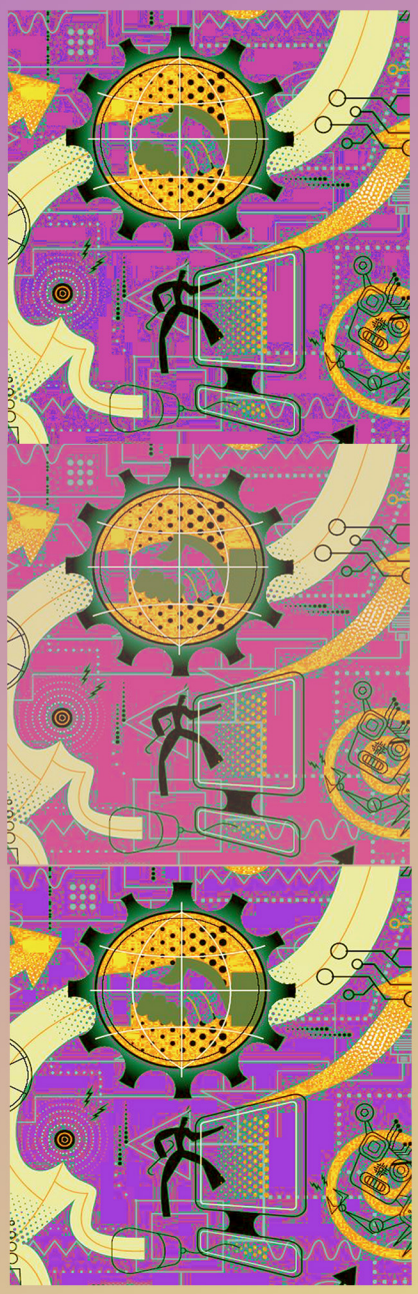
Proceedings of the IADIS International Conference

WWW/INTERNET

Rome, Italy

19 - 22 November

2009



Volume I

Edited by
Pedro Isaías
Bebo White
Miguel Baptista Nunes

iadis
international association for development of the information society

IADIS INTERNATIONAL CONFERENCE

WWW/INTERNET 2009

PROCEEDINGS OF THE
IADIS INTERNATIONAL CONFERENCE
WWW/INTERNET 2009

Volume I

ROME, ITALY

NOVEMBER 19-22, 2009

Organised by
IADIS
International Association for Development of the Information Society

Edited by Pedro Isaías, Bebo White and Miguel Baptista Nunes

Associate Editors: Luís Rodrigues and Patrícia Barbosa

ISBN: 978-972-8924-93-5

TABLE OF CONTENTS

FULL PAPERS

FOREWORD

These proceedings contain the papers and posters of the IADIS International Conference WWW/Internet 2009, which was organised by the International Association for Development of the Information Society, in Rome, Italy, 19-22 November 2009.

The IADIS WWW/Internet 2009 Conference aims to address the main issues of concern within WWW/Internet. WWW and Internet had a huge development in recent years. Aspects of concern are no longer just technical anymore but other aspects have aroused. This conference aims to cover both technological as well as non-technological issues related to these developments

The purpose of this conference is to serve as a forum to gather researchers, practitioners, students and anyone who works or studies the field of WWW/Internet.

Submissions were accepted under the following main tracks and topics:

- Web 2.0
  - Collaborative Systems
  - Social Networks
  - • Folksonomies
  - • Enterprise Wikis and Blogging
  - • Mashups and Web Programming
  - • Tagging and User Rating Systems
  - • Citizen Journalism

- Semantic Web and XML
  - Semantics Web Architectures
  - Semantic Web Middleware
  - Semantic Web Services
  - Semantic Web Agents
  - Ontologies
  - Applications of Semantic Web
  - Semantic Web Data Management
  - Information Retrieval in Semantic Web

- Applications and Uses
  - e-Learning
  - e-Commerce / e-Business
  - e-Government
  - e-Health
  - e-Procurement
  - e-Society
  - Digital Libraries
  - Web Services/SaaS
  - Application Interoperability
  - Web-based multimedia technologies

- Services, Architectures and Web Development
  - Wireless Web
  - Mobile Web
  - Cloud/Grid Computing
  - Web Metrics
  - Web Standards
  - Internet Architectures
  - Network Algorithms
  - Network Architectures
  - Network Computing
  - Network Management
  - Network Performance
  - Content Delivery Technologies
  - Protocols and Standards
  - Traffic Models
- Research Issues
  - Web Science
  - Digital Rights Management
  - Bioinformatics
  - Human Computer Interaction and Usability
  - Web Security and Privacy
  - Online Trust and Reputation Systems
  - • Data Mining
  - • Information Retrieval
  - Search Engine Optimization

The IADIS WWW/Internet 2009 Conference had 347 submissions from more than 44 countries. Each submission has been anonymously reviewed by an average of four independent reviewers, to ensure the final high standard of the accepted submissions. The final result was the approval of 75 full papers, which means that the acceptance rate was below 22%. A few more papers have been accepted as short papers, reflection papers and posters. Best papers will be selected for publishing as extended versions in the *IADIS International Journal on WWW/Internet* (IJWI*)* and in other selected journals.

The conference, besides the presentation of full papers, short papers, reflection papers, posters and doctoral consortium presentations also included a keynote presentation talk from an internationally distinguished researcher. We would like also to express our gratitude to Professor Daniel Schwabe, Department of Informatics, Catholic University in Rio de Janeiro (PUC-Rio), Brazil, for being our keynote speaker.

As we all know, organising a conference requires the effort of many individuals. We would like to thank all members of the Program Committee for their hard work in reviewing and selecting the papers that appear in the proceedings.

This volume has taken shape as a result of the contributions from a number of individuals. We are grateful to all authors who have submitted their papers to enrich the conference proceedings.

Each of the Proceeding books contains a rich experience of the academic & research institutions and the industry on diverse themes related to the Internet and Web. We do hope that researchers, knowledge workers and innovators both in academia and the industry will find it a valuable reference material.

Last but not the least, we hope that everybody will have a good time in Rome and we invite all participants for the next year edition of the *IADIS International Conference WWW/Internet 2010*.

Pedro Isaías, Universidade Aberta (Portuguese Open University), Portugal
Conference Chair

Bebo White, Stanford University, USA
Miguel Baptista Nunes, University of Sheffield, United Kingdom
Program Co-Chairs

Rome, Italy
November 2009

ISBN: 978-972-8924-93-5 © 2009 IADIS

DQ-VORD: A METHODOLOGY FOR MANAGING AND INTEGRATING DATA QUALITY REQUIREMENTS INTO SOFTWARE REQUIREMENT SPECIFICATION

César Guerra-García, Ismael Caballero, and Mario Piattini
Departamento de Tecnologías y Sistemas de Información – Universidad de Castilla-La Mancha
Paseo de la Universidad, 4 – 13071 Ciudad Real, España

ABSTRACT

Providing data with an adequate level of quality, especially for Internet applications (e.g. Web Portals), has been increasingly becoming a very important issue for organizations, because data is a strategic asset for making decisions, or concretizing business. We posit that the raising and assurance of adequate levels of quality for data must be taken into account from the earliest phases of software development. To achieve this goal, we propose to introduce specific data quality requirements as new software requirements, more specific for Web Portals ones. For this goal, we have depicted a strategy consisting on identifying and integrating requirements corresponding to functionalities of Web Portals for the different kind of users (viewpoints), and their specific Data Quality requirements, for developing software enable to manage data with the adequate level of quality. To simplify the underlying complexity, in this paper we present a methodology based on viewpoints, DQ-VORD, which can be used to guide developers to build Data Quality aware - Web Portals.

KEYWORDS

Data Quality, Requirements Elicitation, Portal Web.

1. INTRODUCTION

Over the last years, the number of organizations which have developed Web Portals has grown considerably. The Internet allows easily collecting and sharing data between users (B2C), or between organizations (B2B). In addition, Web Portals provide users with access to different data sources (Mahdavi et al., 2004), as well as to other on-line information and information-related services (Yang et al., 2004). Web Portals also create a suitable working environment for users easily to navigate to find data they actually need to perform their operational or strategic functions (Collins, 2001).

However, problems due to inadequate levels of data quality (DQ) can make ineffective such kind of applications. These problems have caused a negative impact that has an important cost (and not only in economic terms), but also social for the organizations (Eppler and Helfert, 2004) (Laudon, 1986). See for instance, those scenarios identified by Levis in (Levis et al., 2007): ill-defined process, untrained information producer as well as defective data design, redundant databases and defective application design.

Problems due to inadequate levels of DQ can be addressed according to their nature (Caballero et al., 2008a) in the following groups: (1) *Technical* (such as those relating to the implementation of data warehouses (Oliveira et al., 2005)). (2) *Organizational* (such as loss of customers (Redman, 2000), large financial losses (Cai and Shankaranarayanan, 2007) (Eppler and Helfert, 2004) or even dissatisfaction of workers (English, 1999) (Strong et al., 1997a)). (3) *Legal*, because of privacy and/or national legislation (like the Spanish Organic Law for Data Protection from 1999).

As the common trend to think about DQ as errors in data stored in databases fails to resolve complex organizational problems (Strong et al., 1997b), the concept of data quality should not be longer understood only as "zero defects" in the data, but "*fitness for use*" of data. Due to this, it should be kept in mind the importance of users performing a role, because they are the ones who lastly determine the degree of fitness for use of the data they are using: they have the perception of the actual level of quality of the data for each task that they are carrying out. It is a matter of trust on data: if data has enough quality, users would better

trust on it to carry out their tasks. For such tasks, users utilize certain tools software, as web applications or Portal Webs (Caballero et al., 2008a). We posit that, given that applications in general, and Portal Webs in particular, are built to satisfy certain business requirements, if they would be designed to also satisfy DQ requirements since earlier phases of their developing, then, it would be possible to assure to organizations and users that problems previously related could be avoided.

The challenge of research question is how to identify and introduce jointly in the Portal Web development the corresponding software requirements for data quality as a new kind of business ones. In addition, as demonstrated by (Wang and Madnick, 1993), these kinds of requirements are transversal to typical functional requirements, what makes even more complex the related DQ management requirement within a software development project. In order to minimize this complexity when dealing DQ requirements, we thought that some researching in order to develop artifacts which make easier the work is required.

After delving into the related bibliography looking for approaches to our researching question, we have only found few proposals, like the one in (Wang and Madnick, 1993) for the relational model, or that one related to semantics technologies in (Caballero et al., 2008b) or (Missier et al., 2006), but not ones related to deal with DQ management requirements in software development. Due to this reason, we designed a two-step researching strategy aimed at discovering how to specify DQ requirements that could avoid the identified DQ problems for Information Systems as described by (Strong et al., 1997a)) to the specific functionalities of a Web Portal, as described by (Collins, 2001).

To achieve this goal, our first step is to list a generic set of DQ requirements coming trying to fulfill those DQ issues. Bearing in mind the idea of "*fitness for use*", one can suspects that these requirements strongly depend on both the nature of the domain, and the user performing actions with Web Portals functionalities. These suspicions have made that our researching efforts are to take into account the various point of views of each kind of user performing an action might have. So, and this is the second step and the main contribution of this paper, is to build a methodology, that we have named as *DQ-VORD*, and which it is based on VORD (Viewpoints Oriented Requirements Definition (Kotonya and Sommerville, 1996)) aimed to identify, elicit and integrate both kind of requirements for Web portal development: those addressing functionalities (typical software requirements), and those ones oriented to DQ management.

The remainder of the paper is structured as follows: Section 2 reviews the three pillars in which our proposal is based: the one regarding to data and information quality measurement (as an important part of DQ management), the one regarding to functionalities of a Portal Web, and the VORD method. In order to manage a standard knowledge, for the description of the DQ foundations, we will align our explanations to the just-released (ISO-25012, 2008), the international standard for data quality. Section 3 presents our proposal: firstly the list of DQ requirements and secondly, the methodology DQ-VORD. Finally in Section 4, we review some conclusions and future work.

2. RELATED LITERATURE

2.1 Data Quality Measurement

In order to reduce the negative impact of problems (technical, organizational or legal) of inadequate levels of DQ, it is essential 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. This firstly involves the definition of representative DQ measures and corresponding measurement procedures on the data, and secondly the establishment of some valid acceptance ranges for these measurements (Caballero et al., 2008a). With this assessment, users could discriminate those pieces of data which are not good enough for their own use. The DQ is recognized to be a multidimensional concept: 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") (Lee et al., 2006) that characterize DQ requirements. 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 have chosen for our research the generic DQ model proposed in the recently released DQ standard ISO/IEC 25012 (ISO-25012, 2008). This standard brings together fifteen DQ dimensions from two points of view:

- *Inherent*: refers to the extent to which quality characteristics of data have the intrinsic potential to satisfy stated and implied needs when data is used under specified conditions.
- *System dependent*: refers to the extent to which data quality is reached and preserved within a computer system when data is used under specified conditions.

The DQ dimensions contained in the standard ISO/IEC 25012 (ISO-25012, 2008) are: *Accuracy, Completeness, Consistency, Credibility, Currentness, Accessibility, Compliance, Confidentiality, Efficiency, Precision, Traceability, Understandability, Availability, Portability and Recoverability.* Ensuring that the data managed by the functionalities fulfills the DQ dimensions selected of the model, is a way to enunciate a DQ Requirement for Web Portal developers. In section 3.1, we will discuss about which DQ dimensions can be associated to each one of the Web Portal functionalities.

2.2 Web Portal Functionalities

As previously mentioned, our first step is associate the relationships between Web portal functionalities and those DQ dimensions which would best represent the various roles' DQ requirements. So, we must first enumerate and review these functionalities as described by (Collins, 2001). We have reordered them taking as criterion the following: "*the greater probability of utilization of a functionality of a Portal Web, the greater probability of being susceptible of finding inadequate levels of DQ*". It is necessary to highlight that the proposed order is determined in base to our experience and knowledge in both data quality and web development area. Table 1 shows the ordered list of functionalities with a small comment from the developer point of view.

Table 1. Web portal functionalities (Collins, 2001)

Functionalities	Description
Content Management	This function supports content creation, authorization, and inclusion in (or exclusion from) Web Portal collections. The organizations have information and reference material that is available from the Web Portal.
Process and actions	This function enables Web Portal users to initiate and participate in an organizational business process.
Search capabilities	These provide Web Portal users with several services for supporting searches throughout the company, the World Wide Web, and/or in search engine catalogs and indexes.
Administration	This functionality provides two kinds of services: the first one corresponding to maintenance activities and using task associated with the portal web system. The second one, a service that can be configured only for the administrator with the Web Portal system, as well as every user through of the personalization.
Security	This provides a description for the levels of access that each user or groups of users are allowed for each portal application and software function included in the Web Portal.
Data points and integration	These provides the ability to access information from a wide range of internal and external information sources and display the resulting information on the single point-of-access desktop.
Communication and collaboration	This functionality facilitates discussion, the location of innovative ideas, and the recognition of enterprising solutions. These elements allow the employees to work together in a more productive way by means of the creation of a shared virtual environment (contribution), bearing electronic messaging (communication) and adding characteristics of communication and contribution to the process of business (coordination).
Presentation	This provides Web Portal users with the visual experience that encapsulates all the functionalities of the portal.
Taxonomy	This provides the context of information, including the organization-specific categories that reflect and support the organizational business. It provides common terminology used in the organization of fast recognition and improvement the semantic for the users of the portal.
Personalization	This is a critical component when creating a working environment that is organized and configured specifically for each user in the company. The key is achieves a balance between the information and the attributes needed to be consistent and constant in the personalization of the portal, and the unique characteristics required for each employee.
Help features	These provide with help when using the Web Portal. The aid should be referred to the characteristics of the portal and to specific characteristics of the organization.

2.3 Requirements Definition

Requirement Engineering is a systematic process of developing requirements through a joint and iterative process of analysis of the problem, documenting the resultant observations in a representation formats variety, and verifying the precision of the earned understanding (Loucopoulos and Karakostas, 1995). The viewpoint-oriented approach for the Requirements Engineering takes into consideration the different viewpoints of the different roles to structure and organize the requirement elicitation process. 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.

There are three kinds of generic viewpoints (Sommerville, 2000): (1) *Direct viewpoints*: they represent people and other systems that interact with the system directly. This kind of viewpoints can give detailed requirements of the system that cover the characteristics, functionalities and interfaces of the system. (2) *Indirect viewpoints*: they represent those stakeholders having interest in some or all of the services which are delivered by the system, but do not interact directly with it. Is more probable they are able to provide organizational requirements and restrictions of high level. (3) *Domain viewpoints*: they represent the characteristics and restrictions of the domain that have influence in the system requirements.

Sometimes, the initial identification of the relevant viewpoints of a system may be difficult. Analyst should try to identify types of viewpoints in a more specific way (Sommerville, 2000). For any not trivial system, an enormous amount of possible viewpoints can exist, and it could be practically impossible to obtain a full and complete list of requirements for all of them. Therefore, it is important to organize and to structure the viewpoints hierarchically according to their criticalness/priority degree, because likely, different viewpoints in the same branch can share common requirements.

The VORD (*Viewpoints-Oriented Requirements Definition*) method (Kotonya and Sommerville, 1996) has been designed to guide the process of elicitation and analysis of requirements. The main steps of this method are described in Table 2.

Table 2. Main steps of VORD method

Steps	Description
1. VI-1. Viewpoints Identification	That implies to discover the ones that receive the functionalities or services of the system, and identify the specific functionalities that supply to each viewpoint.
2. VS-2. Viewpoints Structuring	That encompasses to classify the viewpoints previously obtained into a hierarchy. The common functionalities are located in the high levels of the hierarchy and the viewpoints of low level are inherited.
3. VD-3. Viewpoints Documentation	Which consist of documenting the viewpoint name, requirements, constraints on its requirement and its source. Viewpoint requirements include a set of required services, control requirements and set of non-functional requirements.
4. VL-4. Viewpoints Layout	That encompasses to identify the objects in an object oriented design utilizing the information of the service encapsulated in the viewpoint.

3. DQ-VORD: A METHODOLOGY FOR THE ELICITATION OF DQ REQUIREMENTS FOR WEB PORTALS DEVELOPMENT

In this section we show the current state of our work. Firstly, we show the results of the analysis carried out after mapping the DQ dimensions to the different Web Portal functionalities. After this, the DQ-VORD methodology for the definition of DQ requirements is explained.

3.1 Mapping DQ Dimensions to Web Portal Functionalities

Once described in section 2.1 the meaning of the most important DQ dimension and listed the functionalities for a Web Portal in section 2.2, we are going to carry out a mapping between the DQ dimensions and the functionalities. Considering this mapping in the moment to development each functionality, it would be possible ensuring that the data that will be stored and manipulated by the functionalities have an acceptable quality level. Therefore, it is necessary to describe the DQ Requirements that can be drawn to avoid or at

least minimize the effect of the problems described by (Strong et al., 1997a). In this sense, the main challenge consists of not only making the mapping, but also discovering how these DQ dimensions can be observed and implemented for each one of the functionalities since the point of view of each role. We have extended the first mapping that Caro et al. performed in (Caro et al., 2008), where a classification of Web Portal Functionalities vs. data quality attributes was carried out according to the data quality model defined by (Strong et al., 1997b). For future developers' convenience, we have rather preferred to align them in our research to the standard ISO/IEC (ISO-25012, 2008). For doing so, we first analyzed each one of the DQ dimensions defined by (Strong et al., 1997b), and we have compared their meaning to those provided by the ISO standard. In this way, we have adapted the "DQ Attributes" axis of the table provided by (Caro et al., 2008) to the new DQ dimensions mentioned in the standard ISO (the mark "X" in Table 3 shows that relationships).

In addition, to the obtained results from our analysis, we have included the remainder of relationships between DQ dimensions provided by the standard ISO and the Web Portal functionalities. As result, the sign "O" represents our new findings. With this mapping, what we intend is to determine a roadmap that analysts can follow to write a Requirement Specification Document, always keeping in mind the DQ dimensions that should be implemented for each functionality. Table 3 shows the whole results of this mapping.

After obtaining an initial set of DQ dimensions that are potentially applicable to the evaluation of DQ in the context of a Web Portal, and in order to gain clarity, below, we describe every one of these new relations signed with the mark "O" (Web Portal Functionalities, DQ dimensions) in Table 4.

Table 3. Relationships between DQ dimensions and Web portal functionalities

DQ dimensions / **Web Portal functionalities**	Accuracy	Completeness	Consistency	Credibility	Currentness	Accessibility	Compliance	Confidentiality	Efficiency	Precision	Traceability	Understan-dability	Availability	Portability	Recoverability
Data points and Integration	X	X	X	X	X				X			X	O		
Taxonomy	O		X	X	X				X			X			
Search capabilities	X	X	X	X	X				X		O	X	O		
Help features			X				X					X			
Content Management	X	X	X	X	X	X	X	O	X		O	X		O	
Process and Action	X	X	X	X	X	X	X		X			X			
Collaboration and Communication	O		X	O	O	X						X	O		
Personalization	X	X			X	X							O		
Presentation	X	X	X	X	X				X			X			
Administration			X			X	O		O	O		X		O	O
Security	X	X	X		X	X		O			O	X			

3.2 DQ-VORD Methodology

The second step is to propose a methodology that can guide developers to identify for each one of selected functionalities, those DQ dimensions that must be assured according to our previous analysis. These DQ dimensions are going to be introduced as new software requirements. In this sense, the VORD method could be taken as a reference since it is one of the main methods for requirements definition. The reason for which we have decided to base our work on this "viewpoints oriented" method is that it allows incorporating the aspects of DQ management during the elicitation process. In this way, DQ requirements can be introduced as normal ones, but taking into account the diverse viewpoints of the different kind of users/stakeholders.

The use of viewpoints as a means of organizing and structuring the requirements engineering activity has been demonstrated to be useful in some scenarios (Nuseibeh and Easterbrook, 2000). Viewpoints are based on the entities whose requirements are responsible for, or may constraint, the development of the intended

system. Different kinds of stakeholders can provide their specific requirements from their own point of view. In this sense, each of the stakeholder proposing requirements is known as "requirements sources". Each requirements source (*viewpoint*) has a relationship with the proposed system based upon its needs and interactions with the system to be developed. Therefore, it is important to assure that techniques to be used can adequately capture and organize, not only global but also the specific requirements for the different viewpoints into a cohesive knowledge structure that must be both complete and visible (e.g. domain requirements together with specific DQ requirements).

Table 4. Description of relations between DQ dimensions and web portal functionalities

Web Portal functionalities	DQ dimensions	Description
Data points and Integration	Availability	The data should be always within reach of the users at the moment that are required.
Taxonomy	Accuracy	The data intrinsically should be syntactic and semantically correct, improving the common terminology utilized in the organizations.
Search capabilities	Availability	The data must be recovered at the moment to execute some type of specific search. It implies that the data is accessible, that the access roads are available and the user has the adequate permissions for its reading.
	Traceability	The system should provide information about the quantity of times that a data has been accessed or modified.
	Confidentiality	The data just should be accessible and interpretable by authorized users.
Content Management	Traceability	The data should provide information about when and who published or distributed them and who will be able to access them.
	Portability	The data will be able to be installed or moved inside any system in the organization.
	Accuracy	The data should be used and understood in the same context for all users (employees), to improve the communication and performance of them.
Collaboration and Communication	Credibility	The data must be considered as truthful by all users, to improve the collaboration among each one.
	Currentness	The data must be updated, helping to carry out creative solutions and innovated in the business process.
	Availability	The data should be available to the authorized users, to help them to discuss and resolve problems, as well as make decisions.
Personalization	Availability	The data should be available to the authorized users, creating a work environment configured especially for each employee.
	Compliance	The data should be adhered to norms or conventions defined by the organization.
	Efficiency	The data can be accessed with an adequate level of operation.
	Precision	The data should be defined accord to the required precision.
Administration	Portability	The data will be able to be installed or substituted by the Administrator of the portal during activities of maintenance.
	Recoverability	The data will be able to be maintained and recovered by the administrator of the portal during activities of maintenance, or in case of some type of failure of the system.
Security	Confidentiality	The data will be just accessed by authorized users.
	Traceability	The data will give information about when and who used them.

Descriptions of the stages of *DQ-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 the VORD method (see Table 2). 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 5).

2. *VS. Viewpoints Structuring*. It encompasses to group the viewpoints related in a hierarchy. The main functionalities are located in the high levels of the hierarchy and are inherited to the viewpoints of low level, besides the DQ dimensions are hierarchize in the same context (see Table 6).

3. *DV .Documentation of the Viewpoints*. It encompasses to refine the description of the viewpoints and the functionalities identified, adding the DQ dimensions (see Table 7).

Table 5. 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. - List of all Web Portal functionalities ((Collins, 2001)).
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 the Table 3.	
Input Product	- List of viewpoints identified being able to propose DQ requirement for the system. - List of chosen functionalities for satisfying requirements of each viewpoint. - List of DQ dimensions (mapped in section 3.1) for each functionality.
Output Product	- List of DQ dimensions associated to the different functionalities. - Document of System Requirements Specification.
Tools and techniques	- Interviews - Work sessions - Brainstorming

Table 6. Artefacts and subactivity for the VS

VS.1. Classification of Data Quality Dimensions (*CDQD*), it consists of classifying the DQ dimensions according to the hierarchy, in base at the priority level that the Web Portal functionalities have.	
Input Product	- List of viewpoints identified in the system. - List of DQ dimensions associated to the different functionalities.
Output Product	- List of classification of DQ dimensions.
Tools and techniques	- Work sessions - Judgment of experts

Table 7. 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. - Document of System Requirements Specification.
Output Product	- Document of System Requirements Specification augmented with DQ Software 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 8).

Table 8. Artefacts and subactivities for the LVS

LVS.1. Modeling of the data quality dimensions (*MDQD*), it consist of modeling the different DQ dimensions in a data model and later on, in a process model.	
Input Product	- Document of System Requirements Specification augmented with DQ Software 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)
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 Software Requirements Specification. - Document of high level design with awareness of data quality.
Output Product	- Final Document approved of "System Requirements Specification augmented with DQ Software Requirements Specification". - Final Document approved of "High level design with awareness of data quality".
Tools and techniques	- Work sessions - Interpersonal negotiation techniques

4. CONCLUSIONS AND FUTURE WORK

In the last decade, Web portals have established as one of the principal information sources in the Internet, and as means for allowing the access to information for all people. Nevertheless, the great majority of users who seek information through portals, needs to be sure that information has the adequate DQ level for the

required use of data. Ideally we wish to imagine an analyst or designer that, using a single software development tool, can be able to development a complete Web Portal, managing it from the early stages of the development all the requirements (functional, non functional, and DQ related), design, testing, until the automatic generation of code. A first approach to this solution, specifically in the stage of requirements specification is shown in this work, in which it is described which DQ dimensions would be presumably implied directly with the different functionalities of a Web Portal. The paper also shows a proposal of a methodology for the elicitation and management of DQ requirements. In a given moment, both *features* would facilitate to the developers to have awareness of the level of data quality that needs to be implemented for each one of the functionalities during all the Web portal development process.

As part of our future work, we pretend to align our progresses to MDA technologies (*Model Driven Architecture*), since due to its abstraction capabilities and characteristics of modeling, we could assure an easier integration of our findings in DQ aware-Web Portals Development with other tools and methodologies related to Software Development.

REFERENCES

Caballero, I., Calero, C., Piattini, M. & Verbo, E. M. (2008a) MMPRO: A Methodology Based on ISO/IEC 15939 to Draw Up Data Quality Measurement Processes. *ICIQ.*

Caballero, I., Verbo, E. M., Calero, C. & Piattini, M. (2008b) DQRDFS:Towards a Semantic Web Enhanced with Data Quality. *Web Information Systems and Technologies.* Funchal, Madeira, Portugal.

Cai, Y. & Shankaranarayanan, G. (2007) Managing data quality in inter-organisational data networks. *International Journal of Information Quality,* 1**,** 254-271.

Caro, A., Calero, C., Caballero, I. & Piattini, M. (2008) A proposal for a set of attributes relevant for Web Portal Data Quality. *Software Quality Journal*.

Collins, H. (2001) *Corporate Portal Definitions and Features.,* New York, NY, USA, Amacom Books.

English, L. (1999) *Improving Data Warehouse and Business Information Quality: Methods for reducing costs and increasing Profits,* New York, NY, USA, Willey & Sons.

Eppler, M. & Helfert, M. (2004) A Classification and Analysis of Data Quality Costs. *International Conference on Information Quality.* MIT, Cambridge, MA, USA.

Iso-25012 (2008) ISO/IEC 25012: Software Engineering-Software product Quality Requirements and Evaluation (SQuaRE)-Data Quality Model.

Kotonya, G. & Sommerville, I. (1996) Requirements engineering with viewpoints. *Software Engineering Journal,* 11.

Laudon, K. C. (1986) Data Quality and Due Process in Large Interorganizational Record System. *Communications of the ACM,* 29**,** 4-11.

Lee, Y. W., Pipino, L. L., Funk, J. D. & Wang, R. Y. (2006) *Journey to Data Quality*, Cambridge, MA, USA, Massachussets Institute of Technology.

Levis, M., Helfert, M. & Brady, M. (2007) Information Quality Management: Review of an Evolving Research Area. *ICIQ'07.* MIT, Cambridge, MA, USA.

Loucopoulos, P. & Karakostas, V. (1995) *System Requirements Engineering,* New York, NY. USA., McGraw-Hill.

Mahdavi, M., Shepherd, J. & Benatallah, B. (2004) A Collaborative Approach for Caching Dynamic Data in Portal Applications. *Proceedings of the fifteenth conference on Australian database.*

Missier, P., Embury, S., Greenwood, M., Preece, A. & Jin, B. (2006) Quality views: capturing and exploiting the user perspective on data quality. *Proceedings of the 32nd international conference on Very large data bases-Volume 32.*

Nuseibeh, B. & Easterbrook, S. (2000) Requirements engineering: a roadmap. *Proceedings of the Conference on The Future of Software Engineering* Limerick, Ireland ACM.

Oliveira, P., Rodrigues, F., Henriques, P. & Galhardas, H. (2005) A Taxonomy of Data Quality Problems. *Second International Workshop on Data and Information Quality (in conjunction with CAISE'05).* Porto, Portugal.

Redman, T. (2000) *Data Quality: The field guide,* Boston, Digital Press.

Sommerville, I. (2000) Software Engineering. 6th Edition.

Strong, D., Lee, Y. & Wang, R. (1997a) Ten Potholes in the Road to Information Quality. *IEEE Computer***,** 38-46.

Strong, D. M., Lee, Y. W. & Wang, R. Y. (1997b) Data Quality in Context. *Communications of the ACM,* 40**,** 103-110.

Wang, R. Y. & Madnick, S. (1993) Data Quality Requirements: Analysis and Modelling. *Ninth International Conference on Data Engineering (ICDE'93).* Vienna, Austria, IEEE Computer Society.

Yang, Z., Cai, S., Zhou, Z. & Zhou, N. (2004) Development and validation of an instrument to measure user perceived service quality of information presenting Web portals. *Information and Management,* 42**,** 575-589.