CWM Extensions for Knowledge and Metadata Integration for Complex Data Warehouse and Big Data

Ralaivao Jean Christian, Razafindraibe Fabrice, Raherinirina Angelo and Rakotonirainy Hasina
University of Fianarantsoa, Madagascar

Keywords: Data Warehouse, Description Logics, Integration, Knowledge, Mapping, Metadata, ODM, Ontology, OWL DL, Transformation, UOP.

Abstract: This document constitutes a continuation of the work carried out in the field of complex data warehouses (DW) relating to the management and formalization of knowledge and metadata. It proposes a methodological approach to integrate two concepts, knowledge and metadata, within the framework of a complex DW architecture. The objective of the work considers the use of the knowledge representation technique by description logic and the extension of the Common Warehouse Metamodel (CWM) specifications. Several essential aspects of this work are expected, including the representation of knowledge in description logics and the declination of this knowledge into coherent UML diagrams while respecting or extending CWM specifications and using XML as a pivot, in particular OWL DL. Furthermore, the coupling between UML Ontology Profile (UOP) and the Ontology Definition Metamodel (ODM), for semantic modeling, integration of ontologies or enrichment of metadata, will be operationalized by transformation of models or by mapping or both simultaneously. As a result, a new extension of CWM metamodel will be developed. This will have performance consequences for a complex DW. The field of application is vast but will be adapted to systems with heterogeneous, complex and unstructured content and requiring a large (re)use of knowledge such as medical data warehouses.

1 INTRODUCTION

We have proposed an architectural framework for a complex data warehouse (Darmont et al., 2005). As Figure 1 illustrates, an important element of this framework is the implementation of domain knowledge and metadata in the three phases of warehousing: ETL (Extract - Transform - Load)/Integration, Administration/Monitoring and Analysis/Usage.

Note that the proposed architecture conforms to the CWM (Common Warehouse Metamodel) recommended by the Object Management Group (OMG). Based on this architecture, we conducted another work on the integration of knowledge and metadata for complex data warehouses. Thus, in (Ralaivao and Darmont, 2007), we discussed the different possibilities of integration and explored the integration of domain knowledge in the form of metadata.

The two other possibilities mentioned in (Ralaivao and Darmont, 2007) remain to be explored, namely:

- integration of metadata in the form of knowledge, in this case, our approach will take us to the field of knowledge-based warehouses ;
- separate knowledge and metadata management.

The architecture we have proposed is the basis of X-WaCoDa, an XML-based approach for online storage and analysis of complex data (Mahboubi et al., 2009). Three trends in XML-based DW (XML Web Warehouses (Vrdoljak et al., 2003; L., 2001; Golfarelli et al., 2003), XML Document Warehouses (Nassis et al., 2005; Rajugan et al., 2005; Baril and Bellahsène, 2003; Zhang et al., 2005), XML Data Warehouses (Pokorny, 2006; Hummer et al., 2003; Rusu et al., 2005; Park et al., 2005; Boussaïd et al., 2006)) were united to acquire a reference model that synthesizes all the work related to this field.

And on the other hand, the community working in the field of data and knowledge management is faced with new approaches such as Big Data and Data Lake (Sawadogo and Darmont, 2020; Sawadogo et al., 2019) whose content is heterogeneous, complex, weakly structured or even unstructured, non-standardized and inconsistent (Sakr and GaberSakr, 2014), adding to these problems the inexistence of methods and/or tools for integrating knowledge and metadata . This will necessarily lead to a review of the architecture and techniques around warehousing (ETL/Integration, Administration/Monitoring, Usage/Reporting) and infer-

Christian, R., Fabrice, R., Angelo, R. and Hasina, R.
CWM Extensions for Knowledge and Metadata Integration for Complex Data Warehouse and Big Data.
DOI: 10.5220/0012689300003690

In *Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024) - Volume 1*, pages 329-336
ISBN: 978-989-758-692-7; ISSN: 2184-4992

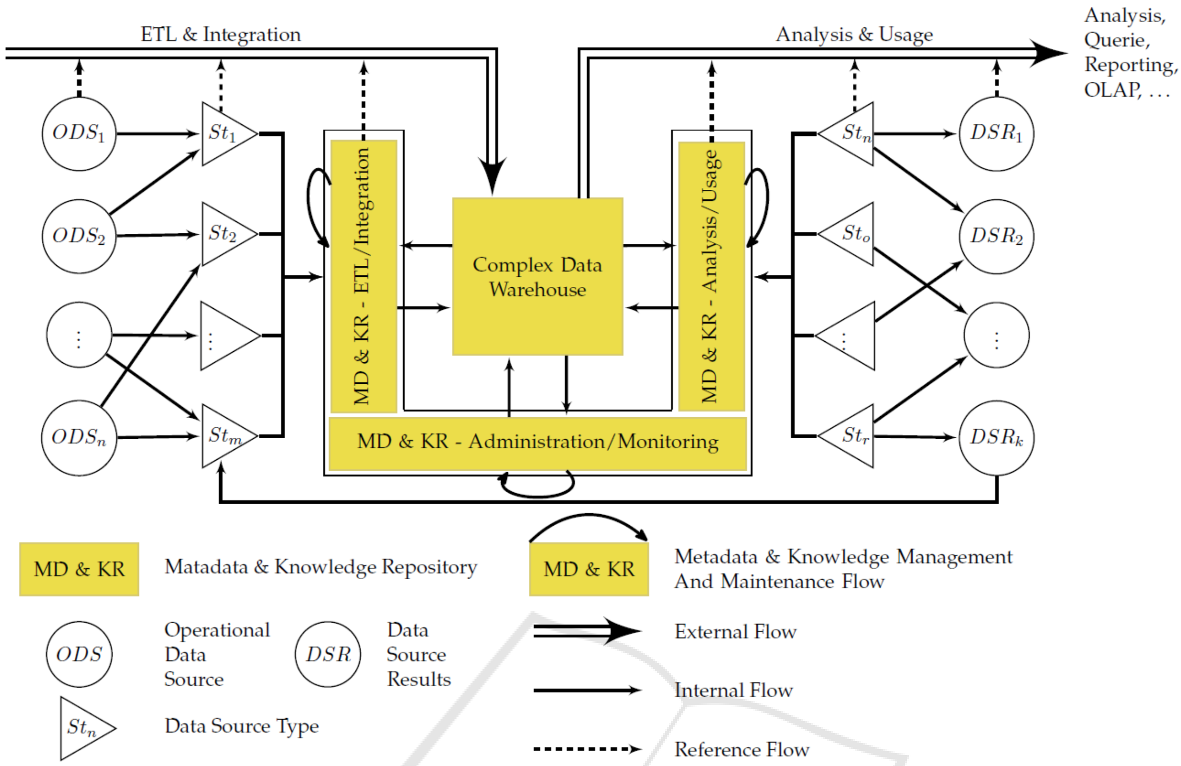

Figure 1: Architecture framework for a complex data warehouse.

ence (Wender, 2017) and especially for the discovery and management of knowledge (Bornschlegl et al., 2016).

We therefore propose to consider and deepen the two axes mentioned above while considering these new approaches.

2 MAIN QUESTION AND HYPOTHESIS

A high-performance complex data warehouse (DW) will integrate metadata of different forms, from its conception to its exploitation through its implementation and its administration. But knowledge of the field is also very important for more performance and flexibility in the various phases of storage. For example, in a complex DW for personalized anticipation medicine (MAP) (Darmont, 2008) the data can come from different operational data sources (ODS) (Biological, cardiovascular, biometric and psychological stores) and can be varied and heterogeneous. Therefore, data exchanges must be standardized by the use of concepts (metadata) and innovative XML technologies as a pivot language.

Would it be interesting to manage metadata and knowledge jointly *vs* separately? Would the integration of knowledge and metadata have an impact on traditional data warehouse architectures and what about the standard that prevails like the CWM of OMG? How to consider the main knowledge mainly formalized by ontology? How CWM will be adapted with the ODM standard?

3 STATE OF THE ART AND RELATED WORKS

(Inmon, 2005) in its definition of the data warehouse uses the terms: *(1) Subject orientation*: constituting an organization of data according to the domain (production, sale, transaction or activity, . . .), *(2) Integration*: structuring phase (Kimball and Ross, 2013), *(3) Historization*: time axis, *(4) Non volatility*: conservation principle, *(5) Aggregation* : availability and accessibility by queries, OLAP, data mining, . . . These terms constitute the main function of a data warehouse.

The DW border breaks free of conventional frameworks by expanding and integrating with the exploitation of Big Data. (Ngo et al., 2019) proposed an agricultural DW architecture for a business intelligence operation. The use of common solutions such as MongoDB, Cassandra or CONSUS DW demonstrates the usefulness of DW in the context of Big Data.

However, there are more dissimilarities than similarities between Big Data and complex DW. The exploitation of Big Data in a DW requires the use of pivot language of which the most answered is XML.

XML documents allow not only to facilitate the standardization of complex data (Darmont et al., 2005) before the loading phase in the DW, but also to constitute the storage core (XML-Based). In this context, research on XML-based storage has been steadily growing.

These different approaches respect the standard imposed by the **CWM** (Poole et al., 2003) metamodel of the OMG. Most DWs are based on metadata. The complexity of the data, however, dictated by the heterogeneity of the sources and the absence of data structures often makes it difficult to integrate them into a DW. The question of performance (Baril and Bellahsène, 2003) takes on a primordial meaning in all phases of storage. In addition, one of the purposes of storage is the extraction of knowledge to facilitate interpretation and decision-making. One of the main questions is: "Will integrating knowledge in addition to metadata improve performance?". It is in this context and following our latest work (Darmont et al., 2005; Ralaivao and Darmont, 2007; Mahboubi et al., 2009) that (Liao et al., 2010) integrates metadata and/or ontologies in the field of semantic Web and (Wu and Hakansson, 2010) uses the Knowledge-based system for integrating metadata into warehousing.

The work of (Srinivasan, 2016) presents an architecture for a business intelligence (BI) which takes into account at the same time the concepts and techniques stemming from artificial intelligence, machine and deep learning and excavations in the advent of Big Data.

The proposed methods take into account three aspects (solutions) of this integration of knowledge and metadata in a complex DW.

4 METHODS

In (Ralaivao and Darmont, 2007), we were able to inventory and classify the metadata as well as to consider the types of knowledge in order to commonly integrate them in complex ED. At the end of these first works, the transformation of knowledge in the form of metadata ($K \rightarrow MD$) is convincing for the domain (Liao et al., 2010; Wu and Hakansson, 2010).

The consideration of ontologies in a DW is essential for the revision of knowledge (interpretation, integration, formulation, ...). A second method therefore proposes the transformation of metadata in the form of knowledge $MD \rightarrow K$. Thus, we will obtain a Knowledge-based Warehouse DW (Nemati et al., 2002).

The two methods will lead us to another perspective which is the hybridization of metadata and knowledge ($MD \rightleftarrows K$) to manage DW. It is based on the theoretical fact that one can transform metadata into knowledge and vice versa. The iterative and/or recursive execution of this transformation constitutes learning at the level of a complex DW.

5 RESEARCH EXPECTATIONS

In addition to the results obtained in the work carried out so far, this work will reinforce the achievements in terms of complex DW architecture and will provide more details on the possibilities of integrating knowledge and metadata. Adding to this we are working on the following integration aspects:

1. the representation of knowledge (Pan, 2020) in description logics (DLs) for a semantic and axiomatic formalization of knowledge of the domain,
2. the mapping of DLs in UML (Dutra, 2002) which will allow use in model-driven engineering and an implementation based on UML diagrams,
3. integration of the UML diagram corresponding to the DLs (De Giacomo, 2010) according to the CWM specifications in order to respect the standards imposed by the OMG in the implementation of DW in general.

5.1 Knowledge Representation in DLs

DLs constitute a formalized set of knowledge representation languages (De Giacomo, 2010; Pan, 2020). Based on two fundamental formal frameworks: the *TBox*[1] and the *ABox*[2], where the DLs formalize a knowledge base K as **1)** description language and **2)** inference services.

(Müller et al., 2011) for example offers an extension of DLs for an integrated framework for representing knowledge in different aspects.

We advance the hypothesis that DLs can be managed or formalized by metadata. So we will have the transformation $K \rightarrow DL \rightarrow MD$ and vice versa.

[1] Set of terminological axioms

[2] Set of assertive axioms

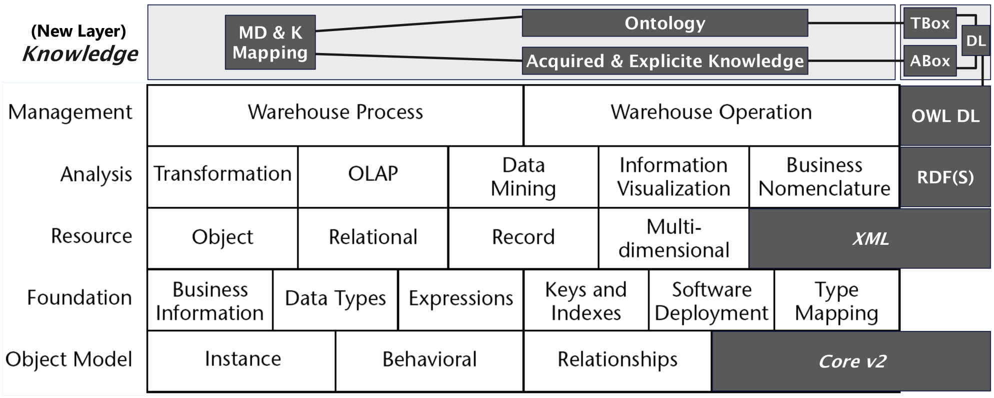

Figure 2: CWM extensions.

5.2 DLs Mapping/Tranformation in UML Diagram and *Vice Versa*

This technique would consist of DLs translation in UML diagram (De Giacomo, 2010; Dutra, 2002). However it is essential to consider inference services as for the UML diagrams semantics (Le Duc, 2008) in order to **1)** check the consistency of a class or a class diagram, **2)** check the subsumption and the equivalence between classes and **3)** explain the logical consequences of the properties.

This would allow an opening or an application of the different existing approaches of knowledge engineering (De Giacomo, 2010).

DL offers the possibility of writing the ontology in OWL or the knowledge base where a concept of DL is a class and a role of DL is a property in OWL. Since RDF/XML is the normative syntax for exchanging information between systems, the OWL DL solution based on RDF(S) offers maximum expressiveness regarding the $\mathcal{SHOIN}$ description logic (Baader et al., 2017; Lutz et al., 2019). The latter supports data values, data types and their properties.

5.3 UML Diagram Integration in CWM Specification

From a standardization point of view, the CWM specifications and their adaptations are crucial in the field of complex DW where the CWM is the champion of standardization. This integration will take as a reference the CWM modeling and specification levels, namely the model, meta-model, meta-meta-model and the Meta Object Facility (MOF) (Poole et al., 2003).

This integration will facilitate the supply of the warehouse, contribute to its administration mainly on the performance aspect, guide users in its operation and finally it will offer the opportunity to re-inject the knowledge extracted from the analysis and excavations in the warehouse itself.

This work will also propose extensions and/or possibly grafts to the specifications of CWM metamodels, while ensuring bottom-up compatibility and eventually propose specific adaptations to Big Data.

Above the five functional layers of CWM will be added a layer for knowledge management.

6 RESULTS

The contributions of this paper are as follows :

1. The CWM extensions with a new layer and somes packages ;
2. OWL DL transformation and/or mapping with UML ;
3. Plugins of package **Core** with a new ODM specification.

6.1 CWM Extensions

Many works to extend CWM (Soler et al., 2008b; Demraoui et al., 2016) have been carried out but none proposed to extend CWM to a new layer in order to manage knowledge and metadata. This constitutes an important advance in the field of complex data warehouses, knowledge integration and its reinjection.

The extension examples proposed by (Demraoui et al., 2016) are available in the works of (Zhao and Huang, ; Soler et al., 2008a; Gomes et al., 2007; Midouni et al., 2009; da Silva et al., 2010; Tavac and Tavac, 2013) and (Thavornun, 2015).

The work of (Thavornun, 2015) is closest to this field of research and advances the following problems:

1. CWM mainly focuses on data warehouse metadata ;

2. CWM does not cover automated knowledge discovery and management system.

The addition of a new layer *Knowledge*, on the shaded box with some integrated packages (MD & KB mapping, Ontology, Acquired and explicite Knowledge, DL) and used packages (XML, RDF(S), OWL DL, Core v2), Figure 2, will solve these problems although it will increase the number of metamodels to consider. It will integrate the simultaneous management of metadata and knowledge. These will arise from OWL DL ontologies, based on the ODM metamodel, a formal and expressive representation of description logic.

The *knowledge* layer, however, uses **XML** metamodels from the *Resource* layer for the OWL DL formalization and **Core** from the Object layer for basic knowledge operations. As a result, a new version of the **Core** package named **Core v2** in CWM extension, including a plugin of it, will be offered subsequently.

The transformation and/or mapping of OWL DL into UML, based on the UOP metamodel, and *vice versa* makes it possible to understand the knowledge in the CWM repository and is proposed in the subsection 6.2.

6.2 OWL DL Transformation and/or Mapping with UML

Figure 3 shows two parallel conceptual Modeling Spaces (MS)[3] at the Metamodel level of the OMG MOF architecture, namely the UOP and the ODM, the respective metamodels of the UML models and the OWL DL ontologies. The latter, located at layer $M1$, model, in parallel, the real world, using modeling languages, in this present research, UML, OWL or other language, defined using different meta-metaconcepts.

RDF(S) as well as three different dialects of OWL, namely OWL Full, OWL DL and OWL Lite, are examples of languages in the $M2$ layer. OWL DL is most suitable in our complex data warehouse context because it promotes maximum expressiveness while maintaining the completeness and decidability of calculations.

At the conceptual level, we establish a transformation from UML to ODM and vice versa at the metamodeling level. An example of a transformation modeling language for such purposes in MOF is **Query-View-Transformation** (QVT) (Gardner and Griffin, 2003; Kruse, 2015; OMG, 2016) or the Atlas Transformation Language (ATL) (Kruse, 2015). And the mapping proposed by (Brockmans et al., 2006) complies with the transformation rules. Figure 4 represents a model for transforming UML into OWL DL and *vice versa*. Two packages formalize the "**QVT transformation**" from UML to OWL and *vice versa* and the "**OWL (de)serialization**".

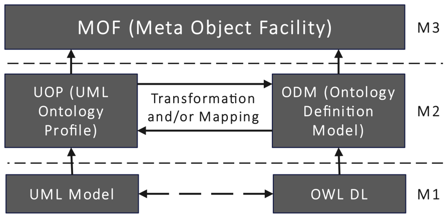

Figure 3: Ontology transformation and/or mapping with UML.

The metamodel proposed in Figure 6 constitutes an isomorphic correspondence, for mapping, with OWL DL using UML concepts.

We propose a QVT transformation model as proposed by (Haasjes,).

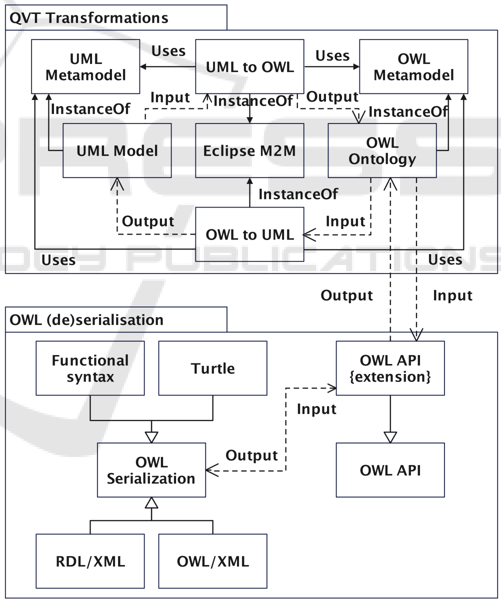

Figure 4: QVT transformation of UML and OWL and (de)serialization of OWL.

6.3 Core Extension

The **AnnotedElement** class, from the Knowledge layer, is extended to the **ModelElement** class of the Core class (Figure 5). This extension will not only ensure ontology integration (plugin of the model in Figure 5), but also compatibility with CWM while leaving all of CWM's functionalities intact.

The extension in question concerns ODM and is

[3]A modeling space (MS) is a modeling architecture based on a particular (meta)metamodel

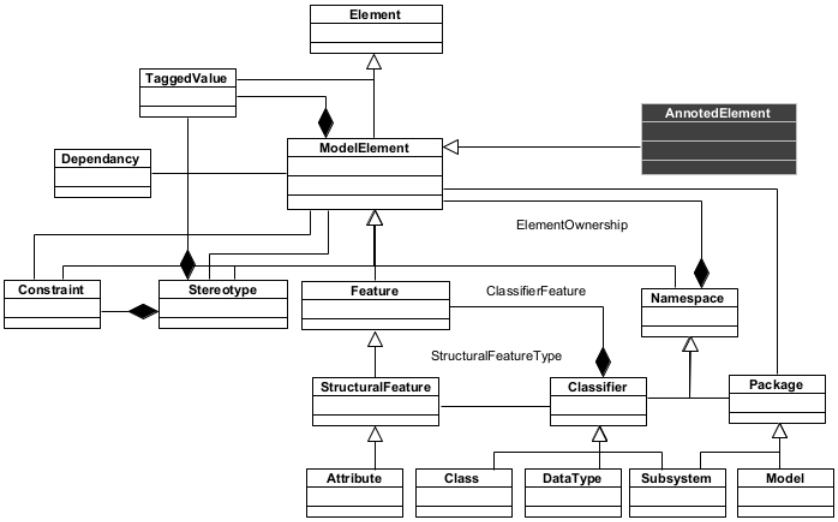

Figure 5: Extension of the ModelElement class of the Core metamodel by the AnnotedElement class.

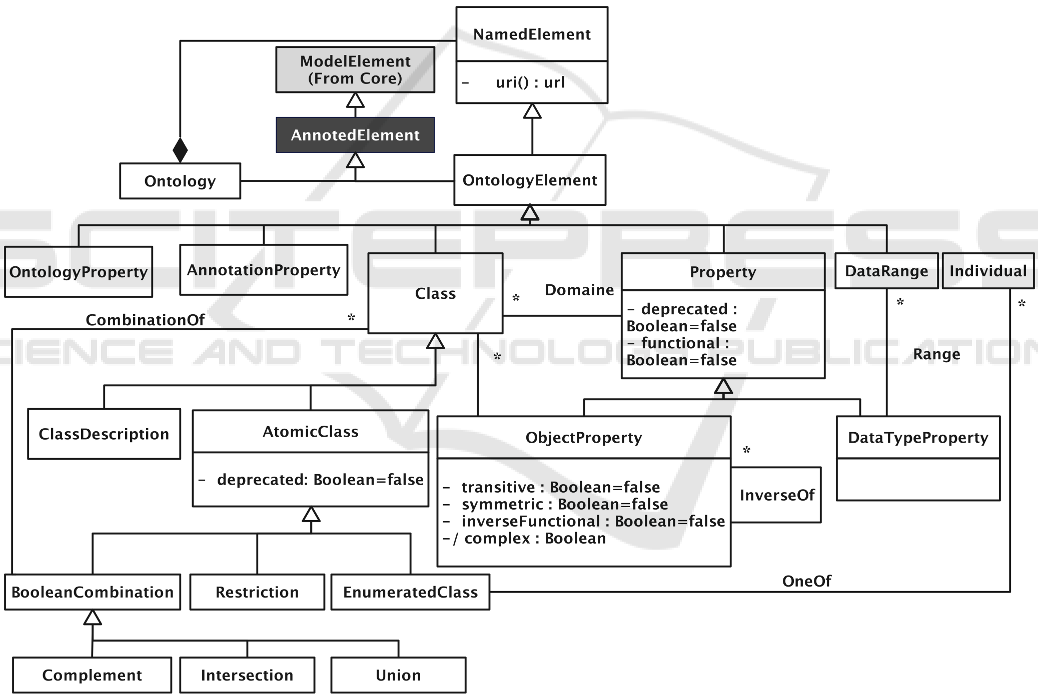

Figure 6: Elements structuring the ODM metamodel.

represented by the Figure 6. It offers all the integrations of the normative syntax between OWL DL and DL (Obitko et al., 2004; Baader et al., 2017; Lutz et al., 2019).

OWL DL axioms and facts (Obitko et al., 2004; Baader et al., 2017; Lutz et al., 2019) about classes such as **EnumerateClass**($A \quad o_1 \ldots o_n$), **SubClassOf**($C_1 C_2$) (Subsumption), **EquivalentClasses**($C_1 \ldots C_n$) (Equivalent Classes), **DisjointClassses**($C_1 \ldots C_n$) (Disjoint Classes) and **DataType**(D); on properties **DataTypeProperty**($range, Symmetric, \ldots, Transitive$) (Data type properties) **SubPropertyOf**($U1U2$) (Subsumption on properties), **EquivalentProperties**($U_1 \ldots U_n$) (Equivalent properties); on annotations, **AnnotationProperty**(S) or on individuals, **SameIndividual**($o_1 \ldots o_n$) and **DifferentIndividual**($o_1 \ldots o_n$) ; are easily represented by the model in Figure 6.

7 CONCLUSION

We recall that the issues to which the work is based were translated such as:

- the difficulty of integrating metadata and even more knowledge into a complex ED;
- the increasing heterogeneity and management of sources and content;
- the performance of existing complex and heterogeneous data management solutions;
- the lack of methods and/or tools for integrating knowledge and metadata.

The integration approach by different representation approaches – DLs, Mapping, Transformation and CWM specifications – should ensure the relevance of the proposed architectural framework as well as its implementation.

The tools that will be developed for experimentation in this research work will be generalized and will be made available to the scientific community working in the field of complex DW and Big Data. Such tools should be used to verify the proposed ideas.

REFERENCES

(2016). *MOF Query/View/Transformation, version 1.3, https://www.omg.org/spec/QVT/1.3/About-QVT/*.

Baader, F., Horrocks, I., Lutz, C., and Sattler, U. (2017). An introduction to description logic.

Baril, X. and Bellahsène, Z. (2003). *Designing and Managing an XML Warehouse. XML Data Management: Native XML and XML-Enabled Database System*. Addison-Wesley.

Bornschlegl, M. X., Engel, F. C., Bond, R., and Hemmje, M. L. (2016). *Advanced Visual Interfaces: Supporting Big Data Applications*. Springer, Information Systems and Applications, gerhard goos, juris hartmanis, and jan van leeuwen edition. Lecture Notes in Computer Science.

Boussaïd, O., BenMessouad, R., Choquet, R., and Anthoard, S. (2006). X-warehousing: An xml-based approach for warehousing complex data. *10th East-European Conference on Advances in Databases and Information System (ADBIS'06), Thessaloniki, Greece*, Vol. 4152 of Lecture Notes in Computer Science. Springer:39–54.

Brockmans, S., Haase, P., Hitzler, P., and Studer, R. (2006). A metamodel and uml profile for rule-extended owl dl ontologies. In Sure, Y. and Domingue, J., editors, *The Semantic Web: Research and Applications*, pages 303–316. Springer Berlin Heidelberg.

da Silva, J., de Oliveira, A. G., do Nascimento Fidalgo, R., Salgado, A. C., and Times, V. C. (2010). Modelling and querying geographical data warehouses. 35:592–614.

Darmont, J. (2008). Entreposage de données complexes pour la médecine d'anticipation personnalisée. *IC-SSHC*.

Darmont, J., Boussaïd, O., Ralaivao, J. C., and Aouiche, K. (2005). An architecture framework for complex data warehouses. *7th International Conference on Enterprise Information Systems (ICEIS'05)*, Miami, USA. INSTICC:370–373.

De Giacomo, G. (2010). Description logics for conceptual data modeling in uml. Technical report, Dipartimento di Informatica e Sistemistica, SAPIENZA Università di Roma.

Demraoui, L., Behja, H., Zemmouri, E. M., and Ben Abbou, R. (2016). A viewpoint based extension of the common warehouse metamodel to support the user's viewpoint approach. 11:137.

Dutra, M. (2002). Using uml for knowledge representation. Technical report, Sandpiper Software, Inc.

Gardner, T. and Griffin, C. (2003). A review of omg mof 2 . 0 query / views / transformations submissions and recommendations towards the final standard.

Golfarelli, M., Rizzi, S., and Vrdoljak, B. (2003). Data warehouse design from xml sources. *4th International Workshop on Data Warehousing and OLAP (DOLAP'01), Atlanta, USA*, ACM Press:40–47.

Gomes, P., Farinha, J. T., and Trigueiros, M. J. (2007). A data quality metamodel extension to cwm. In *Asia-Pacific Conference on Conceptual Modelling*.

Haasjes, R. Metamodel transformations between uml and owl.

Hummer, W., Bauer, A., and Harde, G. (2003). Xcube: Xml for data warehouse. *6th International Workshop on Data Warehousing and OLAP (DOLAP'03), New Orleans, USA*, ACM:33–40.

Inmon, W. H. (2005). *Building the Data Warehouse*, volume 4. Wiley Publishing Inc., robert elliot edition.

Kimball, R. and Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition*. J. Wiley and Sons, Inc., wiley edition.

Kruse, S. (2015). *Co-Evolution of Metamodels and Model Transformations - An operator-based, stepwise approach for the impact resolution of metamodel evolution on model transformations.*

L., X. (2001). A dynamique warehouse for xml data of the web. *International Database Engineering and Application Symposium (IDEAS'05), Grenoble, France*, IEEE Computer Society:3–7.

Le Duc, C. (2008). *Transformation d'ontologies basées sur la logique de description : Application dans le commerce électronique*. Phd thesis, Université de Nice - Sophia Antipolis.

Liao, S.-H., Huang, H.-C., and Chen, Y.-N. (2010). A semantic approch to heterogeneous metadata integration. *Springer- Verlag Berlin heidelberg*, pages 205–214.

Lutz, C., Sattler, U., Tinelli, C., Turhan, A.-Y., and Wolter, F. (2019). Description logic, theory combination, and all that: Essays dedicated to franz baader on the occasion of his 60th birthday. *Lecture Notes in Computer Science N°11560*.

Mahboubi, H., Ralaivao, J. C., Loudcher, S., Boussaïd, O., and Bentayeb, F. (2009). X-wacoda: An xml-based approach for warehousing and analyzing complex data. *Advances in Data Warehousing and Mining*, IGI Publishing(3):38–54. Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction.

Midouni, S. A. D., Darmont, J., and Bentayeb, F. (2009). Approche de modélisation multidimensionnelle des données complexes : application aux données médicales,. In *Journées Francophones sur les Entrepôts de Données et l'Analyse en ligne*.

Müller, J.-P., Rakotonirainy, H. L., and Hervé, D. (2011). Towards a description logic for scientific modeling. *International Conference on Knowledge Engineering and Ontology Development (KEOD)*.

Nassis, V., Rajugan, R., Dillon, T. S., and Rahayu, J. W. (2005). Conceptual and systematic design approach for xml document warehouses. *International Journal of Data Warehousing & Mining 1 (3)*, pages 63–86.

Nemati, H. R., Steiger, D. M., Iyer, L. S., and Herschel, R. T. (2002). Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. pages 143–161.

Ngo, V. M., Le-Khac, N.-A., and Kechadi, M.-T. (2019). Designing and implementing datawarehouse for agricultural big data. In Chen, K., Seshadri, S., and Zhang, L.-J., editors, *Big Data - BigData 2019*, pages 1–17. 8th International Congress Held as Part of the Services Conference Federation, SCF 2019.

Obitko, M., Snásel, V., and Smid, J. (2004). Ontology design with formal concept analysis. In *International Conference on Concept Lattices and their Applications*.

Pan, Y. (2020). Multiple knowledge representation of artificial intelligence. *ELSEVIER*. Institute of Artificial Intelligence, Zhejiang University.

Park, B. K., Han, H., and Song, I. Y. (2005). Xml-olap: A multidimensional analysis framework for xlm warehouses. *7th International Conference on Data Warehousing and Knowledge Discovery (DaWaK'05), Copenhagen, Denmark*, Springer:267–280. Vol. 3589 of lecture Notes in Computer Science.

Pokorny, J. (2006). Xml data warehouse: Modeling and querying. *5th International Baltic Conference (BalticDb&IS'06), Tallin, Estonia*, Institute of Cybernectics at Tallin Technical University:267–280.

Poole, J., Chang, D., Tolbert, D., and Mellor, D. (2003). *Common Warehouse Metamodels*. Wiley Publishing Inc., omg press edition.

Rajugan, R., Chang, E., and Dillon, T. S. (2005). Conceptual design of an xml fact repository for dispersed xml document warehouses and xml marts. *20th International Conference on Computer and Information Technology (CIT'05), Shanghai, China*.

Ralaivao, J. C. and Darmont, J. (2007). Knowledge and metadata integration for warehousing complex data. *6th International Conference on Information Systems Technology and its Applications (ISTA 07)*, Kharkiv, Ukraine. GI-Edition(107):164–175. Lecture Notes in Informatics.

Rusu, L. I., Rahaya, J. W., and Taniar, D. (2005). A methodology for buidlding xml data warehouse. *International Journal of DataWarehousing and Mining 1 (2)*, pages 67–92.

Sakr, S. and GaberSakr, M. M. (2014). *Large Scale and Big Data: Processing and Management*. CRC Press.

Sawadogo, P. and Darmont, J. (2020). On data lake architectures and metadata management. *JIIS*.

Sawadogo, P. N., Kibata, T., and Darmont, J. (2019). Metadata management for textual documents in data lakes. *ICEIS*.

Soler, E., Trujillo, J., Fernández-Medina, E., and Piattini, M. G. (2008a). Building a secure star schema in data warehouses by an extension of the relational package from cwm. 30:341–350.

Soler, E., Trujillo, J., Fernández-Medina, E., and Piattini, M. G. (2008b). An extension of the relational metamodel of cwm to represent secure data warehouses at the logical level. 6:355–362.

Srinivasan, V. (2016). *The Intelligent Enterprise in the Era of Big Data*. WILEY.

Tavac, M. and Tavac, V. (2013). The general algorithm for the design of the mda transformation models. pages 171–176.

Thavornun, V. (2015). Metadata management for knowledge discovery.

Vrdoljak, B., Banek, M., and Rizzi, S. (2003). Designing web warehouse from xml schemas. *5th Intenational Conference on DataWarehousing and knowledge Diskovery (DaWak'03), Prague, Czech Republic*, pages 89–98. Vol. 2737 of Lecture Notes in Computer Science. Springer.

Wender, B. A. (2017). *Refining the Concept of Scientific Inference When Working with Big Data: Proceedings of a Workshop*. The National Academic Press. National Academics of Sciences, Engineering, and Medicine.

Wu, D. and Hakansson, A. (2010). Applying a knowledge based system for metadata integration for data warahouses. In Setchi, R., Jordanov, I., Howlett, R. J., and Jain, L. C., editors, *Knowledge-Based and Intelligent Information and Engineering Systems*, pages 60–69.

Zhang, J., Wang, W., Liu, H., and Zhang, S. (2005). X-warehouse: building query pattern-driven data. *International conference World Wide Web (WWW'05), CHiba, Japan*, ACM:896–897.

Zhao, X. and Huang, Z. A formal framework for reasoning on metadata based on cwm. In *International Conference on Conceptual Modeling*.