Enforcing Data Protection Regulations within e-Government Master Data Management Systems

Federico Piedrabuena, Laura González and Raúl Ruggia
Instituto de Computación, Facultad de Ingeniería, Universidad de la República,
J. Herrera y Reissig 565, Montevideo, Uruguay

Keywords: Master Data Management, e-Government, Data Protection, Information as a Service.

Abstract: The growing adoption of information technology by governments has led to the implementation of e-Government systems which are usually supported by middleware-based integration platforms. In particular, the increasing need of information sharing across government agencies has motivated the implementation of shared Master Data Management (MDM) Systems. On the other hand, these systems have to comply with Data Protection regulations which may hinder an extensive reuse of information in a government context. This paper addresses the issues of enforcing Data Protection (DP) regulations in e-Government MDM systems. In particular, it analyzes the requirements that DP issues pose on these systems and it proposes solutions, which leverage middleware-based capabilities and traditional MDM systems, to enforce these regulations considering different MDM architecture styles.

1 INTRODUCTION

E-Government systems have been increasingly implemented to improve the quality of public services by providing relevant information and self-services to citizens as well as by enabling a more effective inter-organizational coordination among public agencies and partners (Akeroyd, 2009). These systems are usually based on middleware platforms providing capabilities to interconnect agencies (Baldoni et al., 2008)(González et al., 2012).

Information sharing is at the root of e-Government as it is required to achieve a consistent systems' interoperability as well as to promote efficiency by reducing duplication of effort in collecting and storing information (Akeroyd, 2009). This has motivated the inclusion of an "Information dimension" in e-Government architectures. Furthermore, ensuring the quality of shared data and the application of rigorous management practices have motivated the adoption of Master Data Management (MDM) as a key component in e-Government (Boydens, 2011).

Briefly, MDM systems consist of Information Systems on core business data enriched with tools and management practices. In this way, MDM systems provide an integration and quality assurance tier between business components implementing public services and data distributed in organizations connected through the e-Government system (Dreibelbis et al., 2008).

In turn, a major challenge to inter-organizational data sharing concerns the application of Data Protection (DP) and other privacy-related regulations which are nowadays adopted by a large number of countries (Del Villar et al., 2001). This fact may hinder an extensive reuse of information in a government context, as shown in some empirical studies (Lips et al., 2011), because they oblige to perform a number of validations about the origins, purpose and existing consents prior to share data.

This paper addresses the issues of enforcing Data Protection regulations within e-Government MDM systems by analyzing the requirements that these issues pose and proposing enforcement solutions considering the different MDM architecture styles. The proposed solutions leverage capabilities of middleware-based integration platforms (e.g. data transformation) and traditional MDM systems as well as broadly recognized standards (e.g. XACML).

The rest of the paper is organized as follows. Section 2 presents background. Section 3 analyses the requirements for a Data Protection-aware e-Government MDM system. Section 4 proposes a solution approach to address these requirements.

Piedrabuena F., González L. and Ruggia R..
Enforcing Data Protection Regulations within e-Government Master Data Management Systems.
DOI: 10.5220/0005458003160321
In *Proceedings of the 17th International Conference on Enterprise Information Systems* (ICEIS-2015), pages 316-321
ISBN: 978-989-758-098-7

Finally, Section 5 presents conclusions.

2 BACKGROUND

2.1 Master Data Management

Master Data Management (MDM) delivers a consolidated, complete and accurate view of business critical Master Data (MD) in an organization or business area. MD are compose by concepts and attributes such as "person" with "name" and "birthday" or "product" with "color" or "size". Given the strategic nature of MD, quality and reliability must be ensured. Therefore, well defined Data Management and Governance practices (Mosley et al., 2010) are established to MD and constitute the roots of MDM.

Functions to manage MD lifecycle are implemented in MDM Systems (MDMS). Design and maintenance of the MDMS architecture is one of the main MDM activities. This architecture controls the shared access, data maintenance, replication and flow of data with the goal of ensuring their quality (Mosley et al., 2010). Several architecture styles for MDMS are identified and described throughout the literature with different names and characteristics (Dreibelbis et al., 2008) (Loshin, 2010) (Galhardas et al., 2010) (Baghi et al., 2014)(Otto, 2012). This work summarizes these styles in two approaches: Repository and Registry.

The Repository style involves storing the entire data set (i.e. all MD attributes used in any and all software applications) in a single database. Data requests from clients are resolved by a central MDMS without interacting with source systems (application software data provider).This style has three variants depending on where MD is maintained (i.e. created, updated and deleted): i) Consolidation; ii) Centralized; and iii) Coexistence. In the consolidation variant all MD maintenance activities are performed by local source systems and data changes flow from these systems to the MDMS. In the centralized variant all MD maintenance activities are performed by MDMS and data changes flow from the MDMS to target systems. Finally, the coexistence variant is a combination of the previous ones and data changes flow bidirectional.

The Registry style consists of only storing references to actual MD attributes. Data requests from clients are resolved by the MDMS by interacting with the systems to retrieve MD.

These two architectural styles can also be combined to implement hybrid solutions where, for example, a portion of the MD is consolidated in the MDMS and sensitive data is kept in source systems.

Three main stages can be identified in MD lifecycle: i) maintenance (i.e. MD creation, modification or deletion), ii) synchronization (i.e. propagation of MD changes within the MDMS), and iii) request (i.e. clients asks for MD to the MDMS).

2.2 E-Government Platforms

Integration Platforms have become a key tool to support the development of e-Government in many countries. They usually provide infrastructure and capabilities aiming at facilitate the interconnection between agencies as well as provide common services to generate economy of scale and encourage the development of multi-agency services (González et al., 2012). Common capabilities are connectivity, security (e.g. authentication), interoperability (e.g. through the use of standards) and mediation services implementing Enterprise Integration Patterns (EIP), e.g. data transformation (Hohpe and Woolf, 2003). The mediation and interoperability capabilities are usually provided by traditional middleware technologies, such as SOAP Web Services and Enterprise Service Bus (ESB). In addition, security capabilities usually rely on well-stablished standards such as XACML (OASIS, 2013).

2.3 Data Protection Regulations

Personal data handled by governments are often very sensitive (Wu, 2014). Indeed, their analysis can be highly invasive when data is combined and aggregated. As a result, most governments have developed some sort of legislation focusing on Data Protection (DP) with rather different approaches (Akeroyd, 2009)(Wu, 2014). They mainly deal with the re-use of information in other contexts for which it was provided. Particularly, in many countries (González et al., 2012) citizens have to provide explicit consents which allow agencies to use / share their information. As shown in Figure 1, in this kind of laws citizens provide to agencies the consents to

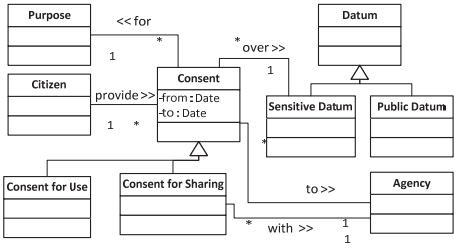

Figure 1: Common concepts in DP Regulations.

use or share, with other agencies, sensitive data for a given purpose and within a given time period.

2.4 Relevant Standards

ISO/IEC 8000-1X0:2009 is a family of standards which deals with the requirement of using a pre-established format for communicating MD, as well as their quality level, between applications (Ismael Caballero et al., 2013). In particular, the ISO 8000-120:2009 part (ISO/IEC, 2009) specifies how information is included in the exchanged messages to describe the MD provenance. For example, attributes are defined to specify the event type performed over data, the date in which the event was performed, the organization which provides the MD and the organization that owns the MD, among others (Ismael Caballero et al., 2013).

XACML (OASIS, 2013) is a specification that describes a language for defining access control policies as well as a language to request and response access control decisions in XML. Usually, a requester tries to run an action on a resource by sending a request to the component that protects it: Policy Enforcement Point (PEP). The PEP performs an authorization request based on attributes of the requester, the resource, the action to be executed and any other relevant information. This request is sent to a Policy Decision Point (PDP) which issues a response indicating if access should be allowed based on the request and policies managed through a Policy Administration Point (PAP). Also, a Policy Information Point (PIP) may be used if additional information for taking the authorization decision is needed. Based on the authorization response, the PEP allows or denies access to the requester.

2.5 Related Work

Some authors have addressed the potentials and challenges of MDM in e-Government and large-scale collaborative systems (Akeroyd, 2009)(Yang and Maxwell, 2011). Also, enforcing DP regulations in cross-organizational data exchanges has been addressed in (Armellin et al., 2010) (Sillaber and Breu, 2012)(Stevovic et al., 2013). Still, to the best of our knowledge, there are neither proposals that jointly address the issues of DP and e-Government MDM Systems nor solutions that leverage middleware-based capabilities to solve them. Finally, although some MDM products deal with data privacy, they focus on solutions of lower level of abstraction (e.g. data encryption) compared to our approach (e.g. handling citizen consents).

3 REQUIREMENTS FOR ENFORCING DP

This section establishes the e-Government MDM context over which this work is based and analyses the requirements for enforcing DP regulations in it.

3.1 E-Government MDM Context

E-Government systems usually have the following characteristics: i) inter-organizational interactions; ii) need to share data of citizens; iii) agreements on data structure of shared information; iv) regulatory bodies (e.g. supervisory agency); and v) use of integration platforms. Figure 2 shows the MDM e-Government context taken as a base for this work.

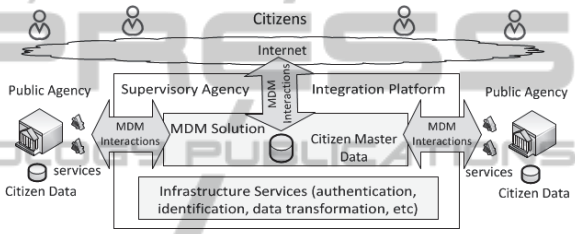

Figure 2: General MDM e-Government Context.

All interactions between public agencies pass through an Integration Platform (InP) maintained by a supervisory agency. This platform provides basic mediation capabilities (e.g. data transformation) and security services (e.g. authentication). It is also responsible for hosting the MDM system which, in particular, handles citizens' data.

MDM interactions that pass through the InP take place when MD is maintained (if this is done at the InP, e.g. with a centralized repository style), synchronized (between the InP and public agencies) or requested (e.g. by other agencies or partners).

Public agencies share citizens' data using a state-wide agreed data model, which is a superset of the MD schema, and following the IaaS approach (i.e. via services). In particular, the MDM system provides / consumes information services, sending messages (e.g. SOAP messages) over the platform, to get / send citizens MD attributes. These messages are compliant with the family of standards ISO/IEC 8000-1X0:2009 (e.g. they include provenance data).

Figure 3 presents a citizens' MD schema, taken as a case study in the rest of the paper, and the structure of messages sent and received through InP to share data. Note that besides citizens' data, the message includes information of its origin, its destination, the sender and provenance data.

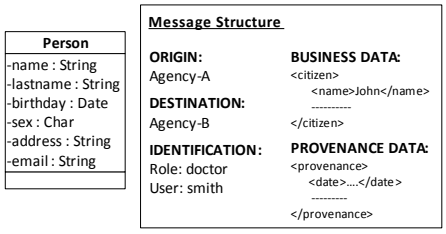

Figure 3: Master Data Schema and Structure of Messages.

3.2 Analysis of Requirements

A solution that allows a supervisory agency to guarantee that all MDM interactions comply with DP regulations in the presented context has the following four general requirements:

RQ1: It has to monitor and enforce DP regulations through components hosted in the InP. This is because: i) the supervisory agency needs full control over the enforcing mechanisms as well as the penalty actions to be taken if a regulation is not satisfied, and ii) some public agencies may not have the infrastructure to host the required components.

RQ2: It has to monitor and enforce all MDM interactions performed within the InP (i.e. when maintaining, synchronizing and requesting MD).

RQ3: It has to allow maintaining the category (i.e. sensitive or public) of each MD attribute, the consents provided by citizens and current DP policies. This is required to know if a given interaction is compliant with the regulations.

RQ4: It has to be able to handle and obtain other required information to verify the compliance of the interactions. This includes: i) information required in the requests (e.g. purpose for which the MD is requested); and ii) information not included in the MD schema (e.g. nationality) that may be required to verify the compliance of the interactions.

4 PROPOSED APPROACH

This section describes a high level architecture of the proposed approach, the process for designing an extended MD and how DP regulations are enforced in the different MDM interactions.

4.1 High Level Architecture

Figure 4 presents a high level architecture of the proposed solution which includes several specialized components hosted in the InP.

First, the Consent Management System (CMS) is in charge of maintaining citizens' consents

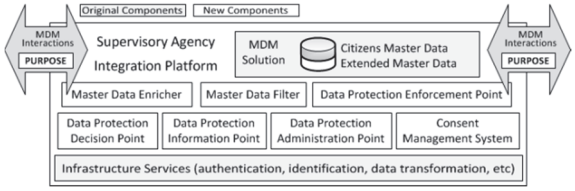

Figure 4: High Level Architecture of the Proposal.

according to the conceptual model of Figure 1. Ideally, it may provide a web-based user interface so citizens can manage their own consents.

Second, various components following the XACML standard are included. The Data Protection Administration Point (DPAP) has the responsibility of managing the DP policies and the category of each MD attribute. The Data Protection Information Point (DPIP) has the function of obtaining additional information required to take an authorization decision; this information may be part of the MD or may have to be obtained from external systems. The Data Protection Decision Point (DPDP) is in charge of taking an authorization decision based on the data included on the requests, current policies and, eventually, additional information obtained from the DPIP. Lastly, the Data Protection Enforcement Point (DPEP) is responsible for enforcing regulations by: i) intercepting all MDM interactions; ii) requesting the DPDP an authorization decision; and iii) acting accordingly (e.g. blocking an interaction).

Third, two specialized components based on well-stablished EIP are also included. The Master Data Enricher (MDE), based on the content-enricher EIP, is in charge of complementing the MD (e.g. at synchronization time) with information required to take the authorization decision. This information, may consist of additional MD attributes (e.g. nationality) or metadata (e.g. provenance). In turn, the Master Data Filter (MDF), based on the content-filter EIP, has the function of taking out MD attributes from a MDM interaction in case they are not allowed to be used or shared. The DPEP may decide to route an interaction to this component based on the response from the DPDP. In addition, the DPDP may decide to route an interaction to other EIP-based components (not included in Figure 4), for example, to log or notify interested parties.

The Extended MD is MD augmented with additional data required to take authorization decisions. This may be new MD concepts or attributes (e.g. nationality) as well as metadata (e.g. provenance). Other required information to take these decisions may not be included in the Extended MD due to it comes from sources outside the MDMS or it is not citizen information.

Lastly, the solution requires MDM interactions to include the purpose for which MD are shared or requested. This can be done with the elements defined in the XACML Privacy Policy Profile.

Note that all components are hosted in the InP which fulfills RQ1. RQ2 is mainly achieved by the DPEP and MDF components and RQ3 is satisfied by the CMS and DPAP components. Lastly, RQ4 is achieved by the DPIP and MDE components.

4.2 Extended MD Design

The process to extend MD takes as input the original MD and DP policies and consists of the following steps: 1) identify required data to take authorization decisions and indicate if they should be included in the MD or not, 2) identify sensitive data in the original MD and in the additional data identified in the previous step, and 3) include provenance attributes for each sensitive attribute.

Figure 5 presents the result of applying the above steps to the example presented in Figure 3. In this example nationality is an additional required attribute and is added to the original MD.

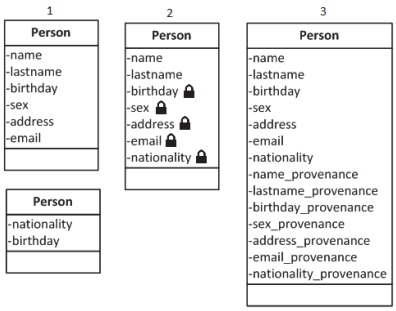

Figure 5: Extended MD Design Steps.

4.3 Enforcing DP in MDM Interactions

This section describes how the proposed solution enforces DP regulations within MDM interactions (maintenance, synchronization and requests) in view of the different architecture styles.

MD maintenance interactions pass through the InP when the coexistence or centralized variants of the repository approach are used. In these cases, the interactions have to include a purpose with the value "MDM-maintain". The DPEP has to allow these interactions if citizens informed consents for using all MD attributes involved in the operation, to the supervisory agency and for the purpose of maintaining (MDM-maintain). Also, if the operation is "insert" or "update", the interaction is routed to the MDE to perform MD augmentation. In these cases, obtaining provenance data is trivial, since MD is maintained at the supervisory agency.

MD synchronization interactions pass through the InP when the repository approach is used. In these cases, the interactions have to include a purpose with the value "MDM-synchronize". The DPEP has to allow these interactions if citizens informed consents for sharing all MD attributes involved in the operation, from the source agency to the target agency (supervisory or others) and for the purpose of synchronizing (MDM-synchronize). Also, if MD is being inserted or updated the interaction is routed to the MDE to perform MD augmentation. In these cases, provenance data is obtained from the messages, given that they are compliant with the standard ISO 8000-120:2009.

MD requests pass through the InP when using both approaches: Registry and Repository. In these cases, the purpose included in the interactions is business-oriented (e.g. it refers to a specific e-Government procedure). The DPEP has to allow these interactions if citizens informed consents for sharing all MD attributes involved in the operation, from their source agency to the requesting agency (client) and for the purpose specified in the request. The source agency of a given attribute is obtained from the provenance data (already stored in the InP if a repository style is used or coming in the MDM response messages if a registry style is used).

In all the interactions, if citizens only provide consents for a subset of the involved MD attributes, the interactions are routed to the MDF to take out the attributes which are not allowed to be used or shared. Also, some attributes may be encrypted if they cannot be shared with the supervisory agency but they can be shared with other agencies.

Finally, other DP-related issues affecting MDM interactions are currently being analyzed. Some of them include: i) DP requirements on the additional information required to take authorization decisions, and ii) variants of different synchronization styles (e.g. notification, request-response).

5 CONCLUSIONS

This paper analyzes requirements to enforce Data Protection in e-Government-based Master Data systems and depicts solutions to guarantee that Master Data interactions comply with these regulations. The overall goal is to carry out efficient DP enforcement in inter-organizational contexts using shared MDMS and e-Government platforms.

To achieve this, the proposed approaches point at

using the e-Government platform as a common component to control DP compliance as well as to enforce regulations by transforming data and operation flows. Such solutions leverage mechanisms of mainstream middleware technologies and of MDM systems in addition to apply recognized standards (e.g. XACML, ISO/IEC 8000).

The main contributions of this ongoing work are: i) identifying requirements of a DP aware MDM e-Government system, ii) specifying a reference architecture to enforce DP in these systems, and iii) proposing enforcement mechanisms (e.g. MD filter). Moreover, this work constitutes a step forward on addressing the issues of regulatory compliance in e-Government systems (González and Ruggia, 2014).

ACKNOWLEDGEMENTS

This work was partially funded by the Comisión Sectorial de Investigación Científica (CSIC), Universidad de la República, Uruguay.

REFERENCES

Akeroyd, J., 2009. Information Architecture and e-Government. INFuture2009"Digital Resour. Knowl. Shar.

Armellin, G., Betti, D., Casati, F., Chiasera, A., Martinez, G., Stevovic, J., 2010. Privacy Preserving Event Driven Integration for Interoperating Social and Health Systems, in: Secure Data Management, Lecture Notes in Computer Science. Springer, pp. 54–69.

Baghi, E., Schlosser, S., Ebner, V., Otto, B., Oesterle, H., 2014. Toward a Decision Model for Master Data Application Architecture, in: 2014 47th Hawaii International Conference on System Sciences (HICSS). pp. 3827–3836.

Baldoni, R., Fuligni, S., Mecella, M., Tortorelli, F., 2008. The Italian e-Government Enterprise Architecture: A Comprehensive Introduction with Focus on the SLA Issue, in: Service Availability, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 1–12.

Boydens, I., 2011. Strategic Issues Relating to Data Quality for E-Government: Learning from an Approach Adopted in Belgium, in: Assar, S., Boughzala, I., Boydens, I. (Eds.), Practical Studies in E-Government. Springer New York, pp. 113–130.

Del Villar, R., de Leon, A.D., Hubert, J.G., 2001. Regulation of Personal Data Protection and of Reporting Agencies: a Comparison of Selected Countries of LA, the US and EU Countries. MIT Press.

Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., Run, P. van, Wolfson, D., 2008. Enterprise Master Data Management: An SOA Approach to Managing Core Information. Pearson.

Galhardas, H., Torres, L., Damásio, J., 2010. Master data management: a proof of concept, in: International Conference on Information Quality.

González, L., Ruggia, R., 2014. Towards a Compliance-Aware Inter-organizational Service Integration Platform, in: On the Move 2014 Workshops, LNCS. Springer Berlin Heidelberg.

González, L., Ruggia, R., Abin, J., Llambías, G., Sosa, R., Rienzi, B., Bello, D., Alvarez, F., 2012. A Service-oriented Integration Platform to Support a Joined-up E-government Approach: The Uruguayan Experience, in: Joint International Conference on Electronic Government, the Information Systems Perspective, and Electronic Democracy. Austria.

Hohpe, G., Woolf, B., 2003. Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions. Addison-Wesley.

Ismael Caballero, Isabel Bermejo, Luisa Parody, María Teresa Gómez López, Rafael Gasca, Mario Piattini, 2013. I8k: an Implementation of ISO 8000-1x0. Presented at the 10th International Conference on Information Quality, USA.

ISO/IEC, 2009. ISO/DIS 8000-120. Master data: Exchange of characteristic data: Provenance.

Lips, A.M.B., O'Neill, R.R., Eppel, E.A., 2011. Cross-agency collaboration in New Zealand: An empirical study of information sharing practices, enablers and barriers in managing for shared social outcomes. Int. J. Public Adm. 34, 255–266.

Loshin, D., 2010. Master Data Management. Morgan Kaufmann.

Mosley, M., Brackett, M.H., Earley, S., Henderson, D., 2010. The DAMA Guide to the Data Management Body of Knowledge: (DAMA-DMBOK Guide). Technics Publications, LLC.

OASIS, 2013. eXtensible Access Control Markup Language (XACML) v3.

Otto, B., 2012. How to design the master data architecture: Findings from a case study at Bosch. Int. J. Inf. Manag. 32, 337–346.

Sillaber, C., Breu, R., 2012. Managing legal compliance through security requirements across service provider chains: A case study on the German Federal Data Protection Act., in: GI-Jahrestagung.

Stevovic, J., Casati, F., Farraj, B., Li, J., Motahari-Nezhad, H.R., Armellin, G., 2013. Compliance aware cross-organization medical record sharing, in: 2013 IFIP/IEEE International Symposium on Integrated Network Management.

Wu, Y., 2014. Protecting personal data in E-government: A cross-country study. Gov. Inf. Q. 31.

Yang, T.-M., Maxwell, T.A., 2011. Information-sharing in public organizations: A literature review of interpersonal, intra-organizational and inter-organizational success factors. Gov. Inf. Q. 28.