An Ontology-Driven Personalization Approach for Data Warehouse Exploitation

Lama El Sarraj
LSIS UMR 7296
Aix-Marseille University
Marseille, France
lama.elsarraj@lsis.org

Bernard Espinasse
LSIS UMR 7296
Aix-Marseille University
Marseille, France
bernard.espinasse@lsis.org

Thérèse Libourel
Espace-Dev UMR 228
Montpellier 2 University
Montpellier, France
therese.libourel@univ-montp2.fr

***Abstract*— Data Warehouses (DW) resources are shared by users' from different backgrounds (e.g., domain, culture, education, profession). Those resources (e.g., OLAP queries, Excel files) are interpreted differently from a user to another. Unfortunately, misinterpreting data could induce serious problems and conflicts. To guarantee relevant interpretation of resources, additional semantic description of resources concepts is necessary. In this context, we present an Ontology-driven Personalization System (OPS) based on three connected ontologies: domain ontology, DW ontology and resources ontology. OPS return a set of personalized resources search based on users' domain and his recurring interests. In addition, resources are enhanced with a semantic description provided by the ontologies. This paper focuses on the methodology used to develop connected ontologies used by OPS.**

Keywords-data warehouse, ontology, personalization, decision support systems, decision making, healthcare institution management.

I. INTRODUCTION

Decision Support Systems (DSS) enables users to analyze and synthesize data according to different perspectives. Big companies need efficient DSS and seek to expand the number of their users. In fact, companies need to have flexible decision tools that include users' requirements and resources (e.g., Excel files, graphs, tables). Resources are shared by users' from different backgrounds (e.g., domain, culture, education, profession). Thus, resources interpretation depends on user backgrounds. We proposed in El Sarraj *et al.* [1] to use an ontology-driven personalization approach to facilitate the exploitation of Data Warehouse (DW) resources.

Generally a DSS uses a collection of Business Intelligence (BI) tools and applications to analyze, query and visualize a big volume of data from heterogeneous sources and domains stored in a DW. DW is the core of most DSS, it is "a subject oriented, nonvolatile, integrated, time variant collection of data in support of management's decisions" [2]. DW uses a multidimensional model that represents facts and their measurements, related to different dimensions, which are the axes of analysis. To facilitate the task of DW analysis and treatment, a subset of the DW is created, called Data Mart (DM). A DM is oriented to a specific business need or a particular user requirement. Most of the times, data mart are organized in a multidimensional structure [3]. Data are represented as a point in a multidimensional space, visualized like a data cube [4]. They give users the possibility to synthetize and analyze data from three (or more) dimensional arrays of values and various granularity levels. Based on this multidimensional model On Line Analytical Processing (OLAP) cubes enable the manipulation of data provided by the DW. In this paper, only the multidimensional table resource is considered.

In the DW field, taking user requirements into account is crucial for the success or the failure of the DW [5], especially when users belong to different domains. The exploitation level of DW, as well as the preliminary conception level, is mainly based and adapted to user requirements [6]. Most research works devoted on DW focuses on the design approach [7], [8], [9]. Even if these approaches are successful at the conceptual level knowledge about the DW resources is still needed. It is important that users understand the semantic of the information they analyze and have a visibility about other resources that could help them make efficient analysis.

Ontologies have already proved their utility to resolve semantic problems in DW domain. Ontologies are widely used in the DSS domain. First, they were used for DW design to facilitate the integration of data from heterogeneous sources. Indeed, DW are considered as data integration systems [10]. Then, researchers in this domain have widely used ontologies in different phases of the DSS, at the conceptual level [11], [12], at the Extract-Transform-Load (ETL) level [13], OLAP cube model [14] and OLAP queries [15].

The goal of this work is to develop an ontology-driven system for DW personalization to support users of various profiles to efficiently exploit a DW using existing DW resources. This paper focuses on the knowledge base component of such a personalization system. This knowledge base is composed of three ontologies: the first one is the domain ontology, the second one presents the schema of an existing DW, and the last one describes existing DW resources of a related DW.

This research concerns an existing DW used in the context of the "Program of Medicalization of Information Systems (PMSI)" to analyze healthcare institutions activity. The PMSI is part of the reform of the French health system. The PMSI is a device that enables quantifying and standardizing the data about the healthcare institutions activity. PMSI data are used to finance healthcare institutions according to their activity. This research has been financed by the public hospitals of Marseilles - Assistance Publique des Hôpitaux de Marseille (APHM).

The paper is organized as follows. Section II presents related works about DW personalization and introduces some elements about ontologies. Section III presents the problematic. First, it introduces the context of this research with a case study from healthcare domain. Then, it presents the aim of this research. Section IV presents the general architecture of the “ontology-driven personalization system” and the uses-cases supported by this system. Section V presents the methodology used to develop the knowledge base of the personalization system. Also, it presents in details the knowledge base component, the type of knowledge concerned, the models in UML and OWL of the three ontologies: Domain Ontology (O_D), Data Warehouse ontology (O_{DW}) and existing resources ontology (O_R). Section VI presents the mapping between knowledge base ontologies. Finally, we conclude and present some perspectives to this work.

II. Related work

This section presents different researches related to DW personalization approaches, which are mainly based on users’ profiles and recommendation techniques. This section also introduces some elements related to ontologies and their use in software development.

A. DW personalization based on user profiles

Researches works, based on user profiles, are usually associated to the "personalization" of DWs. After introducing and defining the concept of personalization in the context of DW, we present various existing approaches related to DW personalization. Then we compare and evaluate their relevance to our problem with the use of a DW.

Personalization is a customized and individualized description of a user or a group of users. Personalization system relies on users’ need, preferences and characteristics [16], and usually on a defined users profiles [17]. Although no consensus exists for the definition of a user profile, but a profile generally includes a set of features that is used to configure or adapt the system to the user. Thereby, the system provides personalized and efficient results [18] adapted to a user profile.

The authors in [19] developed a state of the art about user modeling based on system requirements. Other researches configure or adapt the personalization system to users’ preferences defined in their profiles [20], [6], [21]. These preferences may be related to their *contexts* defining application frameworks, as proposed in some researches in the DW domain [21], [22].

Bentayeb *et al*. [23], characterize the personalization of a DW based on user's profile from two perspectives, the definition of users profiles, which can be explicit or implicit, and the exploitation of these profiles to personalize the DW treatments:

- *Explicit implication* of the user at the profile definition level mainly needs to set parameters related to the recommendation process.
- *Implicit implication* of the user creates automatically a group of users profile based on a learning method and leads to an automatic transformation of the system.

The explicit definition is related to the configuration (customization, user modeling) and the implicit definition is related to adaptation (user profiling). In both cases, the profile may be operated by *recommendation* or by *transformation*, with automatic processes.

Jerbi *et al.* [24] distinguish three main objectives from DW personalization researches:

- *Customizing data sources schema* [22], [23], adapting the data structures to a specific needs of users.
- *Customizing queries visualization* [20], or representation [6], [21], [25].
- *Recommendation of OLAP queries* [26], [27] to assist in the exploration of the DW.

The first two objectives seem to affect data-centric personalization, in the first case by customizing the schema and in the second case by representing customized queries results. The third objective concerns the recommendation of a new method to treat data, queries.

B. DW personalization by recommendation or transformation

The personalization of the DW by recommendation is treated by various works such as [23], [26], [27], [28], [29], [30], [31], [32]. In these works we can distinguish two categories of recommendation methods: methods based on the *content* and methods based on *collaborative filtering*. The methods based on contents recommend similar objects. This recommendation is based on previous user actions while the methods based on collaborative filtering recommends items based on the interest and similar user.

The personalization of the DW by transformation, is mentioned by the authors in [20] that treats personalized visualization of OLAP queries. The authors in [33] propose a solution to evolve the DW schema according to user requirements. This method is based on “if-then” rules. The research work in [34] propose a solution to expand the DW architecture with event/condition/action rules. Finally, the authors in [21] propose customized

OLAP tables, based on users preferences and on analysis context.

To the best of our knowledge, no research uses ontologies to facilitate the exploitation of a DW. However, in our approach, we propose an ontology-driven personalization approach to facilitate the DW exploitation. The aim of our research is presented in details in Section III.

C. Ontologies

Ontologies have been used in the domain of knowledge engineering to facilitate requirements expression and detect incoherencies and semantic ambiguities between users [35]. Description Logic (DL) is a formalism used to build ontologies [36]. In this section, we define and propose a formalization based on DL for ontologies.

The first goal in the expected ontology is to provide resources to achieve automatic process, whether for machines interaction and interoperation with each other or with humans. Ontologies are used in several domains to resolve syntactic and semantic heterogeneity problems. In the software engineering field, ontology had been used first in the field of artificial intelligence systems and knowledge base systems, and then adapted to the problems of information retrieval. The use of ontologies in software engineering adds a wealth of knowledge to the systems.

Ontologies design requires the establishment of processes to extract the knowledge connected to a domain and make it suitable for both information systems and humans. In this context, several definitions of ontologies have been proposed in the field of software engineering. Gruber [37] defines ontology as a specification of a conceptualization "*[...] A conceptualization is an abstract, simplified view of the world that we want to represent*". This definition was extended by [38], which focuses on the formal characteristic of an ontology.

In our work, we consider the definition proposed by Jean *et al.* [39] a definition that characterizes an ontology as a referencing formal representation and consensus of all shared concepts. In this definition, the most important terms are:

- *Formal*: the ontology is interpretable by machines.
- *Explicit*: all concepts and properties of ontology are explicitly specified independently of any particular point of view or implicit context.
- *Referenceable*: any concepts described in the ontology can be referenced in a unique way from any context, in order to clarify the semantics of the referenced item.
- *Consensual*: the ontology is recognized and accepted by all the members of a community.

D. Formalization of the ontology

There's different existing languages to define ontologies. Ontology Web Language (OWL) is the standard language for representing ontologies [40], [41]. W3C consortium recommends OWL to define ontologies. The OMG [41] define the OWL meta-model.

DL language present the formalism underlying OWL language [36]. In DL, structured knowledge is described using concepts and roles. Concepts represent sets of individuals, and roles represent binary relationships between individuals.

A knowledge base described with DL is composed of two components: the Terminological Box (TBox), and the Assertion Box (ABox). The TBox specifies the intentional knowledge of the modeled domain.

In general, terminological axioms have the form of inclusions ($C \sqsubseteq D$) or equivalence ($C \equiv D$) such as (C, D denote concepts or roles).

Based on this definition, the ontology is formalized as 5-uplet [42] as follows:

O: <C, P, ClassPropt, ClassAssoc, Formal> such as:

- *C* represents the classes of the ontological model.
- *P* represents the properties of the ontological model, and P is partitioned into:
  - P_{value}: represents the characteristics properties.
  - P_{fct}: represents domain dependent properties.
- *ClassPropt*: C → 2P relates each class to its property.
- *ClassAssoc*: C → (Opr, Expr (C)) is an expression that associate to each class an operator (inclusion or exclusion) and an expression to other classes.
- *Formal*: is the formalism followed by the ontology model (e.g., RDF, OWL).

To facilitate the creation and the visualization of ontologies there are OWL ontology editors, such as Protégé [43] that manipulates ontologies (e.g., edit, load, define taxonomies). Protégé provides a detailed view for each concept in ontology. There are also visualization tools of ontologies, the most common ones are IsaViz [42], OWLViz [10], Growl [44], Welkin [39].

UML is a standard used to model information systems and software engineering. UML is a semi-formal formalism. UML is a graphical language for visualizing, specifying and building tool components. UML provides different diagrams (e.g., class diagrams). However, UML is not suitable to represent complex reasoning and inferences [46]. One of the major advantages of UML is that it is widely used in the academic environment and even by non-professionals. UML notations facilitate the knowledge visualization, especially of ontologies. Most informatics designers use UML to describe their diagram.

Several studies propose to model ontologies with UML [45], [46], [47], [48]. There are many commonalities between the formal languages of ontologies and UML. A comparison UML/OWL is studied [46], but the only drawback is the lack of semantics in UML. For those reasons, we can consider

UML as an adequate formal model for the representation of ontologies.

The process to create an ontology can be accomplished by: (i) modeling the ontology with UML to have a consensus between different users' experts, (ii) transforming UML model into an OWL model, to reason on the ontology.

E. Conclusion

Even if the personalization of DW is a recent field of research, various studies propose methods to treat this problem. In their study, the authors in [24] compare different works of DW personalization domain, they take in consideration three main aspects: (i) personalization objectives, customized schema or queries (the result or the visualization), (ii) user model type, that has been selected to define the user (rules, scores, preferences, annotations) and his contextualization, (iii) the algorithms implemented for DW personalization. These approaches do not seem totally adapted to our problematic. Indeed, the specificities of data exploited by a big number of users' from different backgrounds, require additional semantic to describe resources provided by DWs. We present in the next section the context and the aim of this work.

III. PROBLEMATIC

This section presents the context of our research introduced by a case study presenting a DW schema. Then we present the aim of our research.

A. Context

The application context of our research concerns the healthcare management applied specifically in the Program of Medicalization of Information Systems (PMSI) supported by the French government. In fact, PMSI is a French adoption of the concept created by the Professor R. Fetter (Yale university, United States of America) to finance hospitals. PMSI specifies the cost of sojourn based on the Diagnosis Related Groups (DRG) that classifies the hospitalization in homogeneous and coherent medico-economic groups. Today, this classification technique is used in France to finance healthcare institutions according to their activity.

To analyze PMSI data, specific DSS have been developed. DSS is mainly dependent on DW, and is used by different profiles of users. We identified two types of users profiles, the first type is related to a medical domain (e.g., doctors, pharmacists, biologists), while the second one does not (e.g., financial affaire managers, computer scientists, human resources).

In this context, in order to illustrate our problematic we consider the DW star schema presented in Fig. 1. This DW contains data concerning "PMSI activity". This DW schema is composed of a fact table, dimensions, and measures:

- Fact table = {F_Activity}
- Dimensions = {D_Time, D_Hospital_Structure, D_International_Classification_Of_Diseases, D_Exit_Mode, D_Diagnosis_Related_Groups D_Age}
- Measures = {Number of patient, Number of beds}

Note that a *pole of activity dimension* "Pole" is a set of medical services units. A *hospital structure dimension* "D_Hospital_Structure" is a set of "Poles".

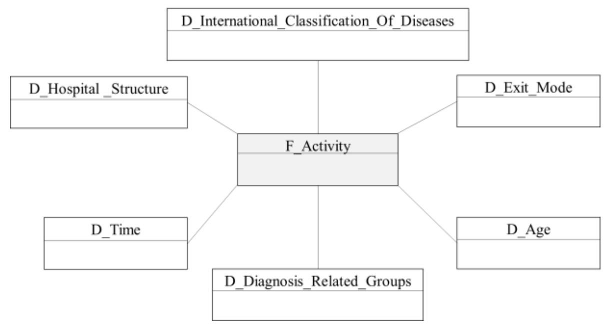

Figure 1: PMSI activity, DW Schema.

In this paper, we take the example of a Multidimensional Table (MT) (MT is defined in Section V.B) is denoted MT = (M, D), where M is the set of measure and D is the set of dimensions. We take an example of a multidimensional pivot table, presented in Fig. 2. For confidentiality issues this table is presented with fictive data:

- D1 = D_Hospital_Structure (dimension level "Pôle").
- D2 = D_Diagnosis_Related_Groups (attributes: DRG, TYPE DRG TITLE).
- M1 = number of patients (calculated measures: total of M1 per "Diagnosis Related Groups", total of M1 per pole, total of M1 for all Diagnosis Related Groups (i.e., DRG) and poles.

Periode : From january to mars

		Pôle 1	Pôle 2	Pôle 3	Total
DRG	TYPE DRG TITLE	288	318	519	1125
1	SURG CRANIOTOMY AGE >17 W CC	253	26	311	590
2	SURG CRANIOTOMY AGE >17 W/O CC	274	520	335	1129
3	SURG CRANIOTOMY AGE 0-17	225	319	212	756
4	SURG NO LONGER VALID	325	215	122	662
5	SURG NO LONGER VALID	125	138	118	381
Total		1490	1536	1617	4643

Figure 2: Example of a MT.

The DW presented in Fig. 1, offers several indicators to respond to users' needs (users from different profiles). In the context of the PMSI, we consider the following indicators:

1) *Offer indicators*: these indicators present the resources according to different dimension levels of a structure "structure", for instance:
   - The beds number of type "Medicine-Surgery-Obstetrics" to indicate the capacity to receive patients.
   - The main specialties by pole, to identify the types of diseases the hospital is able to treat.
2) *Needs and patient flow indicator (care consumption):* these indicators are mainly based either on the patient age or exit mode, for instance:
   - Describes the sojourn, analyze sojourns according to the group of diseases.
   - The main specialties of a pole.
   - Identify the population susceptible to be treated.
3) *Patient flow indicators:* these indicators presents the cause of the hospitalization and patient destination:
   - Where do the mothers come from?
   - What is the destination of the mother after the childbirth?

Various resources have been developed, to compute indicators from data, to analyze, visualize and aggregate data, elaborate dashboards and so forth. These existing exploitation resources are often numerous and of different type: formulas, OLAP requests, excel tables, and so on.

B. Aim of our research

Users have different profiles. In our context, for example, they belong either to the medical domain or to other domains. DW resources are numerous and complex, it is not easy for users from different domains to find relevant resources. In addition, existing resources do not have the same significance for these users from different profiles. In this context, we noticed many difficulties. We identify a *semantic lack related to DW concepts*: dimensions definition, measurements calculation methods and resources sources. Because of this semantic lack, the users cannot understand the usage of the DW resources that may respond to their needs. On the other hand, there is vocabulary *heterogeneity in query expression*: users do not belong to the same domain. They do not have the same vocabulary background. They do not express their need with the same terms (e.g., number of sojourn could be expressed as number of venue). Finally, concerning analysis needs, most of the time, users need to analyze many resources to make a decision. In big institutions, like the APHM, big numbers of resources make this task complicated. Thus, users need to have a global vision about resources responding to his requirements (e.g., calculation date, sources, criteria considered to calculate an indicator).

Consequently, to find, understand and choose relevant resources is a difficult task for users. Our challenge is to support users from heterogeneous domains in the exploitation of the existing resources. To this purpose, we propose to develop a personalization system supporting the users to exploit DW resources. We should note that our proposal is not limited to the healthcare domain. It can be used in other business contexts where users are from heterogeneous domains. In general, this is the case in big institutions.

The Ontology-driven Personalization System (OPS) is dedicated to support users from heterogeneous domains to exploit existing DW resources. This support is based on a knowledge base describing the domain (in this paper, we consider PMSI domain), the DW schema and the resources description. The following section presents the architecture of OPS and three scenarios of user support possibilities.

IV. An Ontology-Driven Approach for DW Personalization

The OPS supports users from heterogeneous domains to exploit existing DW resources. This support is based on a knowledge base that takes in consideration user domain, the DW schema and resources description. In order to provide such a personalization system, we developed an ontology-driven approach. In this section, we present first the general architecture of our ontology-driven personalization system, and then we present some uses-cases supported by OPS system.

A. General architecture

The general architecture of our OPS is illustrated in Fig. 3. OPS take in consideration information's collected from different DW construction levels. For example, at the conceptual level stores the DW schema.

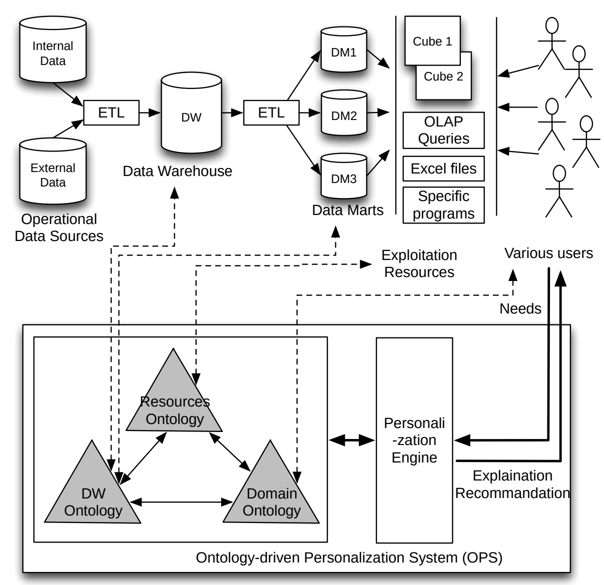

Figure 3: Ontology-driven Personalization System Architecture.

The two main components of our OPS are:

- *Knowledge base*: is an OWL database based on three related ontologies described in order: Domain (Hospital management - PMSI), DW schema (conceptual model), and existing DW resources.
- *Personalization Engine (PE)*: is the sub-system that personalizes users' interactions; the user expresses his needs to the OPS and the system provides semantic explanations or DW resources recommendations. This issue is based on the reasoning of the three ontologies.

B. *Use cases of the Ontology-driven Personalization System*

Several scenarios of user support functionality of OPS have been defined to develop and test the OPS. Each scenario corresponds to a user need expressed by a request addressed to the OPS (input). The OPS responds to the user with an explanation or a recommendation (output) depending on the nature of the expressed need. Examples of use-cases or expressed needs include:

1) ***Use-case 1:***

Entry: DW concept.

Output:

- *Domain concepts* - What are the existing measures to analyze a domain concept?
- *DW schema concepts* - What is the DW related concepts, measures: What are the different measures related to an analysis axe? What are the different analysis axes related to a measure? What are the measures that could be analyzed over a dimension?
- *Resources concept* - What are the existing resources to analyze a measure?

2) ***Use-case 2:***

Entry: Resources concept.

Output:

- *DW schema concepts* - What is the DW that provides a resource?
- *Domain concepts* - What are the existing resources to analyze a domain concept?
- *Resources concept* - What are the existing resources to analyze a measure?

3) ***Use-case 3:***

Entry: Domain concept.

Output:

- *Domain concepts* - What are the related domain concepts?
- *DW schema* – What are the domain concepts related to DW concepts?
- *Resources concept* - What are the resources related to a domain concept?

These use-cases are treated in the OPS by the PE reasoning on one or more ontologies. Fig. 4 illustrates the connection between ontologies and the users.

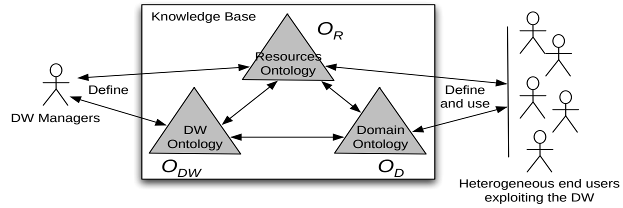

Figure 4: On the use of ontologies.

We distinguish two types of users:

- *The DW manager user*: he is in charge of the DW management and exploitation. He is mainly interested about the O_{DW} and the operational resources of the O_R.
- *The end-users*: they are heterogeneous; they search for resources that respond to their need. They expect resources and recommendations from the OPS to exploit the DW. These end-users express their needs using concepts belonging to the O_D and the conceptual resources, part of the O_R.

In this paper, we focus on the methodology used to develop the knowledge base composed of three ontologies: O_D, O_{DW} and O_R.

V. Knowledge Base Components

This section presents the Knowledge base of our OPS. This knowledge base is composed of three ontologies: O_D, O_{DW} and O_R. We present each of these three ontologies, the knowledge concerned, the methodology used to develop it and the models obtained in UML or in OWL.

To elaborate these ontologies, we use the ontology editor OWLGrEd. OWLGrEd uses a textual syntax OWL Manchester to create, edit and view an ontology [51]. OWLGrEd provides a comprehensive overview of OWL ontology with UML. OWLGrEd visualizes OWL classes as UML classes, data properties as attributes of classes, object properties as associations, individuals as objects and cardinality restrictions, associations between domain classes as UML cardinalities. To visualize other constructors of OWL, OWLGrEd enriched the UML class diagram with new notations [50], [51].

A. *Domain ontology (O_D)*

This sub-section present the description, the elaboration method and some exploitation results of the O_D.

1) Description:

The O_D gathers and streamlines the vocabulary related to a domain. Domain concepts are semantically related and defined in the ontology.

2) Elaboration methodology:

There are two solutions to obtain O_D. We can extract a part of existing O_D or create a new one manually. In the first case, the ontology can be extracted from the existing ontology using ProSé plugin available with Protégé editor, it ensures the completeness of the extracted ontology [52]. As no O_D exists concerning "PMSI domain" we develop a new one.

To develop this O_D we decided to use UML, because this language is more user friendly for domain experts, and makes the validation process of the ontology easier. The methodology used to elaborate this ontology is illustrated in Fig. 5.

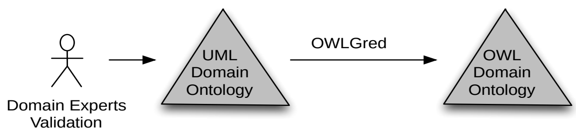

Figure 5: Domain Ontology Development.

Fig. 6 presents the O_D schema with UML. This schema is inspired from the model studied and presented in [53].

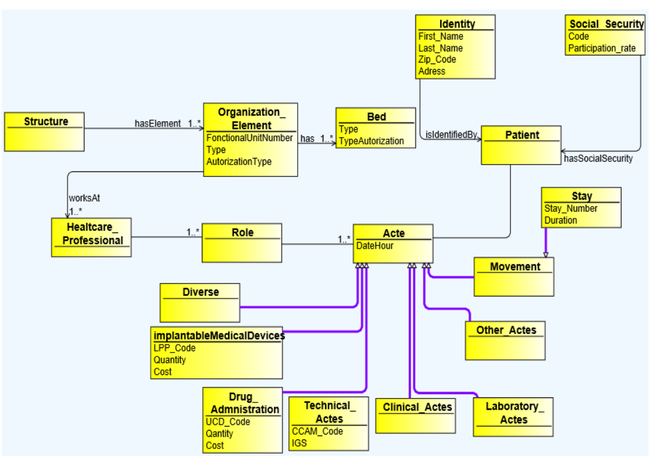

Figure 6: O_D schema presented with OWLGred.

This model is enriched and validated by domain experts. This ontology is presented here with the OWLGred tool.

3) Results:

The schema in Fig. 6 is used to validate the O_D with domain experts. Then, the UML schema is exported to OWL via the OWLGred tool. The O_D in OWL is visualized with Protégé in Fig. 7.

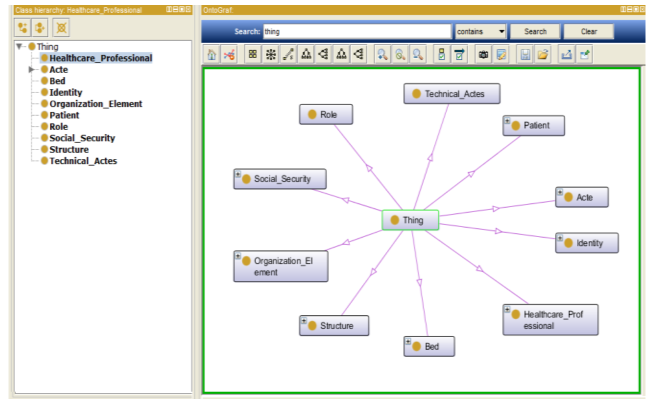

Figure 7: O_D presented with Protégé/OntoGraf.

This O_D is connected to other ontologies with semantic relations. O_D describes existing domains. OPS gives the possibility to visualize ontology concepts and relations between them, either they belong to the same ontology or not.

B. DW ontology (O_{DW})

This sub-section present the description, the elaboration method and some exploitation results of the O_{DW}.

1) Description:

Multidimensional model associated to the DW organizes data into facts and dimension. The O_{DW} concerns the DW conceptual schema. Facts represent the subject of analysis and dimensions represent the axes of analysis. Fact table is the center of the multidimensional model. It stores elementary indicators, called measures. Dimensions can form hierarchies, structured in different granularity levels.

2) Elaboration methodology:

To construct the O_{DW} we use a specific process. The first step of the process starts with the creation/extraction of the ROLAP structure of the DW (metabase) based on the SQL script of the relational data base of the DW. Then we annotate the tables with the multidimensional concepts (e.g., fact, dimension).

The atomization of this transformation from the conceptual model of the DW (the script SQL of the create table) to OWL is based on the research work of Prat *et al.* [54], Fig. 8 presents the O_{DW} development process. The research work of Prat *et al.* [54] defines a multidimensional meta-model, the concepts of OWL-DL, and transformation rules for mapping a multidimensional model into OWL-DL ontology.

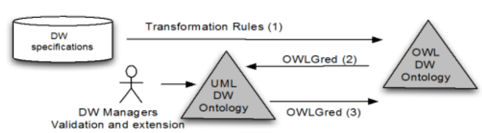

Figure 8: O_{DW} development process.

To generate the O_{DW} in OWL, the transformation rules proposed by Prat *et al.* [54] are adapted to our problematic. To validate and extend the model with DW manager the ontology is presented in UML. OWLGred tool translate the ontology from OWL script to UML. This process is illustrated in Fig. 8.

3) Results:

For the transformation of the O_{DW} from OWL intto UML we used OWLGred tool. Let's take the example of the DW schema Fig. 1, in the O_{DW} the dimension: D_Time is presented as a concept A_Time_Dimension. This concept have different dimension level Day, week, month and year. Those concept are presented with OWLGred (Fig. 9).

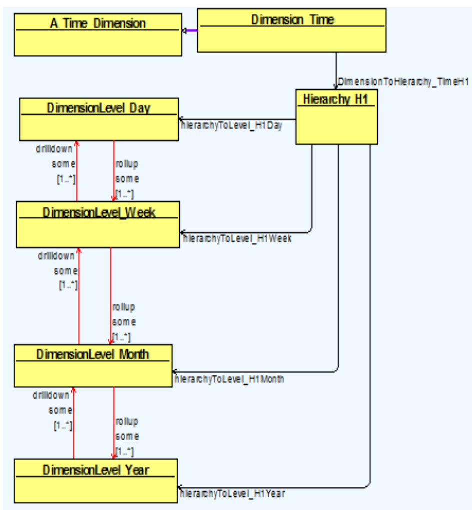

Figure 9: Dimension "D_Time" schema presented with OWLGred.

After the transformation of concepts from UML to OWL, O_D are visualized with Prtoégé/OntoGraf in Fig. 10.

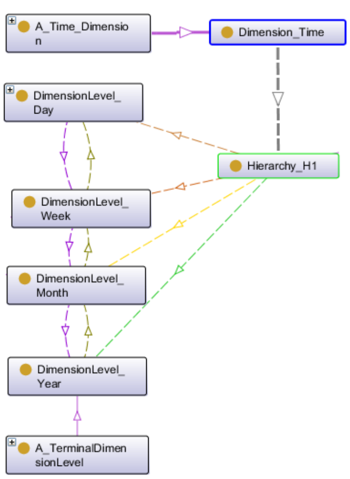

Figure 10: Dimension "D_Time' presented with Protégé/OntoGraf.

The O_{DW} is connected to other ontologies with semantic relations. This ontology presents the DW structure. It is mainly used by the DW manager.

C. Resources Ontology (O_R)

This sub-section present the description, the elaboration method and some exploitation results of the O_R.

1) Description:

Even if the multidimensional model is based on the metaphor of the cube or hypercube, the most common structure of the visualization is the MT presented in Fig. 2, which provides data presented in two axes of analysis [55], [3] enabling the visualization of a slice of the cube. Note that, other visualization possibilities exist to present the DW data (e.g., histograms, graphs).

Resources are related to the DW and are defined by the DW managers. To understand a resources components a user needs to have description information (e.g., calculation method, unit of measure, calculation period, date of creation, date of update, date of validity, objective, definition and the relation with the DW). We identified two types of DW resources:

- *Operational resources:* they concern the direct exploitation of DW, the resources requires an execution before being used for analysis (e.g., OLAP queries). They are used by the DW manager.
- *Conceptual resources:* they are user-oriented, they are resources used by the end-users (e.g., Excel files).

2) Elaboration methodology:

To develop O_R, as for O_D, we use UML for the same reasons. The conceptual resources (user-oriented resources) are validated by domain experts/users, and the DW managers validate the operational resources. The methodology used to elaborate this ontology is illustrated in Fig. 11.

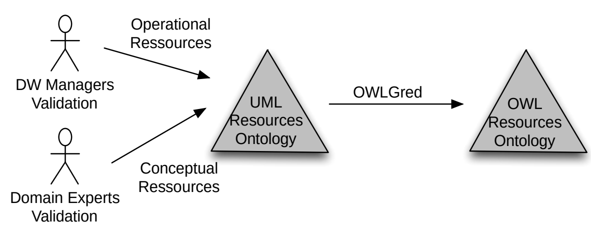

Figure 11: O_D development process.

Once the O_R expressed in UML class diagram, is validated with domain experts, it is transformed it into OWL with OWLGred tool (Fig. 12).

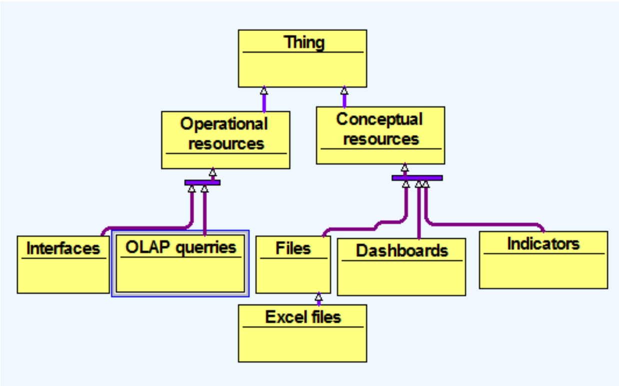

Figure 12: Extract of O_R schema, presented with OWLGRED.

3) Results:

The ontology O_R in OWL is visualized with Protégé tool in Fig. 13.

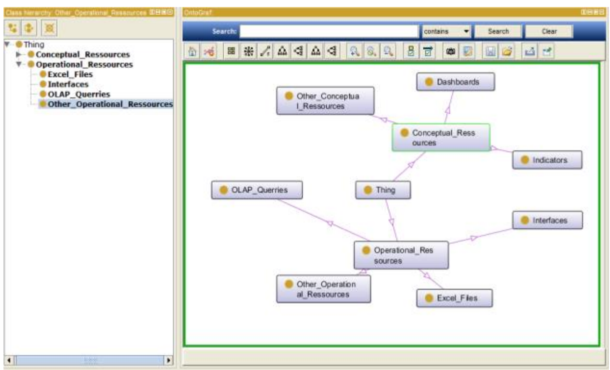

Figure 13: O_R presented with Protégé/OntoGraf .

This O_R is connected to other ontologies with semantic relations. This ontology enhances resources with descriptions. This ontology is mainly used by end-users.

VI. MAPPING ONTOLOGIES

The knowledge base of OPS is composed of three ontologies: O_D, O_{DW} and O_R. We formalize our ontology by the quadruple < O_{DW}, O_R, O_D, Map> where:

- *O_D* is the O_D that provides a schema about the domain.
- *O_{DW}* is a DW schema that describes DW schema.
- *O_R* is a resources ontology that describes the resources related to the DW.
- *Map* is the mapping between O_{DW}, O_R and O_D that establishes the connection between domain concepts, the DW and the resources components.

These mapped ontologies can be used for many purposes with OPS. On the one hand, to give a vision about the relation between DW resources and domain concepts, and on the other hand, to propose to users other related resources to make analysis based on reasoning technologies.

In this section, we focus on the *mapping* of these ontologies permitting this reasoning. We describe the mapping process. Then, we define the mappings between the three ontologies.

A. On the Mapping process

Considering two ontologies O_S and O_T, a *mapping* M between O_S and O_T, is a (declarative) specification of the semantic overlap between O_S and O_T at the concept level (Tbox). This *mapping* can be one-way (injective) or two-way (bijective). In an injective mapping, we specify how to express terms in O_S using terms from O_T in a way that is not easily invertible. A bijective mapping works both ways, i.e., a term in O_T is expressed using terms of O_S and the other way around. In ontology engineering, the following processes are pre-defined [56]:

1) *Ontology Merging* concerns ***creation of one new ontology from two or more ontologies***. In this case, the new ontology unifies and replaces the original ontologies. This often requires considerable adaptation and extension of the ontology.
2) *Ontology Aligning* brings the ontologies into mutual agreement. The ontologies are kept separate, but at ***least one of the original ontologies is adapted***, such as the conceptualization and the vocabulary match in overlapping parts of ontologies.
3) *Ontology Mapping (or relating ontology)* specifies how the concepts in different ontologies are related in a logical sense. This means that the original ontologies had not changed, but that ***additional axioms describe the connection between the concepts***. Leaving the original ontologies unchanged often implies *only a part of the integration*, because major differences may require adaptation of the ontologies.

As each of these ontologies can evolve, we do not choose the merging strategy to limit the impact of evolution changes. We prefer to keep three separate ontologies to limit the changes only to the connection (mapping) between them if necessary. Consequently, in our case, we have opt for *Ontology Aligning* or *Ontology Mapping* processes as defined before.

B. Concerned mappings

In our case, we considered three different mappings connecting these three ontologies two by two, depending on the connection between users and ontologies (Fig. 14).

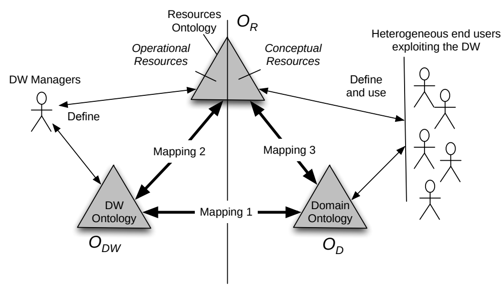

Figure 14: Different mappings between the three ontologies.

Mapping the three ontologies is necessary to facilitate the navigation between them. The Mapping 1 supports the connection between O_{DW} and O_D, Mapping 2 supports the connection between O_{DW} and the operational resources of the O_R, and, finally, Mapping 3 supports the connection between O_D and the conceptual DW resources of the O_R.

Ontology Aligning or *Mapping* processes related to these three mappings concerns: first searching similarities between ontologies, and then specifying mappings between ontologies. In our case, these two tasks are performed in a *manual manner using Protégé.*

1) *Mapping 1: $O_{DW} – O_D$*

This mapping is the first mapping to consider, because it is closely related to the DW design: a concept of the O_{DW} can be related to one or more concept(s) of the O_D, and one concept of the O_D can be related to one or more concept(s) of the O_{DW}.

2) *Mapping 2: $O_{DW} – O_R$*

This mapping can be considered as an extension of the O_{DW} towards operational resources of O_R: a concept of operational resource can be related to one or more concept(s) of O_{DW} (e.g., OLAP Query concept can be related to fact and dimension concepts). On the other side, a concept of the DW schema can be related to one or more concepts(s) of operational resources. For example, a measure can be implied in OLAP Query and Excel file. The O_{DW} concepts and O_R concepts concerned by this mapping are the lower classes of the respective ontology.

3) *Mapping 3: $O_D – O_R$*

Mapping 3 is deduced. The relation between O_D and O_R is identified through a process of deduction based on the transitive relation between O_D and O_{DW}. We present in Table I an example with OWL-DL.

TABLE I. CONCEPTS AND INFERRED CONCEPTS WITH OWL-DL.

Ontology	Concept
O_{DW}	A_Hospital_Structure_Dimension ⊆ A_Dimension
O_D	Structure
O_R	Resources1
O_{DW} - O_D	A_Hospital_Structure_Dimension ≡Structure
O_{DW} - O_R	Resources1ToDimension_Structure T ⊆ ∀Resources1ToDimension_Structure.Structure T ⊆ ∀Resources1ToDimension_Structure⁻.Resources1

This example presents the ontologies and their concepts "O_{DW} concepts", "O_D concept", "O_{DW} and O_D related concepts", "O_R concept" and finally "reasoning result concepts between O_D–O_R".

VII. VALIDATION EXAMPLE

To illustrate our proposal we suggest to respond to "Use-case 3" questions, we'll use OntoGraf [57] to visualize the ontologies' concepts. Fig. 15 shows the results of the search done on the mapped ontologies.

To show the definition and the concepts related of "DRG". The user enters "DRG".

Entry: Domain concept "DRG".

Output:

- *Domain concepts (from O_D):* the concept defining the 'Diagnosis related groups' (the user can access to the concept definition).
- DW schema element *(from O_{DW})*: the concept presenting a dimension 'D_DRG', note that D_DRG is a subclass of Dimension (Dimension ⊆ D_DRG).
- Resources concept *(from O_R)*: the concept identifying a resource 'Resource_Activity_Pole_DRG', this concept describes a multidimensional table representing data about PMSI activity per DRG and per Pole).

The benefits of a connected ontology is the information that it provides to describe a resource. The returned information is not only from O_R, it is also about connected concepts from O_D, and O_{DW}.

This section, presented preliminary test done by DW manager to define and validate the ontologies. However, end-user uses OPS system to search for resources that respond to his needs. OPS is based on O_D, O_{DW} and O_R connected ontologies to visualize the description of each resource.

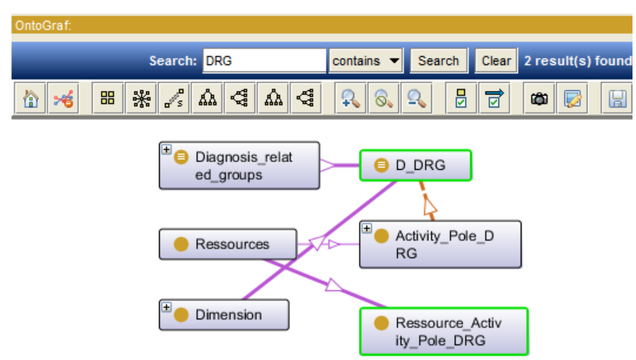

Figure 15: Example, retrieve "DRG" concept from the ontology Protégé/OntoGraf.

Thus, in the real application, OPS returns a resource with a set of information's form ontologies concepts describing the resource. For example, the resource presented in Fig. 2 will be visualized with a set of descriptions, presented in Fig. 16.

		Pôle 1	Pôle 2	Pôle 3	Total
DRG	TYPE DRG TITLE	288	318	519	1125
1	SURG CRANIOTOMY AGE >17 W CC	253	26	311	590
2	SURG CRANIOTOMY AGE >17 W/O CC	274	520	335	1129
3	SURG CRANIOTOMY AGE 0-17	225	319	212	756
4	SURG NO LONGER VALID	325	215	122	662
5	SURG NO LONGER VALID	125	138	118	381
	Total	1490	1536	1617	4643

Objectif : number of patients per DIAGNOSIS, per POLE, per PERIOD.
Resource administrator: X Dupont.
Considered criterion: the considered diagnosis classification is the DRG used in the context of the PMSI, ...
Criterion defined by: Dr. Y Mouri, X Lalio.
Validity date : January to mars.
Update date: 01/04/2014.
Keywords: DRG, PMSI, Pole, Total, number of patients, period.
File name: number_of_patients_per_DRG_pole_period.xls

Figure 16: Example, resource with description from the three ontologies.

OPS have a user-friendly interfaces that offers several functionalities to end-users, for example, resources description or personalized resources retrieval.

VIII. CONCLUSION AND FUTURE WORK

Ontologies are used in several domains to resolve syntactic and semantic heterogeneity problems. They facilitate the management of data, clarify and give a sense to ambiguous concepts. In a healthcare management context based on PMSI, numerous existing DW resources are provided to exploit a DW, they are shared by users from heterogeneous domains. These resources can be interpreted differently from a user to another. In addition, the personalization of specific and relevant resources to user is the aim of this research.

In this recent research field various studies propose different approaches to treat personalization problems, but they appear to be not adapted to our problematic. Indeed, the specificities of data related to healthcare management require semantic resources, in particular to tackle the heterogeneity of the users' profiles and domain complexity.

We have proposed an ontology-driven approach for a DW personalization system, in order to support heterogeneous users to explain or personalize (recommend) some existing DW resources adapted to their needs. This approach is based on a personalization engine using a knowledge base composed of three specific and related ontologies: O_D, O_{DW} and O_R.

In this paper, in progress of our proposition presented in [1], we focused on the elaboration of OPS knowledge base. We introduced the methodology used to develop each of the knowledge base ontologies, and presented the three ontologies models obtained in UML and in OWL languages. Then we have presented the mappings between these ontologies. To illustrate the use of this knowledge base to provide some resources explanations or recommendations to users, we have simulated the personalization engine using Protégé editor. We also queried and visualized ontologies with OntoGraf. We validated our approach by testing it on a simple user-case related to the healthcare domain, characterized by users' heterogeneity and domains complexity. We should note that our approach is not restricted to this domain; it can be applied in others domains.

This work leads to many other tasks. Future works on this research concern first the development of a user-friendly personalization engine of OPS, giving the user a friendly environment to query, provides resource explanation and resource personalization (recommendation). Then a validation process of the OPS has to be performed in a larger context, with DW managers' and end-users. Finally, we expect to study the impact of ontology evolution on OPS.

ACKNOWLEDGMENT

The authors wish to thank Sophie Rodier, Laetitia Malavolti, Bernard Guisiano, Jean-Francois Mirreti, Ghislaine Morvillez and Hélène Lherbet from Assistance Publique des Hôpitaux de Marseille (APHM), for their valuable contribution to this research.

REFERENCE

[1] L. EL Sarraj, B. Espinasse, T. Libourel, and S. Rodier, "Towards ontology-driven approach for data warehouse analysis. Case study: healthcare domain," The Eighth International Conference on Software Engineering Advances (ICSEA 2013) IARIA, Oct. 2013, pp. 426-431, ISSN: 2308-4235, ISBN: 978-1-61208-304-9

[2] W. H. Inmon, Building The Data Warehouse, John Wiley & Sons, 1992.

[3] W. Lehner, "Modelling large scale OLAP scenarios," Advances in Database Technology (EDBT), 1998, pp. 153-167.

[4] A. Bosworth, J. Gray, A. Layman, and H. Pirahesh, "Data cube: a relational aggregation operator generalizing group-by, cross-tab, and sub-total," Data Min. Knowl. Discov., pp. 152-159, 1995.

[5] S. Rizzi, A. Abello, J. Lechtenborger, and J. Trujillo, "Research in data warehouse modeling and design: dead or alive?," The 9th International Workshop on Data warehousing and OLAP (DOLAP 2006) ACM, pp. 3-10.

[6] M. Golfarelli, "From user requirements to conceptual design in data warehouse design – a survey," Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction, IGI Global, 2010.

[7] R. Kimball, and M. Ross, The Data Warehousing Toolkit, John Wiley&Sons, 1996.

[8] N. Prat, and J. Akoka, "From UML to ROLAP multidimensional databases using a pivot model," The 8th Journées Bases de Données Avancées, 2002, pp. 24.

[9] A. Tsois, N. Karayannidis, and T. K. Sellis, "Mac: Conceptual data modeling for OLAP," The 3rd International Workshop on Design and Management of Data Warehouses (DMDW 2001), Theodoratos 2001, pp. 2001.

[10] L. Bellatreche, D. Nguyen Xuan, G. Pierra and H. Dehainsala, "Contribution of ontology-based data modeling to automatic integration of electronic catalogues within

engineering databases," Computers in Industry Journal Elsevier, vol. 57, no. 8-9, pp. 711-724, 2006.

[11] L. Bellatreche, S. Khouri, I. Boukhari, and R. Bouchakri, "Using ontologies and requirements for constructing and optimizing data warehouses," Proc. of the 35th International Convention (MIPRO 2012), pp.1568-1573, May 2012.

[12] N. Prat, I. Megdiche, and J. Akoka, "Multidimensional models meet the semantic web: defining and reasoning on OWL-DL ontologies for OLAP," The 15th International Workshop on Data Warehousing and OLAP (DOLAP 2012) ACM, pp. 17-24, doi: 10.1145/2390045.2390049.

[13] S. Khouri, L. Bellatreche, and N. Berkani, "MODETL: a complete modeling and etl method for designing data warehouses from semantic databases," The International Conference on Management of Data (COMAD 2012), pp. 113.

[14] T. Niemi, and M. Niinimäki, "Ontologies and summarizability in OLAP," The 2010 ACM Symposium on Applied Computing (SAC 2010), pp. 1349-1353, doi:10.1145/1774088.1774378 http://doi.acm.org/10.1145/1774088.1774378.

[15] B. Neumayr, S. Anderlik, and M. Schrefl, "Towards ontology-based OLAP: datalog-based reasoning over multidimensional ontologies," The 15th International Workshop on Data warehousing and OLAP (DOLAP 2012), pp. 41-48, doi: 10.1145/2390045.2390053.

[16] Y. Ioannidis, and G. Koutrika, "Personalized systems: models and methods from an ir and db perspective," The 31st conference in the series of the Very Large Data Bases conferences (VLDB 2005), pp. 1365–1365.

[17] R. R. Korfhage, Information Storage And Retrieval, John Wiley & Sons, 1997.

[18] C. Domshlak, and T. Joachims, "Efficient and non-parametric reasoning over user preferences," User Modeling and User-Adapted Interaction, vol. 17, no. 1-2, pp. 41-69, 2007.

[19] A. Kobsa, "Generic user modeling systems," User Modeling and User-Adapted Interaction archive, Kluwer Academic Publishers Hingham, vol. 11, The Adaptive Web, pp. 49-63, 2001.

[20] L. Bellatreche, A. Giacometti, P. Marcel, H. Mouloudi, and D. Laurent, "A personalization framework for OLAP queries," The Eighth International Workshop on Data Warehousing and OLAP (DOLAP 2005), 2005, pp. 9-18.

[21] H. Jerbi, F. Ravat, O. Teste, and G. Zurfluh, "Management of context-aware preferences in multidimensional databases. ," the Third International Conference on Digital Information Management (ICDIM 2008), pp. 669-675.

[22] I. Garrigos, J. Pardillo, J.-N. Mazon, and J. Trujillo, "A conceptual modeling approach for OLAP personalization," in Conceptual Modeling-ER Verlag Berlin Heidelberg, 2009, pp. 401-414.

[23] F. Bentayeb, O. Boussaid, C. Favre, F. Ravat, and O. Teste, "Personnalisation dans les entrepôts de données: bilan et perspectives," The Fifth Journées sur les Entrepôt de Données et Analyse en ligne (EDA 2009), 2009, pp. 7-22.

[24] H. Jerbi, G. Pujolle, F. Ravat, and O. Teste, "Recommandation de requêtes dans les bases de données multidimensionnelles annotées," Revue des Sciences et Technologies de l'Information, Ingénierie des Systèmes d'Information, vol. 16, no. 1, pp. 133-138, 2011.

[25] D. Xin, J. Han, H. Cheng, and X. Li, A, "Answering top-k queries with multi-dimensional selections: The ranking cube approach," in VLDB, 2006, pp. 463-475.

[26] A. Giacometti, P. Marcel, and E. Negre, "A framework for recommending OLAP queries," Proc. The eleventh international workshop on Data warehousing and OLAP (DOLAP 2008) ACM, Pages 73-80, doi: 10.1145/1458432.1458446.

[27] A. Giacometti, P. Marcel, and E. Negre, "Recommending multidimensional queries," The 16th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2009), 2009, pp. 453-466.

[28] C. Sapia, "On modeling and predicting query behavior in OLAP systems," Proc. Workshop on Design and Management of Data Warehouses (DMDW), 1999, pp. 2.1-2.10.

[29] G. Chatzopoulou, M. Eirinaki, and N. Polyzotis, "Query recommendations for interactive database exploration," The 21st International Conference on Scientific and Statistical Database Management (SSDBM 2009), 2009, pp. 3-18.

[30] H. Jerbi, F. Ravat, O. Teste, and G. Zurfluh, "Applying recommendation technology in OLAP systems," The 11th International Conference on Enterprise Information Systems (ICEIS 2009), 2009, pp. 220-233.

[31] A. Giacometti, P. Marcel, E. Negre, and A. Soulet, "Query recommendations for OLAP discovery driven analysis," The 12th International Workshop on Data Warehousing and OLAP (DOLAP 2009) ACM, 2009, pp. 81-88, ISBN: 978-1-60558-801-8.

[32] E. Negre, Exploration Collaborative De Cubes De Données, François Rabelais Tours, France, 2009.

[33] C. Favre, F. Bentayeb, and O. Boussaid, "Evolution et personnalisation des analyses dans les entrepôts de données. Une approche orientée utilisateur," The 25th Informatique des Organisations et Systèmes d'Information et de Décision (INFORSID 2007), 2007, pp. 308-323.

[34] T. Thalhammer, M. Schrefl, and M. Mohania, "Active data warehouses: complementing OLAP with analysis rules," Data & Knowledge Engineering, vol. 39, pp. 241-269, 2001.

[35] A. Aybuke, and W. Claes, "Engineering and managing software requirements," Springer, 2005.

[36] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, "The Description logic handbook: theory, implementation, and applications," Cambridge University Press, 2003, ISBN:0-521-78176-0

[37] T. Gruber, "A translation approach to portable ontology specification," Knowledge Acquisition, vol. 5, no. 2, pp. 199-220, 1993.

[38] N. Guarino, and P. Giaretta, "Ontologies and knowledge bases, towards a terminological clarification," IOS Press, Amsterdam, pp. 25-32, 1995.

[39] S. Jean, G. Pierra, and Y. Ait-Ameur, "Domain ontologies: a database-oriented analysis," Springer, vol. 1, pp. 238-254, 2007.

[40] G. Antoniou, and F. V. Harmelen, "Web ontology language: OWL," Handbook on ontologies, 2009.

[41] Object Management Group, "OMG formal," http://www.omg.org/spec/, Retrieved 2014.

[42] N. Berkani, S. Khouri, and L. Bellatreche, "Generic methodology for semantic data warehouse design: From schema definition to etl," in the Intelligent Networking and Collaborative Systems (INCoS2012), Sept. 2012, pp. 404-411, ISBN: 978-1-4673-2279-9

[43] "Fact++," http://owl.man.ac.uk/factplusplus/, Retrieved 2014.

[44] F. Arvidsson, and A. Flycht-Eriksson, "Ontologies I," 2008.

[45] G. Pierra, "Chapter context representation in domain ontologies and its use for semantic integration of data,"

Journal on Data Semantics X, vol. 4900, pp. 174–211, 2008.

[46] P. Coret, J. Richard, E. Talavet, and T. Trofimoya, Introduction A OWL, Langage XML D'Ontologies, Technical report, 2006.

[47] S. Khouri, L. EL Sarraj, L. Bellatreche, B. Espinasse, N. Berkani, S. Rodier, and T. Libourel, "CiDHouse: contextual semantic data warehouses," The 24th Database and Expert Systems Applications (DEXA 2013), Springer Berlin Heidelberg, 2013, pp. 458-465.

[48] F. Pinet, C. Roussey, T. Brun, and F. Vigier, "The use of UML as a tool for the formalisation of standards and the design of ontologies in agriculture," Advances in Modeling Agricultural Systems, vol. 25, pp. 131-147, 2009.

[49] R. Liepins, K. Cerans, and A. Sprogis, "Visualizing and editing ontology fragments with OWLGrEd," Proc. CEUR Workshop Proceedings, vol. 932, pp. 22-25, 2012.

[50] J. Barzdins, G. Barzdins, K. Cerans, R. Liepins, and A. Sprogis, "OWLGrEd: a UML style graphical notation and editor for OWL 2," Proc. CEUR Workshop Proceedings vol. 614, 2010.

[51] J. Barzdins, K. Cerans, R. Liepins, and A. Sprogis, "UML style graphical notation and editor for OWL 2," Perspectives in Business Informatics Research, Springer Berlin Heidelberg, vol. 64, pp. 102-113, ISBN:978-3-642-16100-1

[52] B. C. Grau, I. Horrocks, and Y. K. yevgeny, "Modular reuse of ontologies: theory and practice," Journal of Artificial Intelligence Research, 2008.

[53] GMSIH, "Tarification à l'activité (T2A) - Guide à l'usage des établissements de santé," Appui santé et médico-social (ANAP), 2005.

[54] N. Prat, J. Akoka, and I. Comyn-Wattiau, "Transforming multidimensional models into OWL-DL ontologies," The IEEE 8th International Conference on Research Challenges in Information Science (RCIS 2011) IEEE, 2011.

[55] M. Gyssens, and L. V. S. Lakshmanan, "A foundation for multi-dimensional databases," Proceeding the 23rd International Conference on Very Large Data Bases (VLDB 1997), pp. 106–115.

[56] J. D. Bruijn, F. Martın-Recuerda, D. Manov, and M. Ehrig, "D4.2.1 State-of-the-art survey on ontology merging and aligning V1," Semantically Enabled Knowledge Technologies (SEKT 2003), 2003.

[57] S. Falconer, "OntoGraf," In Protégé Wiki, from http://protegewiki.stanford.edu/wiki/OntoGraf, Retrieved 2014.