Modeling a Secure Cloud Data Warehouse with SoaML

Emna Guermazi
Mir@cl Laboratory,
University of Sfax, Sfax, Tunisia
guermaziemna1@gmail.com

Hanêne Ben-Abdallah
King Abdulaziz University, Jeddah,
Saudi Arabia
HBenAbdallah@kau.edu.sa

Mounir Ben Ayed
REGIM Laboratory,
University of Sfax, Sfax, Tunisia
mounir.benayed@ieee.org

***Abstract*—Cloud computing offers generous computing resources needed to deploy data warehouses efficiently. However, security remains a key challenge for a widespread adoption of the Cloud for data warehousing. In this paper, we extend BPMN to provide for modelling secure Extract-Transform-Load processes deployed as web services in a Cloud environment. In addition, following a model-driven engineering approach, we outline how the ETL model can be transformed into a set of services to be deployed in the Cloud.**

Keywords— cloud data warehouse; cloud security; Business Process Management Notation; Service oriented architecture modeling language ; Model-driven

I. INTRODUCTION

A Data Warehouse is a special type of database systems where a large volume of data are stored and queried in an *ad hoc* way for decision making purposes. Given its high computational and storage resource requirements, in-house deployment of a DW may be too costly and time consuming for small and medium size enterprises. This motivated several propositions on how to deploy efficiently a DW in a Cloud environment. This paper deals with the security aspects of a Cloud DW, *i.e.* a DW deployed in a Cloud environment. In particular, it addresses the need for modelling the security aspects of the Extract-Transform-Load processes which are needed to load data into a DW deployed on a Cloud.

Our approach has two merits. First, it adopts a Service Oriented Architecture (SOA) for the implementation of the ETL processes in the Cloud. Secondly, it follows a Model Driven Development (MDD) approach for the derivation of ETL services. These merits offer an agile way to deal with the adaptability of both business intelligence (BI) and security requirements in a DW. Indeed, the BI requirements evolve during the lifecycle of a DW, inducing new DW schemas and security needs. These changes may be accounted for in a more agile way thanks to: *i)* an SOA approach where services can be composed as needed, and *ii)* an MDD approach that reduces the cost of ETL service development by automating their derivation.

More specifically, we propose in this paper an extension of the BPMN language to model the security aspects of ETL processes; this extension builds on the work of [1] where ETL processes have been modeled using the standard BPMN [2], and it incorporates DW security requirements identified in [3]. In addition, we outline how ETL processes modelled with the secure BPMN can be transformed into a Service oriented architecture Modeling Language (SoaML) specification. SoaML supports the modeling of services using a model-driven development approach [4].

The remainder of this paper is organized as follows: Section II presents works relevant to data warehouses in the Cloud. Section III presents our approach. Finally, the last section is devoted to draw the conclusions and suggest future possible issues.

II. RELATED WORK

A. Modeling ETL process with BPMN

An Extract–Transform–Load (ETL) process first extracts pertinent data from heterogeneous source systems. It then transforms the extracted data through a set of conversion and cleaning operations to fit the DW schema. Finally, it loads the transformed data into the target DW. One can consider an ETL process as a workflow where a set of activities (the extraction, transformation, cleaning and loading operations) are orchestrated in a particular way. As such, several researchers examined modeling ELT processes using the Business Process Modeling Natation.

BPMN is an OMG standard [5] proposed to model business processes from different perspectives (informational, functional, behavioral, and organizational). In [6], Berkani et al consider a fragment of BPMN as core elements to describe ETL processes. The fragment BPMN meta-model described by the authors is composed of the following meta-classes: BaseElement, Flow objects, Activity, Task, Event, DataInput, DataOutput, InputSet, OutputSet, Gateway, Inclusive, Exclusive, Parallel, Complex. In addition to these meta-classes, the authors extended this BPMN fragment with the basic ETL operations: Extract, Retrieve, Merge, Union, Join, Filter, Conversion, Aggregation and Store.

In [7,8], the authors also proposed a customization of BPMN for designing ETL processes. In this BPMN customization, the authors propose to model an ETL task as a BPMN activity or a sub-process; each operation of selection,

projection, and type conversion is represented as a simple task, whereas aggregation, sort, pivot, join, union and difference are each represented by either a simple splitting or joining task.

B. Extending BPMN with security requirements

In [9], the authors propose a set of graphical concepts to capture six security requirements within the business process: non-repudiation, attack harm detection, integrity, privacy and access control. Non-repudiation is associated with Message Flow. Attack harm detection is associated with Pool, Lane, Group, Activity, Message Flow and Data object. Integrity is associated with Message Flow and Data Object. Privacy is associated with Pool, Lane and Group. Finally, access control is associated with Pool, Lane, Group and Activity.

C. Cloud data warehouse security

In [10], the authors propose an approach that allows on-line analysis processing on the Cloud while considering data privacy, integrity and availability. They propose to maintain data integrity with cryptography and hash functions [3]. This approach is applied only when many cloud service providers exists.

In [16], the authors develop a novel method for encrypting a data warehouse based on many kinds of encryption over the original value.

Deploying a data warehouse on the Cloud in a secure manner and applying ETL operations with a secure manner is our goal. Generally security requirements of ETL applied for cloud data warehouse are: (1) Data Privacy (confidentiality and anonymity for sensitive data), (2) the possibility to process analytical queries over encrypted data, and (3) data integrity [3]. The adopted security solution by [10,11] is the data encryption, we we adopted this solution but we integrated and we model the security requirements at a high level, in addition the security solution is modeled through services.

D. Using SoaML to specify cloud-based ETL services

Specifying ETL processes as a set of services provides for a better adaptability of the ETL processes face to changes in the BI and/or security requirements. Besides the appropriate specification language to be used, we argue in this section that an automated generation of the ETL service specification is vital for an efficient development.

In terms of service specification languages, SoaML extends UML with a set of stereotypes to model SOA requirements. The new stereotypes are: participants, service interfaces, service contracts, services architectures, service data and capabilities [4]. With these stereotypes, there are three different approaches to specify services using SoaML: simple interface, service contract, or service interface [4].

In [4], Elvesaeter et al use SoaML for the specification of Cloud-based services. They discuss the different ways to specify services and they provide practical modeling guidelines for how the different SoaML service specification approaches can be aligned and used as a baseline for specifying Cloud-based services. The authors suggest that the three service specification approaches can be reduced to two, because the first approach (simple interface) is a degenerate case of the remaining two. They propose to align a service contract with a service interface as refinements, and to align a service contract with a service interface as views.

In terms of ETL service specification, the generation of services based on a model-driven approach from the business process helps in the reuse of knowledge embedded in the model. In this context, the authors in [11] propose solutions for the generation of service oriented development from BPMN (Business Process Modeling). Towards this end, several works define transformations of BPMN models to generate SoaML services in general, e.g. [11].

E. A Service-oriented Architecture for Business Intelligence

Wu and Bartolini [13] propose to break down the ETL process into generic smaller service modules, which provides for the specification of a flexible and scalable ETL process. The authors propose SOA-ITPA system. The first phase of the life cycle of system development is schema management of reporting data store, the second specify services of the data cleaning, mapping and loading, in the third phase reporting data store is implemented, the measurement and presentation of IT key performance is done in fourth phase, finally Business intelligence solution was implemented. "addCube", "addCubeResponse", "populateCube" are an example of implementatition of cube managment service.

TABLE I. EVALUATION OF WORKS

References	Requirements covered				
	C1	C2	C3	C4	C5
[12]	N	N	N	N	N
[13]	N	N	N	Y	N
[4]	Y	N	N	N	N
[6]	N	Y	Y	N	N
[9]	N	N	N	Y	N
[14, 16]	N	Y	N	N	N
[10,11]	N	N	N	N	Y

Table I summarizes the features of the above approaches, with respect to using SoaML to specify Cloud-based services (C1); modeling ETL process with BPMN (C2); ETL specified as a service (C3); Extending BPMN with security requirements (C4); Cloud data warehouse security (C5). In this table, we use "Y" when a requirement is covered and, "N" otherwise.

As summarized in Table I, we notice that none of the existing works models ETL processes with BPMN extended with the security requirements. Like existing works, BPMN is our choice to model the ETL process for two main reasons: the expressive power of BPMN covers the ETL concepts, and the existence of many tools support the conceptual model of BPMN, which can be mapped into many execution languages. However, because of the Cloud context, it is important to

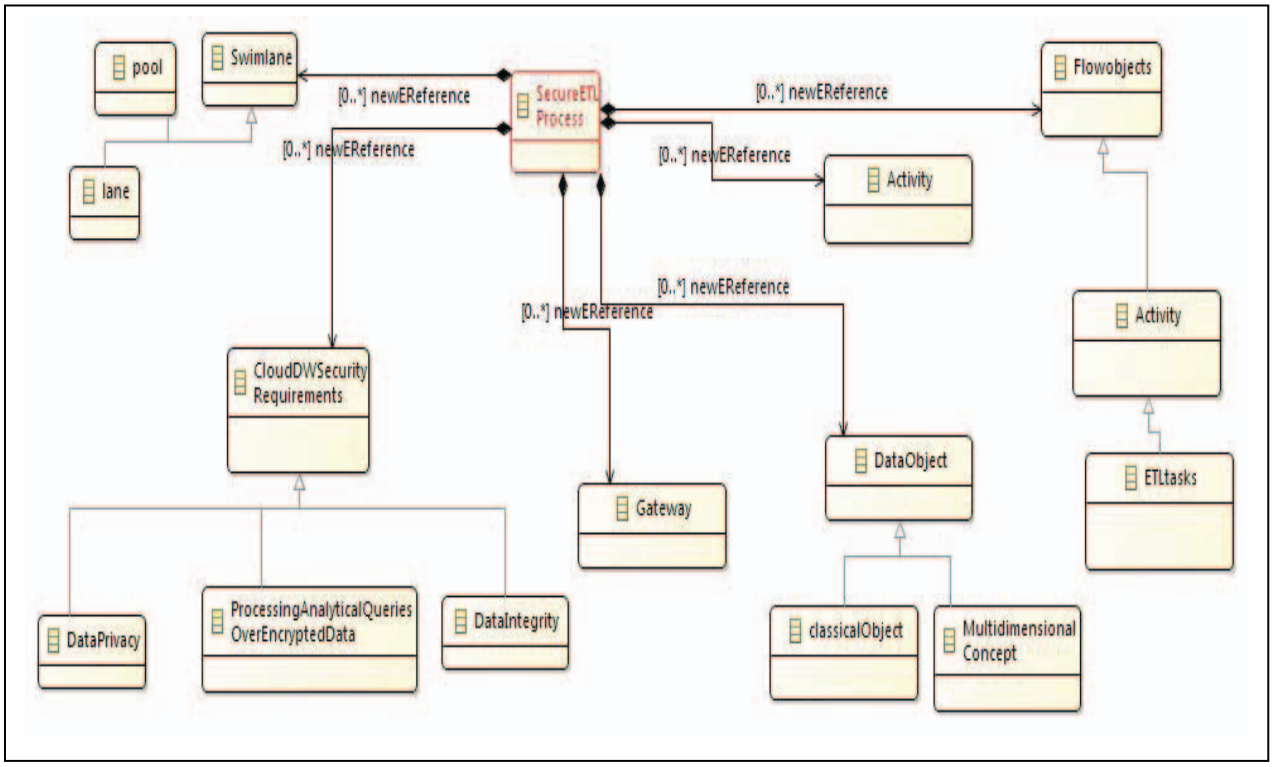

Fig. 1. Meta-model of secure BPMN for specifying secure ETL processes

extend BPMN with concepts to model several DW security requirements [15].

In addition, we note from Table I that none of the existing works dealt with the generation of secure ETL SoaML services for cloud data warehouse based on a model driven approach. These two shortages are the main contribution of this paper.

III. Proposed Approach

1. Defining a secure ETL process: an example

To illustrate our approach to develop a secure ETL process, let us consider the Sales fact example which records the total quantity and revenue for each product and client over a precise period of time. Table II shows the corresponding ETL process written in the OCL language.

To deploy this ETL process in a Cloud environment, let us consider the following three security requirements:

(1) Processing analytical queries over encrypted data which represent the possibility to query encrypted data using analytical queries [3];

(2) Privacy represents both confidentiality and anonymity requirements; and

(3) Data integrity means preventing of any unauthorized information modification, with the help of signatures like digital signature.

Our previous work in [15] shows how the above security requirements can be mapped to the ETL process:

All ETL operations should be mapped to data integrity security requirement. The created multidimensional concepts (fact, measure, dimension, hierarchies) should be mapped to the first security requirements, means encrypt them with homorphic or pseudo-homomorphic algorithms.

2. Modeling a secure ETL process using secure BPMN

We extend BPMN with concepts to model Cloud data warehouse security requirements (see Fig 1): An ETL process is composed of a set of Activity, Gateway and Data Object. Each Activity or a subprocess represents one ETL task. For example, in Fig. 2, the Load day from (order.Date) represents one activity/ETL task. A Gateway controls the flow of both Branching and Merging data flow. A Data Object can be a classical object coming from the source or a multidimensional concept such: Fact, Measure, Dimension, and Attribute.

TABLE II. ETL-OCL Expresions [19]

relative to	ETL-OCL Expressions
the dimension Product	Context DW.P roduit :: CodeP : Integer derive : SR.produit.codeP Context DW.P roduit :: Gamme : String derive : SR.produit.gamme.designationG Context DW.P roduit :: Secteur : String Derive : SR.produit.gamme.secteur.designationS
the dimension Time	Context DW.T emps :: Jour : Date derive : SR.commande.dateC Context DW.T emps :: Mois : Integer derive : SR.commande.dateC.Month() Context DW.T emps :: Annee : Integer

	derive : SR.commande.dateC.Y ear()
the fact Sales	Context DW.V entes :: quantit : Real derive : AGG(SR.LigneCom.quantite → sum(); SR.ligneCom.produit.codeP, SR.ligneCom.commande.client.codeC, SR.ligneCom.commande.dateC;) Context DW.V entes :: montant : Real derive : AGG((SR.ligneCom.quantite □ SR.ligneCom.produit.prixUnit;) → sum(); SR.ligneCom.produit.codeP, SR.ligneCom.commande.client.codeC, SR.ligneCom.commande.dateC;)

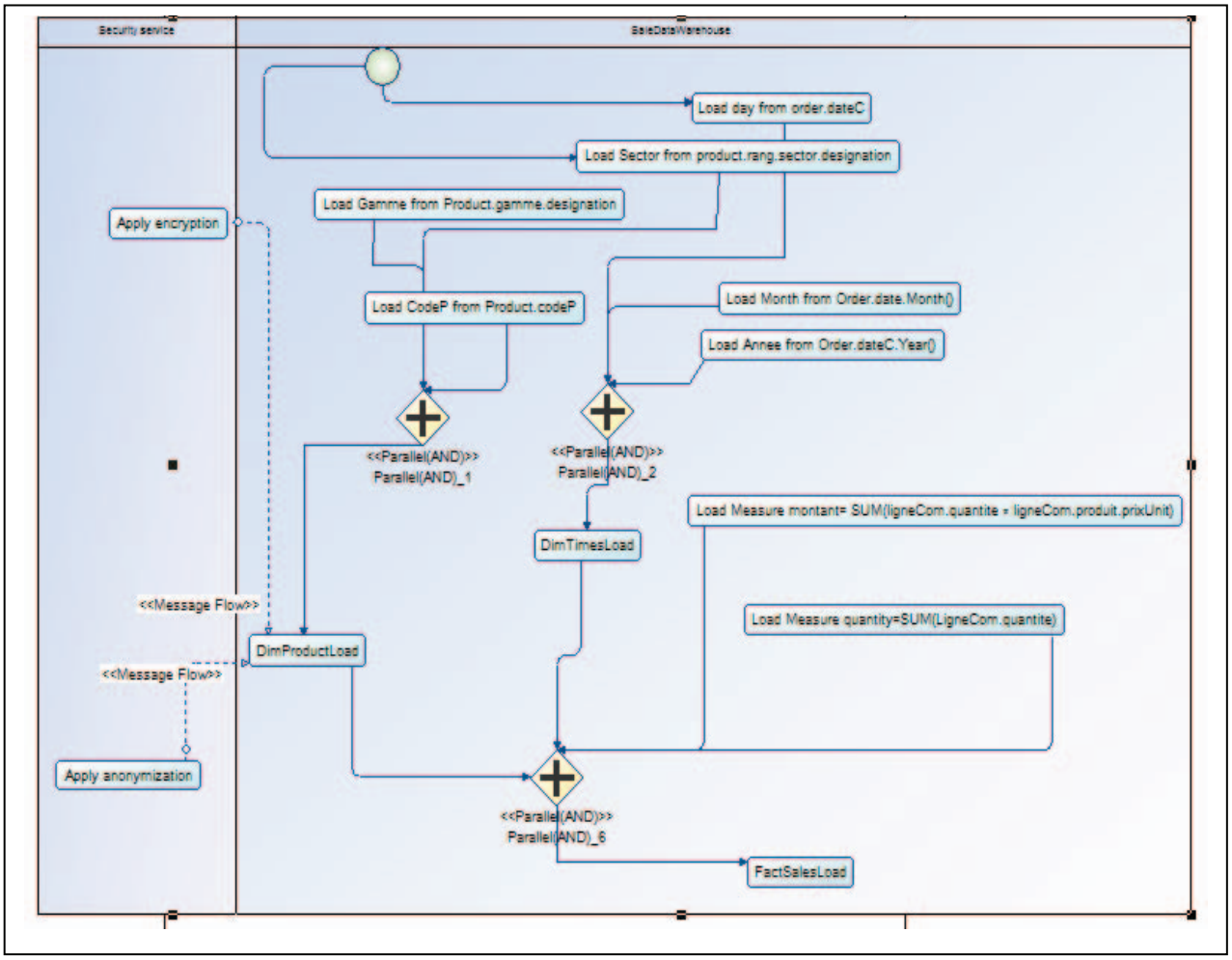

Fig. 2. Extract of the secure BPMN model of the ETL process for the DW Sales

For the Sales fact, its ETL-OCL expressions are transformed into the BPMN diagram illustrated in Fig 3. The secure DW service is specified in terms of two roles, their respective consumer and provider interface type, and the corresponding ports on the participants.

3. Transforming a secure BPMN into a secure SoaML

As discussed in Section II, our approach has the merit of adopting an MDD approach to generate from the secure BPMN model of the ETL process, its corresponding SoaML specification.

To transform a secure BPMN into a secure SoaML, we adapt the transformations proposed in [11,17,18] to deal with our security extensions in BPMN:

- Pool, and Lane are transformed to Participant in SoaML.

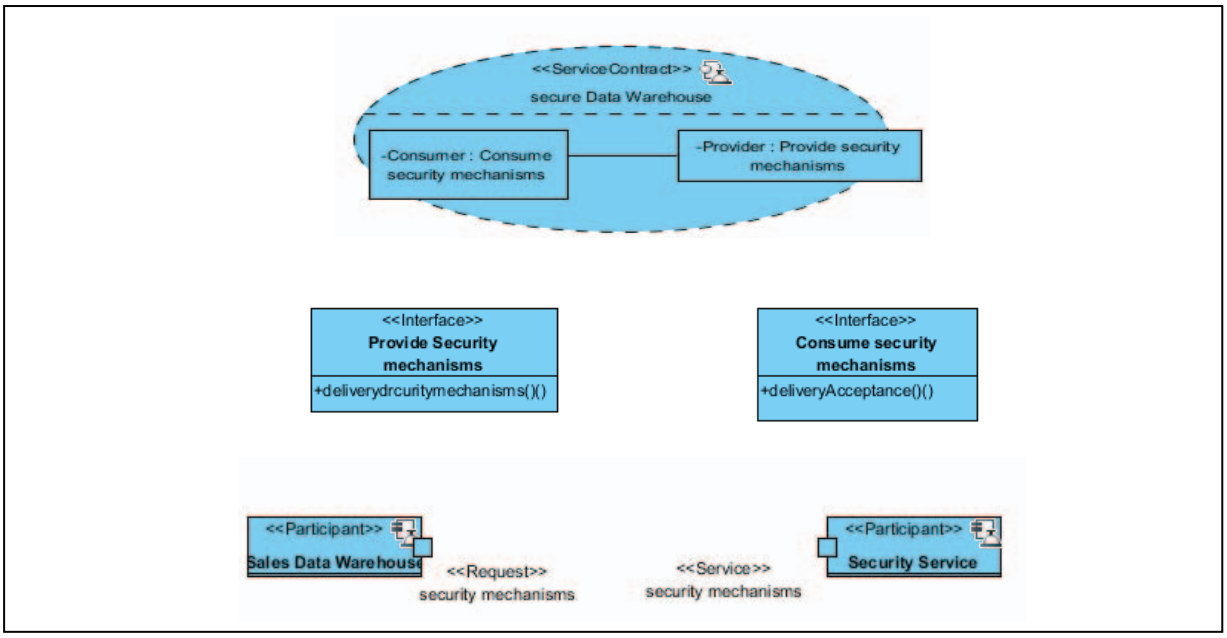

Fig. 3. Specification of the Sales fact in secure BPMN

- Interaction points between the business entities represented trough tasks are transformed to service contracts in SoaML.
- ActivityMessage, and PoolMessage are transformed to ServicePoint.
- Data objects are transformed to message types or data entities in SoaML.

To illustrate our work we introduce an example. In this example we start by loading day from order.datec and by loading Sector from product.rang.sector.designation. All other Multidimensional concepts are modeled in Fig 2. In addition, all multidimensional concepts are encrypted like the product dimension.

The corresponding SoaMl specification is generated based on the service contract approach: In Fig 3, two participants are modeled: Sales Data Warehouse (consumer), and security service (provider), and two interfaces are presented the first provide security mechanisms and the second consume security mechanisms and delivery acceptance.

IV. Conclusion

In this paper, we proposed an approach for the generation of security services for Cloud DW (in particular the ETL process). Our approach facilitates the adaptation of the security solution to address changing security requirements. By adopting Model Driven Development, our approach generates SoaML specifications from ETL modeled as secure BPMN. Security is integrated with BPMN and SoaML to generate security services for DW. The approach was illustrated with a "Sales fact" example.

For the future research activities, we are actively involved in the transformation from the first PIM level to the second PIM level and from the second PIM level to the PSM level. In addition, we are developing tool support for the modeling and transformation activities of our Cloud DW development approach.

Acknowledgment

The research and innovation are performed in the Framework of a thesis MOBIDOC financed by the EU under the program PASRI.
The authors would like to acknowledge the financial support of this wok by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

References

[1] Z. Al-Akkaoui, J-N. Mazón, A. Vaisman, E. Zimányi, " BPMN-Based Conceptual Modeling of ETL Processes", Data Warehousing and Knowledge Discovery Lecture Notes in Computer Science Volume 7448, 2012, pp 1-14.

[2] B. Elvesæter, C. Carrez, P. Mohagheghi, A.-J. Berre, S. G. Johnsen, and A. Solberg, "Model-driven Service Engineering with SoaML", in Service Engineering - European Research Results, S. Dustdar and F. Li (eds.), Wien, Springer, 2011, pp. 25-54. ISBN 978-3-7091-0414-9

[3] E. Guermazi, M. Ben Ayed, H. Ben-Abdallah, "A Survey of Data Warehouse Security," Ingénierie des Systèmes d'Information VOL 19/5 - 2014

[4] B. Elvesæter, A.JBerre., A. Sadovykh, "Specifying services using the service oriented architecture modeling language (SOAML): A baseline for Specification of Cloud-based Services", 1st International Conference on Cloud Computing and Services Science, CLOSER, Noordwijkerhout, The Netherlands, May 2011.

[5] OMG Standard, http://www.omg.org/spec/BPMN/2.0

[6] N. Berkani, L. Bellatreche, S. Khouri, "Towards a conceptualization of ETL and physical storage, of semantic data warehouses as a service,"

Journal Cluster Computing Volume 16 Issue 4, December 2013 Pages 915-931 Kluwer Academic Publishers Hingham, MA, US.

[7] Z. El-Akkaoui, E. Zimányi, “Defining ETL Worfklows using BPMN and BPEL,” DOLAP’09, November 6, 2009, Hong Kong, China. Copyright 2009 ACM

[8] Z. Al-Akkaoui, E. Zimányi, J-N. Mazón, J. Trujillo, "A BPMN-Based Design and Maintenance Framework for ETL Processes," International Journal of Data Warehousing and Mining, Volume 9, Issue 3, 2013.

[9] A.Rodirguez, E. Fernandez-Medina, M.Piattini, “A BPMN Extension for the Modeling of Security Requirements in Business Processes, Journal IEICE - Transactions on Information and Systems Volume E90-D Issue 4, Pages 745-752,March 2007.

[10] V. Attasena, N. Harbi, J. Darmont, “A Novel Multi-Secret Sharing Approach for Secure Data Warehousing and On-Line Analysis Processing,” In the Cloud. International Journal of Data Warehousing and Mining, 2014.

[11] Delgado, A., de Guzmán, I.G.R., Piattini, M.: From BPMN business process models to SoaML service models: a transformation-driven approach. In: 2nd International Conference on Software Technology and Engineering (ICSTE), 2010.

[12] W. H. Utomo, “ Implementation of mda method into soa environment for enterprise integration,” IJCSI International Journal of Computer Science Issues,Vol. 8, Issue 6, No 3, November 2011.

[13] L. Wu, G. Barash,C. Bartolini, “A Service oriented Architecture for Business Intelligence ; Service-Oriented Computing and Applications,” SOCA '07. IEEE International Conference, 2007.

[14] Z. El-Akkaoui, E. Zimányi, “Defining ETL Worfklows using BPMN and BPEL,” DOLAP’09, November 6, 2009, Hong Kong, China. Copyright 2009 ACM

[15] E. Guermazi, M. Ben Ayed, H. Ben-Abdallah, "Adaptive security for Cloud data warehouse as a service”. IEEE/ACIS (The International Association for Computer and Information Science) conference ICIS: International Conference on Computer and Information Science, Las Vegas, 2015.

[16] Z. Al-Akkaoui, E. Zimányi, J-N. Mazón, J. Trujillo, "A BPMN-Based Design and Maintenance Framework for ETL Processes," International Journal of Data Warehousing and Mining, Volume 9, Issue 3, 2013.

[17] B. Elvesæter, C. Carrez, P. Mohagheghi, A.-J. Berre, S. G. Johnsen, and A. Solberg, "Model-driven Service Engineering with SoaML", in Service Engineering - European Research Results, S. Dustdar and F. Li (eds.), Wien, Springer, 2011, pp. 25-54. ISBN 978-3-7091-0414-9

[18] A. Delgado, I.G.R.D. Guzmán, F. Ruiz, M. Piattini, “From BPMN business process models to SoaML service models:a transformation driven approach,” 2nd International Conference on Software Technology and Engineering(ICSTE),2010

[19] F. Atigui, "Approche dirigée par les modèles pour l’implantation et la réduction d’entrepôts de données", thesis, École doctorale Mathématiques, Informatique et Télécommunications (Toulouse, Haute-Garonne) , en partenariat avec Institut de Recherche en Informatique de Toulouse 05 décembre 2013.