



Modeling a Data Warehouse of Annotation of Medical Images using UML profile

Mouhamed Gaith AYADIDepartment of Computer Sciences &
ISG university of Tunis-Tunisia**Riadh BOUSLIMI**Department of Computer Sciences &
ISG university of Tunis-Tunisia**Jalel AKAICHI**Department of Computer Sciences &
ISG university of Tunis-Tunisia

Abstract— *Medicine unlike other scientific disciplines has no single standardized vocabulary. It is favorable to emphasize the diversity of medical terms whose definition is rarely rigorous sometimes ambiguous specially the information from the medical image. Such imprecision is disadvantageous when you want to aggregate data from many medical practitioners. This gave birth of the annotation of medical image process. The diversifications of different sources from the annotations need to be modeled by using of data warehouses. However, classical conceptual modeling does not incorporate the specificity of these annotations. The process of developing a data warehouse starts with identifying and gathering requirements, designing the dimensional model followed by testing and maintenance. The design phase is the most important activity in the successful building of a data warehouse. For this reason, we focus in this paper on presenting the conceptual modeling of the data warehouse by defining a new profile using the StarUML extensibility mechanism.*

Keywords— *Annotation of medical images, Data warehouse, UML extension, UML profile*

I. INTRODUCTION

Thanks to recent advances in technology in recent years, particularly in the field of multimedia and computing. Digital information has become the heart of all sectors and especially the medical field. The doctor faces several constraints only, is to establish the diagnosis adequate and make the best decision. This pushes sometimes to commit malpractice and generate more damage. In fact, according to the latest statistics provided by the foundation Tunisian research health services, the number of malpractice among physicians noted reached in 2011, Tunisia 1687, 1083 in Canada and 14,396 in the United states to reduce this number, we must provide effective collaboration between physicians. Doctors need to annotate medical images and to store the medical images annotation in a database to facilitate the access and taking care for the association of semantics to a medical image. But the storage limits of traditional databases require the passage to a more efficient concept: data warehouse. The success of the warehousing process rests on a good conceptual modeling schema. In fact, conceptual modeling offers a higher level of abstraction while describing the data warehousing project since it stays valid in case of technological evolution. However, no contribution is at the present time standard in term of data semantic models. This finding leads us to propose a new UML profile with user oriented graphical support to represent data and data warehouse conceptual modeling with structural model (class diagram).

This paper is organized as follows. In section 2, we present an overview of research works related to conceptual approaches and extensibility of UML for applications' needs. In section 3, we present the methodology that we adopted to extend the StarUML profile. In section 4, we present the UML profile. In section 5, we present the UML profile realization. In section 6, we summarize the work and we propose some perspectives that can be done in the future.

II. RELATED WORK

In this section, we present different approaches related to the conceptual modeling methodology, then we present research works that extended UML to adopt it to their conceptual modeling needs. In the literature, we can find three categories of conceptual approaches; the top down approach, the bottom up approach and the middle out approach. The difference between those latter is situated in the starting point. In fact, each approach has its own starting point such as users' needs, data marts or both users' needs and data marts. In the literature, different data models [1, 2, 3] both conceptual and logical have been proposed for data warehouse design. These approaches are based on their own visual modeling languages or make use of well known graphical notation like ER model or UML, but to the best of our knowledge, there is no standard method or model that allows us to model all aspects of a DW. Moreover, during our survey we noticed that most of the research efforts in designing and modeling DWs have been focused on the development of MD data models and conceptual design, the interest on the physical design of DWs has been very poor.

The pioneer author in the field of data warehouse design is Juan Trujilio. He has made a major contribution. He proposed the use of UML for the design of data warehouse. He defined four UML profiles for modeling different aspects of data warehouse: the UML profile for Multidimensional Modeling, the Data Mapping profile, the ETL profile and Database Deployment profile. In [4], authors propose an approach that provides a theoretical foundation for the use of OO databases and Object relational databases in DW. This approach introduces a set of minimal constraints and

extensions to UML for representing multidimensional modeling properties for DW. In [5, 6]; authors have proposed a multidimensional profile for the Data warehouse conceptual schema and Client conceptual schema. The author has also shown work in the field of physical schema. He has presented the database deployment profile in [7]. Another author who also has a significant role in the design of data warehouse is Stefano Rizzi. The author in [8] proposed a graphical conceptual model for data warehouses, called Dimensional Fact model, and gave a semi-automated methodology to build it from the preexisting (conceptual or logical)schemes. Then in [3] based on the Dimensional Fact Model (DFM), he gave a general methodological framework for data warehouse design. Then he discussed some issues in Multidimensional modeling for the design of data ware house in [10].After that different authors gave different techniques and models for the design of data warehouse which we have discussed and compared in the next paragraph. In [11] authors proposed a graphical conceptual model for DW called Dimensional Fact Model. They gave a methodology to build it from the preexisting schemes. They used the star schema. This approach can be extended to logical and physical level. In [12] authors proposed a general methodological framework for data warehouse design, based on Dimensional Fact Model (DFM). After analyzing the existing information system and collecting the user requirements, a conceptual design is carried out semi automatically starting from the operational database scheme and the workload is prepared. They used the star schema. This approach can be extended to logical and physical level. In [13] some authors proposed an object oriented approach to accomplish the conceptual modeling of DW, MD databases and OLAP application. They introduced a set of constraints and extensions to UML for showing MD modeling properties. They used the star schema. This approach cannot be extended to logical and physical level. Other work in [14] illustrated the main features of Wand, a prototype CASE tool for data warehouse design. Authors implemented the tool WAND, which helps in structuring a data mart and carrying out the conceptual design in a semi- automatic fashion. This approach can be extended to logical level. Other authors show [15] to manage the representation, manipulation and representation of MD models on the web by means of extensible style sheet language transformations (XSLT). They used object-oriented approach based on the unified modeling language. There is some work too [16] where they present a MD conceptual object oriented model using extension of UML. They used different structures of object oriented model like nodes, arcs, detailed levels, stereotypes. In [5] authors present the development of multidimensional MD models for data warehouse using UML package diagrams. They present design guidelines and explain them with various examples. The authors in [17] present a DW design method based on UML which is used to handle all DW design phases and steps from operational data sources to final implementation. The work is based on object oriented method based on UML. They used MD modeling, MD databases and OLAP support. This approach can be extended to logical and physical level. They proposed in [18] to model the physical design of a DW by using the component diagram and deployment diagram of UML. The work is based on framework of DW with 5 stages (source, integration, DW, customization, and client) and 3 levels (conceptual, logical and physical). Each level is composed of diagrams making a total of 15 diagrams. Their approach reduces the overall development time of a DW and covers all main design phases of DW from conceptual modeling till final implementation. This approach can be extended to logical and physical level.

Authors used a framework for the design of DW back-stage in [19] including transformation rules at the attribute level and modeling of relations hip between source and target in different levels of granularity. Their approach is based on usage of UML packages. This approach can be extended to logical and physical level.

In [20] authors aimed at discussing some open issues in modeling and design of data warehouse. Like issues regarding conceptual models, logical models, method for design... But in [21] authors gave UML profile to represent MD and security aspects of conceptual modeling. They use UML packages to group classes into higher level units. They used the star schema. This work is related to the health system. Other authors in [22] compare various conceptual and logical DW design models and to find which is more suitable for implementing DW.

In addition some authors in [23] gave an UML profile for modeling DW usage on conceptual level. They designed a conceptual UML model and translated it into XML logical model which is later converted to xml document as physical model. They used the star schema. This work can be extended to logical and physical level. In the same year, authors [24] gave anti-standardization technique in DW design. They converted ER model to multidimensional data model. They gave different anti- standardization methods like increasing data redundancy, increasing derived columns... They used the star schema. In the same context, authors proposed a conceptual multidimensional data model called as PIM and transformed it into XML DW as a logical mode called as PSM. They used model driven architecture MDA for developing secure DW. They set of transformation rules to convert conceptual data model to logical model. This approach can be extended to logical level.

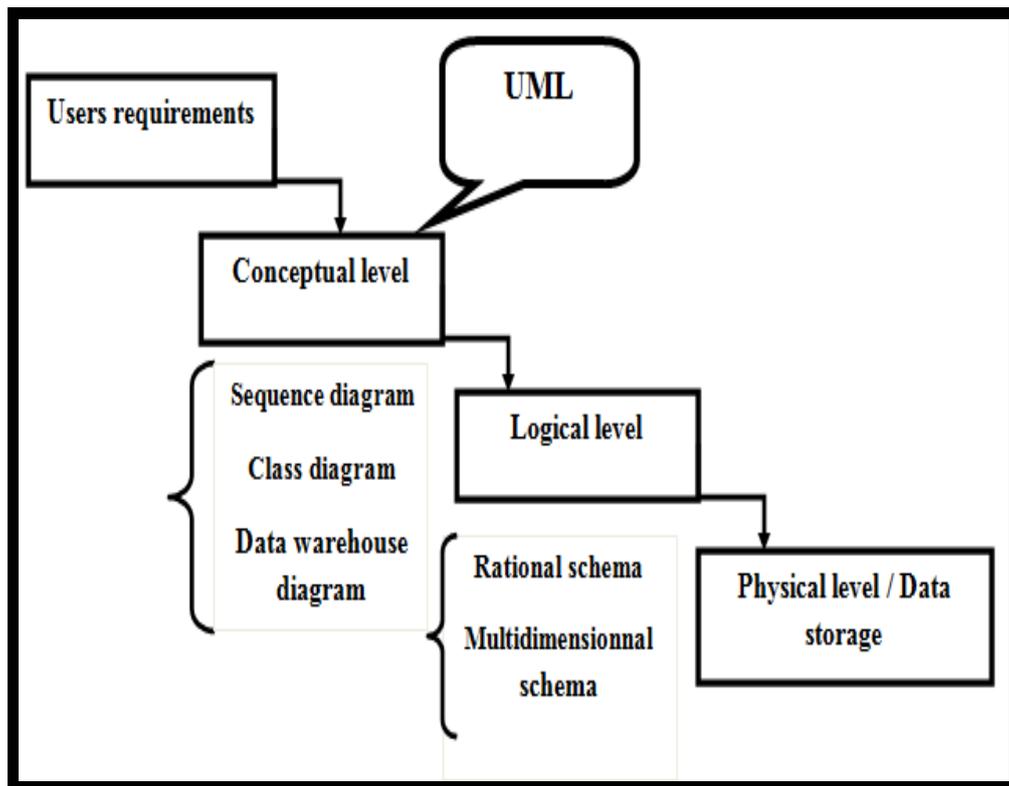
In [26] authors proposed an UML multidimensional model from various data source based on UML schema. They used the conceptual level integration on framework based on UML sources. First, they convert UML schema to UML class diagrams and then they build multidimensional model from it. They based the work on object oriented approach for DW design and mapping rules to convert UML class diagram to multidimensional model. They used the snowflake schema too.

In another work [27], authors used UML to build a DW model. They gave an overview on UML based techniques and tools used in agricultural DW. This approach can be extended to logical and physical level. Others [28] extended this work and introduced seven operators over DW model. They defined object oriented MD data model for description of data. In [29], authors proposed that the use of ontologies will improve several aspects of the design of data warehouses. They described several shortcomings of current data warehouse design approaches and discussed the benefit of using ontologies to overcome them.

Related to this work, authors [3] proposed an Object Oriented framework for data warehouse conceptual design. This framework has many benefits. Firstly, the object oriented multidimensional approach is the best as it satisfies all the criteria required for the data warehouse design and it is more adaptable as the user requirements are constantly changing. Secondly, we have used UML which is easy to learn and can model all real world objects. Thirdly, star and snowflake schemas are more efficient for data warehouse design as they are easy to learn and need fewer joins.

III. ADOPTED METHODOLOGY

There are three methods of conceptual modeling of DWs; the first one is the top down approach [31] that is based on the needs of the users, the second is the bottom-up approach [32] that begins with the operational data sources and finally the mixed approach [33] that combines the two previous approaches. We used the top down approach in our modeling phase because we were interested in user's needs. In term of MDA (Model Driven Architecture) [26] our solution is situated in the CIM (Computation Independent Model) level because the models are not inevitably transformed into code. For the abstraction levels [33] (conceptual, logical and physical) our solution is established to cover the conceptual level. Here is a plan showing the position of our solution:



We also adopted object oriented paradigm because it has several advantages for the multidimensional modeling such as the classification / instantiation, the Generalization / specialization and the Aggregation / decomposition. We chose to adopt the object oriented approach which is based on the UML profiles.

IV. UML PROFILE

An UML profile [32] allows specializing UML in a precise domain, it consists of stereotypes, tagged values and constraints. A stereotype [3] is an element of the model that defines new values, new constraints and a new graphic representation. Its role is to give a semantic representation to an element of the model. A stereotype can be represented as a string character between two quotation marks << >> or with an icon. A marked value specifies a new property attached to an element of the model. It is represented between { } and placed with the name of another element. A constraint can become attached to any element of the model to refine its semantics and prevent an arbitrary use of the various elements.

It can be defined with the natural language and/or with the OCL (object constraint language) [32] which is a declarative language that allows developers to write constraints on the model's objects. Recently, UML profiles have a great progress in the ways for conception of Data Warehouses. We present in this section, a conceptual solution for data warehouses design. We proceeded by an UML profile in order to add stereotypes. Our UML profile contains the Class Diagram.

A. Class diagram

We defined in this diagram stereotypes and icons related to the annotation of medical image. This diagram can be used in each case based on annotation.

Classes Stereotypes: We defined in this table classes's stereotypes used in the modeling.

TABLE I
FONT SIZES FOR PAPERS

Stereotypes name	Class type	Description	Icon
<< Annotation >>	class	This stereotype indicates that the class represents the annotation	
<< Medical image >>	class	This stereotype indicates that the class represents the image for the annotation	
<< Hospital >>	class	This stereotype indicates that the class represents the hospital	
<< Dr. GP >>	class	This stereotype indicates that the class represents the generalist doctor	
<< Specialist doctor >>	class	This stereotype indicates that the class represents the specialist doctor	
<< Patient >>	class	This stereotype indicates that the class represents the patient	
<< Date >>	class	This stereotype indicates that the class represents the date of annotation	
<< Person >>	class	This stereotype indicates that the class represents the person in general	
<< Medical Staff >>	class	This stereotype indicates that the class represents the medical staff	
<< Patient file >>	class	This stereotype indicates that the class represents the patient file	

At the end of this section, we propose the following example of defining an UML Class related to the Class Diagram with an extended stereotype and icon using XML:

```

<?xml version="1.0" encoding="UTF-8"?> <PROFILE version="1.0">
<HEADER> <NAME>Annotation</NAME>
<DISPLAYNAME>Annotation</DISPLAYNAME>
<DESCRIPTION>AnnotationBD conceptual modeling</DESCRIPTION>
</HEADER>
<BODY>
<STEREOTYPELIST>
<STEREOTYPE> <NAME>AnnotationObject</NAME>
<DESCRIPTION>Object view.</DESCRIPTION>
<BASECLASSES>
<BASECLASS>UMLAbstractClass</BASECLASS>
</BASECLASSES>
</STEREOTYPE>
<STEREOTYPE>
<NAME>Annotation</NAME>
<DESCRIPTION>Image Annotation </DESCRIPTION>
<BASECLASSES>
<BASECLASS>UMLClass</BASECLASS>
</BASECLASSES>
<PARENT>AnnotationObject</PARENT>
<ICON>annotation.bmp</ICON>
</STEREOTYPE>

```

B. UML profile realization

To implement our approach we chose the StarUML open source platform that uses the language XML to create the profiles UML. In this section we describe StarUML by showing its stretchable parts, and then we model a DW and their components with our UML profile.

1) *The StarUML platform:* StarUML is a modeling platform with the UML language, conceived to support the MDA (Model Driven Architecture) approach. It is characterized by a strong flexibility and an excellent extensibility of its features. Indeed, besides the predefined functions, StarUML allows the addition of new functions which can be adapted to the user's needs. The inconveniences of this platform are that it does not allow specifying more than a stereotype for an element and it excludes the definition of the constraints. Thus in our work we considered that every element has only a single stereotype.

2) *The implementation of Annotation-UML profile:* An UML profile is one package belonging to the mechanism of extension. This package is stereotypical << Profile >> which is written in XML as we see in the following figure:

```
<?xml version="1.0" encoding="UTF-8"?> <PROFILE version="1.0">
<HEADER> <NAME>Annotation</NAME>
<DISPLAYNAME>Annotation</DISPLAYNAME>
<DESCRIPTION>AnnotationBD conceptual modeling</DESCRIPTION>
</HEADER>
```

In the StarUML platform, we added a profile UML called "Annotation" that contains "class diagram". Indeed, we have created a file XML for the profile. Inside this file we appealed to extensions of notation which allows realizing specific notations that are different from those contained in UML. In this part, we represent the interfaces of our added profile:

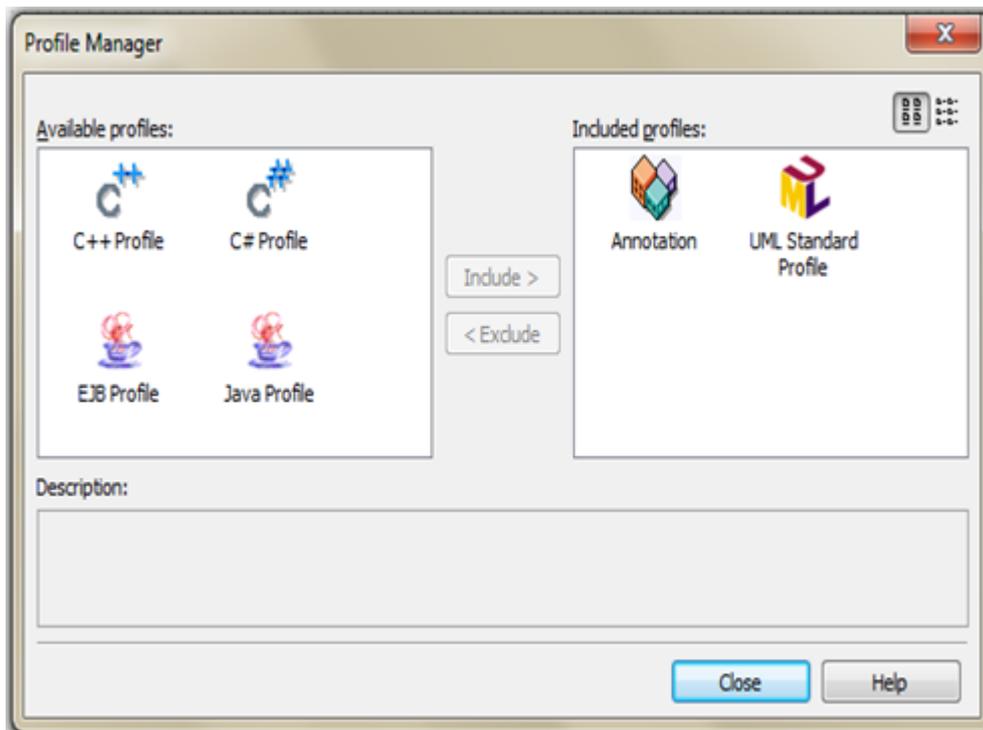


Figure 1. Annotation profile

Now, on the tab **Model Explorer** of upper right select the object "**Untitled**" with a click of the mouse right button and choose the option: **Add-> Design Model** to create a new blank drawing template.

```
Output
[14:45:05] The Profile "annotation1" is loaded successfully.
[14:45:05] The Profile "Cpp" is loaded successfully.
[14:45:05] The Profile "CSharp" is loaded successfully.
[14:45:05] The Profile "EJB" is loaded successfully.
[14:45:05] The Profile "Java" is loaded successfully.
[14:45:05] The Profile "UMLStandard" is loaded successfully.
[14:45:05] The Framework "MFC6.0" is loaded successfully.
```

Figure 2. Loading profit

It is now possible to apply the *Annotation* stereotypes in UML elements of the diagram you created earlier. To test the *Annotation Profile*, add an element **Class** on the diagram. Having the class selected, in the tab **Properties** go into **Stereotype**.

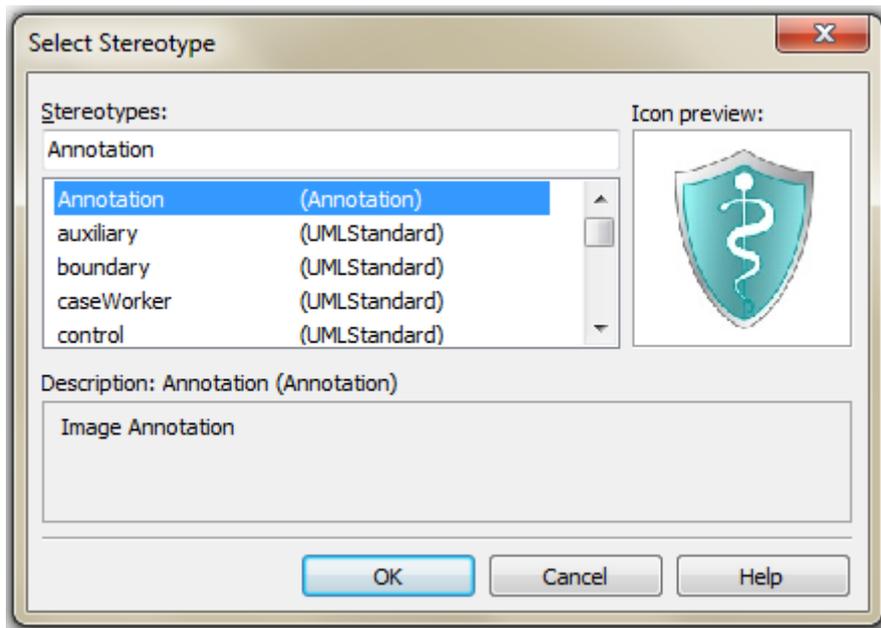


Figure 3. Stereotype of annotation

Note that, when you select a stereotype of *Annotation*, its icon will be shown on the field **Icon Preview**. Select any stereotype and click **OK** to apply it to the class. Note that the name of the stereotype will be displayed between <<...>> at the top of the class. *StarUML* only allows us to add one stereotype per element and allows you to show stereotypes in textual form "Textual", in form of icon "Iconic" or both "Decoration". To change the preview type, select the object and, in the top tab, select the type of view you want in the option **Stereotype Display**. Viewing the stereotype can also be changed by right-clicking on the object in the option **Format option-> Stereotype Display**.

We created a new Class Diagram, in which we added some stereotypes to identify each class (entity). In this diagram, there are some stereotypes and icons that can be used in any application related to the annotation of medical image.

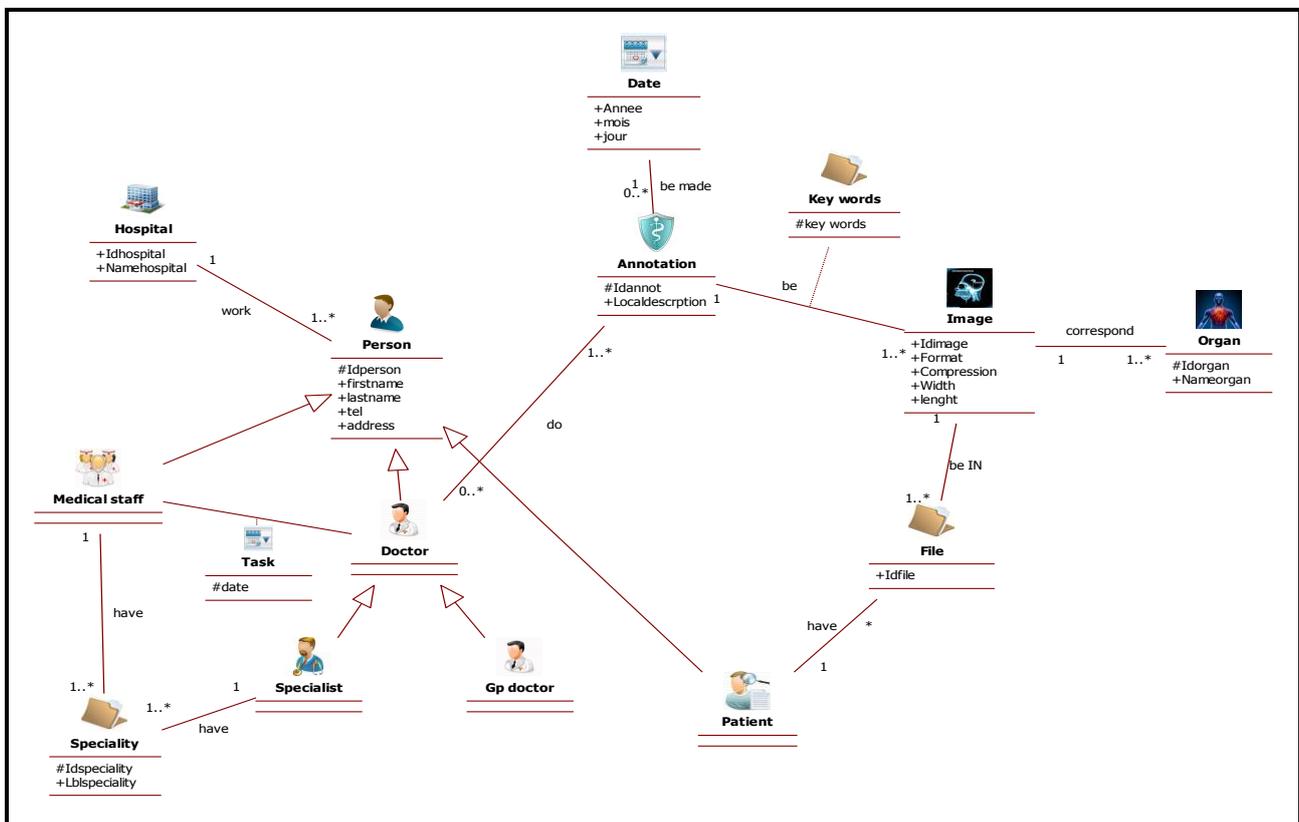


Figure 4. The class diagram

As defined in *Annotation Profile*, stereotypes can also be applied to relationships. Add a class to model and an element Association between them. To apply a stereotype to an element **Association** proceed the same way as described above for elements of type **Class**. Note that, now the only available stereotypes are those previously defined as being of type **UML Association** in the profile.

Remembering that *StarUML* only allows adding and viewing only one stereotype per element, and also that this tool does not support the definition of constraints on OCL language for profiles. Therefore, to verify the validity and consistency of the model is the responsibility of the designer.

V. CONCLUSIONS

In this paper, we described our profile named Annotation. This profile contains the Class Diagram which gives a conceptual representation of the annotation of medical image by specifying relationships between different entities of the diagram. We described the realization of the Annotation profile. To estimate our approach we ended this paper with an experimentation of the class diagram. We propose as future work to represent a model of the component diagram that is based on an UML profile for the physical level.

REFERENCES

- [1] L. Cabibbo and R.Torlone “A Logical Approach to Multidimensional Databases” in Proceedings of 6th International Conference on Extending Database Technology EDBT 98 ,Volume 1337 of LNCS,pp 183-197, Spain, Springer, 1998.
- [2] J. Trujillo, M. Palomar, J. Gómez, I.-Y. Song, “Designing data warehouses with OO conceptual models”, IEEE Comput. 34 (12) (2001) 66– 75,2001.
- [3] R. Jindal and S. Taneja, “Comparative Study of Data Warehouse Design Approaches : A Survey”, Journal: International Journal of Database Management Systems, ISSN 0975-5985, Volume: 4; Issue: 1, 2012
- [4] K. Alaskar¹ and A. Shaikh. (2009) ” Object Oriented Data Modeling for Data Warehousing (An Extension of UML approach to study Hajj pilgrim’s private tour as a Case Study)”, International Arab Journal of e-Technology, Vol. 1, No. 2.
- [5] S.Lujan Mora and I.Song, “Multidimensional Modeling with UML Package Diagrams “ In Proc. of the 21st Int. Conf. on Conceptual Modeling.Lecture Notes in Computer Science pp 199-213,Finland, October 7-11,2002, . Springer-Verlag, 2002.
- [6] S. Luján-Mora, J. Trujillo, and I. Song, “Multidimensional modeling with UML package diagrams warehouses” in Proceedings of 21 st International Conference on Conceptual Modeling, ER 02 ,Volume 2503 of LNCS,pp 199-213, Finland, Springer, 2002.
- [7] M. Blaschka, C. Sapia, G. Höfling, and B. Dinter, “ Finding your way through ultidimensional data models” In Proceedings of the 9th International Conference on Database and Expert Systems Applications DEXA’98, volume 1460 of Lecture Notes in Computer science, pp 198–203, Vienna, Austria, Springer –Verlag, August 1998.
- [8] S. Rizzi, M. Golfarelli. “A Methodological Framework for Data Warehouse Design”. DOLAP 98 Washington DC USA.Copyright ACM 1999 1-581 13-120-8/98, 1998.
- [9] R. T.,Abello. “Research in Data Warehouse Modeling and Design:Dead or Alive?”. DOLAP’06, Arlington, Virginia, USA.Copyright 2006 ACM 1-59593-530-4/06/0011, 2006.
- [10] S. Rizzi, Matteo Golfarelli,D.Maio ”The Dimensional Fact Model: A Conceptual Model for Data Warehouses.” International Journal of Cooperative Information Systems (IJC IS),7(2- 3):215-247,1998.
- [11] K. Alaskar¹ and A. Shaikh. ” Object Oriented Data Modeling for Data Warehousing(An Extension of UML approach to study Hajj pilgrim’s private tour as a Case Study). International Arab Journal of e-Technology, Vol. 1, No. 2, 2009.
- [12] E .Medina and S.L. Mora “A Web Oriented Approach to manage Multidimensional Models through XML Schemas and XSLT “EDBT 2002 Workshops, LNCS 2490, pp. 29–44, 2002. Springer-Verlag Berlin Heidelberg, 2002.
- [13] M. Golfarelli and S. Rizzi ” WanD: A CASE Tool for Data Warehouse Design”. In Demo Proceedings 17th International Conference on Data Engineering (ICDE 2001), Heidelberg, Germany,7-9, 2001.
- [14] S. Rizzi, A. Abelló, J. Lechtenböcker and J. Trujillo “Research in data warehouse modelling and design: dead or alive” DOLAP, ACM, , pp. 3–10, 2006.
- [15] A. Abelló, J. Samos, and F. Saltor “A Framework for the Classification and Description of Multidimensional Data Models” In Proceedings of the 12th International Conference on Database and Expert Systems Applications (DEXA’01), 2001.
- [16] L. Mora and J. Trujilio “A Comprehensive Method for Data Warehouse Design.”in Proceedings of 5th International Workshop on Design and Management of Data Warehouse(DMDW’03), pp 1.1-1.14, 2003.
- [17] J. Trujillo and S. LujánMora ” Physical Modeling of Data Warehouses using UML” DOLAP’04, Washington, DC, USA.Copyright 2004 ACM 1581139772/04/0011, 2004.
- [18] S. Luján-Mora¹, P. Vassiliadis and Juan Trujillo “Data Mapping Diagrams for Data Warehouse Design with UML”in Proceedings of 23 rd International Conference on Conceptual Modeling (ER 04),volume 3288 of LNCS,China, Springer, 2004.

- [19] L. Mora and J. Trujilio "Physical Modeling of Data warehouses by using UML Component and Deployment Diagrams: Design and implementation issues." *Journal of Database Management* 17(1), 2006.
- [20] R. Villarroel, E. Soler, E. Fernández-Medina, J. Trujillo, and M. Piattini "Using UML Packages for Designing Secure Data Warehouses", ICCSA 2006, LNCS 3982, pp. 1024 – 1034. © Springer-Verlag Berlin Heidelberg, 2006.
- [21] D. Mishra, A. Y. Beri and P. Basaran. "A Case study of Data Models in Data Warehousing." 978-1-4244-2624-9/08/ 2008 **IEEE**, 2008.
- [22] K. Alaskar and A. Shaikh "Object Oriented Data Modeling for Data Warehousing (An Extension of UML approach to study Hajj pilgrim's private tour as a Case Study). *International Arab Journal of e-Technology*, Vol. 1, No. 2, 2009.
- [23] H Ma, Y. Yang and F. Zhang "The Anti-standardized Design Research of Data Warehouse". **IEEE**, 2009.
- [24] Fernandez Medina and al. " Model Driven Development of Secure XML Data Warehouses: A Case Study" *EDBT 2010*, Lausanne, Switzerland. Copyright 2010 ACM 978-1-60558-945-9/10/0003.
- [25] P. Pahwa and S. Taneja. "Design of a Multidimensional model using Object Oriented Features in UML." *IARS International journal*, 2011.
- [26] F. Pinet and al. "The use of UML to design agricultural data warehouses." *AgEng 2010*, International Conference on Agricultural Engineering, France, 2010.
- [27] A. Gosain and S. Mann "Object Oriented Multidimensional Model for a Data Warehouse with Operators", *International Journal of Database Theory and Application* Vol. 3, No. 4, 2010.
- [28] J. Pardillo and Jose-Norberto Mazón " Using Ontologies for the Design of Data Warehouses." *International Journal of Database Management Systems (IJDBMS)*, Vol.3, No.2, 2011.
- [29] P. Pahwa, S. Taneja and G. Thakur "Uclean: A Requirement based Object Oriented ETL Framework", *International Journal of Computer Science & Engineering Survey (IJCSES)* Vol.2, No.4, November 2011.
- [30] P. Pahwa and S. Taneja "Design of a Multidimensional model using Object Oriented Features in UML." *IARS International journal*, 2011.
- [31] H. Ma, Y. Yang and F. Zhang "The Anti-standardized Design Research of Data Warehouse". **IEEE**, 2009.
- [32] A. Gosain and S. Mann "Object Oriented Multidimensional Model for a Data Warehouse with Operators", *International Journal of Database Theory and Application* Vol. 3, No. 4, 2010.
- [33] M. Sharma, N. Rajpal and B.V.R.Reddy "Physical Data Warehouse Design using Neural Network". *International Journal of Computer Applications* 1(3):86–94, February 2010. Published By Foundation of Computer Science.