A Survey of Decisional Requirements:

Imprecision study

LARBI Abdelmadjid
Mathematics & Computer Science Department
EEDIS Laboratory
UDL SBA, Algeria
UTMB, Algeria
ENERGARID Laboratory
Amdlarbi@gmail.com

Mimoun M
Computer Science Department
Ecole Supérieure Informatique,
Sidi Bel Abbes, Algeria
EEDIS Laboratory
MMalki@gmail.com

Boukhalfa K
Computer Science Department
USTHB, Algiers Algeria
IS Laboratory
Kboukhalfa@gmail.com

***Abstract*—The success and the failure of a data warehouse (DW) project are mainly related to the design phase according to most researchers in this domain. When analyzing the decision-making system requirements, many recurring problems appear and requirements modeling difficulties are detected. Also, we encounter the problem associated with the requirements expression by non-IT professionals and non-experts makers on design models. The ambiguity of the term of decision-making requirements leads to a misinterpretation of the requirements resulting from data warehouse design failure and incorrect OLAP analysis. Therefore, many studies have focused on the inclusion of vague data in information systems in general, but few studies have examined this case in data warehouses. This article describes one of the shortcomings of current approaches to data warehouse design which is the study of in the requirements inaccuracy expression and how ontologies can help us to overcome it. We present a survey on this topic showing that few works that take into account the imprecision in the study of this crucial phase in the decision-making process for the presentation of challenges and problems that arise and requires more attention by researchers to improve DW design. According to our knowledge, no rigorous study of vagueness in this area were made.**

*Keywords— **Data warehouses Design, requirements analysis, imprecision, ontology***

I. INTRODUCTION

In the literature [1], DW designers follow different directions on the number of design phases. Although the main phases are four [2], studies in [3] proposed three steps compared to the conceptual level, logical level and the physical level, whereas in the study [4] follows the classic division into four stages. Another comprehensive study by [5] defines eight steps to design a DW, these measures include: requirements analysis, analysis and reconciliation, conceptual design, refinement workload, logical design, the design of the staging data, physical design and implementation. According to [6], there are five steps in the DW design: requirements analysis, conceptual modeling, logical modeling, ETL and physical modeling. [7] has stated that the requirements definition phase is paramount and has an impact on almost all decisions in DW project [8]. Most failed of DW projects run poor requirements definition phase and some of them skip this phase in order to focus on design issues such as query performance and database modeling. [9] Among the requirements: What is the granularity scope in each data mart? What constraints (eg, legal terms, OLAP constraints) restrict multidimensional data analysis? How users like to have summary data for each dimension? The answers to this and many other important issues will govern dimensional modeling and must be represented in the overall project objectives. If the decision maker is non-IT and non-expert on patterns design, its requirements expressions will present inaccuracies that can taint the design.

This paper highlights a very important step in the decision process, in terms of requirements analysis. We focus on the study of vagueness in the decision-making requirements and its impact on the whole process. It describes one of the shortcomings of current approaches to data warehouse design which is the inaccuracy study in the requirements analysis phase and especially in the requirements expression and how ontologies can help us to overcome it. This paper presents four sections. The first section presents introduction of the treated problem, the second section present a background of the subject, the section 3 present a survey of requirements analysis. In Section 4, we present the challenges and the open problems of the studied subject then finally a conclusion.

II. BACKGROUND

In this section, we will introduce the basic concepts of this work in detail, which are two concepts: Vagueness or imprecision and requirements analysis of decision-making:

A. Imprecision

In the literature, there are several types of imperfect information [10] which is the vagueness concerning the content of information, uncertainty as to the reality of the information, the inconsistency that characterizes the conflict, the incompleteness extent the absence of part of the information and the ambiguity of the various possible interpretations. The imperfections data in the decision-making systems are due to many causes such as the difficulty to have close models of reality and limits of instruments and procurement. The inaccuracy may be due to the requirements expression ambiguity of the decision maker. Based on fuzzy logic, Lee99 [11] proposed to formulate vague requirements along four dimensions: (i) extend a class by grouping objects with properties similar to a fuzzy category (ii) encapsulating the fuzzy rules in fuzzy class to describe the relationship between attributes, (iii) to evaluate the function of a fuzzy class member by considering both the static and dynamic properties, and (iv) to model the uncertain fuzzy associations between classes. The proposed approach is illustrated using the problem domain of a meeting. Lee03 [12] proposes to model requirements specification in XML format in order to facilitate the modeling of imprecise requirements; Denger03 [13] said that the requirements in natural language have a major drawback, namely the inherent imprecision, ie, ambiguity, incompleteness and inaccuracy, of natural language. Since the requirements document forms the basis of the overall development process. It presents an approach to reduce the vagueness problem in natural language of requirements specifications for the use of natural language models, which allow the development of requirements sentences less ambiguous, more complete and more accurate.

B. Requirements analysis

We will first set the search domain of requirements analysis which is the requirements engineering (RE) The RE is a vast domain of research [6] interested mainly the approaches and techniques to improve the specification phase of a computer system.

• Requirements: According to Abran04 et al. [14], a requirement is a description of one or more properties of the system that must be satisfied. The requirements are expressed in terms of phenomena or objects shared by the system and its real world environment, with an accessible vocabulary for users.

• Often requirements can be classified into two categories: [6]

(i) Functional requirements: Specifying a function as the system or component of the system must be able to function. The authors of this definition emphasize that the system can be considered as a set of components.

(ii) Non-functional requirements defined by an attribute or system constraints such as flexibility, performance, security, etc., which does not affect the function of this system.

• The RE is responsible for defining the process of the corresponding requirements for the mission and goals of the organization. These expressed requirements by policy makers are identified, negotiated, validated and specified requirements listed in the specification documents. In the literature [6,15], four models groups are available to represent the requirements of a Decision Support System (DSS), namely: (i) The entity-relationship model; (ii) use case models (iii) requests model; (iv) goals model . Here is a presentation of some work for each model group:

TABLE I. MODELS GROUPS OF DECISIONAL REQUIREMENTS

Models Groups	**Works – authors & Publication Year**	**References**
E/R	(Cavero2003), (Winter2003), (Jensen2004), (Annoni2007), (Golfarelli2009), (Romero2009)	[14] [5] [12] [13] [16,5] [17]
Uses cases	(Lujan-Mora2006), (Soussi2005), (Soler2008), (Di Tria2012)	[4] [18][19][20]
Queries	(Phipps2002) (Ghozzi2005) (Ravat2007) (Romero2010)	[21][22][23][24]
Goals	(Prakash2005) (Giorgini2005) (El Golli2008) (Mazon2008), (Cravero2013)	[26][27][28][29][30]

In the decision-world, the requirements analysis process usually consists of the following steps: (a) Organization diagnostic, (b) Requirements gathering, (c) Requirements analysis, (d) Requirements specification, (e) Requirements validation, (f) Requirements modelling. In [11], three models are proposed to represent the requirement analysis: (i) the process-oriented analysis: analyzes the requirements by identifying the business processes of "organization. (ii) The goal-oriented analysis: analyzes the requirements by identifying the goals and objectives of the "organization. (iii) The user-oriented analysis focuses on identifying the target users and specify and analyze their individual requirements to integrate into a unified model requirements.

The "decision requirements" are the requirements of all actors' types. "User requirements" can be compared to the defined functional components (power supply, storage, restitution). "Equipment Requirements" are related to the technical aspect of DSS such as existing data sources and decision equipment already deployed [10]. The paradigm used by all the solutions for decisional requirements are: supply-driven; demand-driven; hybrid-driven. The works in this domain represent either approaches, either methodologies or tools. Each group of requirements analysis approach has advantages but also present disadvantages.

III. STATE OF ART

Recent work are interested to DW based ontology in analytical phase operational or [22], but few studies are interested in this type of structure in the design phase. Many studies on the design of the data warehouse based on the ontology [22, 56, 44] are goals-based, but there is less queries-bases studies or uses cases models [37.40].

TABLE II. COMPARATIVE OF DECISIONAL REQUIREMENTS ANALYSIS APPROACHES

Réf.	–	Auto	Types			Requirem.Expression Language	Input	Conceptual Design Level			Treatment
			Onto	TD	M			L	C	P	
07	Kimball1&al,2008			X					X		No
32	Phipps, 2002	Semi	No	X			-	X			No
14	Vrdoljak, 2003			X				X			No
34	Winter & 2003		No	X		Naturel	No	X			No
25	Trujillo et al., 2004			X		UML/XML		X			No
35	Prakash &al.,2005			X		Formal	SQL	X			No
39	Mazon & al.2008		No	X		Natural	No	X			No
26	Prat, 2006		No	X		Object/Constr		X			No
12	Romero & al 2009	Auto	Yes	X		SQL		X			No
37	Jovanovic &al.,	Semi		X		Requêtes		X			No
37	Cavero & al., 2003				X		-	X			No
38	Giorgini &al.2005	Semi	No		X	Natural	Goals	X			No
44	Ghozzi &al.,2005				X	Natural	Query	X			Yes Perspectiv
39	Soussi & al.,2005				X	Natural	Table	X			No
46	Anonni, 2007				X	Table	Decisional Diagram	X			No
41	El Golli, 2008				X	Natural	Goals	X			No
48	Soler & al. 2008				X	Formal		X			No
35	Zapeda2008 & al.	Auto			X	Formal		X			No
40	Romero &al. 2010				X	SQL	Query	X			No
32	Mazón &al,2009	Auto			X			X			No
48	Romero & al, 2011	Semi	Yes		X			X	X		No
36	S. Rizzi P. 2010				X			X			No
43	Atigui & al., 2012				X			X			No
50	Di Tria 2012	Semi	No		X	Formal	BMM Model	X			No
03	Khouri 2012	Semi	Yes		X	Formal	Domain Ontology	X			No
45	Carvero & al. 2013				X	Formal	-	X			No
18	Tenmozhi 2013	Auto	Yes		X	Natural		X			No
49	Di Tria 2014		No		X	Formal	-	X			No
28	Nebot 2009	Semi	Yes	X		Formal	-				Yes

IV. CHALLENGES AND OPEN PROBLEMS:

We note, after having presented this comparative study that the vagueness has been mentioned, but has not been studied carefully by Jensen04, Ghozzi07, Nebot09, Golfarelli09 and pardillo11:

(i) Jensen04 & al [23] have noted that the partial containment has introduced the imprecision in the aggregation. They propose a method to assess the vagueness of these paths. The paper also proposes the transformations in the dimension hierarchy with relations of partial containment to Simple hierarchies, for which the technical pre-existing calculations are applicable. As prospects for this work, it is appropriate to examine how the model can be effectively implemented in the use of specialized algorithms and data structures. It is also interesting to study if the structures in the lattice of the schema can be used directly in the user interface of a tool of the OLAP based on the model. (ii) As a perspective to the work of thesis of Ghozzi10 [58], the management of uncertain data in the model, in the query language and in the methodology which take into account the inaccuracies in the real world. Nebot09 [28] noted that the ontologies used resemble the patterns of database, but they are more flexible in the sense that they give the definitions of data generated incomplete, inaccurate and implicit. As perspectives, it envisages the specifying the schema of multidimensional analysis in terms of the axioms of the ontology. (iii) Golfarelli09 [17] had mentioned that the needs uncertain or little clear can be corrected by the use of ontologies for improving the design of DW. (v) Pardillo11 claims to begin such a discussion, thus posing a starting point for research in the field, by also consider other related aspects, such as the evolution of schema, the quality of the data, or the fusion of data.

V. CONCLUSION

The quality of the DW design is closely linked with the analysis requirements (functional and non-functional) and approaches to realize this design. The incorrect expression, ambiguity, missing or inadequate requirements are the cause of design flaws. This study highlights a very important point, which is the imprecision study and that researchers have not given the importance it deserves. We aim through this study to initiate a discussion on this focal point to improve the DW design and satisfy the various actors in the decision making process. To the best of our knowledge, there has been no rigorous study of vagueness in DW designing. So we have more interest in Nebot09's and Golfarelli09's solutions but we think the fuzzy ontology can evaluate the imprecision better than using a single ontology. Therefore we project propose a solution based on fuzzy ontology representing the requirements of a query-based model and taking into account the inaccuracy of the requirements analysis phase, fuzzy ontology will allow to assess the imprecision in order to improve the DW design.

REFERENCES

[1] Elhaj elamin & al., (2014) "Toward an ontology based approach for data warehousing, State of the art and proposal",

[2] Rizzi S. & al., (2006) "Research in DW modeling and design: deador alive?" Proceedings of the 9th ACM international workshop on Data warehousing and OLAP.

[3] Khouri, S. & al., (2012) "Ontology-based structured web DWs for sustainable interoperability: requirement modeling, design methodology and tool", Computers in industry; Elsevier.

[4] Lujan-Mora S. & al., "Physical modeling of DWs using UML component and deployment diagrams: design and implementation issues". Journal of Database Management (JDM) 17.2 (2006): 12-42.

[5] Golfarelli M. & al. ; (2009) "DW Design: Modern Principles and Methodologies". Book; USA

[6] Boukhari I., (2014), "Intégration et exploitation de besoins en entreprise étendue fondées sur la sémantique", thèse, ENSMA Poitiers, France

[7] R. Kimball, M. Ross & al. (2008) "The DW Lifecycle Toolkit"

[8] F. Abdelhedi & al. (2014) "Conception assistée d'entrepôts de données et de documents XML pour l'analyse OLAP", thèse.

[9] Paim R. F. S. & al. (2003) "DWARF: An Approach for Requirements Definition and Management of DW Systems". Proceedings of the 11th IEEE International Requirements Engineering Conference.

[10] Dubois D. & H. Prade, (1988) "Representation and combination of uncertainty with belief functions and possibility measures". Computational Intelligence, 4:244–264.

[11] J. Lee & al., (1999) "Modeling imprecise requirements with fuzzy objects", Information Sciences 118 101–119.

[12] J. Lee & al. (2003) "Modeling imprecise requirements with XML", Information and Software Technology 45 445–460

[13] Denger C. & al., (2003) "Higher Quality Requirements Specifications through Natural Language Patterns", Proceedings of the IEEE International Conference on Software—Science, Technology & Engineering.

[14] Abran A.&al., (2004) "Guide to the Software Engineering Body of Knowledge" http://www.swebok.org/.IEEE Computer Society.

[15] Sabri, A. ; Kjiri L. (2012) "Processus de traitement et de formalisation des besoins métier d'un SI Décisionnel"

[16] Romero O. & al.; (2009) "A Survey of Multidimensional Modeling Methodologies", International Journal of Data Warehousing & Mining, 5(2), 1-23

[17] Golfarelli M. & al., (2009), "A comprehensive approach to DW testing". Proceedings of the ACM twelfth international workshop on Data warehousing and OLAP. ACM.

[18] Pardillo J.& al., (2011) "Using ontologies for the Design of DWs", IJDMS, 2011

[19] Jindal R. & al., (2012) "Comparative study of DW Design approaches: a survey", IJDMS

[20] Tebourski W., (2013) "Semi-automatic DW Design methodologies: a survey"

[21] Abai N. & al.(2013) "User Requirement Analysis in DW Design: A Review", Procedia Technology

[22] Thenmozhi, M. & al. (2013) "A Tool for Data Warehouse Multidimensional Schema Design using Ontology". International Journal of Computer Science Issues (IJCSI) 10.2.

[23] Jensen C.S. (2004) "Multidimensional data modeling for location-based services". The VLDB Journal.

[24] Song,Il &al., (2007) "SAMSTAR: a semi-automated lexical method for generating star schemas from an entity-relationship diagram". Proceedings of the ACM 10° International workshop on DW and OLAP.

[25] Golfarelli, M. & al., (2002) "WAND: A CASE Tool for Workload-Based Design of a Data Mart.- SEBD", Citeseer

[26] Pinet F. & al., (2009) "A Unified Object Constraint Model for Designing and Implementing Multidimensional Systems", Journal on Data Semantics XIII, Volume 5530 CS pp 37-71,

[27] Carmè A. & al., (2010) "A Model-Driven Heuristic Approach for Detecting Multidimensional Facts in Relational Data Sources",DaWaK, Springer-Verlag Berlin Heidelberg

[28] Di Tria F. & al., (2013) Ontological Approach to Data Warehouse Source Integration. In Proceedings of the 28th International Symposium on Computer and Information Sciences, Springer, 251-259.

[29] Hachaichi Y. (2013) "An automatic method for the design of multidimensional schemas from object oriented databases", International Journal of Information Technology & Decision Making Vol. 12, No. 6 1223–1259

[30] Trujillo J.& al., (2004) "Applying UML and XML for designing and interchanging, information for DWs and OLAP applications", Journal of Database Management, Vol. 15, No.1.

[31] Prat, & al. (2006) "A UML-based DW design method". Decision Support Systems 42.3

[32] Phipps C. and Davis K.C. (2002) Automating DW Conceptual Schema Design and Evaluation. In Proceedings of the 4th International Workshop Design and Management of DWs, 23-32.

[33] Nebot V. & al., (2009) Efficient retrieval of ontology fragments using an interval labeling scheme, information Sciences,– Elsevier

[34] Winter R., (2003) A Method for Demand-driven Information Requirements Analysis in Data Warehousing Projects

[35] Prakash N.&al., (2005) Requirements Driven DW Development, CAiSE Short Paper Proceedings, India.

[36] Mazón J.N& al. (2007) A Model-Driven Goal-Oriented Requirement Engineering Approach for DWs. International Workshop on Ontologies and Information Systems for the Semantic Web.

[37] P. Jovanovic et al. (2014) A requirement-driven approach to the design and evolution of data warehouses. Information Systems 44 94–119

[38] Giorgini P. & al. (2005) “Goal-oriented requirement analysis for DW design”, DOLAP’05. Proceedings of the 8th ACM international workshop on DWsing and OLAP, pages 47–56.

[39] Mazon J.N. & al., (2008) Applying the i* framework to the development of DWs. Proceedings of the 3rd Intern.I* Workshop

[40] Romero, O.; Abelló, A.; (2010) A framework for multidimensional design of DWs from ontologies; Data & Knowledge Engineering;Volume 69, Issue 11, Pages 1138–1157– Elsevier.

[41] Zepeda L., (2008) A Mixed Approach for Data Warehouse Conceptual Design with MDA, ICCSA '08 Proceedings of the international conference on Computational Science and Its Applications

[42] Mazón, J.N. & al. (2009) "A hybrid model driven development framework for the multidimensional modeling of data warehouses". ACM SIGMOD Record 38.2: 12-17.

[43] Cavero J., (2003) A multidimensional DW development methodology. Managing Data Mining Technologies in Organizations

[44] Ghozzi F., (2005) Méthode de conception d’une base multidimensionnelle contrainte. In Revue des Nouvelles Technologies de l’Information - Entrepôts de Données et l’Analyse en ligne (EDA’05), volume RNTI B-1,

[45] Soussi A. & al. (2005) « Approche semi automatisée de conception de schémas multidimensionnels valides ». Revue des Nouvelles Technologies de l’Information. EDA’05

[46] Annoni E., (2007) Eléments méthodologiques pour le développement des systèmes décisionnels dans un contexte de réutilisation, Thèse, Computer Science. Université Paul Sabatier - Toulouse III. France.

[47] El Golli, I.G., (2008) «Ingénierie des exigences pour les Systèmes d’Information Décisionnels : Concepts, Modèles et Processus (la méthode CADWE) », Thèse de Doctorat, Université Paris-Panthéon-Sorbonne, France

[48] Soler E., (2008) Towards Comprehensive Requirement Analysis for DWs: Considering Security Requirements Third Intern. Conference on Availability, Reliability and Security, Spain.

[49] Atigui F., (2012) Modélisation conjointe des données et des processus pour l'implantation de schémas d'entrepôts. Dans : Journal of Decision Systems, Taylor & Francis Group, Vol. 21, N. 1, p. 27-49

[50] Di Tria F.; (2012) Academic DW Design Using a Hybrid Methodology, Computer Science and Information Systems;

[51] Cravero A.; (2014) Goal oriented requirements engineering in DWs: A comparative study; Ingeniería E Investigación vol. 34 N°2.

[52] Romero, O&al. (2007) "Automating multidimensional design from ontologies". Proceedings of the ACM 10° International workshop on Data warehousing and OLAP. ACM.

[53] Thenmozhi M& al., (2012) An Ontology based Hybrid Approach to Derive Multidimensional Schema for DW, International Journal of Computer Applications

[54] Romero, O. &al. (2011) "GEM: requirement-driven generation of ETL and multidimensional conceptual designs". Data Warehousing and Knowledge Discovery. Springer , 80-95.

[55] Di Tria & al., (2014) Academic DW Design Using a Hybrid Methodology, Computer Science and Information Systems

[56] Maté A. & al., (2012) Improving the Maintainability of DW Designs: Modeling Relationships between Sources and User Concepts, Intern. Workshop on DOLAP.

[57] Golfarelli M.&al., (2011) Modern Soft. Engineering Methodologies Meet DW Design: 4WD

[58] Ghozzi , (2008) Conception et manipulation de bases de Donnees dimensionnelles a contraintes, Thèse, université Toulouse III - Paul Sabatier

Giorgini P. & al., (2008) GRAnD: A goal-oriented approach to requirement analysis in data warehouses, Decision Support Systems 45 4 –21