On Privacy-aware Information Lifecycle Management in Enterprises: Setting the Context[♦]

Marco Casassa Mont
Trusted Systems Laboratory
HP Laboratories Bristol
HPL-2006-109
August 1, 2006*

privacy, ILM, identity management, privacy management

This paper aims at setting the context for privacy-aware information lifecycle management within enterprises, i.e. the process of handling the lifecycle of personal and confidential information in a way that is compliant with privacy laws and people's expectations (including data retention, deletion, notifications, data transformation, etc.). Despite the fact that enterprises are already using Information Lifecycle Management (ILM) and Identity Management (IDM) solutions to store and manage various types of data, in terms of "privacy-aware" lifecycle management of information much is still done by means of manual processes that are complex and hard to monitor. This is a green field, open to innovation. We argue that automation can be introduced to address this aspect by leveraging, among other things, existing enterprise ILM and IDM solutions. In this context, we investigate and analyse core privacy requirements and issues that need to be addressed by enterprises along with their implications and impact on existing ILM and IDM solutions. The goal is to create awareness and suggest potential ways to move towards their automation and simplification. We provide an overview of research and work done by HP Labs to develop approaches and technologies that can help enterprises to implement and automate aspects of privacy-aware information lifecycle management.

* Internal Accession Date Only

[♦] ISSE 2006, Rome, Italy, 10-12 October 2006

On Privacy-aware Information Lifecycle Management in Enterprises: Setting the Context

Marco Casassa Mont

Hewlett-Packard Laboratories
Trusted Systems Lab, Bristol, UK
marco.casassa-mont@hp.com

Abstract

This paper aims at setting the context for privacy-aware information lifecycle management within enterprises, i.e. the process of handling the lifecycle of personal and confidential information in a way that is compliant with privacy laws and people's expectations (including data retention, deletion, notifications, data transformation, etc.). Despite the fact that enterprises are already using Information Lifecycle Management (ILM) and Identity Management (IDM) solutions to store and manage various types of data, in terms of "privacy-aware" lifecycle management of information much is still done by means of manual processes that are complex and hard to monitor. This is a green field, open to innovation. We argue that automation can be introduced to address this aspect by leveraging, among other things, existing enterprise ILM and IDM solutions.

In this context, we investigate and analyse core privacy requirements and issues that need to be addressed by enterprises along with their implications and impact on existing ILM and IDM solutions. The goal is to create awareness and suggest potential ways to move towards their automation and simplification. We provide an overview of research and work done by HP Labs to develop approaches and technologies that can help enterprises to implement and automate aspects of privacy-aware information lifecycle management.

1 Introduction

Enterprises collect large amounts of *information* to enable their business processes and interactions. The term "*Information Lifecycle Management*" refers to the processes, mechanisms and solutions that are put in place by enterprises to handle the lifecycle of this information, including its storage, retrieval, usage, prioritization, update, transformation and deletion. Managed information consists of documents, files, records, etc. Enterprises have been investing in *Information Lifecycle Management* solutions to address the above aspects.

Enterprises also collect, store and process *personal information* and *digital identities* to allow them to authorise business transactions and interactions and provide users with more customised and effective services. In this context, enterprises have been investing in "*Identity Management*" solutions to: deal with user provisioning and account management; store and retrieve personal identity information; handle secure access to data and systems/services within their IT infrastructure; use digital identities and profiles for authentication and authorization purposes within an enterprise boundary or across multiple organisations. These solutions provide basic information lifecycle management functionalities – focused on managed digital identities and user profiles.

There is currently a “dichotomy” between *Identity Management* solutions and *Information Lifecycle Management* solutions due to: (a) the different nature of the managed information (identity information vs. more traditional documents/files/records); (b) different business requirements; (c) different information usage patterns (intensive operational usage vs. prioritised storage, off-line retrieval and consultation usage).

Both types of solutions handle *digital material* that might contain personal information. As such, this information must be managed according to privacy laws (e.g. HIPPA, COPPA, SOX, EU Data Protection Law, etc.) [Laur04], privacy guidelines [OECD80] and data subjects’ preferences. This requires keeping into account *privacy rights and permissions* (e.g. privacy-aware access to data based on consent and purpose) and *privacy obligations* (e.g. data retention, data deletion, data transformation and notifications). In particular the lifecycle of this information needs to be managed in a privacy-aware way, according to stated obligations and other privacy constraints. Enterprises need to address all these aspects for regulatory compliance, to satisfy customers’ expectations and to retain good reputation and brand. The increasing number of incidents - including identity thefts, misuses of personal data, data leakages, etc. - shows how complex and hard is to manage confidential documents and personal information in a privacy-compliant way: this involves knowledge of regulation, definition of privacy policies, implementation of good practices and processes, the deployment and usage of technologies to manage and enforce these policies and monitor/audit for their compliance.

In this context, enterprises need to enforce *privacy-aware access control* on stored personal and confidential information: data should be accessed or disclosed based on the enforcement of stated privacy policies. Progress has already been made in this space, as described in [IBM04a, IBM04b, CATB05]. In addition, *privacy-aware information lifecycle management* processes must be put in place by enterprises to effectively manage the lifecycle of personal and confidential information according to privacy requirements - over time and across various contexts and solutions. As anticipated, this includes dealing with data retention, data deletion, satisfying notice requirements, supporting data transformations and management of complex workflows.

This requires a *well-planned*, *systemic* and *ongoing* effort, because: privacy policies and personal preferences can change over time; data and confidential documents can be subject to different privacy and data protection laws depending on geographical and organisational boundaries; data needs to be disposed or transformed over time. The lifecycle of the involved privacy policies must be managed as well.

This paper focuses on the *privacy-aware information lifecycle management* aspect, as it is important for privacy management and still a green field. The dichotomy between “traditional” *Information Lifecycle Management* (ILM) solutions and *Identity Management* (IDM) solutions does not really help enterprises to address, in an integrated and common way, the lifecycle management of information driven by privacy policies. Today most of privacy management work involves human processes that are duplicated, prone to mistakes and subject to high operational costs.

This paper aims at setting the context for *privacy-aware information lifecycle management* within enterprises. It provides an overview of current enterprises’ *ILM* and *IDM* solutions and their limitations. It investigates and analyses core requirements and issues that need to be addressed by enterprises along with properties and features that should be provided by *privacy-aware information lifecycle management* solutions. Related implications for ILM and IDM solutions are discussed. This paper also describes research and work done by HP Labs to develop approaches and technologies that can help enterprises to implement and automate aspects of *privacy-aware information lifecycle management*.

2 Overview of ILM and IDM Solutions

2.1 Information Lifecycle Management Solutions

Information Lifecycle Management (ILM) is a comprehensive approach to manage the flow of an information system's data and associated *"metadata"* from creation and initial storage to the time when it becomes obsolete and is deleted [PETR06]. ILM involves various aspects of dealing with data, starting with user practices, rather than just automating storage procedures, as for example, done by hierarchical storage management (HSM) systems. ILM enables basic criteria for storage management based on data age and frequency of access and includes policy-driven management of data, e.g. [BDJK05]. At the very base, ILM solutions automate the processes of: (1) Organizing data into separate tiers according to specified policies; (2) Data migration from one tier to another based on those criteria.

Newer data, and frequently accessed data, is stored on faster, but more expensive storage media, while less critical data is stored on cheaper, but slower media. ILM solutions provide degrees of support for the following information/data management phases:

- **Assessment**: this is about understanding what data resides on the storage assets in an enterprise environment;
- **Data Analysis**: based on the outcome of the assessment phase, this phase is about analysing and explaining the breakdown of storage asset utilization, data usage patterns and the costs involved;
- **Classification**: depending on how data is used and how critical it is to the business, data is prioritised based on business requirements (mission critical, business sensitive, etc.) and its value determined. This defines where data should be stored through its lifecycle and assist in creating policies to migrate data to the proper storage "class" over time. These classes might keep into account different data properties, such as: type, organisation, value and age;
- **Automation**: once data has been classified, policies must be established to determine on which storage resources data should be located. Tools can automate the migration of data from one storage class to another based on these policies and deal with aspects such as replication, mirroring and back-ups. Disaster recovery and business continuance criteria are also considered in this context to ensure that mission-critical data is always available;
- **Review**: this consists on an ongoing activity of continuously reviewing the usage patterns of storage resources and ensure adherence to policies and procedures.

Privacy-aware information and data management have become increasingly important as businesses face regulatory compliance issues in the wake of privacy legislation. So far, only few privacy aspects (such as data retention/deletion) have been taken into account by ILM solutions. A more comprehensive list of requirements that need to be fulfilled by enterprises is described in the "*Requirements and Open Issues*" section.

2.2 Identity Management Solutions

Enterprise Identity Management (IDM) solutions deal with the management of digital identities, user accounts and user profiles and provide services to enterprise applications and services [CaBP03,DeRo04]. Specifically, they support functionalities such as authentication, SSO, authorization, auditing, user provisioning, data storage, link to legacy systems and data consolidation. They target different types of users and contexts including e-commerce, service providers, enterprises and government institutions.

The main components and functionalities provided by current identity management products and solutions include [DeRo04]:

- **Directory services, meta-directories, virtual directories and databases** deal with the representation, storage and management of identity and profiling information and provide standard APIs and protocols for their access.
- **Authentication, authorization and auditing** functionalities. Authentication ranges from local authentication on a system to complex distributed authentication, including single-sign-on (SSO) within and across organizational boundaries. Authorization can include simple access control management at the OS level, more sophisticated role-based access control - RBAC - up to flexible, distributed, policy-driven authorization, at the application and service levels.
- **Provisioning** components are used by enterprises, organizations and e-commerce sites to deal with the lifecycle management of identities, including the enrolment, customization, modification and destruction of accounts associated to users, employees and customers along with associated identity information (including rights, permissions and access control information). Related functionalities deal with the issuance, certification, management and revocation of digital entitlements and credentials in a secure and trusted way;
- **Self-Registration, Personalization** components provide core functionalities to end-users (i.e. data subjects) in terms of self-registration and management of their personal information and identities.

In particular, the *Provisioning* component, that handles lifecycle management aspects of digital identities and personal information, is usually not integrated with ILM solutions.

3 Privacy-Aware Information Lifecycle Management

Privacy-aware Information Lifecycle Management is the process of ensuring that personal and confidential data - stored and used by enterprises - are managed according to stated privacy policies and laws, people’s preferences and enterprise privacy guidelines. This section describes requirements and open issues; highlights core properties and features that should be provided by *Privacy-aware Information Lifecycle Management solutions*; describes our current work in this space and next steps.

3.1 Requirements and Open Issues

Privacy laws, such as HIPPA, COPPA, EU Data Protection Directives [Laur04] and privacy guidelines, such as OECD [OECD80], dictate key privacy requirements for enterprises that have direct implications for *Privacy-aware Information Lifecycle Management* processes:

- Enterprises should clearly state the purposes for which they collect personal data and should take into account the consent (or lack of consent) given by data subjects (people) to use their data for these purposes;
- People should be enabled to express their privacy preferences on how their personal data should be handled (e.g. consent, retention, notifications) and change them afterwards;

- People should be notified of changes affecting the management of their personal data and they should retain a degree of control over it;
- Personal data should be deleted once its retention is not required anymore;
- Openness and transparency over how data is processed, manipulated and disclosed to third parties are also key requirements;
- Compliance to all these aspects must be monitored and any violation promptly reported and addressed.

As anticipated in the introduction, privacy policies are commonly used to represent and describe these privacy laws and guidelines, in terms of *rights* of data subjects, *permissions* over usage of personal data and *obligations* to be fulfilled. In particular *obligations* [Casa04a,Casa04b] describe expectations and duties on how to handle personal data. They might dictate deletion/data retention constraints, notification requirements, data transformation criteria (encryption, minimisation, etc.) and complex workflow that need to be executed on this data, involving human and computer-based interactions. *Obligations* have direct implications on the "*lifecycle management*" of personal data within enterprises.

As a consequence, the constraints and conditions dictated by obligations (on how to manage personal data) must be kept into account by ILM and IDM solutions, to really enable enterprise-wide *privacy-aware information lifecycle management*.

In this context, *privacy policies*, inclusive of *privacy obligations*, need to be managed as well. Related requirements follow:

- **Lifecycle management of privacy policies:** Privacy policies must be understood, refined and authored by enterprises. Their lifecycle has to be managed;
- **Deployment and enforcement of privacy policies:** privacy policies need to be deployed within enterprises data management processes and IT infrastructures and enforced;
- **Auditing and monitoring of policy enforcement for compliance**.

Enterprises that span across different geographical and organisational boundaries might be subject to different privacy laws and privacy policies.

There are a few important, open issues that need to be addressed to enable effective *privacy-aware lifecycle management* of personal and sensitive information within enterprises:

1. **Lack of Automation**: current enterprise's privacy management practices are mainly based on manual processes, good behaviours and common sense. Not only are human processes prone to failure but the scale of the problem highlights the desire for additional technology to be part of the solution. The trend towards complexity and dynamism in system configurations heightens this need for automation to ensure that privacy and security properties are maintained as changes occur, and in addition to check that privacy is delivered as expected;
2. **Lack of Integration:** the duplication of data management efforts and capabilities – such as the ones provided by ILM and Identity Management solutions - do not help enterprises to deal with privacy matters. Even if in the short/medium term it is hard to envisage a convergence of these solutions into an integrated approach, progress should be made at least to avoid duplications of efforts in the management of privacy. Solutions that enable centralized management of privacy policies and *privacy-aware information lifecycle management* should be leveraged and integrated with both ILM and IDM solutions.

Next sections describe in more details core properties that should be provided by *privacy-aware information lifecycle management* solutions and introduce our R&D work done in this area.

3.2 Core Properties and Functionalities

Our analysis of requirements and open issues has identified a few core properties and functionalities that a *privacy-aware information lifecycle management solution* should provide:

- **Explicit modelling of personal and confidential data**: this solution should specify and use a model of stored personal and confidential data. This model, at least, should describe where this data is stored (database, LDAP repository, etc.), should provide data schema details, properties of data attributes, unique data identifiers, etc.;
- **Explicit definition of privacy policies, in particular obligations**: this solution should explicitly support the representation and authoring of privacy policies, in particular privacy obligations. A format to represent these policies should be defined and used;
- **Integrated lifecycle management of these policies**: this solution should support the overall lifecycle management (i.e. creation, authoring, versioning and disposal) of privacy policies, in particular of privacy obligations;
- **Deployment and enforcement of these policies, potentially by leveraging ILM and IDM infrastructures**: this solution should explicitly deploy privacy policies within relevant enterprise IT systems and enforce them. Part of the enforcement process can be delegated to ILM and IDM solutions (when feasible) to enable an integrated enforcement approach;
- **Integrated Monitoring and checking for compliance to these policies**: this solution should check that privacy policies are fulfilled over time and report any violation.

Current ILM and IDM solutions do not provide most of these functionalities: they basically support data deletion and aspects of data transformation. A *privacy-aware lifecycle management* solution must provide this core set of functionalities and be integrated with relevant enterprise systems/solution that already handle personal and confidential data, including ILM and IDM solutions.

Given the current dichotomy of IDM and ILM solutions, we believe that this *privacy-aware lifecycle management* solution must retain a key role in representing, authoring, managing, deploying and monitoring privacy policies, in particular privacy obligations. Another important role of this solution is also dealing with the lifecycle management of these policies.

3.3 Our Approach

HP Labs have been researching and workings in the space of *privacy-aware information lifecycle management* for the last two years. In this context, our work addresses the problem of automating the management and enforcement of privacy obligations in enterprises and enable *privacy-aware lifecycle management* of personal and confidential data.

Privacy obligations [Casa04a,Casa04b] dictate expectations and duties on how to handle personal and confidential data and deal with its lifecycle management, including: dealing with data deletion, data transformation (e.g. encryption), sending notifications, executing complex workflows, etc.

In our vision, at the core of *privacy-aware information lifecycle management* solutions there is an *Obligation Management Framework* to centralise (within enterprises) the representation and management of privacy obligations and orchestrate their overall enforcement and monitoring by leveraging and extending current enterprises IT solutions – specifically ILM and IDM solutions. This section describes our current approach, results and next steps.

An explicit model of privacy obligations is introduced along with mechanisms to handle privacy obligations and describe the affected data. In our work, we defined an *obligation management model* [Casa04a,Casa04b], where privacy obligations are "first class" entities, i.e. they are explicit entities that are represented and managed. In this model, a privacy obligation is an "object" that includes (at least) the following aspects [Casa04a,Casa04b]: *Obligation Identifier*; T*argeted Personal Data*; *Triggering Events* (e.g. time-based events); *Actions* (e.g. data deletion, sending notifications). Different categories of privacy obligation [Casa04a] need to be managed and enforced by enterprises: *transactional obligations*; *data retention and handling obligations*; *other types of event-driven obligations*. A complementary classification of our managed privacy obligations is based on their activation timeframe and period of validity: *short-term obligations*; *long-term obligations*; *ongoing obligations* [Casa04a].

A related *obligation management framework* [Casa04a,Casa04b] is also introduced to deal with the management of privacy obligations. In this framework data subjects (people, users) can explicitly define privacy preferences (e.g. on data deletion, notifications, etc.) on their personal data at the disclosure time (e.g. during a self-registration process) or at any subsequent time. These preferences are automatically turned into privacy obligations. Enterprise privacy administrators can further define other privacy obligations, for example dictated by laws or internal guidelines. This *obligation management framework* provides the following core functionalities (based on the properties and requirements previously described):

- **Scheduling the enforcement of privacy obligations**: it schedules which obligations need to be fulfilled and under which circumstances (events);
- **Enforcing privacy obligations**: it enforces privacy obligations once they are triggered. Enforcement may range from execution of simple actions (e.g. notifications, deletions) to complex workflows involving human intervention;
- **Monitoring fulfilment of privacy obligations**: it monitors and audits enforced obligations, at least for a predefined period of time, to ensure that the desired status of data is not changed and to report anomalies.

An *Obligation Management System* [Casa04a,Casa04b,CTCB05] has been derived from our privacy obligation model and obligation management framework. A working prototype has been fully implemented in the context of the EU PRIME project [PRIM06], as a proof of concept, providing the specified core functionalities: scheduling, enforcement and monitoring of privacy obligations.

The *Obligation Management System* can be leveraged to enable *privacy-aware information lifecycle management* within enterprises, by:

1. Providing centralized capabilities for representing, authoring, managing and monitoring privacy obligations;
2. Providing centralised modelling and abstraction of managed data;
3. Integrating the deployment, enforcement and monitoring (for compliance checking) of these obligations by interacting with existing IDM and ILM solutions.

Figure 1 illustrates the high-level architecture of this integrated solution.

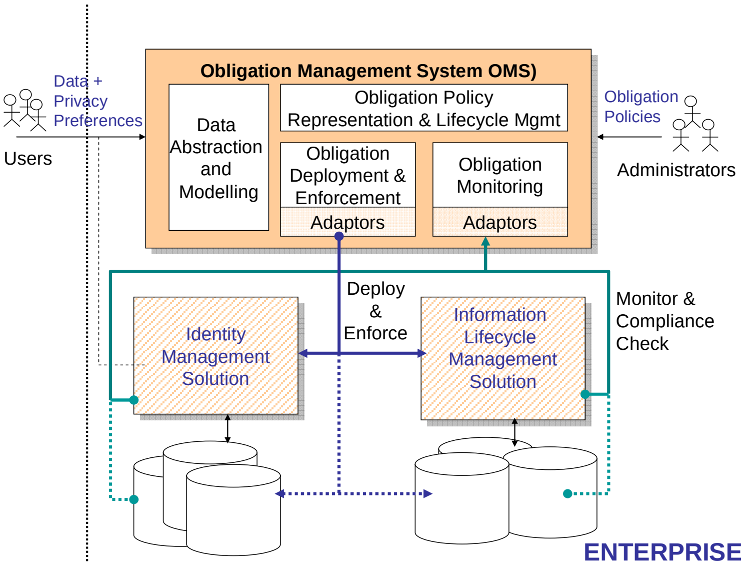
Obligation Management System OMS)
Data + Privacy Preferences
Users
Data Abstraction and Modelling
Obligation Policy Representation & Lifecycle Mgmt
Obligation Deployment & Enforcement
Adaptors
Obligation Monitoring
Adaptors
Obligation Policies
Administrators
Deploy & Enforce
Monitor & Compliance Check
Identity Management Solution
Information Lifecycle Management Solution
ENTERPRISE

Figure 1: High Level Architecture

The *Provisioning* and *Self-Registration* capabilities of current IDM solutions can be used to collect people's privacy preferences on how their personal data should be managed. These privacy preferences can be automatically translated by IDM *add-ons* into privacy obligations and managed by our *Obligation Management System* (OMS). Similarly, privacy administrators within the enterprise can define additional privacy obligations related to personal and confidential information stored and managed by ILM solutions (or any other solutions).

In this context the OMS becomes the *central point of control and orchestration* of *privacy-aware information lifecycle management* in enterprises. Thanks to software adaptors and the usage of various APIs, the OMS can configure IDM and ILM solutions to deal with the constraints dictated by privacy obligations (e.g. deletion preferences, etc.) and enforce them. If the IDM and ILM solutions provide no support for this, the OMS will directly enforce these privacy obligations by directly interacting with the data repositories.

To demonstrate the feasibility of our vision and how this can be achieved in a practical way, we have already integrated the prototype of our OMS system with a state-of-the-art enterprise Identity Management solution [HP05a] specialized in user provisioning and account management capabilities. More details about current results are available in [CTCB05]. We are also exploring integration of this prototype OMS system with a state-of-the-art ILM solution [HP05b] as another proof-of-concept activity. We can anticipate that this is feasible and we are currently working on the implementation details. Results will be published once a first related prototype is implemented.

We believe that an approach based on a centralised management of privacy obligations and its integration with current ILM and IDM solutions (and potentially other solutions handling sensitive data) constitutes a suitable reference model to enable *privacy-aware information lifecycle management* within enterprises. It can be deployed in a variety of contexts, including on enterprise systems with a highly distributed topology. Part of our coming work consists of making experiments that leverage and use a heterogeneous set of IDM and ILM solutions. We are keen in getting feedback and engaging with "lighthouse customers" for technological trials.

4 Conclusions

This paper sets the context for *privacy-aware information lifecycle management* within enterprises. Core privacy requirements and issues that must be addressed by enterprises have been analysed along with their implications and impact on existing ILM and IDM solutions. Key properties and functionalities of *privacy-aware information lifecycle management solutions* have been introduced.

Research and work done by HP Labs to develop approaches and technologies to automate aspects of *privacy-aware information lifecycle management* have been discussed. In this context, we described our current work on an *Obligation Management System* – a central point of control to author and manage privacy obligations, trigger their enforcement and check for their compliance – and how it can be leveraged and integrated with current ILM and IDM solutions. Further research and work is going to be done to refine our concepts and make experiments, both in the context of HP Labs and the EU PRIME project.

References

[BDJK05] Beigi, M., Devarakonda, M., Jain, R., Kaplan, M., Pease, D., Rubas, J., Sharma, U., Verma, A.: Policy-based information lifecycle management in a large-scale file system. Policies for Distributed Systems and Networks, 2005, Sixth IEEE International Workshop on, 6-8 June 2005, 2005

[CaBP03] Casassa Mont, M. Bramhall, P., Pato, J.: On Adaptive Identity Management: The Next Generation of Identity Management Technologies. HP Labs Technical Report, HPL-2003-149, 2003

[Casa04a] Casassa Mont, M.: Dealing with Privacy Obligations in Enterprises. HP Labs Technical Report, HPL-2004-109, 2004

[Casa04b] Casassa Mont, M.: Dealing with Privacy Obligations: Important Aspects and Technical Approaches. TrustBus 2004, 2004

[CaTB05] Casassa Mont, M., Thyne, R., Bramhall, P.: Privacy Enforcement with HP Select Access for Regulatory Compliance. HP Labs Technical Report, HPL-2005-10, 2005

[CTCB05] Casassa Mont, M., Thyne, R., Chan, K., Bramhall, P.: Extending HP Identity Management Solutions to Enforce Privacy Policies and Obligations for Regulatory Compliance by Enterprises. HP Labs Technical Report, HPL-2005-110, 2005

[DeRo04] De Clercq, J., Rouault, J.: An Introduction to Identity Management. HP Reports, http://devresource.hp.com/drc/resources/idmgt_intro/idmgt_intro.pdf, 2004

[HP05a] Hewlett-Packard (HP): HP OpenView Select Identity: Overview and Features. http://www.openview.hp.com/products/slctid/index.html, 2005

[HP05b] Hewlett-Packard (HP): RISS Software Development Kit. http://h18006.www1.hp.com/products/storageworks/riss/sdk.html, 2005

[IBM04a] IBM: The Enterprise Privacy Authorization Language (EPAL). EPAL 1.2 specification, http://www.zurich.ibm.com/security/enterprise-privacy/epal/, 2004

[IBM04b] IBM Tivoli Privacy Manager: Privacy manager main web page, http://www-306.ibm.com/software/tivoli/products/privacy-mgr-e-bus/, 2005

[Laur04] Laurant, C.: Privacy International: Privacy and Human Rights 2004: an International Survey of Privacy Laws and Developments. Electronic Privacy Information Center (EPIC), Privacy International, http://www.privacyinternational.org/survey/phr2004/, 2004

[OECD80] OECD: OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. http://www1.oecd.org/publications/e-book/9302011E.PDF, 1980

[PRIM06] PRIME Project: Privacy and Identity Management for Europe. European RTD Integrated Project under the FP6/IST Programme, http://www.prime-project.eu/, 2006

[PETR06] Petrocelli, T.: Data Protection and Information Lifecycle Management. Prentice Hall, Chapter 8, 2006

Keywords

Privacy, Information Lifecycle Management, Identity Management, Obligations, Policy Enforcement, Policy Management, Regulatory Compliance