# ICEIS 2012

## 14th International Conference on Enterprise Information Systems

Wrocław, Poland
28 June - 1 July, 2012

http://www.iceis.org

iceis.secretariat@insticc.org

### Table of Contents



Sponsored by: INSTICC

In Cooperation with: AAAI, EIC, SWIM, acm In-Cooperation, sigmis, SIGCHI

Co-organized by: Wroclaw University of Economics

Proceedings will be submitted for indexation by: THOMSON REUTERS CONFERENCE PROCEEDINGS CITATION INDEX, Inspec, dblp .uni-trier.de Computer Science Bibliography, Ei

# ENASE 2012

## 7th International Conference on Evaluation of Novel Approaches to Software Engineering

Wrocław, Poland 29 - 30 June, 2012

http://www.enase.org

enase.secretariat@insticc.org

### Table of Contents



Sponsored by: INSTICC

Co-organized by: Wroclaw University of Economics

Proceedings will be submitted for indexation by: THOMSON REUTERS CONFERENCE PROCEEDINGS CITATION INDEX, Inspec, dblp .uni-trier.de Computer Science Bibliography, Ei

# MSVVEIS & WEBI 2012

Costin Badica, George Eleftherakis,
Grzegorz J. Nalepa, Manuel Isidoro Capel-Tuñón,
Khaled Ghédira and Valérie Monfort (Eds.)

# Modelling, Simulation, Verification and Validation of Enterprise Information Systems & Web Intelligence

Proceedings of MSVVEIS / WEBI 2012
10th International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems
1st International Workshop on Web Intelligence
In conjunction with ICEIS 2012
June 28th - July 1st, 2012
Wrocław, Poland

Costin Badica
George Eleftherakis
Grzegorz J. Nalepa
Manuel Isidoro Capel-Tuñón
Khaled Ghédira and
Valérie Monfort (Eds.)

# Modelling, Simulation, Verification and Validation of Enterprise Information Systems & Web Intelligence

**Proceedings of the**
**10th International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems**
**MSVVEIS 2012**
**&**
**1st International Workshop on Web Intelligence**
**WEBI 2012**

In conjunction with ICEIS 2012
Wrocław, Poland, 28 June - 1 July, 2012

**SciTePress**
***Portugal***

Volume Editors

Costin Badica
University of Craiova
Romania

George Eleftherakis
CITY College, International Faculty of the University of Sheffield
Greece

Grzegorz J. Nalepa
AGH University of Science and Technology
Poland

Manuel Isidoro Capel-Tuñón
University of Granada
Spain

Khaled Ghédira
Soiie
Tunisia

and

Valérie Monfort
SOIE Tunis
Tunisia

10th International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems &
1st International Workshop on Web Intelligence
Wrocław, Poland, 28 June - 1 July, 2012



Printed in Portugal

ISBN: 978-989-8565-14-3
Depósito Legal: 344494/12

# Foreword

The joint workshop MSVVEIS/WEBI 2012 includes: i) the tenth edition of the International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems (MSVVEIS 2012), and ii) the first edition of International Workshop on Web Intelligence (WEBI 2012). The joint workshop was held at the 14th International Conference of Enterprise Information Systems (ICEIS-2012) in Wroclaw, Poland.

SCOPE OF MSVVEIS 2012

One of the most important recurrent problems in any enterprise is how to ensure the reliability and correctness of the core processes and systems the company relies on. This event focuses on the provision of methods and tools that can increase the level of confidence on Enterprise Information Systems (EIS).

Procedures to increase the quality of the outcome for an EIS can be exercised at different levels and this forum considers methodologies that can help, either at an organizational or at a software development level, to increase the level of confidence in the IS used and produced. The complexity of modern companies, which are usually geographically distributed and support online simultaneous operations from many customers around the world, is reflected in complex operational procedures as well as in the sophisticated software that is needed to realize that operational structure. Several methodologies have been developed to analyze and develop processes that whilst reflecting the complex operational contexts of modern companies are also reliable. Modelling, Simulation, Verification and Validation are particularly connected with the responsible production of systems and quality assurance testing. They can be connected to each other in order to explore the behaviour of a system under development and to evaluate how it relates to the intended implementation.

Continuous advances in the complexity of systems produced around the world relentlessly push ahead the boundaries uncovering new challenges as new application domains are considered and new technologies are combined or created. Some characteristic problems faced by software developers in the area of EIS are the use of distributed resources interacting via synchronous or asynchronous communication, consistency of data, security and performance issues, as well as integrating web services and agent technologies, to mention a few.

At higher levels of granularity in the various IS co-existing in a company the fundamental procedures used to operate a business can be also subject of rigorous analysis and refinement to increase the quality and reliability on the overall business process, for example through the analysis of workflows. In recent years an emerging trend claims that, to cope with this complexity, systems should focus on the customer essential needs, those needs that really deliver value to the customer. This trend is sometimes called agile or lean.

After several decades of sustained effort, many techniques and associated tools are now available to industry and business-related professionals to rigorously scrutinize the core processes and products of their operation. Still the problems are numerous as systems grow and new technologies are considered. EIS are a continuous source of interesting challenges and to contribute to the progress of this area our workshop is annually organized in order to stimulate the exchange of ideas/experiences of practitioners, researchers, and engineers interested in the elaboration of more reliable systems. MSVVEIS 2012 topics of interest included but were not limited to:

Formal Methods - Combination of verification systems - Deductive systems - Finite-state abstractions of infinite-state systems - Model checking - Petri nets - Process algebra - Reuse of specifications and proofs - Rule-based modelling - Semantics of modelling notations

Modelling notations - Business and software process modelling, simulation, analysis and design - Information systems modelling and design - Integration of modelling and specification - Modelling application integration (Web services, agents, a.o.) - Modelling business services - Modelling guidelines - Modelling software architecture - Modelling using objects, components and agents - Notation standards (BPMN, UML, ontologies, XML-based, etc.) - Organization modelling for EIS - Requirements specifications

Quality control and assurance - Modelling & Simulation to increase software reliability - Modelling & Simulation, Verification & Validation as part of the software lifecycle - Testing - Validation and certification - Workflow modelling, simulation and verification, and quality assessment

Applications and case studies - Applications of objects, components and agents - Business / IT alignment - Business and industry applications - Consistency checking and data integrity - Large scale component based development - Safety critical systems - Technical

frameworks and tool support - Use cases - Working product evaluation

SCOPE OF WEBI 2012

Web Intelligence (WI) has been recognized as a new direction for scientific research and development to explore the fundamental roles as well as practical impacts of Artificial Intelligence (AI) [E.g., knowledge representation, planning, knowledge discovery and data mining, intelligent agents, and social network intelligence) and advanced Information Technology (IT) (E.g., wireless networks, ubiquitous devices, social networks, semantic Web, wisdom Web, and data/knowledge grids) on the next generation of Web-empowered products, systems, services, and activities. It is one of the most important as well as promising IT research fields in the era of Web, agent intelligence and virtual communities.

WEBI 2012 topics of interest included but were not limited to: - Intelligent Human-Web Interaction and Virtual community - Web Information Filtering & Retrieval, linked data - Electronic Commerce - Conversational Systems - Browsing and Exploration - Adaptive Web - User Profiling/Clustering - Negotiation Systems - Security, Privacy and Trust - Web Mining and Farming - Social Networks Mining - Intelligent e-Technology - Semantic Web - Web Agents - Knowledge Grids & Grid Intelligence - Ubiquity, pervasive systems - Grid Computing, Cloud Computing, Web Services, Search in Networks - Industrial feedback and prototypes

PAPERS

Four MSVVEIS 2012 papers were selected for presentation and were included in the workshop proceedings.

The first paper, by Maha Naceur, Lotfi Majdoub, and Riadh Robbana, is in the area of test generation. In their paper, the authors introduce a method for test generation for duration systems using coverage criteria. The system is modelled as duration variables timed graphs with inputs-outputs. The system model is approximated with a digitization method. The authors present a coverage generating test tree algorithm and prove the soundness of the resulting test cases.

The second paper, by Manuel Isidoro Capel-Tuñón, is in the area of safety–critical systems. In his paper, the author employs the Formal Compositional Verification Approach to facilitate the sound design

of safety–critical systems. The method is exemplified on a real-life application for mobile phone communications.

The third paper, by Bogumila Hnatkowska and Mateusz Grzegorczyn, is in the areas of requirements engineering. In their paper, the authors present the results of a controlled experiment concerning the influence of the formal notation onto the comprehensibility of the software requirements specification. The authors consider three notations: natural language, use case diagrams, and activity diagrams. The results showed that activity diagrams received the highest score concerning the correctness of the interpretation. Nevertheless, they require longer interpretation time than natural language, and shorter than use case diagrams.

The fourth article, by Damián Adalid, Alberto Salmerón, María del Mar Gallardo, and Pedro Merino, is in the area of program testing. In their paper, the authors present an approach for testing reactive and concurrent Java programs which combines model checking and runtime monitoring. The method is implemented in the TJT tool using SPIN for model checking and Java Debug Interface for runtime monitoring.

Six WEBI 2012 papers were selected for presentation and were included in the workshop proceedings.

The first paper, by Fatma Zohra Lebib, Hakima Mellah, and Youssef Amghar is in the area of multimedia Web services. The authors introduce the Multimedia as a Service – MaaS model through which multimedia content providers expose their content. Validation experiments were carried out in the sport domain.

The second paper, by Amel Benna, Hakima Mellah, Islam Choui, and Ali Oualid, is in the area of social tagging. The authors propose the use of k-means clustering for the hierarchical classification of user-defined annotations of resources. The evaluation results obtained on the social bookmarking system del.icio.us improved the social search process.

The third paper, by Olatz Arbelaitz, Ibai Gurrutxaga, Aizea Lojo, Javier Muguerza, Jesus M. Pérez, and Inigo Perona is in the area of Web usage mining. The authors propose the use of machine learning and collaborative filtering to generate interesting navigation links based on users' preferences.

The fourth paper, by César Guerra-García1, Ismael Caballero,

Rodrigo Testillano, Rafael Llamas, and Mario Piattini is in the area of requirements engineering. The authors propose a method that takes into account data quality during the process of Web portals' requirements elicitation.

The fifth paper, by Alexander Gromoff, Julia Stavenko, Kristina Evina, and Nikolay Kazantsev is in the area of knowledge management. The authors propose the method of expertise search during the process of business process management.

The sixth paper, by Syrine Ben Meskina, Valérie Monfort, and Achraf Ben Miled is in the areas of social networks. The authors propose a method for enhancing social networks with ontologies and data mining to promote women of emergent low income countries.

Concluding, we would like to thank the authors who submitted papers to the joint MSVVEIS/WEBI 2012 workshop, as well as the reviewers for reviewing the submissions. We would like to thank the organizers of the ICEIS 2012 conference for accepting and hosting the joint MSVVEIS/WEBI 2012 workshop.

June 2012,

**Costin Badica**
University of Craiova, Romania

**George Eleftherakis**
CITY College, International Faculty of the University of Sheffield, Greece

**Grzegorz J. Nalepa**
AGH University of Science and Technology, Poland

**Manuel Isidoro Capel-Tuñón**
University of Granada, Spain

**Khaled Ghédira**
Soiie, Tunisia

**Valérie Monfort**
SOIE Tunis, Tunisia

## MSVVEIS Workshop Co-chairs

Costin Badica
University of Craiova
Romania

George Eleftherakis
CITY College, International Faculty of the University of Sheffield
Greece

Grzegorz J. Nalepa
AGH University of Science and Technology
Poland

and

Manuel Isidoro Capel-Tuñón
University of Granada
Spain

## MSVVEIS Program Committee

Costin Badica, University of Craiova, Romania
G. Botterweck , Lero - The Irish Software Engineering Research Centre, Ireland
David Bustard, University of Ulster, United Kingdom
Manuel Isidoro Capel-Tuñón, University of Granada, Spain
Oana Captarencu, Äl. I. CuzaÜniversity, Romania
Vincent Chapurtlat, EMA - Ecole Des Mines D'alès, France
Jorge C. A. de Figueiredo, Universidade Federal de Campina Grande, Brazil
Laura Dillon, Michigan State University, United States
George Eleftherakis, CITY College, International Faculty of the University of Sheffield, Greece
Anthony Finkelstein, University College London, United Kingdom
Juan Garbajosa, Technical University of Madrid, Spain
Klaus Havelund, Nasa/Jet Propulsion Laboratory, United States
Bogumila Hnatkowska, Wroclaw University of Technology, Poland
Michael Köhler-BuSSmeier, University of Hamburg, Germany
Adam Meissner, Poznan University of Technology, Poland

Daniel Méndez Fernández, Technische Universität München, Germany
Daniel Moldt, University of Hamburg, Germany
Grzegorz J. Nalepa, AGH University of Science and Technology, Poland
David Sanan, Singapore University of Technology and Design, Singapore
Alexei Sharpanskykh, Vrije Universiteit Amsterdam, Netherlands
Gheorghe Stefanescu, University of Bucharest, Romania
Lone Leth Thomsen, Aalborg University, Denmark
John Van Der Poll, University of South Africa, South Africa
Bozena Wozna-Szczesniak, Jan Dlugosz University, Poland
W. Lok Yeung, Lingnan University, China

## WEBI Workshop Co-chairs

Khaled Ghédira
Soiie
Tunisia

and

Valérie Monfort
SOIE Tunis
Tunisia

## WEBI Program Committee

Lamjed Bensaid, SOIE, Tunisia
Juri Luca De Coi, Université Jean Monnet, France
Faiez Gargouri, Miracl Laboratory, Tunisia
Khaled Ghédira, SOIE, Tunisia
Wasif Gilani, SAP Research, United Kingdom
Jean Pierre Lorre, PetalsLink, France
Pierre Maret, Université de Saint Etienne, France
Valérie Monfort, SOIE Tunis, Tunisia
Pravin Amrut Pawar, University of Twente , Netherlands
Christoph Stahl, DFKI, Germany
Jan-Willem van 't Klooster, University of Twente, Netherlands
Laurent Vercouter, LITIS, France

# Elicitation of Specific Requirements of Data Quality during the Web Portal Development

César Guerra-García[1,2], Ismael Caballero[2], Rodrigo Testillano[2], Rafael Llamas[1] and Mario Piattini[2]

[1] Department of Information Technologies and Telemathic,
Polytechnic University of San Luis Potosí, Urbano Villalón 500, San Luis Potosí, México
{Cesar.Guerra, Rafael.Llamas}@upslp.edu.mx
[2] Alarcos Research Group, Institute of Technologies and Information Systems,
University of Castilla-La Mancha, Paseo de la Universidad 4, Ciudad Real, Spain
{Ismael.Caballero, Rodrigo.testillano, Mario.Piattini}@uclm.es

**Abstract.** Data is one of the most important assets for making decisions and concretizing business in organizations. So, providing data with adequate levels of quality, especially for Internet applications is a very important issue. However, most of developers of these applications do not take in account the incorporation of artifacts to the necessary management of data quality (DQ) from the early stage of development. Due to that, we have elaborated a strategy with two approaches: methodologic and technologic. The first one is aimed to identification the requirements corresponding to the Web portal functionalities for the different kind of users and their specific DQ software requirements. For the technologic approach, an UML profile to model the DQ software requirements is shown; it embraces aspects considered basics to integrate in the specification and modeling of these kinds of requirements. The final objective is developing applications that satisfy the different DQ software requirements specified by each user, at the moment to perform a function with the system.

## 1 Introduction

During the last years, the number of organizations and enterprises which have developed Web portals have increased considerably [1]. This applications enable to users the access to large amounts of data and information on line [2], through different data resources [3]. Web portals have provided users with a more intuitive and simple work environment, allowing users to find the data they need to perform their tasks in a better way. However, the apparition of problems due to inadequate levels of data quality (DQ) has been proven to negatively affect the tasks performed by people, and consequently the performance of organizations. These problems can cause a negative impact with a substantial cost, and not only in economic terms but also social in the organizations [4]. The concept of DQ should not be longer understood only as "zero defects" in the data, but "*fitness for use*" of data for a task for a specific user, that is, the ability of a data set to satisfy user´s requirements [5]. In this sense, a DQ software requirement supplied by a user can be specified, in which indicates the characteristics of DQ required or needed to some data when they are

used in certain specific task. The focus of this research is centred on how to elicit and introduce in the stage of requirements analysis, the corresponding software requirements for the management of DQ as a new kind of requirements. With this in mind, in this paper we describe the work made. Firstly, it was carried out a systematic review of the literature related with the purpose of getting a list of works done [6], which could be presumably related both methodologic focus as technologic into the area of DQ requirements specification. We have only found few proposals, as shown by [7], [8] and [9] for the relational model, or these related to semantic technology showed in [10] and [11], but none specifically related to deal with DQ requirements management. Due to this reason, as part of the methodologic focus of the research, it was designed a strategy of work in which the first step was relate the DQ problems (*potholes*) identified in Information Systems (IS) described by Strong et al. in [12] to the specific Web portal functionalities defined by Collins in [13]. In order to identify the DQ characteristics that could be critiques at the moment to implement each one of Web functionalities. Within this strategy, we got a list with a generic set of DQ software requirements that any development team would like to include into a System Requirements Specification (SRS) document. These requirements should guide to the analyst in the identification of software requirements related to DQ from the viewpoint of each role performing a task. Due to the above, the first objective of this work is *DAQUA-VORD*, a methodology aimed to identify and elicit both kinds of software requirements for Web portal development: those focused to functionalities, and those ones oriented to DQ management. The second objective is included as part of the technological approach, in which a UML profile is proposed in order to model in a clear way, all DQ requirements related to each one of Web functionalities.

The remainder of the paper is structured as follows: Section 2 reviews the three pillars in which our proposal is grounded: data quality measurement, Web portal functionalities and the requirements elicitation method. Section 3 presents our proposal: firstly the list of DQ software requirements, immediately the methodology DAQUA-VORD, and finally the UML profile. In Section 4 an example of application is shown. Finally in section 5 we introduce some conclusions.

## 2 Revision of Related Areas

### 2.1 Data Quality

In order to reduce the negative impact of problems (technical, organizational or legal) due to inadequate levels of DQ [14], it is paramount that companies can have a quantitative perception of their actual importance. So, they must assess how good their organizational data resources are for the tasks at hand. Organizations have to deal to the DQ, both in subjective perceptions by individuals that use the data, as objective measures based on a set of data. An assessment of DQ in a subjective way can reflect the needs and experiences of users with a set of data [8]. If the users assess the quality of data as poor, their tasks could be influenced by this assessment [15]. As mentioned, the most accepted definition for the concept "Data Quality" is "*fitness for use*" [16]. This means that a user typically evaluates the quality of a set of data for a particular task, which it is done in a specific context according to a set of criteria or

dimensions of DQ. An user performing a role within a IS can specify for a piece of data different DQ software requirements as be necessary, specifying the DQ dimensions that better represent this kind of requirements for a determined task. So, *the perception about the DQ level of a set of data could be different for diverse tasks, even for the same user performing different roles*. For measuring the level of DQ of a piece of data, it is necessary to identify several DQ dimensions (known the set as "*DQ model*") which can characterize the DQ requirements in a better way. Although there exists many DQ models most of them are quite domain dependant, which diminishes their applicability. In order to get a broader perspective as possible, we chose for our research the generic DQ model proposed in the standard ISO/IEC 25012 [17]. This international standard brings together fifteen DQ dimensions from two points of view: *Inherent* and *System dependent*.

### 2.2 Web Portal Functionalities

In stated in the introduction, our first step is to associate the relationships between Web portal functionalities and those DQ dimensions, which would best represent the various roles' DQ software requirements. So, we must first enumerate and review these functionalities as described by Collins [13]. We have reordered them in base to our experience and knowledge in both data quality and web development areas: *Content Management, Process and actions, Search capabilities, Administration, Security, Data points and integrations, Communication and collaboration, Presentation, Taxonomy, Personalization* and *Help features*. This reordering was taking as criterion the following: a greater probability of using of a Web functionality, a greater probability of being susceptible of finding inadequate levels of DQ.

### 2.3 Requirements Elicitation Method

The Requirements Elicitation is perhaps the activity most often regarded as the first step in the Requirements Engineering process, this activity is responsible to identify the stakeholders of the system and discover the requirements from them [18]. The viewpoint-oriented approach takes into consideration the different viewpoints of the different roles to structure and organize the requirement elicitation process [19]. The key point of the viewpoint-oriented analysis takes into account the existence of several perspectives and provides a framework to discover conflicts between the requirements proposed by different viewpoints. The viewpoint can be used as a form to classify the stakeholders. The VORD (*Viewpoints-Oriented Requirements Definition*) method proposed in [19] was designed to guide the process of elicitation and analysis of requirements having into account the different point of views of a system. The steps of this method are: *VI-1.Viewpoints Identification, VS-2.Viewpoints Structuring, VD-3.Viewpoints Documentation* and *VL-4.Viewpoints Layout*.

# 3 A Methodology for the Elicitation of DQ Software Requirements

## 3.1 Relation between DQ Dimensions and Web Portal Functionalities

Once presented in section 2.1 the DQ dimensions, and listed the Web functionalities in section 2.2, we performed an analysis in both areas, it getting a matrix of relationship between the DQ dimensions and Web portal functionalities. Considering these relations at the moment to develop every one of Web functionality, it would be possible to ensure that the data that will be stored and manipulated by the functionalities have an acceptable level of DQ. Therefore, it is necessary to describe the DQ requirements that can be drawn to avoid or minimize the effect of the common source of problems, as those described by Strong et al. in [12]. So the main challenge is not only to specify the relations itself, but also to express them through of the specification of DQ software requirements. Once defined these kinds of requirements the analyst will be able to specify the DQ dimensions that should be observed and implemented for each one of the Web functionalities. In this sense, we made an analysis about what kind of problems (defined by Strong et al. in [12]) could be related to each one of the Web portal functionalities, it getting as result the next matrix (see **Table 1**). Once completed the matrix, it was performed an analysis and comparison of each one of DQ dimensions described both in the model proposed by Wang and Strong [20] as in the standard ISO/IEC 25012. The aim of this comparison was to resolve possible conflicts in the description of the different DQ dimensions, either the existence of dimensions with the same name and different meaning, or dimensions with different name but the same meaning.

**Table 1.** Matrix of relationship between web functionalities and problems identified by [12].

| Problems (*potholes*) / *Functionalities* | Multiple sources | Subjective production | Production errors | Too Much information | Distributed systems | Nonnumeric information | Advanced analisys requirements | Changing task needs | Security requirements | Lack of resources |
|---|---|---|---|---|---|---|---|---|---|---|
| Content Management | √ | √ | √ | √ | | √ | √ | √ | √ | √ |
| Process and Action | | | | √ | | | | √ | | |
| Search capabilities | | | | √ | √ | √ | | | √ | √ |
| Administration | | | | | | | | √ | | |
| Security | | | | | | | | | √ | |
| Data points and integration | | | | √ | √ | | | | √ | |
| Collaboration and Communication | | | | √ | √ | | √ | | | √ |
| Presentation | | | | | | | | √ | | |
| Taxonomy | | | | | √ | | | | | |
| Personalization | | | | | | | | √ | √ | |
| Help features | | | | | | | | | | |

Finally, taking as reference the research published by Strong et al. in [12], where the DQ dimensions that affect each one of the problems ("*potholes*") were classified based on their model [20], it was obtained the next matrix of relation (see Table 2). In this matrix, the DQ dimensions established in the Wang´s model were changed by their similar described in the standard ISO/IEC 25012.

**Table 2.** Matrix of relationship of web functionalities and DQ dimensions.

| *DQ Dimensions (ISO 25012)* / *Web Portal functionalities* | Accuracy | Completeness | Consistency | Credibility | Currentness | Accesibility | Compliance | Confidentiality | Efficiency | Precision | Traceability | Understandability | Availability | Portability | Recoverability |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Content Management | | √ | √ | √ | √ | √ | √ | √ | | | | | | | |
| Process and Action | | √ | | | √ | √ | √ | | | | | | | | |
| Search capabilities | | √ | √ | | √ | √ | √ | √ | | | | | | | |
| Administration | | √ | | | | | √ | | | | | | | | |
| Security | | | | | | √ | √ | √ | | | | | | | |
| Data points and integration | | √ | √ | | √ | √ | √ | √ | | | | | | | |
| Collaboration and Communication | | √ | √ | | √ | √ | √ | | | | | | | | |
| Presentation | | √ | | | | | √ | | | | | | | | |
| Taxonomy | | | √ | | √ | | √ | | | | | | | | |
| Personalization | | √ | | | | √ | √ | √ | | | | | | | |
| Help features | | | | | | | | | | | | | | | |

### 3.2 DAQUA-VORD Methodology

As one result of this research, the DAQUA-VORD methodology is proposed, it can guide developers in the specification of DQ requirements, it identifying for each one of the functionalities selected, those DQ dimensions that have to be considered (and implemented) according to previous matrix (Table 2). The specification of these DQ dimensions from different perspectives (*viewpoints*) of the users performing a specific task should be introduced as new software requirements. The reason why we decided to use the "*VORD*" method as reference is that it allows the incorporation of DQ management aspects during the requirement elicitation process. In this way, DQ software requirements can be introduced as normal ones, but always taking into account the diverse viewpoints of the different kind of users performing a task. It is important assuring that techniques to be used can adequately capture and organize all kind of requirements (e.g. functional requirements together with specific DQ requirements). Descriptions of the stages of *DAQUA-VORD* methodology, mapped from those one from *VORD*, as well as its subactivities, its input and output products, and techniques/tools related will be next shown.

1. *IWPV. Identification of the Web Portal Viewpoints*. This stage is analogous to the step VI-1 of VORD method. It implies to discover the different viewpoints that will receive the functionalities of the Web Portal, besides the identification of the Web Portal functionalities together with the DQ dimensions associated (see Table 3).

2. *VS. Viewpoints Structuring*. It is aimed at grouping the viewpoints related in a suitable hierarchy. The main functionalities are located at the top levels of the hierarchy, once done that, these functionalities are inherited to the viewpoints of low level, besides the DQ dimensions are hierarchized in the same context (see Table 4).

**Table 3.** Artefacts and subactivities for the IWPV.

| *IWPV.1. Identification of the Web Portal Functionalities (IWPF) to be implemented,* it implies to identify the specific functionalities that are provided to each viewpoint. | |
|---|---|
| Input Product | - List of identified viewpoints being able to propose software requirements for the system.<br>- List of all Web Portal functionalities [13]. |
| Output Product | - List of chosen functionalities for satisfying requirements of each viewpoint. |
| Tools and techniques | - Interviews - Study of documentation - Questionnaire - Brainstorming |
| *IWPV.2. Identification of the Data Quality Dimensions* (*IDQD*), it implies to identify the different DQ dimensions related to each one of the functionalities described for each viewpoint, taking as base the matrix of Table 2. | |
| Input Product | - List of viewpoints identified being able to propose DQ requirement for the system.<br>- List of chosen functionalities for satisfying requirements of each viewpoint.<br>- List of DQ dimensions (see Table 2) for each functionality. |
| Output Product | - List of DQ dimensions associated to the different functionalities.<br>- Document of System Requirements Specification. |
| Tools and techniques | - Interviews - Work sessions - Brainstorming |

**Table 4.** Artefacts and subactivity for the VS.

| *VS.1. Choose a DQ Model* (*CDQM*), it consists of classifying the DQ dimensions according to the hierarchy, in base at the priority level that the Web Portal functionalities have (listed in section 2.2). | |
|---|---|
| Input Product | - List of viewpoints identified in the system.<br>- List of DQ dimensions associated to the different functionalities. |
| Output Product | - List of classification of DQ dimensions (*DQ Model*). |
| Tools and techniques | - Work sessions - Judgment of experts |

3. *DV. Documentation of the Viewpoints*. It encompasses the refinement of the description of the viewpoints and the functionalities identified, adding the DQ dimensions (Table 5).

**Table 5.** Artefacts and subactivity for the DV.

| *DV.1. Documentation of the Data Quality Dimensions* (*DDQD*), it consists of documenting or modeling if possible, the DQ dimensions identified (e.g. through use cases diagram). | |
|---|---|
| Input Product | - List of classification of data quality dimensions.<br>- Document of System Requirements Specification. |
| Output Product | - Document of System Requirements Specification (SRS) augmented with DQ Requirements Specification. |
| Tools and techniques | - Work sessions - Judgment of experts - Tools like Word processors - Modeling tools for UML |

4. *LVS. Layout of the Viewpoints of the System.* It encompasses identifying the main objects in an object-oriented design using the information of the functionality encapsulated in the viewpoints (see Table 6).

### 3.3 UML Profile to Management of DQ Software Requirements

In this section we show the proposal for modeling of DQ software requirements, by using a UML profile. Unlike of the proposals founded in the systematic review [21], this profile is focused in modelling DQ software requirements from the perspective of each user (*viewpoint*) at the moment to perform a specific task. The motivation of this

**Table 6.** Artefacts and subactivities for the LVS.

| | |
|---|---|
| *LVS.1. Modeling of Data Quality Requirements* (*MDQR*), it consists of modeling the different DQ requirements (DQ dimensions) in a data model and later on, in a process model. | |
| Input Product | - Document of SRS augmented with DQ Requirements Specification. |
| Output Product | - Document of high level design with awareness of data quality (data model and process). |
| Tools and techniques | - Object oriented modeling tools (Rational Rose, Visual Paradigm, Poseidon, ArgoUML, etc.). |
| *LVS.2. Validation of Model* (*VM*), it consists of validating the complete model with the stakeholders. | |
| Input Product | - Document of System Requirements Specification augmented with DQ Requirements Specification.<br>- Document of high level design with awareness of data quality. |
| Output Product | - Final Document approved of "System Requirements Specification augmented with DQ Requirements Specif".<br>- Final Document approved of "High level design with awareness of data quality". |
| Tools and techniques | - Work sessions - Interpersonal negotiation techniques |

proposal appears from the necessity of allowing analysts and designers specifying in a more clear way, which DQ dimensions (related each Web functionality) should be implemented from the specification of user requirements. For this reason, both functional requirements (*information requirements*) as DQ software requirements should be considered from the earliest stages of development, because it will allow to designer to model all the requirements through the convenient extensions (e.g. use case diagrams). This UML profile specifies how the concepts of the Web functionalities and the DQ dimensions are related and represented, through stereotypes of UML language. The package that contains the stereotypes defined into the profile (see Fig. 1) is represented with a extended class diagram of UML2 [22].

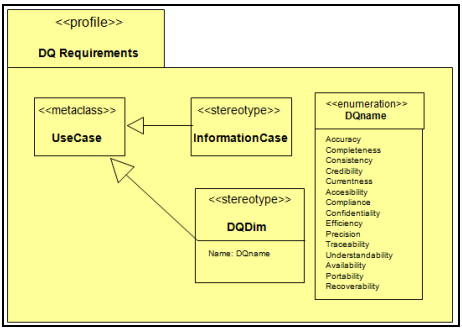


**Fig. 1.** Profile to specification of DQ requirements.

## 4 Example of Application

In Table 7, a typical problem statement for developing a Web portal is showed.

**Table 7.** Document of problem statement.

| |
|---|
| The ACME Realtors company would like to create an e-development solution that will replace the home listing catalogs that are printed on a monthly basis. The new system will allow to any *user* doing search in the property´s database for current listings or find a Realtor, but only users registered (prospective *buyers*) will be able to initiate the loan process. *Realtors* will be able to list their properties on the ACME Realtor system and update the pictures of every property. A prospective *buyer* will be able to log on to the system and set up a personal profile. This profile will allow the *buyer* to enter a set of personal preferences and search requirements. *Buyers* will also be able to bookmark properties to the personal planner for easy reference the next time they log on. After a *buyer* has logged on to the system they may choose to search for a home, find a Realtor, or apply for a mortgage loan. The *buyer* and *Realtor* should be able to search for a home in a geographic area by city, zip code, or the Multiple Listing Service (MLS) number. The *buyer* should be able to further narrow their search through a series of filter criteria until they find a number of homes they are interested in. Any *user* and *buyer* should be able to view a picture of the home and see a full text description on all the amenities and features that the home has to offer. Finally, if the buyer is interested in receiving more information on the home, the *buyer* will be able to send an e-mail to the listing broker. The prospective *buyer* has the option to apply for a mortgage loan using the ACME Realty System. ACME Realtors has an existing Loan System that communicates with a number of partner lenders to gain loan pre-qualification approvals. This system should continue to be used for sending loan requests to potential lenders. The Realty System will ask the prospective buyer a series of questions about their current financial standing. After the prospective *buyer* has answered all questions, the system will send the data to the Loan System and receive a list of possible offers for a loan. If the *buyer* chooses to select one of the pre-qualification offers, the system will inform the customer that a credit report must be generated. The *Administrator* will be responsible to generate the Credit Reporting. *Realtors* subscribe to a Credit Reporting service, and the existing interface to this system should be used to provide this service. The *buyer* should be allowed to view the broker's personal profile that may contain any type of information that the broker enters and also a summary of all the properties that the broker currently has listed. *Realtors* must be able to access the on-line system to modify their personal profiles that are displayed to buyers. The *Administrator* will be responsible to create a new listing of properties and assign them to every Realtor. Besides, the *Administrator* will assign the nominal fee of each property and he will be able to update some pictures. |

Once shown the problem statement, we can begin with the application of the methodology as follows:

1. *IWPV. Identification of the Web Portal Viewpoints*. One of main output product of this stage is identifying the viewpoints, which are: (a) *Buyer*, (b) *Realtor*, (c) User, and (d) *Administrator*.

*IWPV.1. Identification of the Web Portal Functionalities (IWPF) to be implemented*. The output product consists of a list of requirements and functionalities identified (see Table 8).

*IWPV.2. Identification of the Data Quality Dimensions (IDQD)*. List of DQdimensions identified for each one of the web functionalities (see Table 9). This listwill be useful to analyst, since they will be able to select whatever of them.

*2. VS. Viewpoints Structuring.* The level of importance of each proposed requirement, taking into account in this example the number of times that every requirement is related to each viewpoint, it is as follow: *(1) Login to the system, (2) Search of properties, (3) Send an email, (4) Update pictures of properties, (5) Subscribe to Credit Reporting Service, (6) See full description of properties, (7) Find a realtor, (8) View a picture, (9) Initiate a loan process, (10) Setup a personal profile, (11) Permit to mark properties to the personal planner, (12) Respond questions about financial standing, (13) Choose a pre-qualification offer, (14) View the broker´s personal profile, (15) View summary of properties assigned to realtors, (16) List their properties assigned, (17) Modify personal profile, (18) Generate a credit report, (19)*

*Create a new list of properties, (20) Assign the nominal fee of each property.* Taking as basis the importance level of each requirement, we can hierarchize the viewpoints in the next order: *1. Realtor, 2. Buyer, 3. Administrator, 4. User.*

*VS.1. Choose a DQ Model (CDQM).* The output product is a hierarchized list of DQ dimensions identified (taking as base Table 4): *1. Accessibility, 2. Compliance, 3. Confidentiality, 4.Completeness, 5.Consistency, 6.Currentness, 7.Credibility.*

**Table 8.** Identification of the Web Portal Functionalities (IWPF).

| *Viewpoint* | *Functional Requirement* | *Web functionality described by [13]* |
|---|---|---|
| Buyer | FR1. Search of properties. | Search capabilities |
| | FR2. Initiate a loan process. | Process and actions |
| | FR3.Login to the system. | Security |
| | FR4. Setup a personal profile. | Content Management |
| | FR5. Permit to mark properties to the personal planner. | Personalization |
| | FR6. Find a realtor. | Search capabilities |
| | FR7. View a property´s picture. | Search capabilities |
| | FR8. See full description of properties. | Presentation |
| | FR9. Send an email. | Collaboration & Communication |
| | FR10. Respond questions about financial standing. | Process and actions |
| | FR11. Choose a pre-qualification offer. | Process and actions |
| | FR12. View the broker´s personal profile. | Collaboration & Communication |
| | FR13. View summary of properties assigned to realtors. | Search capabilities |
| Realtor | FR14. List their properties assigned. | Search capabilities |
| | FR15. Update pictures of properties. | Content Management |
| | FR16. Subscribe to Credit Reporting Service. | Process and actions |
| | FR17. Modify personal profile. | Content Management |
| | FR9. Send an email. | Collaboration & Communication |
| | FR3. Login to the system. | Security |
| | FR1. Search of properties. | Search capabilities |
| Administrator | FR18. Generate a credit report. | Administration |
| | FR19. Create a new list of properties. | Content Management |
| | FR20. Assign the nominal fee of each property. | Administration |
| | FR3. Login to the system. | Security |
| | FR9. Send an email. | Collaboration & Communication |
| | FR16. Subscribe to Credit Reporting Service. | Process and actions |
| | FR15. Update pictures of properties. | Content Management |
| User | FR1. Search of properties. | Search capabilities |
| | FR7. View a property´s picture. | Search capabilities |
| | FR8. See full description of properties. | Presentation |
| | FR6. Find a realtor. | Search capabilities |

3. *DV. Documentation of the Viewpoints*. We use the following templates to conveniently document the different viewpoints and requirements. The results are gathered in Tables 10 and 11 (due to pages restriction of the paper, we only describe some of them). This documentation is a key part of a *System Requirement Specification document augmented with DQ Requirements Specification.*

**Table 9.** Identification of DQ dimensions.

| *Web functionality* | *DQ dimensions related* |
|---|---|
| Administration | Completeness, Compliance. |
| Security | Accessibility, Compliance, Confidentiality. |
| Process and actions | Completeness, Currentness, Accessibility, Compliance. |
| Search capabilities | Completeness, Consistency, Credibility, Currentness, Efficiency, Traceability, Understandability, Availability. |
| Personalization | Completeness, Accessibility, Compliance, Confidentiality. |
| Collaboration and Communication | Completeness, Consistency, Currentness, Accessibility, Compliance. |
| Presentation | Completeness, Compliance. |
| Content Management | Completeness, Consistency, Credibility, Currentness, Accessibility, Compliance, Confidentiality. |

**Table 10.** Specification of viewpoint "Buyer".

| | |
|---|---|
| Reference | Buyer. |
| Focus | Viewpoint of the Buyer, he performs the main business functionalities of the application. |
| Attributes | Name, address, telephone, email, salary. |
| Requirements | Search of properties, Initiate a loan process, Login to the system, Setup a personal profile, Permit to mark properties to the personal planner, Find a realtor, View a property´s picture, See full description of properties, Send an email, Respond questions about financial standing, Choose a pre-qualification offer, View the broker´s personal profile, View summary of properties assigned to realtors. |
| Web functionalities | Search capabilities, Process and actions, Security, Content Management, Personalization, Presentation, Collaboration and Communication. |
| Exceptions | None. |
| History | No alterations. |

**Table 11.** Requirement “Setup a personal profile”.

| | |
|---|---|
| Reference | Setup a personal profile. |
| Description | This requirement is related to manage and update of all the buyer personal information. |
| Data | Name, address, email, salary. |
| Viewpoints | Buyer. |
| Non-functional requirements | None. |
| DQ requirements | Completeness, Consistency, Credibility, Currentness, Accessibility, Compliance, Confidentiality. |

*DV.1. Documentation of the Data Quality Dimensions* (*DDQD*). As part of the output product of this stage based on the Table 5, and with the goal of documenting and modeling the DQ dimensions, we apply in this point the UML profile proposed previously, this profile will permit us modeling DQ requirements (*DQ dimensions*) associated to the different functionalities that the system will provide, taking as basis this profile we can model an “*Information case diagram*”, which is much more explicit that a common use case diagram. In this “*Information case diagram*” (see Fig. 2) we can see the requirements previously referred: “*FR4. Set up a personal profile”* and *“FR20. Assign the nominal fee of each property”*, which can be modelled like “*Information cases” (IC)*, they maintain a relation of type “*include*” with the use cases stereotyped like “*DQDim*”, it means that data managed for each one of *Information cases* should satisfy the DQ dimensions specified. Thus, the developer

will have to consider the DQ dimensions at the moment of implementing the different functionalities of the application. In this diagram, the Information Case "*Set up a personal profile*" (associated with the Web functionality "*Content Management*") manages mainly the following pieces of data: *name*, *address, email* and *salary*. It means that these data should be compliant with the DQ dimensions of *Completeness, Consistency, Credibility, Currentness, Accessibility, Compliance* and *Confidentiality*. In this specific case, the analyst has chosen modeling only three of them. Similarly, the Information case "*Assign the nominal fee of each property*" (associated with the Web functionality "*Administration*") will manage the following pieces of data: *ID property, nominal fee, Realtor in charge* and *address*. So, these data should be compliant with the DQ dimensions of *Completeness* and *Compliance*.

4. *LVS.1. Modeling of Data Quality Requirements (MDQR).* The output product of this stage based on Table 6, it consists mainly in getting an object-oriented design, it should contains the main classes responsible for providing the functionalities of the Web portal, as well as the classes responsible for implementing the DQ dimensions. These diagrams are part of a *document of high-level design with awareness of DQ*.

*LVS.2. Validation of Model* (*VM*). Finally, the main documents obtained once applied the methodology ("*Document of System Requirement Specification with DQ Requirements Specification"* and "*Document of high-level design with awareness of DQ*") should be validated with the client.

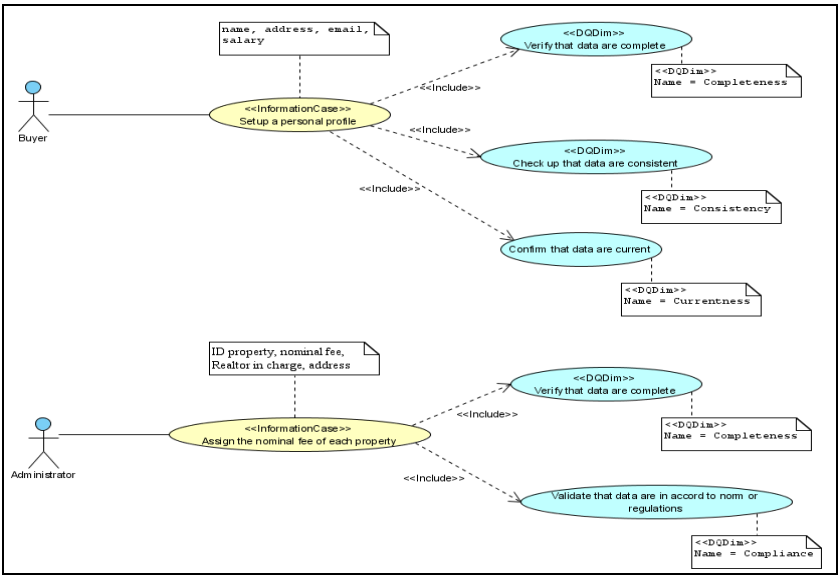


**Fig. 2.** Information case diagram.

## 5 Conclusions

At present, data and information are fundamental assets of any organization. In the last years the Web portals has established as one of the main information sources in

Internet, and as means for allowing the access to information for all people. Nevertheless, the great majority of users who seek information needs to be sure that it has the adequate DQ level for the use that they require. A first solution to this problem is showed in this paper, where it is described which DQ dimensions are presumably related with the different Web functionalities. Immediately, we show a Methodology to elicit and define DQ requirements (*DAQUA-VORD*), besides a UML profile in order to modeling these kinds of requirements. Thus, we are able to encompass both approaches: methodological and technologic. These approaches could facilitate to the analyst and developers getting awareness about the DQ level that need to be implemented for each one of the functionalities during all Web development process.

## Acknowledgements

This research is part of the PEGASO-MAGO (TIN2009-13718-C02-01), and IQMNet (TIN2010-09809-E) projects, supported by the Spanish Ministerio de Educación y Ciencia. ALTAMIRA project (PII2I09-0106-2463) supported by JCCM Consejería de Educación y Ciencia.

## References

1. Calero, C., A. Caro, and M. Piattini, An Applicable Data Quality Model for Web Portal Data Consumers http://dx.doi.org/10.1007/s11280-008-0048-y World Wide Web 2008. 11 (4 ): p. 465-484
2. Yang, Z., et al., Development and validation of an instrument to measure user perceived service quality of information presenting Web portals. Information and Management, 2004. 42(4): p. 575-589.
3. Mahdavi, M., J. Shepherd, and B. Benatallah. A Collaborative Approach for Caching Dynamic Data in Portal Applications. in Proceedings of the fifteenth conference on Australian database. 2004.
4. Eppler, M. and M. Helfert. A Classification and Analysis of Data Quality Costs. in International Conference on Information Quality. 2004. MIT, Cambridge, MA, USA.
5. Cappiello, C. and M. Comuzzi. Efficient Allocation of Quality Improvement Efforts to Support the Definition of Data Service Offerings. in 12th International Conference on Information Quality. 2007. Cambridge, MA.
6. Guerra-García, C., I. Caballero, and M. Piattini, A Survey on How to Manage Specific Data Quality Requirements during Information System Development. Lecture Notes in Computer Science, 2011(Evaluation of Novel Approaches to Software Engineering).
7. Wang, R. Y. and S. Madnick. Data Quality Requirements: Analysis and Modelling. in Ninth International Conference on Data Engineering (ICDE'93). 1993. Vienna, Austria: IEEE Computer Society.
8. Wang, R. Y., A Product Perspective on Total Data Quality Management. Communications of the ACM, 1998. 41(2): p. 58-65.
9. Becker, D., W. McMullen, and K. Hetherington-Young. A Flexible and Generic Data Quality Metamodel. in International Conference on Information Quality. 2007.
10. Caballero, I., et al. DQRDFS:Towards a Semantic Web Enhanced with Data Quality. in Web Information Systems and Technologies. 2008. Funchal, Madeira, Portugal.

11. Missier, P., et al., Quality views: capturing and exploiting the user perspective on data quality. Proceedings of the 32nd international conference on Very large data bases-Volume 32, 2006.
12. Strong, D., Y. Lee, and R. Wang, Ten Potholes in the Road to Information Quality. IEEE Computer, 1997: p. 38-46.
13. Collins, H., Corporate Portal Definitions and Features.2001, New York, NY, USA: Amacom Books.
14. Caballero, I., et al. MMPRO: A Methodology Based on ISO/IEC 15939 to Draw Up Data Quality Measurement Processes. in ICIQ. 2008.
15. Pipino, L., Y. Lee, and R. Wang, Data Quality Assessment. Communications of the ACM, 2002. 45(4): p. 211-218.
16. Ge, M. and M. Helfert. A Review of Information Quality Research. in International Conference on Information Quality. 2007. MIT, Cambridge, MA, USA.
17. ISO-25012, ISO/IEC 25012: Software Engineering-Software product Quality Requirements and Evaluation (SQuaRE)-Data Quality Model. 2008.
18. Sommerville, I., Integrated Requirements Engineering: A Tutorial. IEEE Softw., 2005. 22(1): p. 16-23.
19. Kotonya, G. and I. Sommerville, Requirements engineering with viewpoints. Software Engineering Journal, 1996.
20. Wang, R. and D. Strong, Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems; Armonk; Spring 1996, 1996. 12(4): p. 5-33.
21. Guerra-García, C., I. Caballero, and M. Piattini. A Systematic Literature Review of How to Introduce Data Quality Requirements into a Software Product Development. in 5th. International Conference on Evaluation of Novel Approaches to Software Engineering, ENASE. 2010. Athens, Greece.
22. OMG. Unified Modeling Language: Superstructure. Versión 2.0. 2005; Available from: <http://www.omg.org/docs/formal/05-07-04.pdf%3E.