Edvard Tijan, univ. mag. ing.
University of Rijeka
Faculty of Maritime Studies
Studentska 2
51000 Rijeka
Croatia

Review article
UDK: 004.773
656.615
Received: 18th October 2009
Accepted: 10th November 2009

DATA CASSIFICATION AND INFORMATION LIFECYCLE MANAGEMENT IN PORT COMMUNITY SYSTEMS

The purpose of this paper is to research the role of the Data Classification (DC) and Information Lifecycle Management (ILM) processes in Port Community Systems (PCS). Issues that are likely to be encountered in the implementation of such systems are defined by the complexity of the PCS stakeholders' own ICT and organizational systems and are therefore clearly outlined. Furthermore, several data classification models are proposed, which are derived from the standard methodology of information systems security. The paper identifies possible key issues during the DC/ILM system implementation and finally proposes suggested implementation paths in order to overcome operational difficulties.

Key words: *Port Community Systems, Data Classification, Information Lifecycle Management, data retention and disposal*

1. INTRODUCTION TO PORT COMMUNITY SYSTEMS

Port Community Systems (PCS) are complex logical and organizational systems built in order to integrate and coordinate the execution of business activities in large ports and their surroundings [5]. Their complexity is derived from organizational requirements oriented towards the promotion of efficiency, ensuring the flawless transport of goods through port systems, compliance with local legislation and providing stability and information flow security and business continuity while each entity included in the PCS maintains its own autonomy and self-sufficiency [4], yet providing the overall system with data requested for proper and timely inclusion of other entities in the business process execution. The PCS's are usually governed by one entity that coordinates

and concentrates efforts and investments trying to provide a superior service to all stakeholders, internal and external [10, 7-30].

Critical points in the creation of efficient PCS are a proper selection of hardware and frontend application software, database system and public key infrastructure (PKI), creation of information security framework, inclusion of all stakeholders in decision making and dedication of PCS to the achievement of goals of excellence. However, at operative level it is easy to overlook the importance of the Data Classification (DC) and Information Lifecycle Management (ILM) during the inclusion and integration of the respective ICT systems into unique PCS. Their synergy enables the data consolidation and provides a solid base for the implementation of best practices in data management.

2. DATA CLASSIFICATION AND INFORMATION LIFECYCLE MANAGEMENT FEATURES

The DC and ILM are two mutually intertwined activities. The ILM is a sustainable storage strategy suited to balance the costs of information storage and governance with its business value that changes as the process is being developed and implemented [8]. Therefore, ILM provides a practical methodology to align the costs of information retention and storage with real business priorities. The DC paradigm is a process that defines the access, recovery and discovery characteristics of an enterprise's (in this case, PCS's) different sets of data, grouping them into logical categories to facilitate business objectives.

Once the data are properly classified, the rules for data management can be applied selectively to the different data/information categories. The rules for the DC do not differ significantly from the rules for object classification in a classic network domain where objects or users are grouped and then a specific set of rules is applied to that group. The major goal of classification is to customize the approach to data management by grouping the data into classes that have similar traits and therefore similar governance requirements.

The DC tasks are generally very demanding and complex due to several contributing factors. Primarily, the introduction of DC refers not only to the already existing but also to new information that will enter the PCS (after PCS is fully deployed). It is much easier to merge new data to the already existing DC system than to apply the DC to information that already exist. The reason is very simple: new information can be easily classified and stored according to the set framework, while information that already exist are often stored in such a way that do not allow easy manipulation (e.g. diverse database structure, diverse applications used, diverse hardware/application layer, diverse owners, etc).

Due to regulatory and legislative demands, the PCS often store data in a non-structured way; therefore it is a challenge to properly classify it. The DC may refer to the PCS contained information regardless of the storage form it assumes in the final deliverable: it can refer to data in a paper form, in digital form stored on the user's workstations, centralized servers, transaction systems used throughout the PCS and other databases. It can also refer to e-mails, data contained on palmtop devices etc. So, the PCS should consider the application of the DC onto the already archived data in paper and electronic form and remote data [9].

It is of paramount importance to set up the DC procedures in a way which will follow the PCS business processes. The DC procedures must be closely suited to the adopted organization form, needs, goals and certified quality systems. Finally, well introduced system of DC should be able to provide PCS the competitive edge in knowledge management.

Moreover, DC enables a coherent ILM; without DC consistent ILM would not be possible. ILM enables cost effectiveness in the physical layer, namely it optimizes investments and operative expenses of hardware used to store and process information, it enhances information security and supports business goals. ILM is therefore a strategy adopted to counterbalance the cost of storing and managing information with its current business value, determined by business processes and data maintenance policy.

3. DATA CLASSIFICATION MODELS

The DC taxonomy is a basic process that consists of giving stored information (or information currently in transit) the value attributed. One often used classification model was devised by Bell and LaPadula – it relies on classic concepts of integrity, availability and confidentiality [6, 2]. Other different models exist also, usually as a part of information security science, that could be looked into and adopted in order to utilize them in a standalone manner, or as a blend during the development of the PCS information classification [2,249].

3.1. Graham-Denning model

The Graham-Denning Model [12] could be used to ensure secure creation and deletion of objects that will consequently enter into the DC scheme. Through this model, adequate addressing of specific rights assignment is also achieved. The usage of this system is beneficial in a case when access to resources should be controlled across distributed systems, which is usually the case in a typical PCS scenario. The usage of the Graham-Denning model describes and implements a definition of a basic right that outlines how authorized subjects

can execute security functions on an object. The model has eight basic protection rules (actions) that describe:

- How to securely create an object or subject.
- How to securely delete an object or subject.
- How to securely provide the read access right of subject on object.
- How to securely provide the grant access right to subject on object.
- How to securely provide the delete access right of subject on object.
- How to securely provide the transfer access right to subject on object.

To ensure consistency, every object has an owner who has special rights regarding the object, and every subject has another subject (top-level controller) with special rights regarding the subject. It is quite customary to set up Document Management Systems (DMS) using the Graham-Denning model.

3.2. Discretionary access control

The discretionary access control (DAC) [11] is defined as a way to limit access to a certain object based on the identity of the subjects and groups to which they belong. The controls used are discretionary in the sense that a subject with certain access permission may pass that permission to any other subject. The exception occurs if mandatory access control denies such a right.

The discretionary access control is usually an opposed concept to the mandatory access control (or non-discretionary access control). Sometimes, the term "discretionary" access control is used to emphasize that a certain information system is not fully compliant with the rules of the mandatory access control. However, mandatory and discretionary access controls can be applied and transferred within the same system. In such a system, the discretionary access control refers to those controls that subjects can transfer among them, while the mandatory access control refers to another category of access controls that imposes limitations upon the discretionary set of rules.

3.3. Mandatory access control

The mandatory access control [13] is a type of access control in which the document management, operating or application system constrains the ability of a subject (or initiator) to access or generally perform some sort of operation on a given object or target. In broader sense, the subject is usually a process or thread; objects can be technical constructs such as data files, directories, databases or transaction systems, while in the scope of this paper, it is in fact any information element contained within the PCS. Subjects and objects each have an assigned set of security attributes. When a subject attempts to access a specified object, embedded authorization rule enforced by the system examines these security attributes and decides whether the access will be granted. Any operation by any subject on any object will be tested against the set of

authorization rules to determine if the operation is allowed. This set of rules is usually called a policy.

3.4. Clark-Wilson model

The Clark-Wilson integrity model [16] provides a basis to specify and analyze an integrity policy of a database or data processing system. This model deals with the formalization of information integrity which is maintained by preventing occurrences of data corruption due to any cause (usually system or user error or intentional corruption). The integrity policy used in this model details how the data items in the system should be kept in order to maintain its validity during transition within the system and specifies the capabilities of various transaction elements. This model's enforcement and certification rules define data items and processes that are the sole basis for the creation of an integrity policy. Therefore, the core of this model is based on the notion of a transaction – the very element so generously used throughout the PCS's system implementation. Addressing the transaction integrity, this model uses a series of operations that change the system state from one to another, common characteristics of both being their consistency. In this model, the transaction certifier and implementer are usually different entities, thus ensuring the separation of duties.

3.5. Multilevel security

The multilevel security model [14] is represented by the application of a computer system to process information with different sensitivities at different security levels, allowing simultaneous access by users with different security permissions and consequently preventing users from obtaining access to information for which they lack appropriate authorization. It allows an easy access to less-sensitive information by individuals with higher levels of clearance, and it allows higher-cleared individuals to easily share documents with less-cleared individuals who are authorized to access them. Furthermore, this model can be used to create sanitized documents and information that does not contain the information which less-cleared individuals do not have the permission to access.

3.6. Biba integrity model

The Biba Integrity Model [15] was developed by Kenneth J. Biba in 1977. It is a formal state transition system of the computer security policy that describes a set of access control rules designed to ensure data integrity. Data and subjects are grouped into hierarchical levels of integrity. The subject may not change data in a level ranked higher than itself, or be corrupted by data from a lower level than the subject.

Generally speaking, the preservation of data integrity aims to prevent unauthorized data modification by authorized or unauthorized parties and to maintain internal and external data consistency. This security model is directed towards data integrity rather than confidentiality, in contrast to the Bell-LaPadula model which is oriented more towards confidentiality and availability. In the Biba model, the users can only change information at or below their own integrity level. The users can only view the content which is at or above their own integrity level.

It is clear that all of the PCS data does not have the same importance to the PCS core business and functions. Some of the data are mission critical; some of them are of limited importance, limited duration or are transient in nature. As already outlined, a process to align the value of data with the cost of storing and managing should be set up in place. This process starts with a clear understanding of the business uses of the PCS data and provides a mechanism for its storage and retention according to the requirements established by business priorities. Optimizing the relationship between the cost of storing and managing data and the delivery of service levels for data access, recovery and discovery is the objective of implementing the ILM strategy. The foundation of an ILM implementation is therefore the taxonomy established to classify data.

The derived conclusion is that the DC initiative must be (as all information related activities should be) sponsored from the very top of the organization and the ICT function [1,82]. Also, those who are in the position to contribute should be involved immediately, at the very beginning. Considering the liaison and relation in the PCS between the PCS governing body and the stakeholders, it should probably be the very top of the PCS governing body management.

- The suggested ILM deployment process has several key steps:
- Classification and/or categorization of data
- Relating PCS business rules and functions to data classes
- Determining service levels including data retention/disposal, access, recovery, discovery and cost objectives
- Establishing tiered services by developing a horizontal view across all existing service level agreements to establish a “catalogue” of standard service offerings. These (for example) might include an offering that provides 100 percent access/availability, sub-second recovery, multi-year retention with cross-enterprise, one-day discovery on any topic at a fixed service cost etc.
- Selection of appropriate products, including management tools and infrastructure, suited to particular PCS - once standard services are in place, DC and rules development is completed, establishing of the underlying infrastructure and management processes to meet service level goals is the next logical step

The existing state of the DC in the PCS is usually a byproduct of hierarchical storage management implementations, where the classification criterion is the age of the data, adopted from the existing systems of the PCS participants and stakeholders.

The data are generally classified based on access or availability and recovery requirements, and incurred cost. Therefore, additional compliance requirements for the PCS contained data, their retention, discovery and requirements for changeability are changed and must be taken into consideration. The future methods of the DC will involve a much broader perspective, determined by business requirements. This delimitation should serve several needs (including ILM, PCS content management, PCS Intranet, PCS Extranet portal and connectivity, business to business – B2B, legal compliance, data mining and top management decision support, security requirements, etc) and should emerge as the basis of the enterprise information asset management.

The top-level goal of the DC for the ILM is assigning business value to different categories of data. This is best accomplished in a highly developed IT environment where service level agreements have been well adopted and transferred into standard service offerings. Nevertheless, the service level agreements require the benefit of an initial classification model to provide that they contain the right attributes in the first place.

4. KEY CONSIDERATIONS DURING DC AND ILM IMPLEMENTATION

Considerations and issues that should be resolved during the creation of a robust ILM model should likely include [7, 2-8]:

1. Access rights and initial placement of data within the PCS which define the performance of the data in production applications

The key questions are: Where is the data initially placed within the PCS and where will the new data be placed? How quickly is the requested information accessible and what conditions need to be met? What are the initial performance requirements for data reach? What is the trigger that changes the status of the data or current classification?

2. Recovery and protection considerations which define what follows after the PCS critical or reference data are destroyed

The key questions are: What is the accessibility attribute: in the event of a failure or disaster of the primary data availability, how quickly can the access be restored? What methods, controls and levels of data protection are needed to satisfy the predefined recovery goals?

3. Discovery, retention and disposal considerations which define the service characteristics of the PCS data that have been archived

The key questions are: How quickly can the archived data be accessed? How broadly can contextual search capability be applied? For how long and how will the data be stored? What is the trigger event or time frame that initiates data disposal? How is disposal accomplished? What is the procedure for audit trail management?

4. Security considerations which define the overall management of the data from the perspective of an unauthorized use

The key questions: What access control, physical protection and encryption will be employed? Will there be a formal certification of the ISMS[1] implemented to ensure consistency in information protection? How will this change as the status of the data changes?

In combination, these "groups" of criteria define the key characteristics of the data lifecycle, hence through the implemented PCS ILM.

5. CONSEQUENCES OF DATA CHANGE IN IMPLEMENTED DC AND ILM SYSTEMS

Opposite to common thinking, data is dynamic in its nature, therefore data classification attributes change. This means that as the information moves between different user groups, its value attributes can be changed, resulting in the change of classification. So, it is wrong to look at information within the PCS's as having once set and never changed attribute. The need for relevant change management certainly calls for the adoption of a consistent, best-practice change management and service implementation system, similar in function to the ITIL[2].

To set up an effective DC system, change management procedures must already be in place, as it is a single point of reference repository for all changes to the system that will also reflect on the information contained within the system. The PCS usually have a very complex organizational structure and dissemination of the adopted DC/ILM system towards all stakeholders is a critical factor in roll out of the service.

The DC is rarely static; it changes with its use [3,3]. The goal of the ILM implies that the used service levels will also change. The dynamic DC (DDC) is a coined term used to describe the inherent need for adaptive classifications

[1] ISMS – Information Security Management System

[2] The Information Technology Infrastructure Library (ITIL) is a set of concepts and policies for managing the Information Technology (IT) services (ITSM), developments and operations.

and classification taxonomy. There are many possible levels of change, three of them being most prevalent in large systems (PCS's are by definition large systems), driven by predetermined trigger events, external forces or time that can make the classification of data difficult. They should be categorized as follows:

1. Changes embodied in a classification or a service level

The classifications and agreed service levels within the PCS themselves will sometimes contain a time-bound or event-triggered change to the handling of data. Additionally, the business timeline creates management aspects that are temporal, related to monthly processing cycles, quarterly actions executed in the system or end-of-year processing. These also impact service level requirements, and should be represented within the classification system.

2. Changes to the purpose or usage of the data

As the usage of data changes, the data may be subject to a different classification. This generally happens when the business process using the data objects changes, and it is usually triggered by a different service level agreement (SLA) or requirements defined in agreements. In the PCS scenario, if business intelligence model is applied and input data is migrated in a suitable format from transaction system to Datawarehouse, this would be a typical change. However, it is possible that data becomes archived within the transaction system, but later becomes critical. Such an example would be the archived data from a PCS participant who at a certain point in time becomes defunct (e.g. an agent of freight forwarder who temporarily goes out of business), but is later reintroduced to the port community (and therefore its "old" data becomes again business critical).

3. Changes to the classification taxonomy

It is realistic to expect that external events (such as changes to regulations, changes of the PCS structure and in the number of the involved parties, technology breakthroughs or changes in business strategy) might change the basic structure of the DC system. The change-management system implemented during the development of the classification taxonomy is critical to its long-term viability. In the PCS, it would be expected that such changes occur in case of the change of the legal framework that encompasses the functioning of the PCS.

The higher possibility of change indicates an elevated need for the adaptive and flexible DC. The process should account for change, and the initial project should focus on uncovering the need for an adaptive classification by executing the analysis of the PCS real life business workflows. The need for solid change management processes associated with the management of the taxonomy is also indicated. The ICT and the users of the PCS must be fully aware of the fundamental changes in regulations in their respective areas,

business strategies and structure, to ensure that the prescribed taxonomy is adapted. A formal mechanism and management review team (such as the one present within classic quality systems management scheme) should be implemented to provide that the data taxonomy is aligned with the internal business environment and external contributing factors.

6. CONCLUSION

The proper usage of the top down methodology during the introduction of the PCS DC and ILM model starts with establishing a clear goal for the project and ongoing DC process ownership and stewardship, followed by outlining what is exactly going to be classified (classification scope) and clearly delimiting the information important for the PCS particular goals of classification. Furthermore, it is critical to set a realistic scope (expectations and resources), meaning to create a cross-department team which will be in charge of the DC process and introduction.

The next stage of the process is establishing a cross-functional project team which will involve the PCS business process owners and IT department in charge of the ICT governance within the PCS. In this stage it is possible to introduce external consultants, but due to the nature of the process, the project ownership should be internal. Generally speaking, the DC project should be kept as close as possible to the organization (it should not be considered a "consultant process").

The implementation of the DC and ILM requires a snapshot of the situation or an audit of current DC practices both from a logical (relating business data objects to PCS business processes) and physical (current policy, placement and technology) perspective. This task engages the team in a broad effort to document the current state from two perspectives: a physical perspective, classifying business data objects based on their current technology and policy, and a logical perspective, relating business data objects to business processes. This should provide a model which will connect the business processes to the business data objects.

The final step in the creation of the DC and ILM is the classification of data objects and building a repository of the classification information and their connections to the PCS business processes. The implementation of the DC in a stepwise fashion provides that the process meets expectations and performs to set specifications. Monitoring the performance of the PCS data taxonomy and classification in production, assessing its implementation and application of the appropriate rules and the business impact of the classification should enable a constant improvement.

BIBLIOGRAPHY

[1] Aksentijevic, S., Recent developments in Croatian ICT security, 2nd conference of Croatian security managers, Biograd na Moru, Croatia, 21-23 Oct 2008.

[2] Hlaca, B., S. Aksentijevic, E. Tijan, Influence of ISO 27001:2005 on the Port of Rijeka security, Pomorstvo,22 (2008), 2, 245-257.

[3] Panagiotis G. I., /et al./, Modeling and Managing Changes in Text Databases, CDER Working Papers, http://www.cs.columbia.edu/~gravano/Papers/2007/tods07a.pdf, 23.06.2006.

[4] Rodon, J., J Ramis-Pujol, Exploring the Intricacies of Integrating with a Port Community System, 19th Bled eConference, Bled, Slovenia, June 5 - 7, 2006.

[5] Srour, F. J., /et al./, Port Community System Implementation: Lessons Learned from International Scan, Transportation Research Board 87th Annual Meeting, Washington DC, 2008.

[6] Stamp, M., Information security principles and practice, Wiley-Interscience, 2006.

[7] Best practices in data classification for information lifecycle management, SUN Microsystems, October 2005, http://whitepapers.zdnet.com/abstract.aspx?docid=129141

[8] Information Lifecycle Management Vision, TechRepublic whitepaper, http://whitepapers.techrepublic.com, May 2005.

[9] Tips for Implementation of Ohio IT Policy ITP-B.11, oit.ohio.gov/IGD/Policy/SupplementalMaterials/DataClassificationTips.doc, 29.04.2009.

[10] Yearbook 2007, Port of Rijeka Authority, 2007.

[11] Curphey, M., /et al./, A Guide to Building Secure Web Applications, The Open Web Application Security Project (OWASP), 2002, http://www.cgisecurity.com/owasp/html/ch08s02.html, 20.04.2009.

[12] Smith, R., Introduction to Multilevel Security., http://www.cs.stthomas.edu/faculty/resmith/r/mls/index.html, 20.04.2009.

[13] http://www.freebsd.org/cgi/man.cgi?mac_biba, 11.04.2009.

[14] http://ou800doc.caldera.com/en/SEC_admin/IS_DiscretionaryAccCntlDAC.html, 11.04.2009.

[15] http://phoenix.goucher.edu/~kelliher/f2006/cs325/oct27.html, 29.03.2009.

[16] Blake, S. Q., The Clark-Wilson Security Model, http://www.lib.iup.edu/comscisec/SANSpapers/blake.htm, 20.04.2009.

Sažetak

KLASIFIKACIJA PODATAKA I UPRAVLJANJE ŽIVOTNIM CIKLUSOM INFORMACIJA U LUČKIM INFORMACIJSKIM SUSTAVIMA

Rad istražuje ulogu klasifikacije podataka (Data Classification) i upravljanja životnim ciklusom informacija (Information Lifecycle Management) u objedinjenim sustavima informatičko-telekomunikacijske razmjene podataka i poruka u lučkom okruženju (Port Community Systems). U radu se analiziraju problemi koji nastaju prilikom implementacije takvih sustava, a uzrokovani su kompleksnošću pojedinih informacijskih i organizacijskih sustava sudionika u procesu. Predlaže se nekoliko modela klasifikacije podataka, izvedenih iz standardne metodologije sigurnosti informacijskih sustava. Identificiraju se ključni elementi potrebni za uvođenja sustava za klasifikaciju podataka i upravljanje životnim ciklusom informacija te se predlažu implementacijski putovi kojima bi se prebrodile moguće poteškoće u funkcioniranju sustava.

Ključne riječi: *lučki informacijski sustavi, klasifikacija podataka, upravljanje životnim ciklusom informacija, zadržavanje i uklanjanje podataka*

Edvard Tijan, mag. ing.
Sveučilište u Rijeci
Pomorski fakultet u Rijeci
Studentska 2
51000 Rijeka
Hrvatska