Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 159 (2019) 756–763

Procedia Computer Science

www.elsevier.com/locate/procedia

23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Specification of the data warehouse for the decision-making dimension of the Bid Process Information System

Manel Zekri[a*], Sahbi Zahaf[b], Sadok Ben Yahia[c]

[a] *University of Tunis El Manar, Faculty of Sciences of Tunis, Department of Computer Science El Manar II 2092, Tunisia*
[b] *MIRACL Laboratory, Higher Institute of Computer and Multimedia, Sakiet Ezzit Technopole, Sfax 242-3021, Tunisia*
[c] *Tallinn University of Technology, Department of Software Science Akadeemia tee 15a, room 649, Tallinn 12618, Estonia*

Abstract

In order to enhance the business process, the Bid Process Information System (BPIS) should be made more performant by ensuring greater flexibility and interoperability between devices. Moreover, the specication of this system has to deal with "three fit" problems. To this end, four dimensions have been identified in order to cope with potential failures: operational, organizational, decision-making, and cooperative dimensions. In this paper, we focus on the decision-making dimension of the BPIS and propose an approach for representing data warehouse schema based on an ontology that captures the multidimensional bid-knowledge.

Keywords: Bid Process, Information System, Data Warehouse, Conceptual Data Model, Multidimensional Design, Ontology;

1. Introduction

The bid process which corresponds to the conceptual phase of the lifecycle of a product is a particular environment for the exploitation of business process [8]. It interacts upstream and involves other processes being design process. It aims to examine the feasibility of the bid before negotiating any contract with any owner following a pre-study carried out before a project launch.

The IS (Information System) [7] that allows to run the bid process (Bid Process Information System or BPIS) must be [20] [19]: integrated, flexible and interoperable. Nevertheless, the enterprise architecture approach [7], on which we rely to implement this system, has to deal with "three fit" problems: vertical, horizontal and transversal fit (Fig. 1). Such problems handicap the exploitation of these three criteria [20]. The "vertical fit" represents the problems of

* Tel.: +216 98 277 018
E-mail address: manel.zekri@fst.utm.tn

1877-0509 © 2019 The Authors. Published by Elsevier B.V.

10.1016/j.procs.2019.09.231

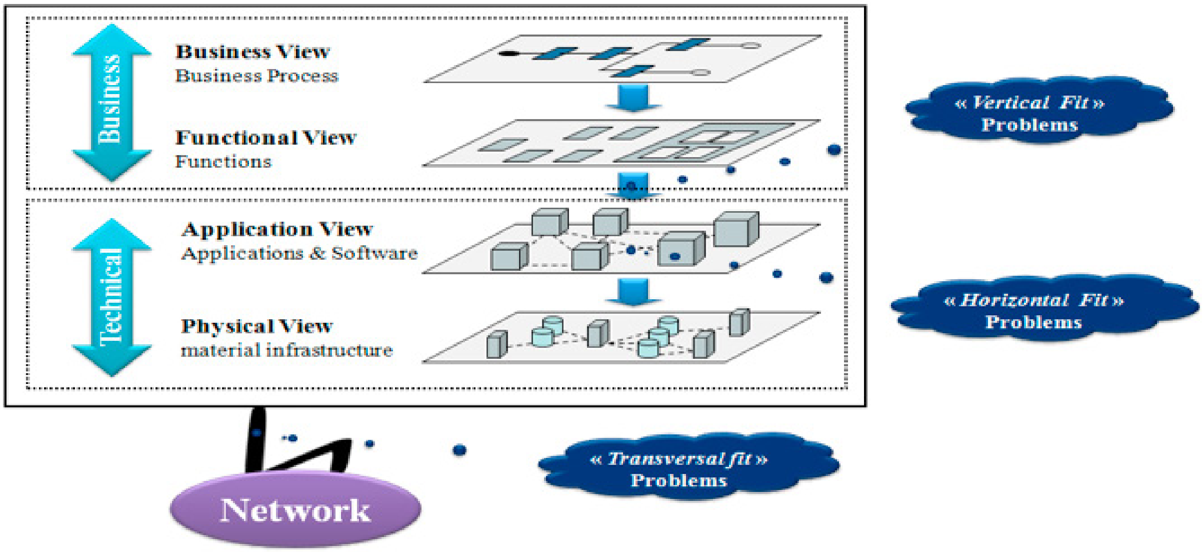

Fig. 1: Enterprise Architecture I.S reference model [7]: three fit problems [19] [20].

integration from a business infrastructure, which is abstract, to a technical infra-structure, which represents implementations. The "horizontal fit" translates not only the softwares problems of identification (induced by the "vertical fit" problems) that can cover the entire infrastructure of the company's business, but also the intra-applicative communications problems (internal interoperability); the purpose being to ensure the interactions between softwares of the same technical infrastructure in the company. The "transversal fit" translates the inter-applicative communications problems (external interoperability carried out dynamically through a network). Therefore, we define an IS that supports four dimensions which are operational, organizational, decision-making, and cooperative dimensions [20]. As for the decision-making dimension of the BPIS, data warehouse is mainly used for making better decisions that can improve the business of organizations and optimize bid process [9], [2]. Hence, building a data warehouse is a complex task that aims at satisfying the needs of decision makers. One of the key points to the success of a bid data warehousing project is the design of the multidimensional schema [1][6]. In this paper, we propose a method that uses ontologies for multidimensional design of bid data warehouses from an operational data source [5]. In addition, we present an ontology-based method for data modeling schema that eventually covers different phases of the data warehouse lifecycle, and takes into account the users by considering their personalized needs as well as their bid knowledge. This work is organized as follows.The second section is a description of the solution to the"three fit" problems. The third section shows the related work: extending Multidimensional Ontology [11]. The fourth section describes the design approach. then we end this work by a conclusion and some prospects of future works.

2. Resolution of the "three fit" problems: BPIS integrated, flexible and interoperable

We have identified four dimensions to deal with "three fit" problems [20],[19]:

- the operational dimension that serves to specify the bid exploitation process by undertaking a specific project;
- the organizational dimension which allow the management and saving of the set of skills and knowledge that the company acquired in previous bid auctions;
- the decision-making dimension which aims at optimizing the right decision-making by the company in bid auctions; and

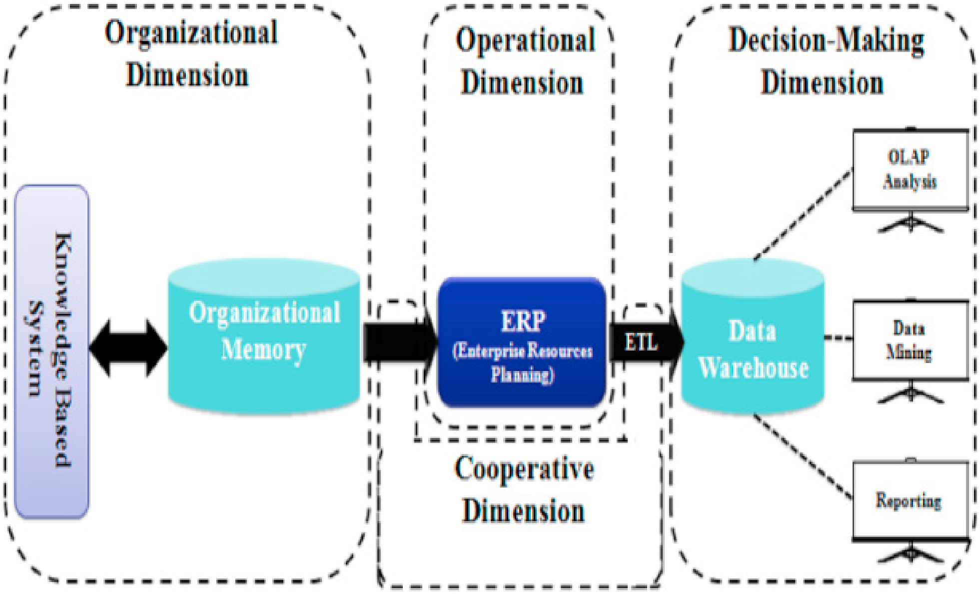

Fig. 2: The Bid Process Information System (BPIS) [20].

- the cooperative dimension which aims at ensuring intra-entreprise communication (internal interoperability) and at planning the inter-entreprise communication on demand, in order to achieve a common goal (dynamic interoperability).

We suggest fulfilling these dimensions while relying on the following hypothesis [20],[19] such as shown in Figure 2. The ERP (Enterprise Resources Planning) aims at build the techno-economic proposal of an offer in order to cover the operational dimension. In addition, the organizational memory covers the organizational dimension [10]. Moreover, a data warehouse will evaluate a set of solutions that make this proposal operational [16] to cover the decision-making dimension[20].

3. Related Work: Extending Multidimensional Ontology

In data warehouse design, different modeling techniques are used to represent the multidimensional concepts extracted from data sources, as well as the sources themselves. It can be ER diagram, UML diagram or graphs [8], etc. Unlike ontologies, which are ready for computing, these techniques are conceptual formalizations intended to graphically represent the domain and not used for querying and reasoning. This work is a continuation to a previous research [8], [22] and it aims at integrating ontologies in the data warehouse design process [13]. Therefore, the starting point was the meta-model of data warehouse scheme [22]. The multidimensional ontology [21] is a representation of knowledge dedicated to decision-making dimension [20], [19]. It specifies the multidimensional concepts and their semantic and multidimensional relations[14]. Its use covers different steps of the data warehouse lifecycle. During these steps, it assists the designer to solve problems of data sources heterogeneity. In the OLAP requirement specification step, the decisional ontology helps to validate the multidimensional concepts (fact, measure, dimension, etc.) and relations between these concepts. It, also prevents associations between concepts not semantically associable (e.g. associating fact-to-fact, dimension-to-dimension, hierarchy-to-fact, etc.). Various approaches were proposed to guide creating ontologies [1] [6], we mainly based the multidimensional ontology construction process on the approach proposed in [3]. While other techniques (mentioned above) for representing multidimensional knowledge are an appropriate choice for their respective approaches, their use ends when the task of designing the data warehouse is accomplished. On the other hand, the multidimensional ontology can still be useful. It can cover the various phases of the life cycle of a decision-making dimension [20], [19] , that is to say, from the requirements specification, the design of data warehouse and data marts [4], until the exploitation and evolution phases. This is convenient because the recent DBMSs allow storing ontologies alongside data in the same database structure [2]. Such database is called OBDB (ontology-based database). Ontologies are scalable and extendable, and they showed their effectiveness for IS and requirements specification [3]. More concretely and in the same way, as ontologies are used for clarifying the semantics of data sources, they are used to identify and manage semantic conflicts between concepts. This allows us to add as many extensions as needed to the multidimensional ontology to cover the various phases of the life cycle of a decision-making dimension [20], [19] . To demonstrate this aspect, we have proposed an extension that represents the operational data source conceptual schema, in this case, an ER diagram [22].

4. The data warehouse lifecycle: ontology-based method for data modeling schema

In this section, we will present an overview of the first phase of the approach which we adopted. It is progressive and iterative. The assistance of the designer throughout this construction is optional. The approach can be executed autonomously, but the intervening of a designer during validation steps is recommended and will result in better output.

A domain ontology contains knowledge that is a semantic representation of the multidimensional concepts. This representation is often complete because of its generality. Hence, multidimensional concepts which could be useful for the decision-makers will be contained in the domain ontology. We design the steps of the data warehouse lifecycle which is an ontology-based method for data modeling schema in Figure 3.

4.1. Multidimensional Relationships

After defining the concepts of the multidimensional ontology, we need to specify the relationships that exist between them. Each relationship is of the form Relation (X, Y), where Relation is a binary predicate, and X and Y are concepts. We define the relationships described below through the schema given in Figure 4.

Both multidimensional concepts and relationships are presented in the bellow:

- Is_Fact_ID (***FID, F***) where Fact_ID (***FID***), Fact (***F***) and ***FID*** is the Id of ***F***.
- Is_Measure (***M,F***) where Measure (***M***), Fact (***F***) and ***M*** is a Measure of ***F***.
- Is_Dimension (***D, F***) where Dimension (***D***), Fact (***F***) and ***D*** is a Dimension of ***F***.
- Is_Dimension_ID (***DID, D***) where Dimension_ID (***DID***), Dimension (***D***) and ***DID*** is the Id of ***D***.
- Is_Hierarchy (***H, D***) where Hierarchy (***H***), Dimension (***D***) and ***H*** is a Hierarchy of ***D***.
- Is_Level (***L,H***) where Level (***L***), Hierarchy (***H***) and ***L*** is a level of ***H***.
- Is_Attribute (***A, L***) where Attribute (***A***), Level (***L***) and ***A*** is an Attribute of ***L***.
- Is_Finer_Than (***Li, Lj***) where Level (***Li***), Level (***Lj***), ***Li*** and ***Lj***are from the same Hierarchy and ***Li*** has a finer granularity that ***Lj***.

With the aim of ensuring the availability of data, we consider the data sources that are represented in a conceptual data model for the production base. In the next step, we extract the multidimensional concepts. It is divided into three stages that are repeated for each multidimensional concept. Thus, we determine a set of potential multidimensional, using extraction rules.

4.2. Fact extraction

Facts describe the daily activities of the bid companies. These activities result in transactions and produce transaction objects. A transaction object is an entity registering the details of an event such as the payment of the bid proposition, etc. These entities are the most interesting for the data warehouse and are the basis for the construction of the fact tables. However, they are not all important. Thus, we must choose those that have an interest in the decision making. Usually a transaction object is a complex object containing multiple pieces of information. Therefore

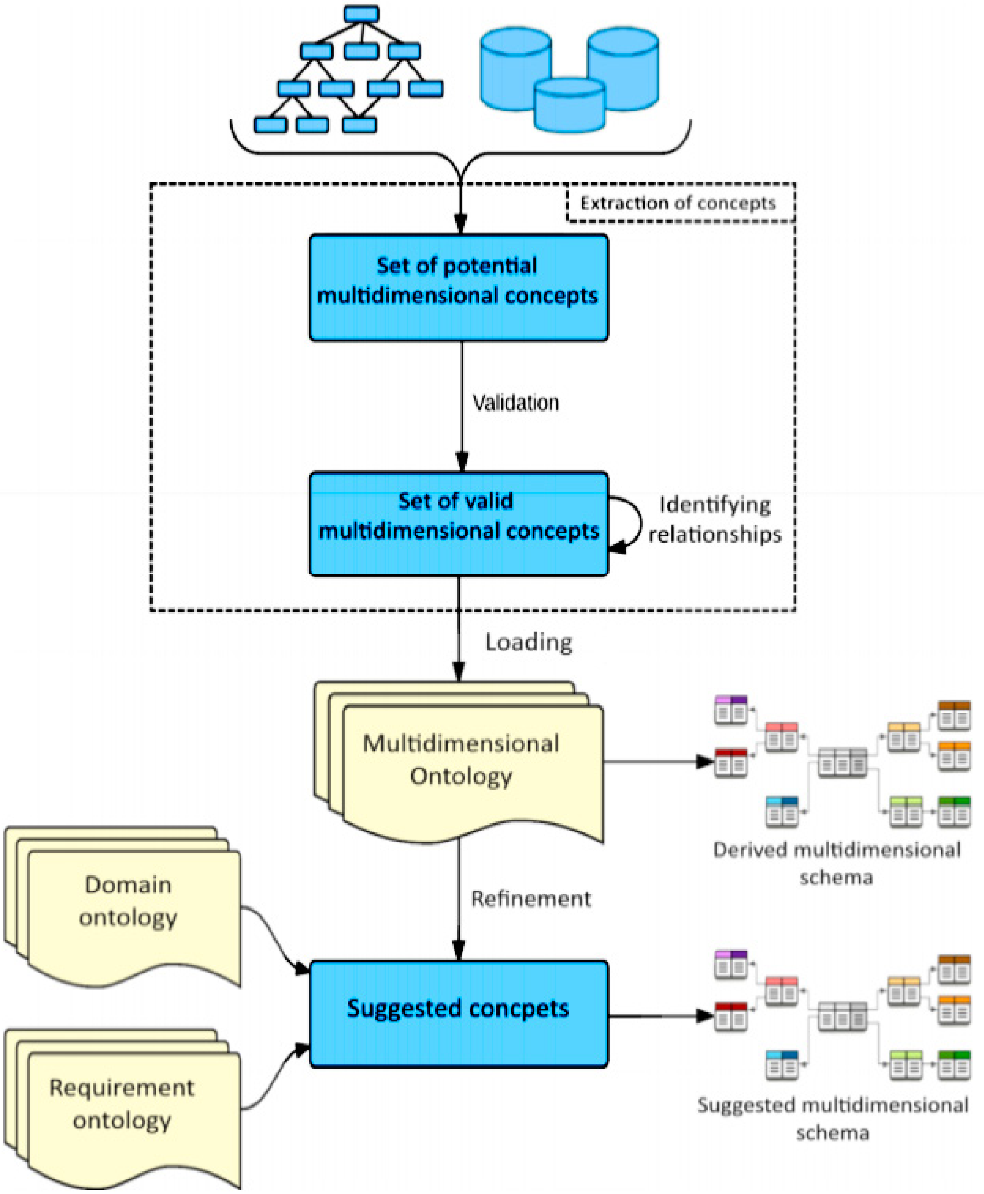

Fig. 3: The data warehouse lifecycle: ontology-based method for data modeling schema [21].

its modeling requires its decomposition into associated sub-objects. In the ER model, a transaction object may be represented in one of two forms:

- An entity connected to an association;
- Two entities linked by an association

In order to determine "***Fp***" (set of potential facts), we define the following heuristic:
HF: All transaction objects are potential facts. For each identified transaction object identified, we associate a more descriptive name, which will be the name of the fact. These facts are necessarily all pertinent, thus a validation phase where the designer may intervene is essential to retain a subset of valid facts (***Fv***).

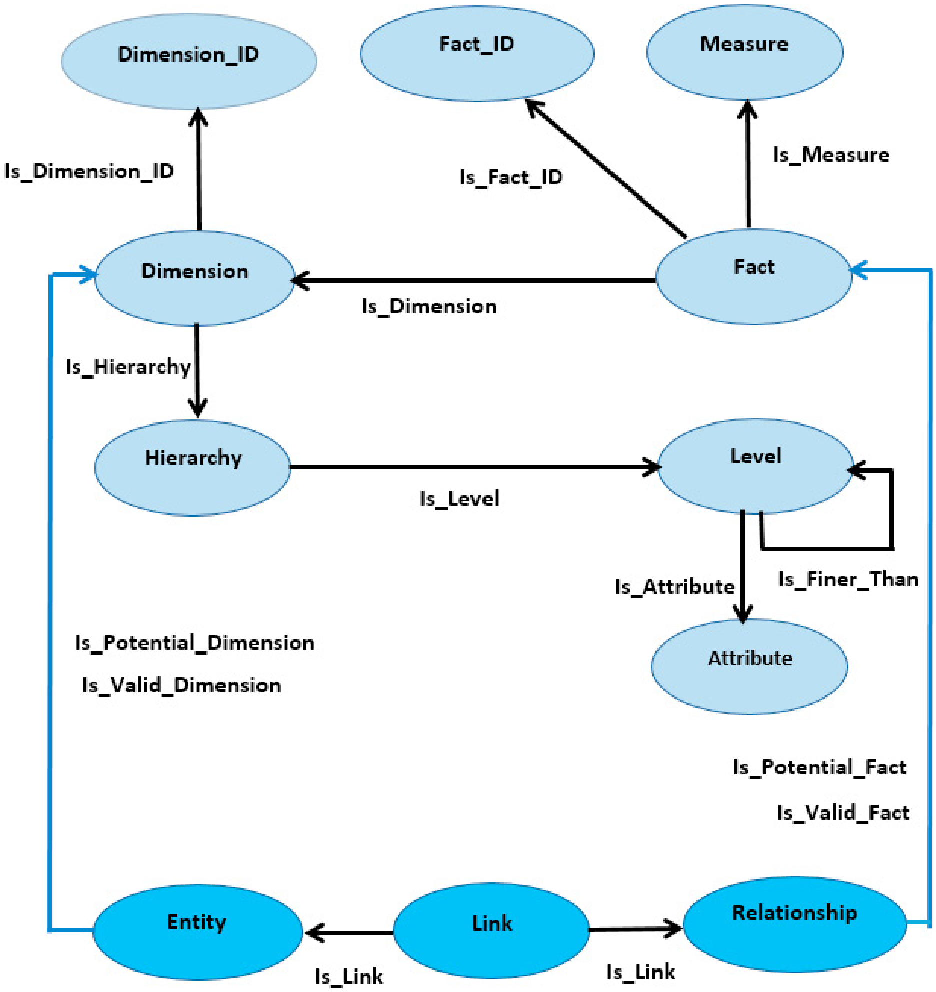

Fig. 4: Graphical representation of multidimensional relationships of ontology.

4.3. Measure extraction

As we previously stated, a transaction object is the result of the bid companies activities. Accordingly, the attributes may be measurements of a fact that are encapsulated in the transaction object. The following heuristics determine the potential measures:
Hm1: Mp (***fv***) contains the non-key numeric attributes belonging to the transaction object representing "***fv***".
Hm2: If the attribute is a Boolean, we add to Mp (fv) the number of instances corresponding to the valuesTrue and False of this attribute.

- Fv: a valid fact from the previous step;
- Mp (***fv***): the set of potential measures of "***fv***";
- Mv (***fv***): the set of measures of "***fv***" approved by the designer, which is a subset of "Mp (***fv***)".

The extraction of measures is also followed by a validation step by the designer in order to determine Mv (***fv***) which elements satisfy the following assertion:

$$\forall fv \in Fv, \forall m \in Mv(fv) \rightarrow Is\_measure(m, fv) \tag{1}$$

4.4. Decision extraction

Extraction of dimensions is based on a second type of object called base object. A base object determines the details of an event by answering the questions "who", "what", "when", "where" and "how" related to a transaction object. For example, the Bid Project event is defined by several components such as (Owner: who bought) and (Bid Proposition: we sold what). A base object completes the meaning of the event represented by a transaction object thus providing additional details. Each object corresponding to one of these questions, and directly or indirectly linked to a transaction object is a potential dimension for the fact representing the transaction object. The extraction of dimensions consists of determining the name, **ID** and hierarchy(s) through these heuristics:
Hd1: Any base object directly or indirectly connected to the transaction object of "***fv***)" is a potential dimension of "***fv***".
Hd2: All **IDs** of a base object obtained by **Hd1** is an "***id***" of "***d***".

- Dp (***fv***): the set of potential dimensions of "***fv***";
- Dv (***fv***): the set of valid dimensions of "***fv***"; which is a subset of "Dp (***fv***)";
- d: ***a*** dimension;
- idd: the "***id***" of a dimension "***d***",

The validation step produces two subsets **Dv** (***fv***) and **IDDv** (***fv***), satisfying the following assertion:

$$\begin{aligned} &\forall fv \in Fv, \forall d \in Dv(fv), \exists dimension\_id(did) \\ &\rightarrow Is\_dimension(d, fv) \wedge Is\_dimension(idd, d) \end{aligned} \tag{2}$$

4.5. Attributes extraction

We define the following heuristics to determine the potential attributes:
Ha1: Any attribute belonging to the base object containing "***idd***" is a potential attribute of "***d***".
Ha2: Any attributes belonging to a base object that generated a valid dimension "***dv***" is a potential attribute of "***dv***".
The validation step produces a set of valid attributes "***Av***", satisfying the following assertion:

$$\forall av \in Av(dv) \rightarrow Is\_attribute(av, dv) \tag{3}$$

Each extracted element becomes an individual (i.e. instance) of the concept that represents its role.

5. The utility of applying this approach in the knowledge of the bid process

Working on a specific bid implies the intervention of several collaborators. Certainly, these contributors exchange knowledge and information flows. However, its environmental differences lead to various representations and interpretations of knowledge ("horizontal fit" problems). Such failures are described in terms of five conflicts: the syntactic conflicts are the results of different terminologies used by stakeholders on the same domain. The structural conflicts are related to different levels of abstraction which aim at classifying knowledge within a virtual company (bid staff). The semantic conflicts concern the ambiguity that emerges due to the stakeholders reasoning in the development of the techno-economic proposal. Heterogeneities conflicts are due to the diversity of data sources. The contextual conflicts are mainly from environmental scalability problems. Thus, stakeholders can evolve in different environments. In this context, we can deduce that the multidimensional schemas permit to overcome the "horizontal fit" problems with these various conflicts and manage the knowledge of bid process.

6. Conclusion and Perspectives

In this work, we tried to define the characteristics of the decision-making dimension of the BPIS. Thus, we have presented an approach for representing data warehouse schema based on an ontology that captures the multidimensional knowledge. We discussed one possible for extending the multidimensional ontology to eventually cover different phases of the data warehouse life cycle. We focused on the design phase, and showed how the use of the multidimensional ontology combined with an extension can be beneficial. In the future we intend to continue to explore the possibility of extending ontologies by considering bid ontologies as extensions that could be used to improve the resulting data warehouse schema; in addition to real cases of the implementation of the approach.

References

[1] Abran, Cuadrado, Garca-Barriocanal, Mendes, Snchez-Alonso, and Sicilia. (2006) "Engineering the ontology for the SWEBOK: Issues and techniques. In Ontologies for software engineering and software technology." Springer, Berlin, Heidelberg: 103-121.

[2] Bellatreche, Valduriez, and Morzy. (2018). Advances in Databases and Information Systems. Information Systems Frontiers, 20(1): 1-6.

[3] Bentayeb, Maz, Mahboubi, Favre, Loudcher, Harbi and Darmont. (2012) "Innovative Approaches for efficiently Warehousing Complex Data from the Web. In Business intelligence applications and the web: Models, systems and technologies." IGI Global: 26-52

[4] Bonifati, Cattaneo, Ceri, Fuggetta Paraboschi.(2001). "Designing data marts for data warehouses. ACM transactions on software engineering and methodology." 10(4): 452-483.

[5] Borgo, Hitzler and Kutz (Eds.). (2018) "Formal Ontology in Information Systems": Proceedings of the 10th International Conference: IOS Press.(Vol. 306).

[6] Calero, Ruiz, and Piattini (Eds.). (2006) "Ontologies for software engineering and software technology." Springer Science Business Media.

[7] Fourrier-Morel, Grojean, Plouin, and Rognon. (2008). "SOA The SI architecture guide". Dunod, Paris.

[8] Gallinucci, Golfarelli, and Rizzi. (2018) "Schema profiling of document-oriented databases." Information Systems, 75: 13-25.

[9] Golfarelli, and Rizzi. (2009)" A survey on temporal data warehousing". International Journal of Data Warehousing and Mining, 5(1): 1-17.

[10] Gronwald, (2017). "Integrated Business Information Systems: A Holistic View of the Linked Business Process Chain ERP-SCM-CRM-BI-Big Data." Springer.

[11] Gruber.(1993). " A translation approach to portable ontology specifications." Knowledge acquisition, 5(2): 199-220.

[12] Guarino. (1998). "Some ontological principles for designing upper level lexical resources." arXiv preprint cmp-lg/9809002.

[13] Khouri, Boukhari, Bellatreche, Sardet, Stéphane, and Michael. (2012)."Ontology-based structured web data warehouses for sustainable interoperability: requirement modeling, design methodology and tool." Computers in industry 63(8): 799-812.

[14] Mazon, Lechtenbrger, and Trujillo. (2009)." A survey on summarizability issues in multidimensional modeling." Data Knowledge Engineering 68(12): 1452-1469.

[15] Ovchinnikova and Khnberger. (2006). "Aspects of automatic ontology extension: Adapting and regeneralizing dynamic updates'.' 72: 51-60.

[16] Phipps, and Davis. (2002). "Automating data warehouse conceptual schema design and evaluation." In DMDW (2): 23-32.

[17] Ralph Kimball. (1996). "The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses."

[18] Xavier Fournier-Morel, Pascal Grojean. (2018). "SOA le Guide de lArchitecture du SIG.P.C.R." Paris.

[19] Zahaf and Gargouri.(2014). "ERP Inter-enterprises for the Operational Dimension of the Urbanized Bid Process Information System." 6th Conference on ENTERprise Information Systems, Vol. 16. Journal of Procedia Technology, pp. 813-823, Elsevier Ltd, 15-17, Troia, Portugal.

[20] Zahaf and Gargouri. (2017). "Business and technical characteristics of the Bid-Process Information System". 16th International Conference on Information and Knowledge Engineering, CSREA Press, ISBN: 1-60132-463-4, pp. 52-60, 17-20, CSREA Press, Las Vegas, Nevada, USA.

[21] Zekri. (2015). "Automatisation de la conception et la mise en œuvre d'un entrepôt de données générique." Ph.D. Thesis. University of Tunis El Manar, Tunisia.

[22] Zekri, Marsit and Abdellatif. (2011). "A New Data Warehouse Approach Using Graph." IEEE ICEBE:65-70