Linking Context to Data Warehouse Design

Aadil Bouchra[1], Kzaz Larbi[2], Ait Wakrime Abderrahim[3], Sekkaki Abderrahim[4]

LIAD, Faculté des Sciences Ain Chock, Université HASSAN II, Casablanca, Maroc[1, 4]
Higher Institute of Commerce and Business Administration, Casablanca, Maroc[2]
Institut de Recherche Technologique Railenium, Famars, France[3]

Abstract—**Data warehouses are now widely used for analysis and decision support purposes. The availability of software solutions, which are more and more user-friendly and easy to manipulate has made it possible to extend their use to end users who are not specialists in the field of business intelligence. The purpose of this article is to provide an approach that assists non-expert users in the data warehouse design process and integrates their contextual data. As well as to provide a method that assists non-expert users in data warehouse design process while incorporating their contextual data. Our proposal consists of a context model and a comprehensive Data Warehouse construction method that attaches the context to data warehouses and uses it to produce customized data marts adapted to the decision makers context.**

Keywords—Business intelligence; data warehouse; context; data mart

I. INTRODUCTION

Recent developments in Business Intelligence (BI) area have been marked by the availability of numerous software solutions combining functional richness, user-friendliness and easiness to use by end-users. Today BI software solutions provide, in addition to their basic functions such as data extraction, transformation and load, a rich and interactive catalog of data processing and visualization features. This has led businesses to enlarge the use of BI solutions at various levels of responsibility and to cover plenty of functions and work positions. Moreover, recent developments in Linked Open Data area [1], offer new opportunities to improve market trends watching and monitoring capabilities, by allowing access to vast, structured and semantically enriched external data sources.

However, implementing a Data Warehouse, which constitute the stone corner of any BI solution, remain the domain of experts and require lengthy steps and intensive analysis and design efforts. We consider that Data Warehouse design methods must include features that assist non-expert users during the Data Warehouse construction process. In addition, contextual data on users participating in data warehouse projects have not been considered by researchers nor by BI solutions providers. We consider that taking into account user context is a crucial issue and could allow producing personalized and context user adapted Data Warehouse. The purpose of this work is to address these two issues, thus we provide a method that assist non-expert users while integrating their contextual data into Data Warehouse design process to produce contextualized data marts. The outline of the paper is as follows. In Section II, we present the related works in data warehouse design, and more specifically

those including context in their approach. Section III presents our method of Data Warehouse design; we first define our model of context, and then outline the steps of our approach. Section IV illustrates our proposal by an example from waste management field using, among others, open data sources. Finally, we conclude and propose some tracks for our future work.

II. RELATED WORKS

A. Data Warehouse Design Approaches

The concept of Data Warehouse appeared about three decades ago; [2] considered as the founder of this concept, defines it as "A subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process". According to [3] "Data Warehouse are databases dedicated to analytic processing; they are used to support decision-making activities in most modern organizations".

Building up Data Warehouse is an arduous and tedious task; it requires efforts of analysis, understanding and identifying end-users needs; it also requires locating appropriate data sources, extracting and integrating data, in order to meet the needs of the decision-makers.

Data Warehouse design approaches are generally classified into two categories [4], data driven approaches and requirements driven. The first approaches starts with an in depth analysis of data stored in internal and/or external databases and derives the Data Warehouse multidimensional scheme [5] Requirement-driven approaches start with an earlier requirements step, which focuses on modeling user analysis [6] Reconciliation between data sources and requirements is done in a later step [7]. Mixed-driven approaches are also proposed [8]. Data sources and requirements are analyzed and satisfied while taking into account available data sources. [9] Provides a survey of the literature related to these design steps and points out pros and cons of the different approaches.

One of the major issues studied intensively in the Data Warehouse field is the integration of heterogeneous data extracted from different sources. The use of ontologies developed in the context of semantic web, [8], [10], [11], [12], [13], [14] and [15], is considered as to be the best tool to solve semantic conflicts and integrate data in the Data Warehouse.

B. What is Context

The notion of context is universal; it refers to all the elements that can influence the understanding of a particular situation. This notion was initially introduced in several disciplines such as psychology, philosophy or linguistics [16]

and [17] and [18]; it is only as from the 90's that it appeared in computer research fields. It is now commonly used in fields such as Artificial Intelligence [19], Information Retrieval [20], Databases [21], Ubiquitous or Pervasive Computing [22], and Recommendation Systems [23], Hence, computer science literature have proposed several definitions; they all describe the context as a set of information associated with something whose nature depends on the application field. The set of information attached to the context strongly depends on how they are used exploited in a specific application field. Thus, according to [24], context is a "Collection of relevant conditions and surrounding influences that make a situation unique and comprehensible". According to [25], context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves. This latter definition is generic; it is widespread and widely used by different research communities. The author in [26] suggests a context model that is a set of couples: attribute name, attribute value associated to each contextual information. A classification of all definitions allows distinguishing three categories. The first category is based on enumerating context attributes. The authors in [27] and [28] use location, time and user identity attributes to define the context. Another context definition refers the context by its synonymous, such as User Environment, or User Situation [29] and [30]. The third category gives specific and application domain dependent [31] and [32].

C. Data Warehouse and Context

Using the concept of context in Data warehouse field is not a new idea; indeed many authors have used it with different purposes. Thus, [33] proposes a query-rewriting algorithm that considers context while loading data warehouse relationships. In [34], the author combines data warehouse with a document repository to build a contextual data warehouse, which helps to produce Data marts characterized by two dimensions: Relevance and Context. The first dimension measures the relevance of facts in the context of analysis, while the second links each fact to unstructured documents stored in the contextual data warehouse which explain and define the associated context. The author in [35] puts forward a multidimensional model that includes user analysis contexts and preferences. The author in [36] proposes a data warehouse design approach to obtain user-specific personalized OLAP models. The suggested approach relies on: (i) A user model representing context information that is relevant to user-personalization, and (ii) A set of personalization rules specifying the required personalization actions. The author in [37] proposes a rewriting-queries algorithm that makes use of contextual hierarchies available in a data warehouse. The author in [38] puts forward a comprehensive contextualized DW design approach by integrating a generic context model that take in consideration concepts as well as properties.

D. Ontologies and Data Warehouse

Several works have used ontologies in different steps of Data Warehouse building process, these works can be classified according to the concerned step: thus [10] and [11] rely on ontologies in data extraction step to ensure an efficient

data selection. Authors in [12] and [13] propose ontologies based models, associate semantics to the extracted terms from data sources and enrich the Data Warehouse with a semantic layer, and thus help users when formulating queries they submit to Data Warehouse [12]. Use ontologies in requirement expression and analysis steps [9]. Points out the shortcomings of different Data Warehouse design approaches and presents the potential benefits of using ontologies to address them. In [14] and [15], the authors transform user requirements into ontologies, align them, produce a global ontology and generate automatically the schema of the Data Warehouse multidimensional model.

III. PROPOSAL OF A CONTEXTUALIZED DW DESIGN METHOD

It results from the above that the notion of context is highly dependent on the field of application and the required goals. We are interested in this work in the issue of DW design; our goal is to help non-expert end users obtaining personalized data mart cubes adapted to their context. In the following, we propose a model and some methods in the use of contextual data. We then present a process of building DW and producing cubes.

A. Proposal of a Context Model

Current software solutions give non-expert end users the opportunity to build and manipulate their autonomously DW. They allow them to take over the different phases of the DW Building process, since the requirement expression phase to data visualization, report and dashboard delivery. Considering users contexts is therefore an important factor in personalizing DW, making them more adapted to end users needs and contexts. In our work, the decision-maker user is the entity for which it is necessary to model, capture and store contextual data. We present in the next section a model of the context and contextual data capture operations.

B. Context Model

The context model of a decision-maker, who is involved in DW building activity, is the cornerstone of our proposal; it consists of six-tuple of attributes defined as follows:

- I: User Identity, this attribute allows the identification of the decision maker inside the organization.

- H: Hierarchical position of the decision maker. Examples: General Manager, Unit Director, Head of Department, Master Officer, Executing Officer, etc.

- F: The Function performed in the context of which the user or the decision-maker pursues his activity. Example: Marketing, Finance, Human Resources, Manufacturing, etc.

- R: An expression of the decision-maker requirement, this is usually formulated as a question asked to understand a management issue. Examples: What is the best medium for promoting a given product? What is the best price to be competitive in the coming months? How the evolution of exchange rates will affect our supplies costs?

- L: the Level, or the Scope, of the decision, as the user perceives it; it could be for example: Strategic, Tactical, or Operational.

- P: Business Process underlying the requirement. Example: Pricing process, launching a promotional campaign Process, etc.

The model we have proposed combines two definitions, the first is the "definition by enumeration" proposed in [27] and [28], the second is the one proposed by [26] who defines a model of context as a list of couples of elements {(Attribute, Value)}.

Since the context is described as a set of couples of data, the model can be easily implemented and makes possible the integration of contextual data into the DW model. It also allows navigating through context lists according to several dimensions. Indeed, each attribute of the model can be turned into a DW dimension. In addition, some of attributes, which characterize the context, contain hierarchy. A dimension hierarchy describes a logical structure using sorted levels to organize and aggregate data. The date dimension for example has often three or more hierarchies that go from day to week, month, quarter, and year. Dimension hierarchy is a very powerful tool that allows user to aggregate data thanks to roll up (drill up) roll down (drill up) OLAP operations.

Some attributes of the proposed model such as R, P, L, and H can be declared as hierarchical dimensions. This allows users to make analysis by aggregating and disaggregating data according to these attributes and to generate corresponding data mart cubes.

In this respect, a global business requirement, designated in the model by the attribute R, can be decomposed in a sequence of elementary requirements. The same logic can be applied to P attribute; indeed, macro processes are decomposable in to sub processes. As well, a user at a level L in the hierarchy of the organization has subordinates who can be involved in building DW too. The organizational structure and its mapping process can be used to establish the hierarchies of L and P attributes

C. Context Collection

Context collection consists of assigning values to the context attributes [26]. This task is to be carried out during two steps:

- During the DW design process and more precisely at the requirement expression step. The collection is done once per need expressed and per decision maker. This corresponds to step I of the process described in the next section (Fig. 1).

- When generating data mart cubes. The contextual data is then captured each time a decision maker wishes to use and manipulate the DW. The attribute values of the context will be used to personalize the generated data mart cubes. This corresponds to step IV of the process described in the following section (Fig. 1).

Examination of the attributes of the context model brings out two categories, explicit attributes that the decision maker have to introduce, and those that can be derived from databases of the organizational structure and its mapping processes. Explicit attributes are User Identity, Requirement expression, a reference or a description of the process to which the requirement is attached as well as its scope. The values of the other attributes: user hierarchical position as well as his function is deductible from the organizational structure. Hierarchy relationships of level and process attributes are also deductible from organizational structure and its mapping processes.

D. Context Collection

Remember that our aim is to provide a method that assists non-expert users to build DW and to produce personalized data marts adapted to their context. To achieve this, we take the method that we developed in our previous works [14] and [15]. The method allows automatic generation of the DW multidimensional schema, so we will extend it by integrating contextual data.

It should be noted that DW building methods, regardless data or requirement centric they are, include the following tasks:

- User requirement collection.

- Determination of available data sources that can meet needs.

- Extraction of data from data sources

- Data transformation.

- Data warehouse multidimensional model implementation and data load.

- Data mart delivery.

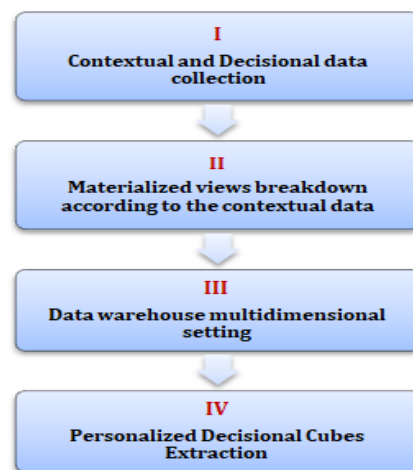The method we provide include all of these tasks and organize them in four steps:
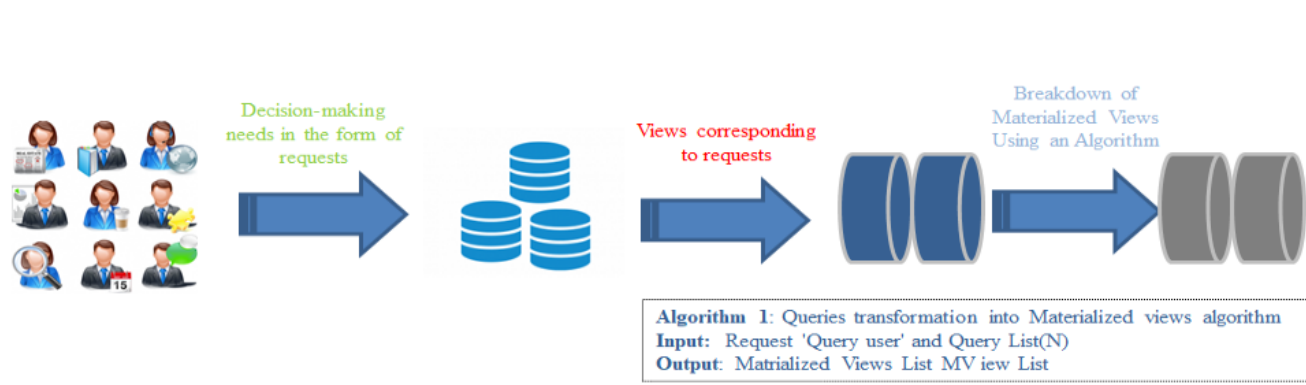

Fig. 1. Example of a Figure Caption.

Fig. 2. Generating Materialized Views Attached to Contexts.

1) Contextual and decisional data collection: This step is to achieve, for each decision maker involved in DW building project and for every need expressed, the following sequence described in Fig. 2:

a) Collect and assign the values to explicit attributes (I, R, P) and deduce implicit attributes values (H, F, l) from the organizational structure and mapping processes data bases. User requirement is a key attribute in the model, it can be expressed either in a natural language or in a specific formal one [39]. Let us note this operation ContextCollect() and C, the resulting context:

C ← contextCollect()

b) Extract data from data sources previously selected and believed relevant by the decision maker. This operation, which use the context C as a parameter, is expressed directly with SQL queries or via appropriate interfaces. Lets us note ExtractData() the operation and Req the set of resulting queries:

Req ← extractData(C)

c) Attach the context to queries, by adding context attribute values stored in the parameter C, to each query of the set Req. Let us note this operation as follows:

Req ← attachContext (C,Req)

d) Launch the queries Req and produce corresponding materialized views. According to [40], a materialized view is a database table that contains the execution result of a query. After this, we obtain a set of materialized views corresponding to the needs of a given decision maker. We note Vi the set of the materialized views corresponding to the ith need expressed by a decision maker. Let us note this operation as follows:

Vi ← materializeView(Req)

2) Group and then breakdown each view Vi according to the context attributes values. This step is described in Fig. 3 and takes place as follows:

a) Carry out the union of all sets of materialized views, be V this set

V ← viewsUnion (V1, V2, V3, …Vi, …. ,Vn)

Where Vi refers to the set of materialized views corresponding to the ith need, and n refers to the total number of needs expressed by all decision-makers who participated in the step I. Note that the same decision-maker can express several needs, on the other hand the same need can be expressed by different decision-makers, of course, with distinct contexts, insofar as identity values are necessarily different.

b) Partition the set V on the basis of a criterion C expressed by the attributes of the context. Let us note Pc the resulting partition, each element of Pc contains a set of materialized views having the same context C. Note the corresponding operation as follows:

PC ← partition (V, C)

Where Vi refers to the set of materialized views corresponding to the itch need, and n refers to the total number of needs expressed by all decision-makers who participated in the step I. Note that the same decision-maker can express several needs, on the other hand the same need can be expressed by different decision-makers, of course, with distinct contexts, insofar as identity values are necessarily different.

c) Examples :

i. Partition (V, I): Partitions V of materialized views by the decision-maker identity. Each element of the partition will contain all views related to a given decision maker.

ii. Partition (V, R): Partitions V according to the requirements expressed by the decision makers. Each element of the partition will contain all views related to the same requirement eventually expressed by different decision makers having different contexts.

iii. Partition (V, R and P): This example allows grouping materialized views having the same requirements and related to the same Business Process.
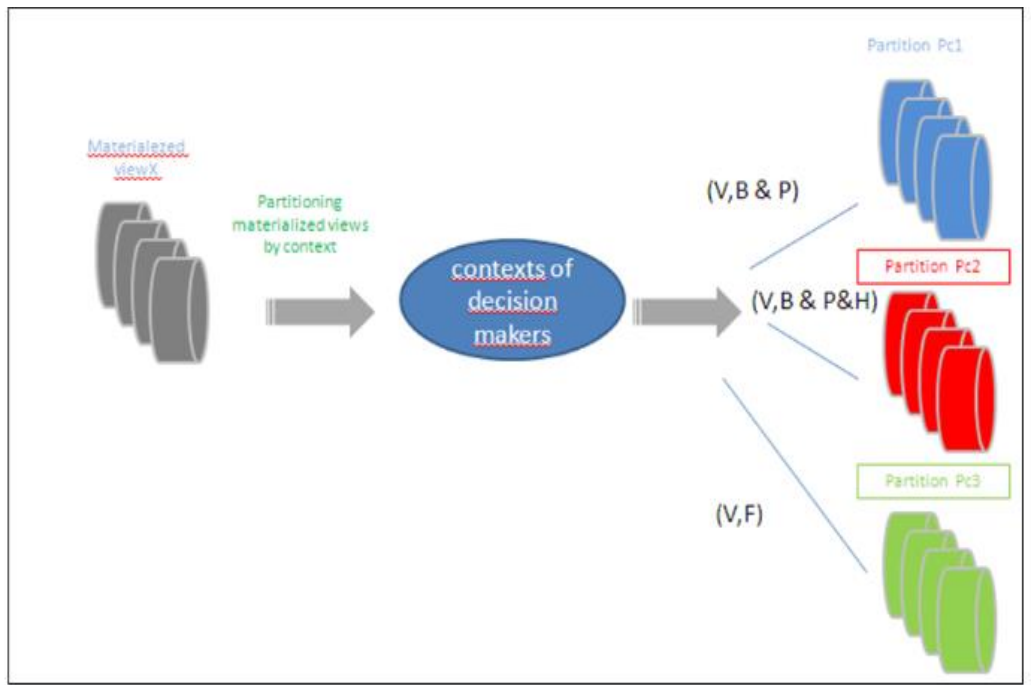
Fig. 3.   Generating Materialized Views Attached to Contexts.

3) Generating the data warehouse model: This step was built on the results of our previous works [14] and [15]; it ends with automatic achievement of Data Warehouse model. Its starting point is a partition Pc consisting of a set of mono contextual materialized views; each materialized view is composed of context attributes as well as attributes extracted from multiple data sources. At this point we face the classic problem of integrating heterogeneous data. It is widely established that the use of ontologies is the best way to resolve semantic heterogeneity and ensure data integration [41]. The sequence of following operations is to be achieved:

a) Transform each materialized view belonging to Pc into an ontology. Each view is to rewrite using an ontology description language like OWL, or RDF. The following operation achieve this transformation, and produces a set of ontologies which we note O_c.

O_C ← Ontologies (P_C)

b) Integrate and merge the set O_c of ontologies.Ontologies belonging O_c may likely have both syntactic and semantic heterogeneities; this is due to the source of their contents. Indeed, context attributes are expressed by different decision makers, as well as data are extracted from different sources leading thus to Integrating and merging all ontologies creates a global ontology. We note Go_c the resulting ontology and mergeIntegrate() the corresponding operation:

GO_C ← mergeIntegrate (O_C)

c) GO_c is a mono contextual global ontology; it describes the whole concepts and terms present in the set V_c of views related to the context C.

d) Generate and produce the Data Warehouse model. This operation is entirely automatized. The algorithm achieving this operation is presented in our work [14] and [15], It delivers a specific and context depending data warehouse DW_c.

DW_C ← generateDW (GO_C)

4) Data mart delivery: Data marts are subsets of data extracted from a global DW [3] Decision-makers use it whenever they need to solve a decision-making problem. They play a key role in understanding, analyzing situations and more broadly supporting decision-make processes. Data marts are used by BI software tools to visualize data and to produce reports and dashboards. The extraction of relevant data marts which matches with the context of the decision maker is a crucial issue to provide efficient support to decision makers. It is the phase that highlights the contribution of our proposal. This step is achieved by the sequence of the two following operations described in Fig. 4:

a) Input and assign the context values of a given decision maker to the corresponding attributes of context C. The list of couple (Attribute, Value) form the partitioning criterion. Let us note Val(C) the function that collects and assign values to C attributes.

b) Extract the data mart from the DW_c taking into account the inputted values of context Val(C). The operation is noted as follows:

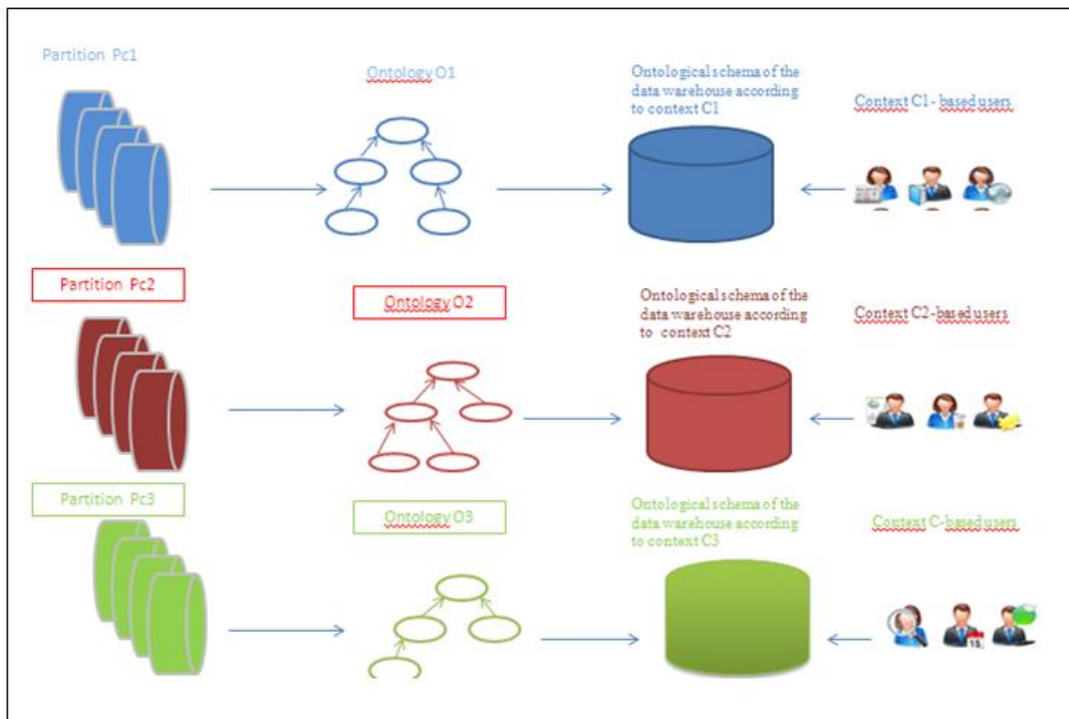DataMart ← extractDataMart (DW_C, Val(C))

Fig. 4.   Production of Data warehouse Schema.

IV.   SUPPORTING EXAMPLE

To illustrate our proposal, we consider an example from the field of "waste management". This area is central in any environmental protection policy of modern cities and is quite complex because of the multitude of stakeholders and the diversity of their concerns, priorities, constraints and data sources. So building relevant and adapted DW is essential to support decision makers. DW should provide data to develop a coherent waste management policy that includes the concerns and visions of the different stakeholders. In addition, this field relies on large volume and variety of data such as types of waste, quantities, levels of danger, their composition, etc. Different stakeholders are also involved such as local government officials in charge of designing and implementing waste management policies in respect of the environment, urban architects, companies emitting waste, citizens etc. Moreover, some data such as the characteristics of industrial waste, sanitary or environmental standards are published and now available in Open Data. Building a DW that meets decision-makers needs requires taking into account quite diverse and heterogeneous data.

We consider four stakeholders A, B, C and D involved in the management and treatment of waste in a certain city. A is an urban architect of the urban commune, working on a strategic project concerning the development of the city in respect of environmental constraints. B is a team leader of the "Environment Centre" of a municipality; he is in charge of coordinating and monitoring environmental projects and monitoring waste production. The stakeholder C is responsible for "Standard and Quality" in a company that plans to set up a new production unit in the perimeter of the municipality; he is interested in the evaluation and the control of waste treatment

processes and compliance with the standards and constraints imposed by local authorities. D is "expert consultant of polluted sites and soils"; he carries out soil pollution diagnosis based on documentary studies and site investigations. Given all this elements, we intend now to apply the method to design a DW that meets stakeholder's needs.

A.   Contextual and Decisional Data Collection

The actor A occupies the position of "head of department" within the municipality, he performs the function "urban architect", his needs are about "waste treatment", and are at a "strategic" level and falls within the scope of the "Setting up an integrated waste system" process.

Given these data, the context of the actor A is defined as follows:

$$C\ (A) \leftarrow contextCollect()$$

C (A) ← (« ID_A », « Head of department», « Urban Architect », « Waste Treatment », « Strategic », « Setting up an integrated waste system»)

The actor B, occupies "Team Leader" position within his organization "city council", he exercises the function "Environment management"; its decision-making need is related to "waste treatment" and is part of the "Operational" level and is part of "Setting up an integrated system for the treatment of waste" process. The context of B is then defined as follows:

$$C\ (B) \leftarrow contextCollect\ ()$$

C(B) ← (« ID_B », « Team Leader », « Environment Management », « Waste treatment », « Operational », « Setting up an integrated waste system»)

The actor C, holds the position of "head of department" within a private company, he performs the function "Standards and quality"; its decision-making need concerns the "treatment of industrial waste", falls under the "Tactical" level and is part of "Company's waste treatment and recycling" process. The context of the actor B is:

C (C)← contextCollect ()

C(C) ← (« ID_C », « head of department », « Standards and quality», « treatment of industrial waste», « Tactical», « Company's waste treatment and recycling »)

The last actor in our example, occupies the "Engineer" within a private company, he is Consultant of polluted sites; his needs are about "Setting a procedure to clean up a given polluted site located in the area of the municipality", are at the "Tactical" level, and are part of "Company's waste treatment and recycling" process. Collection of the context of this actor is defined as follows

C (D)← contextCollect ()

C(D) ← (« ID_D », « Engineer », « Consultant of polluted sites», « Setting a procedure to clean up a given polluted», « Tactical», « Company's waste treatment and recycling »)

Once the data context has been defined, the task now is to extract data from sources that for each actor considers relevant. Thus for the actor A, it is a question of identifying:

- The production of waste by Kg / People / day.

- The recycling rate of products from household waste.

- The recycling rate by branch.

To achieve this, the actor A submits to the data source whose schema is given in Fig. 5, the following queries {Q1, Q2, and Q3}:

- Q1: SELECT Title FROM Wastes GROUP BY CodeArea.

- Q2: SELECT Weight FROM Production GROUP BY IdProducer.

- Q3: SELECT Weight FROM Production GROUP BY CodeArea.

Actor A completes its needs with data on standards available on the website of the Ministry of the Environment. This website publishes several waste indicators by city and type, and provides data on the standards applied by activity sector and waste type. The corresponding query is:

Q4: SELECT * FROM Standards

Among the needs of actor B we can give by way of illustration:

- Quantification of household waste by neighborhood.

- Evaluation of environmental impacts related to each product consumed.

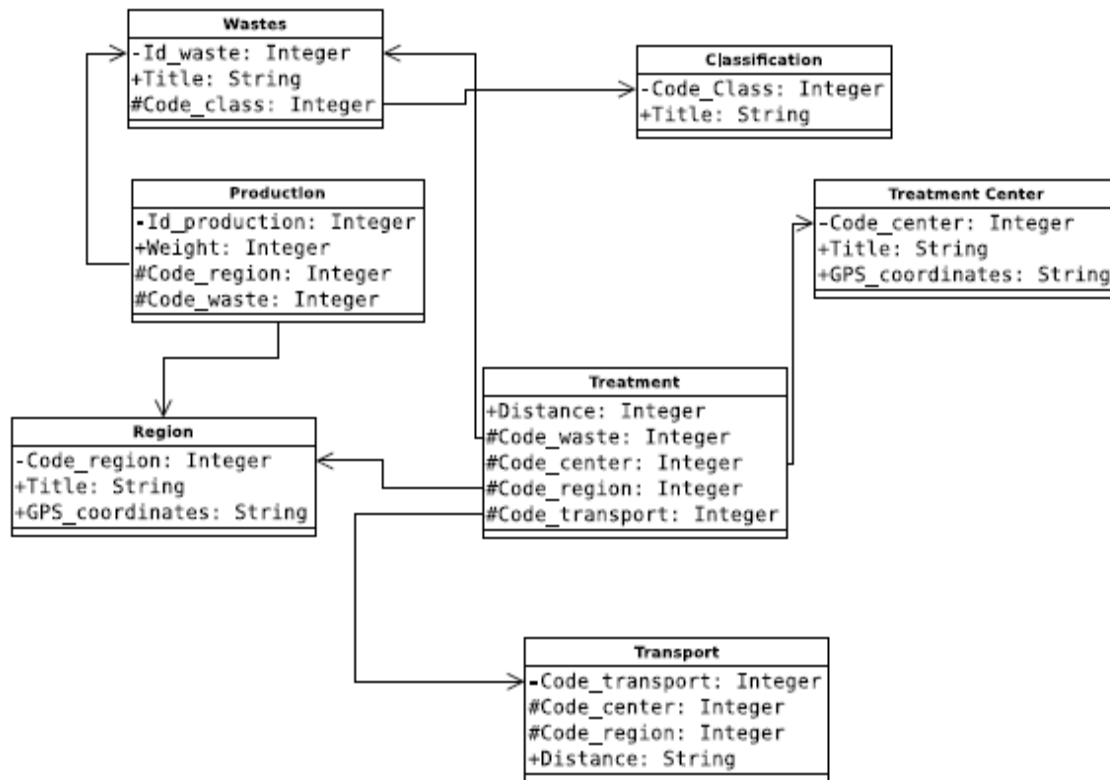- Identification of recovery rates by type of waste.

Fig. 5. Excerpt of the Schema of the Data Source used by Actors A and B.

Actor B extracts data by submitting queries {Q5, Q6} to data source whose schema is given in {Fig. 5}

Q5: SELECT Wastes.Title FROM Wastes WHERE Wastes.CodeClass= Classifcation.CodeClass AND Classifiation. Title =Household

Q6: SELECT Weight FROM Production WHERE Production.CodeWaste=Wastes.IdWaste AND Wastes.CodeClass= Classifcation.CodeClass AND Classification. Title =Houeshold

Actor C needs the following indicators:

- Waste rate during the manufacture of a product

- Cost of waste management

- List of industrial wastes

He extracts necessary data by submitting queries {Q7, Q8} to the data source whose schema is given in {Fig. 6}:

Q7: SELECT Wastes.Title FROM Wastes AND Classification WHERE Classification.Title=Industrial AND Wastes.codeclass=Classification.codeclass;

Q8: SELECT Production.codewatse FROM Wastes AND Classification WHERE Classification.Title=Industrial

Q9: SELECT * FROM Standards WHERE Sector=industrial AND Wastes.codeclass=Classification.codeclass GROUP BY Production.coderegion;

He completes his needs by submitting the query Q9 to an open data source on standards available on the web site of the ministry of environment:

Q9: SELECT * FROM Standards WHERE Sector=industrial

The actor D needs the following indicators:

- Identification of potential resources of pollution.

- Identification of types of pollutants.

- Definition of the characteristics of different sites for each region.

- Define sites potentially emitting CO2.

D extracts data by submitting queries {Q10, Q11} to the data source which schema is provided in {Fig. 6}

- Q10: SELECT nature FROM Pollutant;

- Q11: SELECT * FROM Site;

The whole queries are expressed by using the formalism described in previous section, as follows:

$\{Q_1, Q_2, Q_3, Q_4\} \leftarrow$ extractData(C(A)) ;

$\{Q_5, Q_6\} \leftarrow$ extractData (C(B)) ;

$\{Q_7, Q_8, Q_9\} \leftarrow$ extractData (C(C)) ;
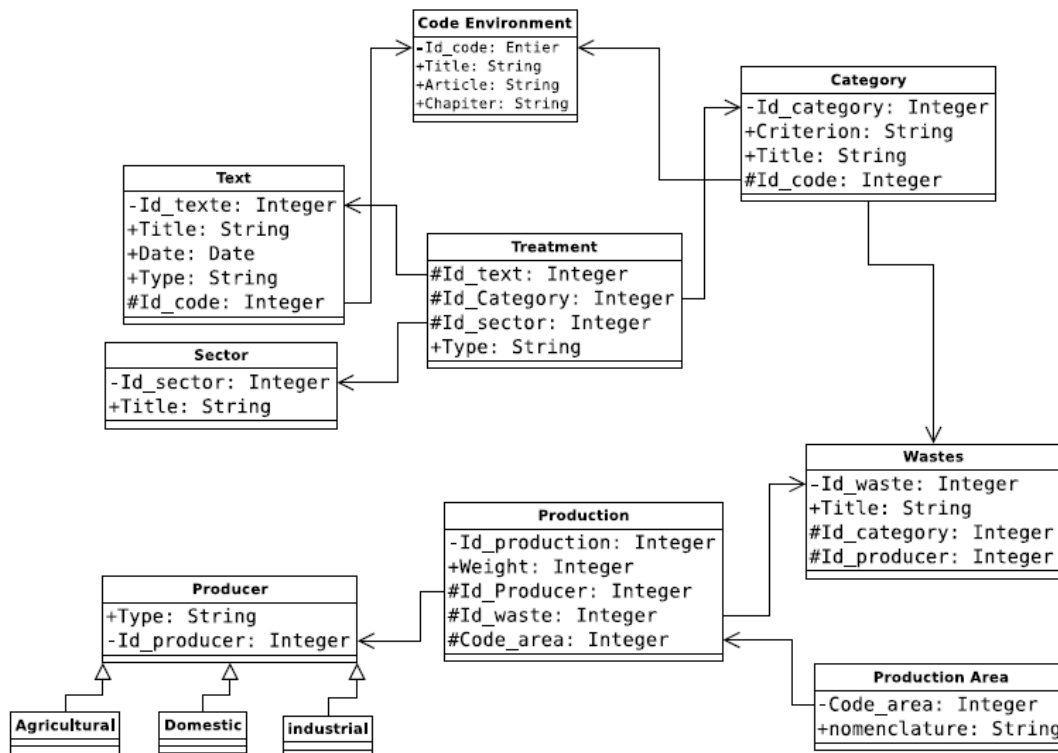
$\{Q_{10}, Q_{11}\} \leftarrow$ extractData (C(D));

Fig. 6. Excerpt of the Schema of the Data Source used by Actors C and D.

Then contexts are attached to each query:

$R_i \leftarrow$ attachContext (C(A), R_i)) i=1,..4

$R_i \leftarrow$ attachContext (C(B), R_i)) i=5..6

$R_i \leftarrow$ attachContext (C(C), R_i)) i=7..9

$R_i \leftarrow$ attachContext (C(D), R_i)) i=10..11

Transform queries into materialized views:

$V_i \leftarrow$ materializeView (R_i)) i=1..11

B. Partitioning Materialized Views According to the Context

This step takes place in several stages; we first unite the 11 obtained views:

$V \leftarrow$ viewUnion (Vi) i=1..11

Then we partition the set V on the basis of a given criterion. Note that at this point, there is various expressions to partition the set V, each expression leads to a specific partition and consequently to specific data warehouse:

- Partitioning on the basis of the level L of decision criterion is achieved by:

$P_N \leftarrow$ Partition (V, L)

This operation produces a partition comprised of three subsets of views, each subset corresponds to a single level, strategic for the views belonging to actor A, operational for view of actor B, and tactical for C and D views:

$P_L = \{\{V_1, V_2, V_3, V_4\}, \{V_5, V_6\}, \{V_7, V_8, V_9, V_{10}, V_{11}\}\}$

- Partitioning on the basis of the business process criterion P:

$P_P \leftarrow$ Partition (V, P)

This produce two subsets, the first contains A and B views which are related to the business process Setting up an integrated waste system, the second subset gathers C and D view also related to the business process Company's waste treatment and recycling

$P_P = \{\{V1, V2, V3, V4, V5, V6\}, \{V7, V8\ V9, V10, V11\}\}$

- Partitioning on the basis of the business process P and decision maker requirement R criterion. This case illustrates the power of our proposal; it shows how it is possible to combine context attributes to express the Partitioning criterion and then to obtain numerous data warehouses depending on user contexts.

$P_{R\ and\ P} \leftarrow$ Partition (V, R and P)

This partition gives three subsets of views. The first contains the views of actors A and B who have the same requirements R and work on the same business process P. The second contains the view belonging to actor C; the latter works on the same business process as the actor D but has different requirement; so C and D have different subsets of views.

$P_{B\ and\ P} = \{\{V1, V2, V3, V4, V5, V6\}, \{V7, V8\ V9\}, \{V10, V11\}\}$.

C. Generating Data Warehouse

This step consists first in transforming each partition P_c, containing mono contextual views, into ontology. Take for instance, the partition P_p based on the business process criterion and that contains two subsets of mono contextual views; the two corresponding ontologies are obtained by this operation:

$O_p \leftarrow$ Ontologies (P_p)

The obtained ontologies are then integrated and merged into a global one by applying the following operation:

$GO_p \leftarrow$ integateMerge (O_p)

- GO_p describes all materialized views related to a business process criterion and belonging to the set V_p. Finally the DW is generated from the global ontology by the following operation:

$DW_p \leftarrow$ generateDW (GO_p)

D. Generating Data Marts

This step, consists of generating personalized data marts adapted to the decision maker contexts. Consider, for example, the partition that corresponds to the level of decision, each decision maker can have his specific data mart depending on his hierarchy. The corresponding data marts are then obtained using the sequence of following operations:

$DM_{actor\ A} \leftarrow$ extractDataMart($\mathcal{DW}_p$, « Strategic»)

$DMt_{actor\ B} \leftarrow$ extractDataMart($\mathcal{DW}_p$,« Operational»)

$DM_{actor\ C} \leftarrow$ extractDataMart($\mathcal{DW}_p$,« Tactical»)

$DM_{actor\ D} \leftarrow$ extractDataMart($\mathcal{DW}_p$,« Tactical»)

V. CONCLUSION

We presented a data warehouse designing solution that is intend to support non-expert users while taking into account and integrating their contexts into the data warehouse. For this purpose, we proposed a model of context representation and contextual data collection and a method that generate the DW multidimensional model. Resolution of conflicts related to the heterogeneity of both context attributes and data extracted from data sources was achieved by using ontologies. The use of contextual data makes it possible to automatically produce customized cubes adapted to the context of the decision maker. This work has also allowed us to further refine the phase of automatic generation of the data warehouse schema and to adapt it to the consideration of contexts. An example from the waste management field in smart cities, which use different data sources including open data, was used to test our proposal. We plan to continue this work by implementing the whole process and integrating databases that describe organizational structure and business processes map.

REFERENCES

[1] Bizer, Christian, Heath, Tom and Berners-Lee, Tim, Linked Data - the story so far, International Journal on Semantic Web and Information Systems, 5 (3), 122 (2009)

[2] W. H. Inmon, Building the Data Warehouse. New York, NY, USA: John Wiley & Sons, Inc., 1992.

[3] NGELA BONIFATI Politecnico di Milano FABIANO CATTANEO Cefriel STEFANO CERI Politecnico di Milano ALFONSO FUGGETTA Politecnico di Milano and Cefriel and STEFANO PARABOSCHI Politecnico di Milano Designing Data Marts for Data Warehouses, ACM Transactions on Software Engineering and Methodology, Vol. 10, No. 4, October 2001.

[4] R. Winter and B. Strauch, A method for demand-driven information requirements analysis in data warehousing projects , in System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on. IEEE, 2003, pp. 9ñpp

[5] M. Golfarelli and S. Rizzi, ìA methodological framework for data warehouse design,î in Proceedings of the 1st ACM international workshop on Data warehousing and OLAP. ACM, 1998, pp. 3ñ9.

[6] R. Kimball and M. Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd ed. New York, NY, USA: John Wiley & Sons, Inc., 2002.

[7] P. Giorgini, S. Rizzi, and M. Garzetti, ìGoal-oriented requirement analysis for data warehouse design,î in Proceedings of the 8th ACM international workshop on Data warehousing and OLAP. ACM, 2005, pp. 47ñ56

[8] L. Zepeda, M. Celma, and R. Zatarain, ìA mixed approach for data warehouse conceptual design with mda,î in Computational Science and Its ApplicationsñICCSA 2008. Springer, 2008, pp. 1204ñ1217

[9] J.-N. Mazon, J. Trujillo, M. Serrano, M. Piattini et al., ìDesigning data warehouses: from business requirement analysis to multidimensional modeling,î REBNITA, vol. 5, pp. 44ñ53, 2005

[10] M. G. Taylor, K. Stoffel, and J. A. Hendler, ìOntology-based induction of high level classification rules.î in DMKD, 1997, pp.

[11] L. Brisson and M. Collard, ìAn ontology driven data mining process,î in International Conference on Enterprise Information Systems, France, 2008, pp

[12] A. Nabli, J. Feki, and F. Gargouri, ìAn ontology based method for normalisation of multidimensional terminology,î in Advanced Internet Based Systems and Applications, ser. Lecture Notes in Computer Science, E. Damiani, K. Yetongnon, R. Chbeir, and A. Dipanda, Eds. Springer Berlin Heidelberg, 2009, vol. 4879, pp.

[13] D. Awad and A. REVEL, ìOntology-based solution for data warehousing in genetic neurological disease,î in Proceeding of the world congress on Engineering, 2012

[14] B. Aadil, A. A. Wakrime, L. Kzaz and A. Sekkaki, "Ontological approach for Data WareHouse design," 2015 International Conference on Protocol Engineering (ICPE) and International Conference on New Technologies of Distributed Systems (NTDS), Paris, 2015, pp. 1-7

[15] B. Aadil, A. A. Wakrime, L. Kzaz and A. Sekkaki, "Automating Data warehouse design using ontology," 2016 International Conference on Electrical and Information Technologies (ICEIT), Tangiers, 2016, pp. 42-48

[16] Matthew Chalmers. A historical view of context. Computer Supported Coope- rative Work (CSCW), 13(3-4):223ñ247, 2004

[17] C Bastien. Contexte et situation. HoudÈ, O., Kayser, D., Koenig, O., Proust, J. & Rastier, F., Dictionnaire des Sciences Cognitives. Paris: PUF, 1998

[18] Thomas KLandauer,PeterWFoltz,and Darrell Laham. An introduction to la- tent semantic analysis. Discourse pro- cesses, 25(2-3):259ñ284, 1998

[19] Patrick BrÈzillon. Context in problem solving: a survey. The Knowledge Engi- neering Review, 14(01):47ñ80, 1999

[20] Lev Finkelstein, Evgeniy Gabrilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppin. Pla- cing search in context: The concept re- visited. In Proceedings of the 10th in- ternational conference on World Wide Web, pages 406ñ414. ACM, 2001

[21] Cristiana Bolchini, CA Curino, Gior- gio Orsi, Elisa Quintarelli, Rosalba Ros- sato, Fabio A Schreiber, and Letizia Tanca. And what can context do for data? Communications of the ACM, 52(11):136ñ140, 2009

[22] Bravo, J., Fuentes, L. Lopez de Ipina, D. Theme issue : ubiquitous computing and ambient intelligence, Personal and Ubiquitous Computing, 2011, vol. 15, n∞ 4, pp. 315-316

[23] F. Ricci, L. Rokach, B. Shapira, and K.B. P., ìRecommender systems handbook,î Recommender Systems Handbook, 2011, pp.

[24] Patrick BrÈzillon. Context dynamic and explanation in contextual graphs. In Modeling and Using Context, pages 94ñ 106. Springer, 2003

[25] Dey, A., Abowd, G., and Salber, D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Human-Computer Interaction, 16(2)

[26] Schilit, B., Theimer, M., and Welch, B. (1993). Customizing mobile applications. In The USENIX Symposium on Mobile and Location-Independent Computing, pages 129

[27] Peter J Brown, John D Bovey, and Xian Chen. Context-aware applications: from the laboratory to the marketplace. Per- sonal Communications, IEEE, 4(5):58ñ 64, 1997

[28] Nick Ryan, Jason Pascoe, and David Morse. Enhanced reality fieldwork: the context aware archaeological assistant. Bar International Series, 750:269ñ274, 1999

[29] Peter J Brown. Triggering informa- tion by context. Personal Technologies, 2(1):18ñ27, 1998

[30] ndy Ward, Alan Jones, and Andy Hop- per. A new location technique for the active office. Personal Communica- tions, IEEE, 4(5):42ñ47, 1997

[31] Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In First Workshop on Mobile Computing Systems and Applications. WMCSA 1994, pages 85. IEEE, 1994

[32] Anind K Dey, Gregory D Abowd, and Andrew Wood. Cyberdesk: A fra- mework for providing self-integrating context-aware services. Knowledge- Based Systems, 11(1):3Ñ13, 1998

[33] Diego Calvanese, Giuseppe De Gia- como, Maurizio Lenzerini, Daniele Nardi, and Riccardo Rosati. A prin- cipled approach to data integration and reconciliation in data warehousing. In DMDW, volume 99, page 16, 1999

[34] Juan Manuel Pérez, Rafael Berlanga, María José Aramburu, and Torben Bach Pedersen. A relevance-extended multidimensional model for a data warehouse contextualized with documents. In Proceedings of the 8th ACM international workshop on Data warehousing and OLAP, pages 19–28. ACM, 2005

[35] Houssem Jerbi, Franck Ravat, Olivier Teste, and Gilles Zurfluh. Management of context-aware preferences in multidi- mensional databases. In Third Interna- tional Conference on Digital Informa- tion Management, 2008. ICDIM 2008, pages 669ñ675. IEEE, 2008

[36] Irene GarrigÛs, Jesˋs Pardillo, Jose- Norberto MazÛn, and Juan Trujillo. A conceptual modeling approach for olap personalization. In International Confe- rence on Conceptual Modeling, pages 401ñ414. Springer, 2009

[37] Yoann Pitarch, CÈcile Favre, Anne Laurent, and Pascal Poncelet. Context- aware generalization for cube measures. In DOLAP 2010, ACM 13th Internatio- nal Workshop on Data Warehousing and OLAP, Toronto, Ontario, Canada, Oc- tober 30, 2010, Proceedings, pages 99ñ 104, 2010

[38] Barkat, O., Khouri, S., Bellatreche, L., Boustia, N., Bridging Context and Data Ware- houses through Ontologies. In Proceedings of the 32nd ACM SIGAPP Symposium On Applied Computing (ACM SAC), Marrakech, Morocco, April 03-07, 2017

[39] Bargui, Fahmi and Feki, J and Ben-Abdallah, HanÍne. (2008). Vers une expression des besoins dÈcisionnels en langage naturel.

[40] H. Gupta and I. Mumick, ìSelection of views to materialize in a data warehouse,î Knowledge and Data Engineering, IEEE Transactions on, vol. 17, no. 1, pp. 24ñ43, Jan 2005

[41] H. Elasri, A. Sekkaki, and L. Kzaz, ìAn ontology-based method for semantic integration of business components,î in New Technologies of Distributed Systems (NOTERE), 2011 11th Annual International Conference on, May 2011, pp. 1N°8