and security.

**Control** defines the general controls necessary to enforce policy.

Its components are a Privacy Requirements Process, the Information Asset Classification and Control, a Compliance Enforcement Process, a definition of the Organizational Roles and Responsibilities as well as an Employee Education Program.

**Practices** defines the incorporation of policy into business processes.

This represents the level of an enterprise's privacy program that translates privacy policy obligations into the general processes, programs and activities that will implement them. Its components are a Privacy Statement declaring the enterprise policy, a Customer Preference Program for defining opt-in and opt-out choices, an Individual Participation Process that enables customers to access their data, a Dispute Process, an External Communication Program that advertises the privacy efforts of an enterprise, and Information Access Controls that protect the enterprises' data and resources.

## 2.3 Privacy Agreements Framework

The Privacy Agreements Framework models the transaction level management of privacy at the points where enterprises use personal information within business processes. This includes processes that connect the individual to the enterprise, processes linking people and departments within the enterprise, and processes linking the enterprise with third parties. This model can then be used to identify privacy agreements that are required between the players involved. The main parts of the model are players, data, and rules:

**Players** The players are the entities that interact while processing collected data. Basic players are data subjects (persons about whom data is collected) and different data users (enterprises or employees using the data). The player model uses an object-oriented modeling technique to identify the players, their operations on the data, as well as the interactions among the players. The result is documented using UML [1] class and collaboration diagrams.

**Data** The data model identifies the data needed for the processes. Besides identifying the fields collected in forms, it classifies data into at least three categories:

- Personally Identifiable Information (PII) is the most sensitive kind of information that can be linked to a real-world identity. Examples include a tuple name/surname or a U.S. social security number.
- Depersonalized Information is PII where the identifying information has been replaced by a pseudonym. Even though this data is less sensitive, some parties are able to repersonalize it by replacing the pseudonym with the identifying information. Examples include the age with a customer number.
- Anonymized Information contains no identifying information or pseudonyms. It is the least sensitive kind of information that can be obtained by removing all PII from a set of data. For anonymized data, it is required that identifying the data subject given the data is virtually impossible. Examples include the town of residence or an age in years on its own (i.e., without any other information that may enable identification of the data subject).

**Rules** The rules model identifies the rules that govern the usage of data by players and their operations. It defines what player may perform which operation for what purpose. In addition, rules may impose conditions and may define obligations that result from performing an operation.

## 2.4 Technical Reference Architecture

To guarantee that an enterprise provides sufficient privacy to its customers, privacy-enforcement needs to be deployed on an enterprise-wide scale. All applications that handle PII need to make sure that the handling adheres to the promised policies. An enterprise-wide privacy-management system uses at least three types of systems (see Figure 2):

**The Policy Management System** enables the administrators of the system to define, change and update privacy policies. It distributes the privacy policies to the privacy enforcement systems.

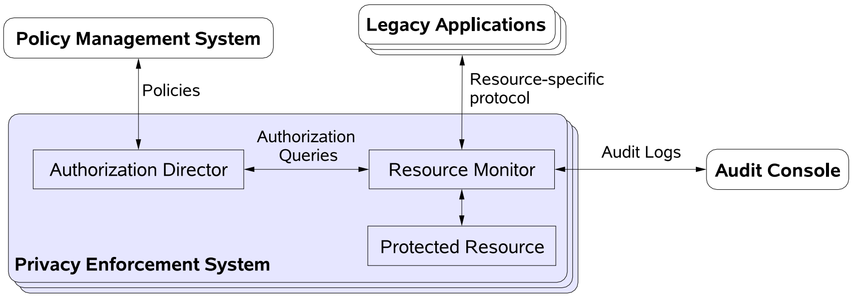

Figure 2: Components of a Enterprise Privacy Enforcement System

**The Privacy Enforcement System** enforces the privacy protection for each individual resource that stores privacy-relevant data. It obtains policies from the policy management system and offers auditing data to the audit console. The privacy enforcement system is usually split into two parts: A resource-specific resource monitor shields the resource and a resource-independent authorization director evaluates the policies and decides whether requests are granted or not.

The authorization director authorizes operations on the collected data. After evaluating the policy, the authorization director returns whether the request is authorized or not and whether an authorized request implies any privacy obligations.

Each kind of protected resource (database, CRM system, . . . ) uses a corresponding resource monitor. This monitor shields the resource from direct access. Each incoming request is translated into a call to the authorization director. Only if the authorization director authorizes the request, the request is forwarded to the resource. The resource monitor records audit data and tracks pending privacy obligations.

**Audit Console** This system enables the Privacy Officer to review the audit information stored in the enforcement nodes and the policies distributed by the policy management systems.

## 3 The Platform for Enterprise Privacy Practices (E-P3P)

The Platform for Enterprise Privacy Practices (E-P3P) describes a refinement of the technical reference architecture. E-P3P enforces the enterprise-internal privacy practices that were identified by the privacy agreements framework.[1] Its core is an authorization scheme that defines whether certain operations are allowed or not.

E-P3P is a concept for privacy-enabled data management. It includes several independent building blocks:

[1] These privacy promises can be formalized using P3P [2]. Note that P3P is coarser than the enterprise privacy practices of E-P3P: While P3P may promise “we do not disclose data”, E-P3P may define the exact flows that are authorized in an enterprise.

**Identify PII** As a prerequisite to PII management, an enterprise needs to identify systems that store PII and channels through which PII enters an enterprise. This defines the domain of any Enterprise Privacy Management System. This is an integral output of the privacy agreements framework.

**Formalize Privacy Policies:** For each type of PII (e.g., defined by application or other entry point where data enters the enterprise), the privacy agreements framework outputs an natural-text description of the applicable privacy policy. The policy needs to reflect legal guidelines as well as the customer's and company's expectations. For automated enforcement in E-P3P, this policy needs to be formalized using a machine-readable privacy policy language. The natural text version can then be inspected by the customers while the machine-readable version will be used for enforcement.

**Identify and Formalize Policy Options** To cover a broad range of customers, the privacy policy can identify certain opt-in as well as opt-out choices or options that depend on the collected data (e.g., whether the given data belongs to a child or not). These options enable a company to use a global policy while still leaving some freedom of choice to the customer.

**Manage Customer Consent** A privacy policy can be seen as a contract between the individual customer and an enterprise. As a consequence, a customer needs to authorize the applicable policy as well as any applicable opt-in and opt-out choices that the policy offers. Another consequence is that consent needs to be recorded on a per-customer basis.

**Enforce Policies and resulting Obligations** Given collected PII and its policy, the policy needs to be enforced. This includes privacy-enabling access control that allows only actions that are authorized by the applicable policy. In addition, resulting obligations (such as ”we delete collected data if consent is not given within 15 days”) need to be enforced as well.

**Audit Compliance** The handling of PII should comply to the privacy policy in an auditable way. This enables a Privacy Officer to later verify that the data was handled properly.

**Provide Customer Privacy Services** Proper handling of PII implies certain additional services to the customers that are mandated by existing privacy regulations A customer should be enabled to inspect and update the data and usage logs stored about it. In addition, an enterprise may offer the option to delete the PII.

## 3.1 Policies and Separation of Duty

Privacy and security authorization in an enterprise involves at least four types of players:

**Data Subject** The parties about whom personally identifiable information (PII) is collected. The most common data subjects are the customers of an enterprise. Other data subjects are employees or customers of cooperating enterprises.

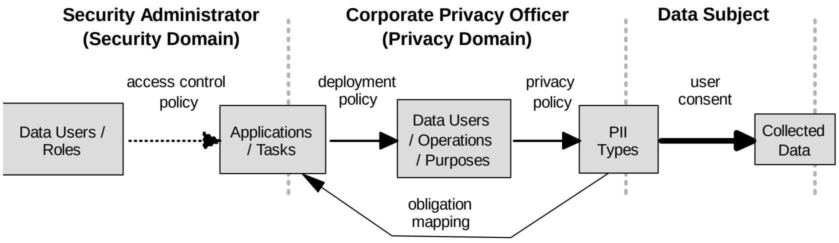

Figure 3: Separation of Duty for Privacy Authorization.

**Data Users** The parties within an enterprise that use the collected data. Using data is done by executing tasks of applications that in turn access fields of the collected data.

**Privacy Officer (PO)** The party defining privacy policies of an enterprise.

**Security Officer (SO)** The party defining access control policies of an enterprise.

E-P3P introduces the following intermediate abstractions and the corresponding policies (see Figure 3):

1. *PII is collected in fields that are grouped in forms.* Each field has a certain PII type such as “medical record”, “address data”, or “order history”. The fields of one data subject that belong together from a data subject’s point of view are associated with each other using a form. A “customer data” form can, e.g., group the fields “name”, “street”, and “town”. Example forms are “customer data”, “purchase history”, or “financial information”.

2. *The PO defines a privacy policy.* A privacy policy defines for each PII type, what *operations* for which *purpose* by which *data user* can be performed on a given *PII Type*. In addition, the policy may define certain privacy obligations like “delete my data after 30 days unless parent consent has been given”. The “marketing department” may be allowed to “read” the data for “e-mail marketing”. In addition, a privacy policy defines the opt-in and opt-out choices of the data subjects.

3. *The PO and the SO define a deployment policy.* The deployment policy maps legacy applications and their tasks onto the privacy-specific terminology used by the privacy policies. This mapping is specific to each enterprise. E.g., in one enterprise a CRM system performing “product notification” as well as a printer for mass-mailings may be mapped onto the action “read” for purpose “marketing”. In addition, another enterprise may use a legacy application instead of using the off-the-shelf CRM system. Again, this legacy implementation would be mapped onto “read” for “marketing”.

4. *The PO defines a collection catalogue.* The collection catalogue identifies the sources where data is collected. For each source, the catalogue defines the collected data, its PII Types, and a default policy. This information is associated with a form for each particular collection point.

5. *The PO defines an obligation mapping.* The obligation mapping translates application independent obligations of the privacy policy (such as "delete") into specific implementations. E.g., a delete may be translated into an "unsubscribe" of the mailing list.

6. *The SO defines an access control policy.* An access control policy defines the roles and users inside an enterprise. In addition, it defines which users or roles can execute which tasks of which applications.

After defining the policies described above a protected resource (see Section 2.4) can store data collected from the data subject. The data subject enters certain data in the form for a given collection point. The form contains the fields and PII types of the entered data as well as a default policy. The data subject may choose opt-in and opt-out options for this policy and then consents or denies the policy and the selected choices. Once consent is given, the form containing choices, fields with data and PII types, and the policy is then stored for further processing.

## 3.2 Granting or Denying an Access

The form associated with the collected data, the privacy policy and the selected options are used to decide whether an access shall be granted or not. This authorization is done in two levels. While access control focuses on restricting access of employees to enterprise applications, privacy control restricts access to collected data.

**Access Control:** An employee acting as a user with certain roles requests to perform a task of a application. The access control policy is used to verify that the user with the given roles is in fact allowed to perform the requested task. If this is the case, the task is executed. This access control system is independent of the privacy authorization.

**Privacy Control:** Once a running task of a corresponding application requests access to certain fields of collected data, the privacy enforcement system retrieves the form and uses it to allow or deny the given request as follows:

1. The privacy enforcement system obtains the request identifying the task of an application as well as the fields to be accessed. At collection time, the accessed fields have been associated by a form with other fields, the PII type of each field, the choices, and the applicable policy.

2. The privacy enforcement system uses the deployment policy to map the task onto a privacy-relevant operation and a purpose.

3. The privacy enforcement retrieves the PII types of the requested fields from the form.

4. Using the consented privacy policy and the data subject's choices, the system decides whether the operation for this purpose is allowed on the given PII type or not. If the privacy-related operation for this purpose is denied, the privacy enforcement rejects access for the given task on the given fields. If the operation is allowed, the task may access the fields. In this case, privacy obligations that may be specified by the policy need to be processed as well.

5. The obligations mapping is used to map each returned obligation onto a concrete implementation that is scheduled for execution.

6. If the operation is allowed, the task can be performed on the requested fields.

### 3.3 A Note About Consent Management

An important aspect of this privacy enforcement system is the management of the data subject's consent on a per-person basis. Consent management includes the management of the consented policy as well as the management of the users opt-in and opt-out choices.

The core of our notion of consent management is the sticky policy paradigm: When submitting data to an enterprise, the user implicitly consents to the applicable policy and to the selected opt-in and opt-out choices. The form then associates the opt-in and opt-out choices as well as the consented policy with the collected data. This holds even if the data is sent to another enterprise.

Note that policy management on a per-user-basis is useful once consent and different sources need to be considered. Examples are managing data of different policy versions (e.g., due to different collection times), different user roles (e.g., premium and users funded by advertising), or users from different legislation (e.g., Europe and US).

## 4 Acknowledgments

We would like to thank Kathy Bohrer, Nigel Brown, Jan Camenisch, Calvin Powers and Els Van Herreweghen for valuable comments. In addition, we would like to thank all members of the EPA team for the productive cooperation.

## References

[1] Grady Booch: Object Oriented Analysis and Design. Benjamin Cummings, 1994.

[2] The Platform for Privacy Preferences (P3P), W3C Candidate Recommendation, http://www.w3.org/TR/2000/CR-P3P-20001215, 2000.