L. El Sarraj, B. Espinasse, T. Libourel (2014), « An Ontology-Driven Personalization Approach for Data Warehouse Exploitation »,
in: International Journal on Advances in Software (IJAS), vol 7 no 1 & 2, year 2014, http://www.iariajournals.org/software/

An Ontology-Driven Personalization Approach for Data Warehouse Exploitation

Lama El Sarraj, Bernard Espinasse
LSIS UMR 7296
Université d'Aix-Marseille,
Marseille, France
{firstname.lastname}@lsis.org

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

***Abstract* — Understanding, reusing, and maintaining Data Warehouse (DW) resources is a key challenge for users. DW resources are shared by different groups of users that belong to various domains and profiles. DW resources, like files, graphs, multidimensional tables and so on, are interpreted differently from a user to another. Unfortunately, misinterpreting data could induce serious problems and conflicts. To guarantee relevant interpretation of DW resources, additional information is necessary. To tackle these challenges we propose to use ontologies to help the users in the exploitation of DW. In this paper we propose an ontology-driven approach for DW personalization. This work is situated in the healthcare context. In this context we present an Ontology-driven Personalization System (OPS) based on three specific and related ontologies: domain ontology (O_D), DW ontology (O_{DW}) and resources ontology (O_R). This paper focuses on the methodology used to develop each of these three ontologies.**

Keywords* — *Data Warehouse; Ontology; Decision Support Systems; Business Intelligence, Decision Making; Healthcare Institution Management

I. INTRODUCTION

Several surveys proved that big companies need efficient Decision Support Systems (DSS) and seek to expand the number of users over their users. In fact, researchers found that companies need to have flexible decision tools that include users' requirements and domain resources. DSS enable users to analyze and synthesize data according to different perspectives. Generally a DSS uses a collection of Business Intelligence (BI) tools and applications permitting to analyze, query and visualize a huge volume of data stored in a Data Warehouse (DW).

DW is the core of most DSS, it's "a subject oriented, nonvolatile, integrated, time variant collection of data in support of management's decisions" [1]. DW uses a multidimensional model, that represent facts and their measures 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 [2]. Data are represented like a point in a multidimensional space, visualized like a data cube [3]. They give users the possibility to synthetize and analyze data from three (or higher) dimensional arrays of values and various granularity levels. Based on this multidimensional model On Line Analytical Processing (OLAP) cubes enable to manipulate 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 [4], 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 [5]. Most research works devoted for DW focus on the design approach [6], [7], [8]. 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 analyses and have a visibility about other resources that could help them to 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 [9]. Then researchers in this domain have widely used ontologies in different phases of the DSS, at the conceptual level [10] [11], with ETL [12], OLAP cube model [13] and OLAP queries [14]. Summing up, ontologies have proved their efficiency in the DSS fields.

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 present the schema of an existing DW, and the last one describes existing DW resources of the related to the DW.

This research concerns an existing 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, France (APHM).

The paper is organized as follows. Section II presents the problematic. First it presents the context, and then the problematic and the aim of our research. Section III presents different approaches for DW personalization and introduces some elements about ontologies and their use in software development. 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. We also present 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). Finally we conclude and present some perspectives to this work.

II. PROBLEMATIC

In this section we present first the context of our research through related to healthcare domain. Then we present the aim of our research.

A. Context

Our research concerns the healthcare management specifically applied 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 specify the cost of sojourn based on the "Diagnosis Related Groups" (DRG) that classify the hospitalization in homogeneous and coherent medico-economic groups. Today this concept is used in France and several countries over the world like United States of America, England, etc. to finance healthcare institutions according to their activity.

To analyze PMSI data, specific decision support system (DSS) have been developed. DSS is mainly centered on data warehouses (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 (doctors, pharmacists, biologists, etc.), and the second does not (financial affaire managers, computer scientists, human resources, etc.).

In this context, to illustrate our problematic we consider a DW with its star schema, 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, …}

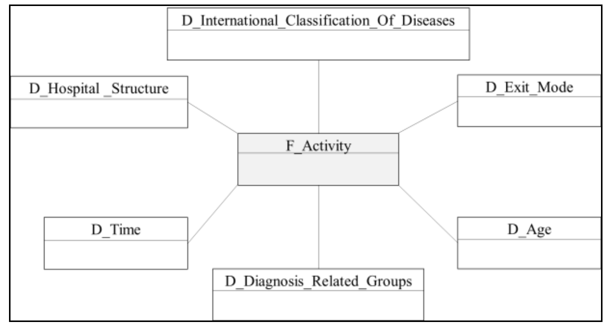

Figure 1: Data Warehouse Schema.

The multidimensional table (MT) is denoted MT = (M, D), where M is a set of measure and D is a set of dimensions. We take an example of a multidimensional pivot table "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, MCD, 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 (DRG) and poles.

Periode : From january to mars

			Pôle 1	Pôle 2	Pôle 3	Total
DRG	MDC	TYPE DRG TITLE	288	318	519	1125
1	01	SURG CRANIOTOMY AGE >17 W CC	253	26	311	590
2	01	SURG CRANIOTOMY AGE >17 W/O CC	274	520	335	1129
3	01	SURG CRANIOTOMY AGE 0-17	225	319	212	756
4	01	SURG NO LONGER VALID	325	215	122	662
5	01	SURG NO LONGER VALID	125	138	118	381
		Total	1490	1536	1617	4643

Figure 2: Example of a Multidimensional Table.

The DW "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 of a medical unit (Hospital_Structure) to receive patients.
   - The main specialties by pole (Pole_Structure).

2) *Needs and patient flow indicator (care consumption):* these indicators are mainly based either on the patient age or on the exit mode, for instance:
   - Describes the sojourn, analyze sojourns according to the group of diseases.
   - The main specialties of a pole (Pole_Structure).
   - Identify the population susceptible to be treated.
3) *Patient flow indicators:* these indicators present the is the cause of the hospitalization (from where the patients come) and their destination (where they go):
   - 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 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 don't 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, measures calculation methods and their sources. Because of this semantic lack, the users cannot understand the usage of the DW resources that may respond to their needs. On other hand, there is vocabulary *heterogeneity in query expression*: users don't belong to the same domain. They don't have the same vocabulary background. They don't express their need with the same terms. For example: number of sojourn could be expressed as number of venue. Finally concerning analysis needs, most of the times, users need to analyze many resources to make a decision. In big institutions, like the APHM, big number of resources makes this task complicated. Thus, users need to have a global vision of the DW structure to visualize the possibilities or existing resources that can help them in making a decision, for example show the resources that analyze the measure of his actual resource that responds to his needs with other existing axis of analysis of the DW and so forth.

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 background concerning various approaches concerning DW personalization, and some elements related to ontologies.

III. BACKGROUND

This section presents different approaches related to DW personalization, 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 personnalisation based on user profiles

Research works based on user profiles are usually associated with 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 try to compare and evaluate their relevance to our problem with the use of our DW "Fig.1".

Personalization is a customized and individualized description of a use or a group of users. Personalization system relies on users' need, preferences and characteristics [15], and usually on a defined users profiles [16]. For example, in our research, we can define user profiles for doctors, computer scientists, department chief, managers, etc. Although no consensus exists for the definition of a user profile, 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 [17] adapted to a user profile.

In their research related to DW, the authors in [18] developed a state of the art about user modeling based on system requirements. Usually a user profile is defined by a set of preferences used to configure or adapt the system to the user [19], [20], [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 profile from two perspectives, the definition of user 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:

- *Customizing data sources schema* [23], [22] adapting the data structures to a specific needs of users.
- *Customizing queries visualization* [19], or representation [20], [25], [21].
- *Recommendation of OLAP queries* [26, 27] to assist in the exploration of the ED.

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 new method to treat data, queries.

B. DW personalization by recommendation or transformation

The personalization of the DW by recommendation can be associated to various works such as [28], [29], [30], [23], [26, 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.

In the domain of transformation for DW personalization issues, we would mention the work of [19] that treats personalized visualization of OLAP queries. The authors in [33] propose a solution to evolve the DW schema according to of 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, basing on users preferences and on analysis context.

C. Ontologies

Ontologies have been used in the domain of knowledge engineering to facilitate requirement expression and detect incoherencies and semantic ambiguities between users [35]. Description Logic (DL) is a formalism used to build ontologies with OWL [36]. In this section we propose a formalization of the ontology basing on DL. We begin by ontologies definition.

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, first ontology has been used 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 systems.

Ontologies design requires the establishment of processes to extract the knowledge connected to a domain and make it handled, at the same time by information systems and humans. In this context, several definitions of ontologies have been proposed in the field of software engineering. Gruber [37] an 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] 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 the machine.
- *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

The ontology can be formalized as 5-uplet [40] as follows O: <C, P, ClassPropt, ClassAssoc, Formal> that concerns the TBOX. We consider:

- *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 like RDF, OWL, PLIB, etc.

Different languages are proposed to define ontologies. Ontology Web Language (OWL) is the standard language for representing ontologies [41] [44]. W3C consortium recommends OWL to define ontologies. The OMG [15] define the OWL meta-model. OWL is originally defined as an extension of RDF Graphs. So OWL is a set of triple subject-predicate-object. This format is very uniform, which makes it easy to analyze and store by the machine, but it is totally unusable by humans [14].

To facilitate the creation and the visualization of ontologies there are OWL ontology editors, such as Protégé [42] to manipulate ontologies (edit, load, define taxonomies, etc..). Protégé provides a detailed view for each concept in ontology. There are also visualization tools of ontologies, the most common are IsaViz [40], OWLViz [9], Growl [43], Welkin [39], etc.

UML is a standard used to model information systems and software engineering. UML is a graphical language for visualizing, specifying and building tool components. UML is a semi-formal formalism. UML provides different diagrams such as class diagrams, etc. However, UML is not suitable to represent complex reasoning and inferences [45] from which new knowledge can be deduced. One of the major advantages of UML is that it is widely used in the academic environment and even by non-professionals. Users prefer UML notations that formal ontologies languages. Most informatics designers use UML to describe their diagram.

Several studies propose to model ontologies with UML [46 , 47 , 48 , 49 , 50, 51]. There are many commonalities between the formal languages of ontologies and UML. A comparison UML/OWL is studied [49], but the only drawback is the lack of semantics. Ontology can be modeled with UML to arrive to a consensus between experts to present knowledge. In a second step, using a transformation tool, the UML model can be translated into (RDF, OWL, etc.). For those reasons, we can consider UML as an adequate formal model for the representation of ontologies.

E. Conclusion

Even if the customization 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 don't seem totally adapted to our problematic. Indeed, the specificities of data related to healthcare management system require additional semantic on DW resources. In particular, to treat the problem of the variety of users' profiles and domains complexity. We propose in the next section an approach that will provide help in the exploitation process of the DW, this approach is *driven by ontologies.

IV. AN ONTOLOGY-DRIVEN APPROACH FOR DW PERSONALIZATION

The Ontology-driven Personalization System (OPS) support users from heterogeneous domains to exploit existing DW resources. This support is based 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.

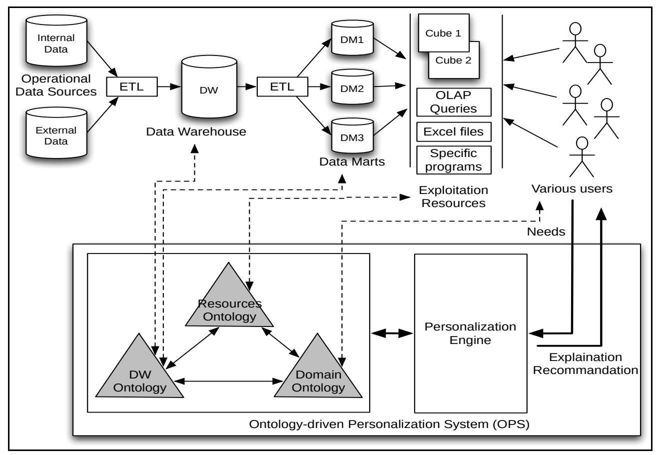

Figure 3: Ontology-driven Personalization System Architecture.

The two main components of our OPS are:

- *Knowledge base*: is an OWL database based on three related ontologies describing 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 express 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 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* - Which is the DW (data mart) related to this domain concept?
- *Resources concept* - What are the resources to analyze a domain concept?

These use-cases are treated in the OPS by the PE reasoning on one or more ontologies. The following "Fig. 4" illustrates the connection between ontologies and the users. 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. The PE components are not presented in this paper.

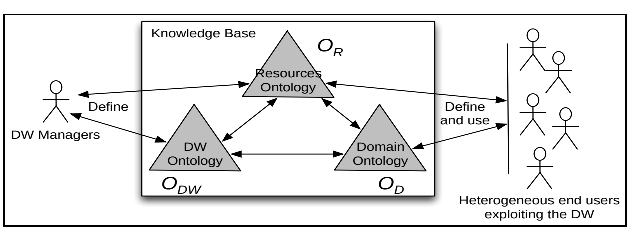

Figure 4: On the use of ontologies.

V. KNOWLEDGE BASE COMPONENT

This section presents the Knowledge base of our Ontology-driven Personalization System (OPS). This knowledge base is composed of three ontologies: Domain Ontology (O_D), Data Warehouse ontology (O_{DW}) and existing resources ontology (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 OWLGrEd tool, an OWL ontology editor. OWLGrEd uses a textual syntax OWL Manchester to create, edit and view an ontology [52]. OWLGrEd provides a comprehensive overview of OWL based on 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 OWL, OWLGrEd enriched the UML class diagram with new notations [53, 54].

A. Domain ontology (O_D)

1) Description:
Domain ontology (O_D) concerns concepts of the domain and the connection between them. A decision is based on one or many indicators. In the analysis processes the user check the informations that he already knows. However, most of the times user needs additional indicators to make his analysis. The domain description provides the information about the connection between domain concepts.

2) Elaboration methodology:
There are two solutions to obtain domain ontology. We can extract a part of existing domain ontology 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 [55]. As no domain ontology exists concerning "PMSI domain" we develop a new one.

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

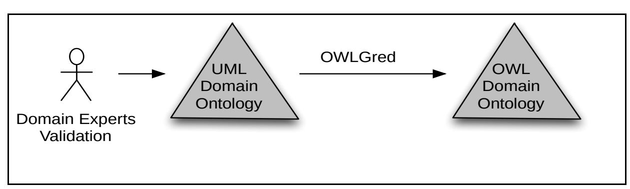

Figure 5: Domain Ontology Development.

The Domain Ontology expressed in a UML class diagram “Fig.6”, inspired from the model presented in [56]. This model was enriched and validated by domain experts. This ontology is presented here with the OWLGred tool.

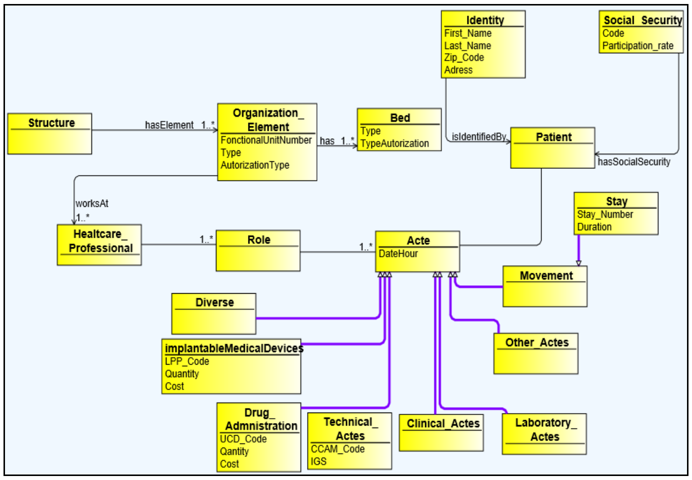

Figure 6: Domain Ontology validated in UML.

3) Results:

Domain Ontology in UML “Fig. 6” validated with domain experts transformed to OWL via the OWLGRED tool. Domain ontology in OWL is visualized with Protégé in Fig. 7.

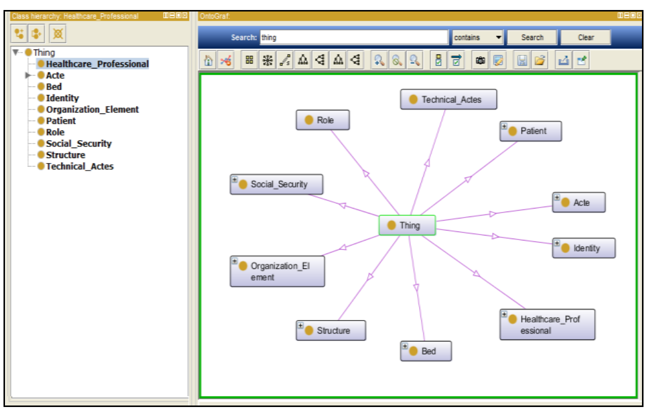

Figure 7: Domain Ontology validated in OWL (with Protégé).

B. DW ontology (O_{DW})

1) Description:

Multidimensional model associated to the DW organizes data into facts and dimension. The DW Ontology (O_{DW}) concerns the DW. We consider only DW implemented in ROLAP (Relational OLAP) technology. 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 DW ontology (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 (fact, dimension, etc.).

The atomization of this transformation “Fig.8” 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.* [57]. The research work of Prat *et al.* [57] defines a multidimensional meta-model, the concepts of OWL-DL, and transformation rules for mapping a multidimensional model into OWL-DL ontology.

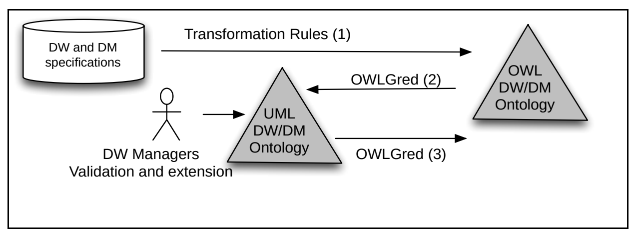

Figure 8: DW/DM Ontology Development.

The transformation rules proposed by [57] are adapted to our problematic to generate the DW ontology in OWL. 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 :

The transformation of the DW ontology from OWL to UML with the OWLGRED tool is presented in Fig. 9. The ontology O_{DW} in OWL is presented with Protégé tool in Fig.10.

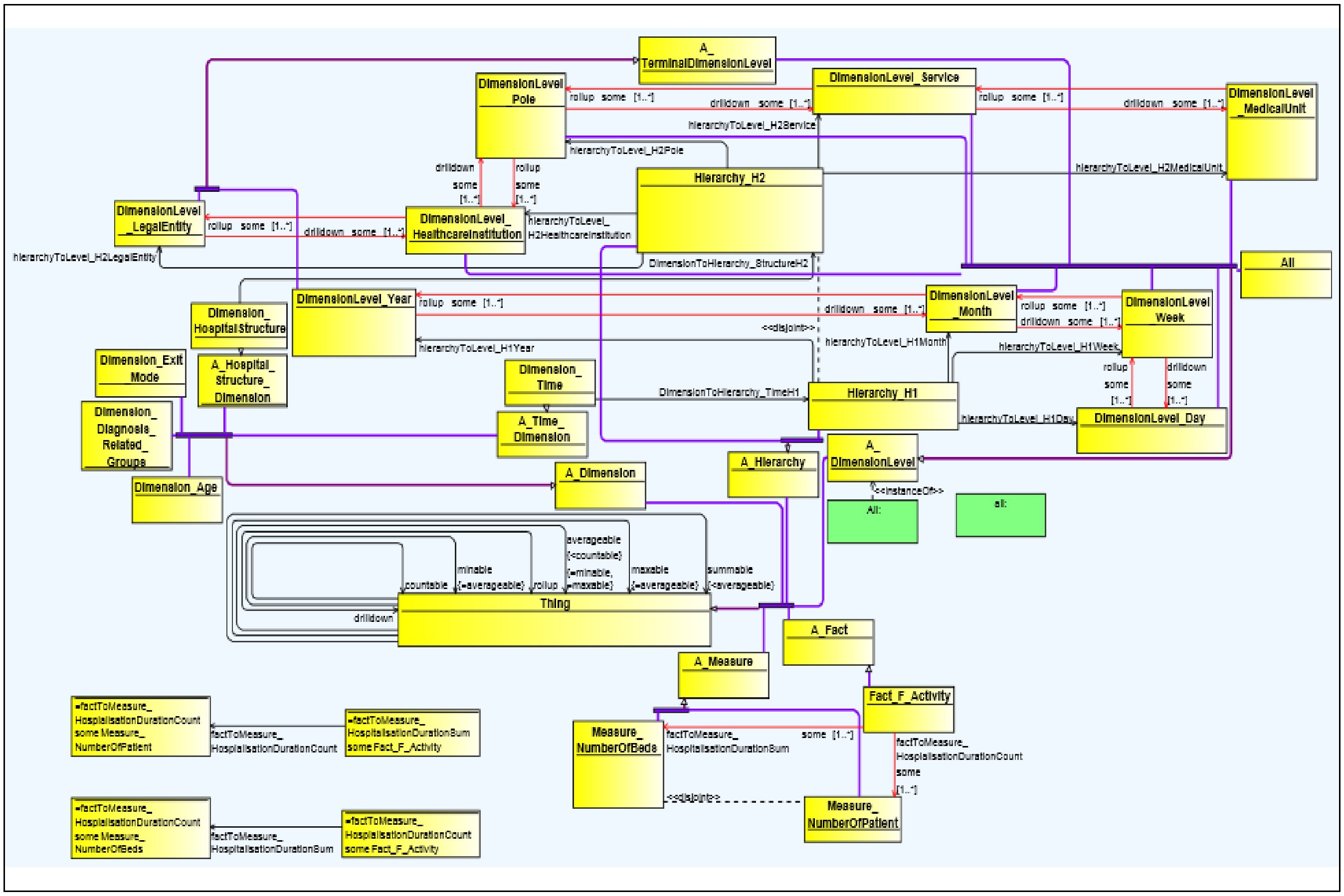

Figure 9: Data Warehouse Ontology validated and extended in UML.

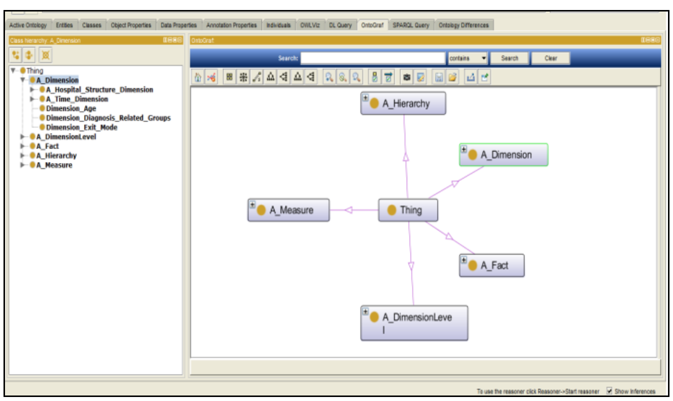

Figure 10: Data Warehouse Ontology translated in OWL (with Protégé).

C. *DW resources Ontology (O_R)*

1) Description:

DW resources can be of different nature, for example, histograms, graphs, etc. So even if the multidimensional model is based on the metaphor of the cube or hypercube, the most common structure of the visualization is Multidimensional Table (MT), Fig.2", which provides data presented in two axes of analysis [58] [2] enabling the visualization of a slice of the cube.

Resources are related to the DW/DM and are defined by the DW managers. To understand a resources components users' needs to have: calculation method, unit of measure, calculation period, date of creation, date of update, date of validity, objective, definition and the relation with the DW/DM.

We identified two types of DW resources:

- *Operational resources:* they are DW/DM oriented, the resources requiring an execution before being used for analysis. For example, OLAP queries, Excel files or arrays and so on.
- *Conceptual resources:* they are user-oriented, they are resources used by the final users. For example, indicators, dashboards, etc.

2) Elaboration methodology:

To develop this DW resources ontology, as for Domain Ontology, 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 the Fig. 11.

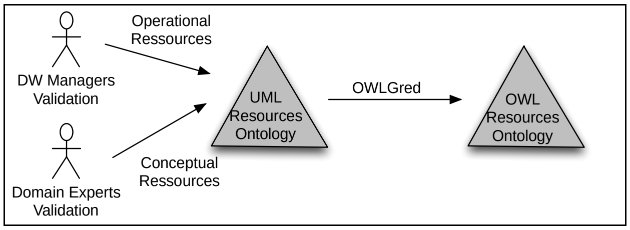

Figure 11: Domain Ontology Development.

Once the resources ontology expressed in UML class diagram, is validated with domain experts, we transform it into OWL with OWLGred tool, Fig. 12.

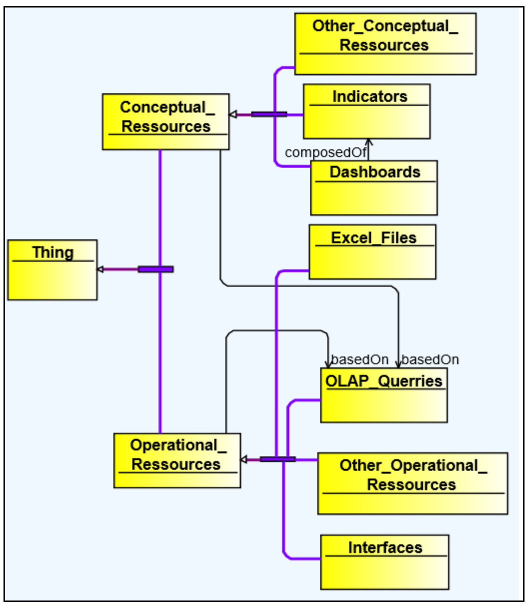

Figure 12: DW resources Ontology validated in UML.

1) Results:

The transformation of the resources ontology from OWL to UML with the OWLGRED tool is presented in Fig. 12. The ontology O_R in OWL is visualized with Protégé tool in Fig.13.

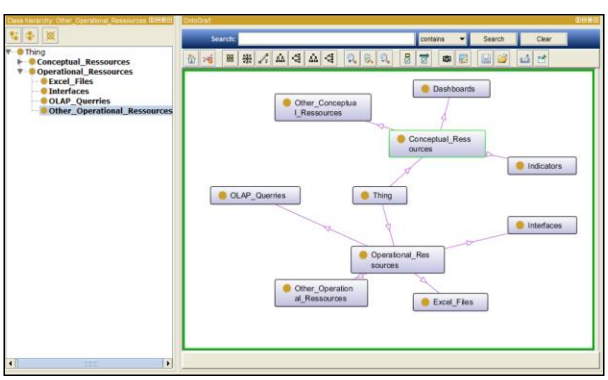

Figure 13: DW resources Ontology validated in OWL.

VI. Mapping ontologies

The knowledge base of our Ontology-driven Personalization System is composed of three ontologies: Domain Ontology (O_D), Data Warehouse (DW) ontology (O_{DW}) and existing Resources Ontology (O_R). We formalize our ontology by the quadruple < O_{DW}, O_R O_D, Map> where:

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

These mapped ontologies can be used for many purposes with OPS, on one hand, to give a vision about the relation between DW, resources and domain concepts, and in 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. First we define the mapping process, then we introduce the different mappings concerned, and then we illustrate these mapped 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 pr-defined [59]***:***

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 have 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 we opt for *Ontology Aligning* or *Ontology Mapping* processes as defined before.

B. Concerned mappings

In our case we have to consider three different mappings relating these three ontologies two by two, depending on the connection between users and ontologies, Fig. 14:

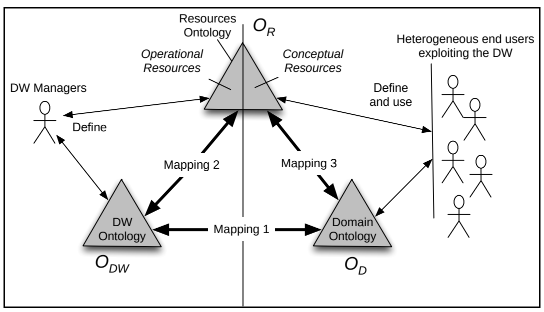

Figure 14: Different mappings between the three ontologies.

The model of the connected ontologies O_D-O_{DW}-O_R is presented in Fig. 15.

The Mapping 1 supports the connection between DW ontology (O_{DW}) and Domain ontology (O_D), Mapping 2 supports the connection between DW ontology and the operational resources of the Resources ontology (O_R), and finally Mapping 3 supports the connection between the (O_D), and the conceptual DW resources of the Resources ontology (O_R).

Ontology Aligning or *Mapping* processes related to these three mappings concern 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}. For example, 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 Tab.1 an example with OWL-DL.

Table 1: Concepts and innfered 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”.

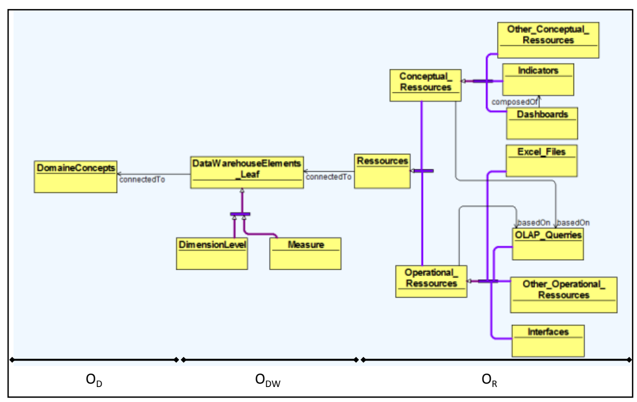

Figure 15: Mapping the three ontologies O_{DW} - O_D - O_R

VII. VALIDATION EXAMPLE

To illustrate our proposal we suggest to respond to “Use-case 3” questions, we’ll use OntoGraph [60] to visualize the ontologies’ concepts. Fig. 16 shows the results of the search done on the mapped ontologies. The concept entered is “DRG”. The result is the concepts that contain “DRG” and the related concepts, for example, resources concepts, domain concepts or DW schema concepts.

Entry: Domain concept “DRG”

Output:

- *DW schema element*: Dimension and D_DRG, because D_DRG is a subclass of Dimension (Dimension ⊆ D_DRG)

- *Resources concept: Resource_Activity_Pole_DRG (a multidimen-sional table contaning PMSI activity per DRG and per Pole)*
- *Resources concept: Resource_Activity_Pole_DRG (a multi-dimen-sional table contaning PMSI activity per DRG and per Pole)*

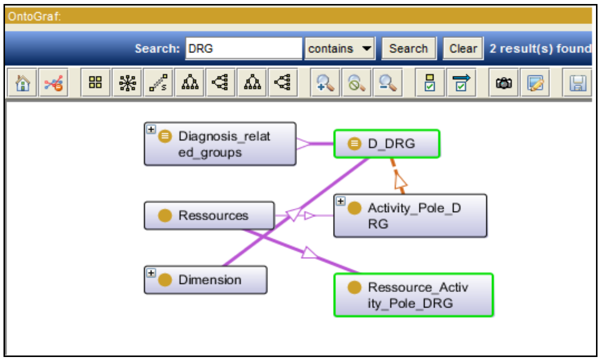

Figure 16: Example, retrieve “DRG” concept from the ontology.

VIII. Conclusion

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 data warehouse (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: Domain ontology (O_D), DW ontology (O_{DW}) and existing resources ontology (O_R).

We have focused in this paper on the knowledge base. We introduced the methodology used to develop each of these 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 OntoGraph. 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 could be applied for 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 the Ontology-driven Personalization System (OPS), giving to user a friendly environment to query, provides resources explanation and resources 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 ontologies evolutions on our 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] Inmon, W.H., *Building the data* warehouse, ed. Second. Vol. 0471141615. 1992, New York, NY, USA.: John Wiley & Sons.

[2] Lehner, W., *Modelling Large Scale OLAP Scenarios*, in *Advances in Database Technology (EDBT)*. 1998, Intl. p. 153-167.

[3] Bosworth, A., et al., *Data Cube : A Relational Aggregation Operator* Generalizing *Group-By, Cross-Tab, and Sub-Total.* Data Min. Knowl. Discov., 1995: p. 152-159.

[4] Rizzi, S., et al. *Research in data warehouse modeling and design:* dead *or alive?* in *DOLAP*. 2006. Arlington, VA, USA

[5] Golfarelli, M., *From user requirements to conceptual design in data warehouse design – a survey.* 2009.

[6] Kimball, R. and M. Ross, *The data warehousing toolkit*. 1996, New York: John Wiley&Sons.

[7] Prat, N. and J. Akoka, *From uml to rolap multidimensional databases* using *a pivot model*, in *8èmes Journées Bases de Données Avancées*. 2002. p. page 24.

[8] Tsois, A., N. Karayannidis, and T.K. Sellis, *Mac : Conceptual data modeling for olap*, in *3rd International Workshop on Design and Management of Data Warehouses (DMDW 2001)*. 2001: Theodoratos p. 5.

[9] Bellatreche L., Nguyen Xuan D., and D.H. Pierra G., *Contribution of ontology-based data modeling to automatic integration of electronic catalogues within engineering databases.* Computers in Industry Journal Elsevier, 2006. **57**(8-9): p. 711-724.

[10] Bellatreche, L., et al. *Using Ontologies and Requirements for Constructing and Optimizing Data Warehouses*. in *30th IEEE International Convention MIPRO*. 2012. Opatija, Croatia.

[11] Prat, N., I. Megdiche, and J. Akoka. *Multidimensional Models Meet the Semantic Web: Defining and Reasoning on OWL-DL Ontologies for OLAP*. in *DOLAP*. 2012. Hawaii, USA.

[12] Khouri, S., L. Bellatreche, and N. Berkani, *MODETL: A complete MODeling and ETL method for designing Data Warehouses from Semantic Databases*, in *International Conference on Management of Data (COMAD)*. 2012.

[13] Niemi, T. and M. Niinimäki. *Ontologies and summarizability in OLAP*. in *Proc. of SAC'10*. 2010. Sierre, Switzerland.

[14] B. Neumayr, S. Anderlik, and M. Schrefl. *Towards ontology-based olap: datalog-based reasoning over multidimensional ontologies*. in *Proceedings of the fteenth international workshop on Data warehousing and OLAP (DOLAP)*. 2012. New York, NY, USA: ACM.

[15] Ioannidis, Y. and G. Koutrika, *Personalized Systems : Models and Methods from an IR and DB Perspective*, in *VLDB*. 2005. p. 1365–1365.

[16] Korfhage, R.R., *Information Storage and Retrieval*, ed. J. Sons. 1997.

[17] Domshlak, C. and T. Joachims, *Efficient and Non-Parametric Reasoning over User Preferences.* User Modeling and User-Adapted Interaction, 2007. **17**((1-2)): p. 41-69.

[18] Kobsa, A., *Generic User Modeling Systems*, in *The Adaptive Web*. 2007. p. 136-154.

[19] Bellatreche, L., et al., *A personalization framework for OLAP queries*, in *8th International Workshop on Data Warehousing and OLAP, DOLAP'05*. 2005: Bremen, Germany. p. 9-18.

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

[21] Jerbi, H., et al., *Management of Context-Aware Preferences in Multidimensional Databases.* , in *ICDIM* 2008. p. 669-675.

[22] Garrigos, I., et al., *A Conceptual Modeling Approach for OLAP Personalization*, in *Conceptual Modeling-ER* 2009, Springer: Verlag Berlin Heidelberg. p. 401-414.

[23] Bentayeb, F., et al., *Personnalisation dans les entrepôts de données : bilan et perspectives*, in *Entrepôt de Données et Analyse en ligne (EDA)*. 2009.

[24] Jerbi, H., et al., *Recommandation de requêtes dans les bases de données multidimensionnelles annotées.* Revue des Sciences et Technologies de l'Information, Ingénierie des Systèmes d'Information, 2011. **16**(1): p. p.133-138.

[25] Xin, D., et al., *Answering top-k queries with multi-dimensional selections: The ranking cube approach*, in *VLDB*. 2006. p. 463-475.

[26] Giacometti, A., P. Marcel, and E. Negre. *A Framework for Recommending OLAP Queries*. in *DOLAP*. 2008. Napa Valley, California, USA.

[27] Giacometti, A., P. Marcel, and E. Negre, *Recommending Multidimensional Queries* in *DaWaK*. 2009. p. 453-466.

[28] Sapia, C., *On Modeling and Predicting Query Behavior in OLAP Systems*, in *DMDW*. 1999. p. 2.1-2.10.

[29] Chatzopoulou, G., M. Eirinaki, and N. Polyzotis, *Query Recommendations for Interactive Database Exploration*, in *SSDBM*. 2009. p. 3-18.

[30] Jerbi, H., et al., *Applying Recommendation Technology in OLAP Systems* in *ICEIS*, Springer, Editor. 2009, LNBIP. p. 220-233.

[31] Giacometti, A., et al. *Query recommendations for OLAP discovery driven analysis.* in *DOLAP*. 2009. Hong Kong, China.

[32] Negre, E., *Exploration collaborative de cubes de données*. 2009, François Rabelais Tours, France.

[33] Favre, C., F. Bentayeb, and O. Boussaid, *Evolution et personnalisation des analyses dans les entrepôts de données. Une approche orientée utilisateur*, in *INFORSID*. 2007. p. 308-323.

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

[35] Aybuke, A. and W. Claes, *Engineering and Managing Software Requirements.* Springer, 2005.

[36] Baader, F., et al., *The description logic handbook: theory, implementation, and applications.* Cambridge University Press 2003.

[37] Gruber, T., *A translation approach to portable ontology specification.* Knowledge Acquisition, 1993. **5**(2): p. 199-220.

[38] Guarino N. and Giaretta P. *Ontologies and knowledge bases, towards a terminological clarification*. in *Towards very large knowledge bases: knowledge building and knowledge sharing*. 1995. IOS Presse.

[39] Jean S., Pierra G., and Ait-Ameur Y., *Domain Ontologies : a Database-Oriented Analysis.* Springer, 2007. **1**: p. 238-254.

[40] N. Berkani, S. Khouri, and L. Bellatreche. *Generic methodology for semantic data warehouse design: From schema definition to etl*. in *Intelligent Networking and Collaborative Systems (INCoS)*. 2012.

[41] Antoniou, G. and F.V. Harmelen, *Web Ontology Language: OWL*. S.Staab and R.studer (eds), ed. Springer. 2009: Handbook on ontologies.

[42] *Fact++*. Available from: http://owl.man.ac.uk/factplusplus/.

[43] Arvidsson, F. and A. Flycht-Eriksson, *Ontologies I* 2008.

[44] Object Management Group. *OMG Formal*. Available from: http://www.omg.org/spec/.

[45] F. Baader, et al. *The Description Logic Handbook : Theory, Implementation, and Applications*. 2003. Cambridge University Press.

[46] Wikipedia. *Metamodeling in Domain Ontologies*. Available from: http://wiki.opensemanticframework.org/index.php/Metamodeling_in_Domain_Ontologies.

[47] Wikipedia. *Ontology Best Practices*. Available from: http://wiki.opensemanticframework.org/index.php/Ontology_Best_Practices.

[48] Pierra., G. *Chapter Context representation in domain ontologies and its use for semantic integration of data*. in *Journal of data semantics (jods)*. 2008. Berlin, Heidelberg: Springer-Verlag.

[49] Coret, P., et al., *Introduction a owl, langage xml d'ontologies* Technical report, 2006.

[50] Khouri, S., et al., *CiDHouse: Contextual SemantIc Data WareHouses*, in *Database and Expert Systems Applications*, H. Decker, et al., Editors. 2013, Springer Berlin Heidelberg. p. 458-465.

[51] Pinet, F., et al. *The use of UML as a tool for the formalisation of standards and the design of ontologies in agriculture*. in *Advances in Modeling Agricultural Systems*. 2009. New York: Springer.

[52] Renars Liepins, Karlis Cerans, and A. Sprogis. *Visualizing and Editing Ontology Fragments with OWLGrEd*. in *I-SEMANTICS 2012*. 2012.

[53] Barzdins, J., et al. *OWLGrEd: a UML Style Graphical Notation and Editor for OWL 2.* . in *In Proc. of OWLED 2010*. 2010.

[54] Barzdins, J., et al. *UML Style Graphical Notation and Editor for OWL 2.* in *BIR'2010,*. 2010. Springer 2010.

[55] Grau, B.C., I. Horrocks, and Y.K. yevgeny, *Modular Reuse of Ontologies: Theory and Practice.* Journal of Artificial Intelligence Research, 2008.

[56] GMSIH, *Tarification à l'activité (T2A) - Guide à l'usage des établissements de santé.* 2005.

[57] Prat, N., J. Akoka, and I. Comyn-Wattiau, *Transforming multidimensional models into OWL-DL ontologies*, in *RCIS*. 2011.

[58] Gyssens, M. and L.V.S. Lakshmanan. *A foundation for multi-dimensional databases*. in *Conf. on Very Large Data Bases (VLDB)*. 1997. Intl.

[59] Jos de Bruijn, et al., *D4.2.1 State-of-the-art survey on Ontology Merging and Aligning V1.* Semantically Enabled Knowledge Technologies (SEKT), 2003.

[60] Falconer, S., *OntoGraf.* 2010.