ACS/IEEE International Conference on

Computer Systems and Applications

Tunis, Tunisia
14-18 July, 2003

ArAb Computer Society

IEEE

IEEE Catalog Number: 03EX722
ISBN: 0-7803-7983-7
Library of Congress: 2003106612

Click anywhere for the
MAIN MENU

Conference Chairs

General Chair

Salim Hariri, The University of Arizona, USA
hairiri@ece.arizona.edu

Program Committee Chair

Albert Y. Zomaya, The University of Sydney, Australia
zomaya@it.usyd.edu.au

Program Committee Vice-Chairs

Algorithmics and Theory - Stephan Olariu, Old Dominion University, USA
Architectures - Hassan Diab, American University of Beirut, Lebanon
Networking and Wireless - Mohan Kumar, University of Texas at Arlington, USA
Software - Ali E. Abdallah, South Bank Univ., U.K.
Applications - Rafik Braham, Université du Centre, Tunisia

Tutorials and Workshops Chairs

Kamel Barkaoui, CEDRIC, CNAM, France
barkaoui@cnam.fr
Ali Mili, New Jersey Institute of Technology, USA
mili@oak.njit.edu

Registration Chair

Ramzi Haraty, Lebanese American University, Lebanon
rharaty@lau.edu.lb
Yahya Slimani, Université de Tunis El Manar, Tunisia
yahyaslimani@fst.rnu.tn

Publicity Chairs

Belhassen Zouari, Université de Tunis El Manar, Tunisia
Belhassen.Zouari@ati.tn
Riadh Robbana, Université de Carthage, Tunisia
Riadh.Robbana@fst.rnu.tn

Organizing Chair

Mohamed Moalla, Faculté des Sciences - Tunis, Université de Tunis El Manar
Mohamed.Moalla@fst.rnu.tn

Program Committee

A. Bouzerdoum, Edith Cowan University, Australia
A. El Saddik, University of Ottawa, Canada
Abdelhamid El Iraki, PRICAM, Morocco
Abdelmajid Ben Hamadou, Southern University, Tunisia
Abdelsalam Helal, University of Florida, USA
Abdullah M. Al-Dhelaan, King Saud University, Saudi Arabia
Adel M. Alimi, University of Sfax, Tunisia
Ahmed Bouajjani, University of Paris 7, France
Aiman H. El-Maleh, King Fahd University of Petroleum & Minerals, Saudi Arabia
Ali Jaoua, University of Qatar, Qatar
Amal El Fallah-Seghrouchni, Université Paris 13, France
Amir Zeid, American University of Cairo, Egypt
Azzedine Boukerche, University of North Texas, USA
Badache Nadjib, USTHB University, Algeria
Bechir Ayeb, Université du Centre, Tunisia
Bhanu Prasad, Georgia Southwestern State University, USA
David Everitt, The University of Sydney, Australia
David Rine, George Mason University, USA
Edward Moreno, Euripides Foundation of Marilia, Brazil
El Mostafa Daoudi, University of Oujda, Morocco
El-ghazali Talbi, LIFL, University of Lille, France
Fadi Busaba, IBM, USA
Faouzi Ghorbel, ISIMA · University of Manouba, Tunisia
Fethi A. Rabhi, University of New South Wales, Australia
Fikret Ercal, University of Missouri-Rolla, USA
Franciszek Seredynski, Polish Academy of Sciences, Poland
Habib Youssef, Universite du Centre, Tunisia
Hany Ammar, West Virginia University, USA
Henda Ben Ghezala, ENSI, University of Manouba, Tunisia
Hesham El-Rewini, Southern Methodist University, USA
Hossam ElGindy , University of New South Wales, Australia
Hossam Hassanein, Queens University, Canada
Hossein Saiedian, University of Kansas, USA
Jean Paul Haton, INRIA-Loria, Nancy, France
Jelel Ezzine, Ecole Nationale d'Ingénieurs de Tunis, Tunisia
Joseph Davis, The University of Sydney, Australia
Kassem Saleh, American Univ of Sharjah, UAE
Khaled Bsaies, Tunis El Manar University, Tunisia
Khalid Ghedira, ENSI, Tunisia
Mansour Jaragh, Kuwait University, Kuwait
Marcelo Campo, UNICEN University, Argentina
Marcus Fontoura, IBM Almaden Research Center, USA
Mohamed Abdel-Mottaleb, University of Miami, USA
Mohamed Ahmed-Nacer, USTHB University, Algeria
Mohamed Bettaz, Philadelphia University, Jordan
Mohamed Ould-Khaoua, University of Glasgow, U.K.
Mohamed Tounsi, Ecole des Mines de Nantes, France
Mohammed A. Almulla, Kuwait University, Kuwait
Mohsen Guizani, University of West Florida, USA
Muhammad Sarfraz, King Fahd University of Petroleum and Minerals, Saudi Arabia
Osama A. Abulnaja, King Abdulaziz University, Saudi Arabia

Steering and Organizing Committees

Steering Committee

Hany Ammar (Chair), West Virginia University, USA
Jarallah AlGhamdi, King Fahd University of Petroleum and Minerals, Saudi Arabia
Mohamed Fayad, University of Nebraska, Lincolin, USA
Pierre Flener, Uppsala University, Sweden
Nashat Mansour, Lebanese American University, Lebanon
Ali Mili, New Jersey Institute of Technology, USA

Local Organization Committee

Webmaster
Iheb Beji, Faculté des Sciences - Tunis
Iheb@Medianet-tunisie.com

Hotel Reservations
Moncef Temani, Faculté des Sciences - Tunis
Moncef.Temani@fst.rnu.tn

Treasurer
Salah Hamami, Institut Supérieur de Commerce et d'Administration d'Entreprise - Manouba
Salah.Hamami@iscae.rnu.tn

Publication
Wahid Nasri, Ecole Supérieure des Sciences et Techniques - Tunis
Faten Ben Hassine, Faculté des Sciences - Bizerte
Hamouda El Hemissi, Faculté des Sciences - Tunis

Members
Narjes Doggaz, Faculté des Sciences - Tunis
Lilia Cheniti-Belcadhi, Institut Supérieur d'Informatique et des Technologies de la Communication - Hammam Sousse
Yosr Slama, Faculté des Sciences - Tunis

Table of Contents

To Use this Table of Contents: Scroll down or use the bookmarks in the left-hand frame to move to a new location in this index. Click on a **blue paper title** to view that paper. To return to this index after viewing a paper, click the PREVIOUS MENU bookmark in the left frame.

This CD-ROM was created using supplied PDF files. Viewing and printing of the text and graphics will depend largely on the accuracy with which each file was created.

Architecture

Networks

Mobility Impact on Mobile Ad hoc Routing Protocols
Nadjib Badache, Djamel Djenouri, Abdelouahid Derhab

Building Wireless Internet Services: State of The Art
Qusay H. Mahmoud

Advanced Research on Wireless LANs Operating at EHF Frequency Band
Larbi Talbi, Ahmed T. Denidni

Seamless Intra-Domain Handoff
D. Tandjaoui, N. Badache, A. Bouabdallah

Adaptive Quality of Service Resource Management in IP Networks
Salim Hariri, Kartik Dev.A.J, Seung Chan Oh, Jang Geun Ki, Kyou Ho Lee, Gil Young Choi

Software

Metamodeling and Measurement for the Software Process Improvement
Félix García, Francisco Ruiz, Mario Piattini

Distributed Software Process Models: An Integration Methodology
N. Lardjane, M. Ahmed-Nacer

The Mobile Unified Modeling Language for An Agent-based Bartering Software Application
Christo El-Morr, Kassem Saleh

Dynamic Software Development: Towards Continuous Growth
Darren Dalcher, Walaa Mohamed Bakry

Evaluating Software Architectures: Development, Stability, and Evolution
Rami Bahsoon, Wolfgang Emmerich

Integrity Enforcement in Relational Systems with Active And Deductive Capabilities
Laura Mota-Herranz, Matilde Celma-Giménez

Definition of a new kind of UML Stereotype Based on OMG MetaModel
Narayan Debnath, Daniel Riesco, Germán Montejano, Alicia Grumelli, Alcides Maccio, Paola Martellotto

Distributed Framework for Real Time Web Based collaboration: M7Tool CASE
Mehdi Snene, Michel Leonard

Including Workflow Concurrent Modeling in an Extension of the UML Activity Diagram Metamodel
Narayan Debnath, Daniel Riesco, Germán Montejano, Edgardo Acosta, Marcelo Uva

A Graphical Language for Defining Views in Object Oriented Databases
Marguerite Sayah, Andre Flory, Elias Choueiri

Timestamping Electronic Documents and Signatures
Nathanael Cottin, Maxime Wack, A. Sehili

BizArk: A Reusable E-Business Architecture based on Pattern-Oriented Technology
Ameur Touir, Hassan Mathkour, Tariq Al-Naeem

Advanced Management of Documents Integrity in an Intranet Environment
Abraham Alvarez, Richard Chbeir, Youssef Amghar

Data Quality Management Improvement
Ismael Caballero, Muñoz-Reja, Mario Piattini

A Support Toolset for the Development in a Temporal Database Environment
Mohamed Mkaouar, Rafik Bouaziz

Data Mining Techniques in Single-Column Indexes
Ying Wah Teh, Abu Bakar Zaitun

On Performance Analysis of Web Oriented Systems
Ines Chouaieb, Rana Ezzine, Kamel Barkaoui

A Taxonomy of Transformation Methods for Structured Documents
Nouhad Amaneddine , Youakim Badr

Techniques for Implementing Efficient Java Thread Serialization
Sara Bouchenak, Daniel Hagimont, Noel De Palma

Chameneos, a Concurrency Game for Java, Ada and Others
Claude Kaiser, Jean-Francois Pradat-Peyre

Specification Fragments with Defined Semantics to Support SW-Evolution
Andreas Bollin, Roland T. Mittermeir

Security Based Approach to Data Consistency in a Virtual Enterprise
Roger Atsa Etoundi , Marcel Fouda Ndjodo

Quantifying Architectural Attributes
D. M. Nassar, W. Abdelmoez, M. Shereshevsky, N. Gradetsky, H.H. Ammar, Bo Yu, S. Bogazzi, A. Mili

Comparative Evaluation of Object Request Broker Technologies
Fadi L. Nammour, Nashat Mansour

Adaptive Electronic Technical Manuals (AETMs)
Rabih F. Kraidli, Hany H. Ammar, D. Reynolds, G. Copen

Methodology for Architecture level Hazard Analysis: a Survey
A. Hassan, K. Goseva-Popstojanova, Hany H. Ammar

Theory

On the Embedding of MDG Specification Languages in HOL
Rabeb Mizouni, Sofiene Tahar, Paul Curzon

Symbolic Model Checking for Real-Time Systems Using Difference Typed Decision Graphs
Syrine Ayadi, Riadh Robbana

A Provably Correct Functional Programming Approach to the Prototyping of Formal Z Specifications
Ali E. Abdallah, Jonathan, P. Bowen, Alexandra Barros, Jose B. Barros

Constraint Propagation versus Local Search for Incremental Temporal Constraint Problems
Malek Mouhoub

Formal Model for Flat Role-Based Access Control with Delegation
Etienne J. Khayat, Ali E. Abdallah

Applications

Data Quality Management Improvement

Ismael Caballero Muñoz-Reja
ALARCOS Research Group,
Escuela Superior de Informática,
Universidad de Castilla – La Mancha
Paseo de la Universidad 4, 13071 Ciudad Real, España
Email: Ismael.Caballero@uclm.es

Mario Piattini Velthuis
ALARCOS Research Group,
Escuela Superior de Informática,
Universidad de Castilla – La Mancha
Paseo de la Universidad 4, 13071 Ciudad Real, España
Email: Mario.Piattini@uclm.es

Abstract:
Nowadays, more and more organizations are realizing of importance of their data, because it can be considered as an important asset present in nearly all business organizational processes. Traditional point of view of quality in Information System has been focused on software quality. The basis of our proposal is to consider information as a result of a Data Management Process, which can be supported by the running software at the Information System. Our aim is to optimize the Data Management Process (DMP) in order to assure data quality. For this, we have just drawn a framework based on maturity levels –as CMMI's ones- with specific and generic data quality goals, which are achieved by executing the corresponding activities.

KEYWORDS:
Software, Data Quality, Data Management Process (DMP), Data Quality Management, and Maturity Levels

INTRODUCTION

Contemporary organizations need Information Systems for supporting their business activities. Many efforts have been made in order to assure the quality of the running of the software, although not of the data. However this situation is changing due to the new status of data: finally, organizations have realized that data is one of the most important assets for them ([24]) Indeed, made tactic, strategic, or operational decisions are based on real data ([2,52]) If data does not have enough quality, these decisions can generate certain mistakes, which will make a negative impact on global efficiency of organizations ([56]). Because of this, it is a matter of management group to take care of data quality, by making commitments in order to improve the quality of existent data and data yet to come ([43]) All of these commitments must be oriented to avoid some kind of potholes which can generate serious problems like data not used, barriers to data accessibility or data utilization difficulty ([52]) These problems may translate into important consequences at a technical level – as in data warehouse implementation ([8])- an organizational level – loss of customers ([47]), important financial losses[i] ([40, 52]) or unsatisfied data workers ([12])- or even at a legal level, because of privacy regulations.

Many researchers agree these problems could be avoided or become resolved with adequate management of data quality by assuring it on all life cycle phases of Information Systems development and exploitation. But there is no methodology which enables a way of discovering how organizations manage their data when manufacturing a product, or giving a service ([36]) Some approaches have been proposed like TDQM – *Total Data Quality Management* – by Madnick and Wang ([41]), which is focused on conceptual and theoretical issues and DWQ – *Data Warehouse Quality* –by Jarke and Vassiliou ([34]), which focuses on technical and specific issues. Despite this fact, it still lacks an integral data quality framework

This paper discusses how to improve data quality of not only data products, but also entire DMP by walking up trough data quality maturity levels. Having a data quality maturity model means it can be known what organization is making good, what organization will not make bad any more and how an organization can reply in a special situation, that is to say: how to prevent defect, improve efficiency a predict the behaviour of the information systems.

The remainder of this paper is structured as follows: data quality matters are discussed in second section; in the third section a data quality model based on maturity levels, which highlight data quality goals for each level in the fourth section; fifth and sixth section deals about some issues related to data quality metrics referred along the works; finally, seventh section is reserved for our conclusions about the work and for our future lines of research.

FROM DATA TO DMP

Although it can seem obvious in this context, it is thought to be necessary to begin by remembering the concept of information, because all of the related work is based on it. Information is an added value plus data obtained by somebody (typically named stakeholder), when a data product is used ([14]) A data product is, as Wang ([57]) states, the result of a data transformation (or/and management) process. Here, data is considered as raw material for the manufacturing process. For instance, data regarding date of the birth, place of birth, names of parents, … collected from a citizen, is raw material for the process named "*Elaborating a Birth Certification*" which produces a data product named "*Birth Certification*" ([50]) If someone uses this product to develop a major work, then some value is added to the data written on a paper entitled "*Birth Certification*". The idea is: the more elaborate a data product, probably, the better the work achieved. It should be pointed out that quality concept is implicit in this last sentence. Therefore, another way of saying the previous: "*the more quality a data product has, the better made work achieved with it*

0-7803-7983-7/03/$17.00 ©2003 IEEE

is". For a product which has been said to have quality, it must verify all customer requirements ([10]) or it must be fit for use ([35]) By analyzing Deming's Fourteen Quality Points ([11]), it can be seen that for ensuring the quality of a product, it is necessary that the workers do their best, the systems work correctly and the raw material is suitable and in a perfect state. Translating these terms into data quality matter, it is possible to say that for a data product to adequately turn into information, it is necessary that this data is recollected, stored, consulted and presented according to several data quality rules. In order for data transformation processes to work to generate products which satisfy customers requirements it is also necessary that information workers have enough qualifications to manage data and run processes in a consistent way.

As one progresses through these ideas, new doubts arise, like which characteristics raw data and data product must present to be said to enable quality, who certifies that data transformation processes work satisfactorily, and finally who and how can somebody ensure a data worker is good enough for developing a data work. Many researchers have worked on these data quality issues, trying to explain them from several and different point of views. For instance, [47] tried to explain data quality through a data life cycle point of view, [24] and [59] tried to obtain some conclusion from achieved researches and [12] has just studied how data is used. Nevertheless, a common trend discovered in covered literature, consists of accepting data life cycles as a way of structuring data quality researches: data is collected, stored (before, it can be pre-processed), queried (before, it can be pre-processed) and presented as data products as required. It is nearly other raw material. But the data is unique due to its own specific nature. [2] identify several intrinsic characteristics belonging to data, which make the processes of collection, storage and usage the data be totally different from the same processes for other raw material, coming into true several potholes analyzed by [53] like the fact that multiple sources of data can generate different values, systematic errors can generate information losses, large amounts of data can be unmanageable for an application in a reasonable time, distributed and heterogeneous systems can generate inconsistent formats, values or definitions, and so on.

Having taking into account all of these issues, some authors like [33] or [55] have proposed to map all data quality requirements (including user ones) into the quality goals for software characteristics showed in ISO 9126 ([31]): functionality, usability, maintainability, efficiency, portability and reliability. Traditionally, these quality goals are known in a data quality environment as data quality dimensions ([47, 58]) and can be broken down again into minor and particular data quality dimensions for making a study easier. Through literature, one can find how many authors have tried to explain the meaning of all relevant dimensions from several point of views ([1, 12, 19, 23, 36, 40, 46, 47, 52]) All of these people have tried to identify a standard set of data quality dimensions valid for any data product; but as [24] states, it is nearly impossible due to different nature of any data environment. Despite this a common schema broadly used is the one proposed by [52], in which, data quality dimensions are grouped into four categories: intrinsic, data accessibility, contextual and data representation.

Having in mind this conception of information being the result of a data transformation process from raw material, some authors, implicitly or explicitly (like [5]) want this data transformation process to be modeled and managed from an engineering point of view in order to optimize data quality for required applications. This implies a lot of efforts, which must be made by all data quality researches. Among these efforts, one can find a metamodel for entirely representing a data environment, a set of procedures for managing data quality requirements, metrics for a quantitative assessment, methods for analysis, … general speaking a framework for managing data quality.

As previously mentioned, many of these efforts have been made, but there is no common criteria for identifying concepts and elements that are named differently by some authors compared with others or a universal data quality dimension set that has been defined.

As more situations and bibliography have been studied, so we think more about data quality needs of an overall DMP. Our idea is to optimise data quality in two ways: first, by controlling and monitoring the DMP and all minor processes (this it to say, environment) and secondly by introducing specific data quality requirements into the DMP.

This paper deals with a way, not only, of assessing data quality, but also of determining the exact point of the road to excellence in which a DMP is. This work aims to draw the entire road by putting in order all of the concepts regarding data management and data quality, and addressing these issues through data product life cycle.

A DATA QUALITY MODEL

As said before, some researches have been aimed at providing a data quality measurement and/or assessment methodology in a data environment. However this is not sufficient, because the only possible thing to be done is to measure and/or to assess, although more is required: knowing how an organization works, and developing the ability to identify major problems or standardize a data quality culture by implying leaders ([25]) None of the strategic, tactic or operative perspectives are drawn by only measuring: it is necessary to assess DMP so that its future behaviour can be predicted or possible sources of its variability determined.

By studying different models of software assessment processes, like CMM ([25, 45]), CMMI ([49]), ISO 9001 ([9]), BootStrap ([4]) or SPICE ([44]), it can be found out that none of them take data quality into account ([36]).

To fill this area of lacking information, a data quality assessment environment is needed. The main purpose of this work is to provide guidance for improving DMP so that organizations become increasingly able to manage all related data processes such as data acquisition processes, data product manufacturing processes and data maintenance processes more efficiently. This will be done by addressing several issues of main process areas

like process management, project management, support and engineering process.

Indeed, a data quality team can:

- Determine data quality status by addressing a maturity level
- Propose improvements to DMP for achieving higher data quality maturity levels

Our work is based on maturity levels proposed by CMMI ([49]). As is known, CMMI offers two models: continuous and staged. The staged model is preferred more than the continuous model because it appears easier to work with a well-defined sequence of improvements (which can cover basic management project principles to complex data quality management issues) As higher levels are reached, some benefits are achieved such as improvements in data process behavioural predictions, greater control of them and improvements of DMP efficiency.

The following elements have been taking into account:

- A set of data management maturity levels
- A set of data quality goals defined for each maturity level
- An assessment method

It is important to realise that an Information System, from our point of view of maturity levels, is defined as a set of DMPs that can share several resources and some kinds of minor processes. So, for determining the global data quality maturity level in which an Information System is, all DMPs must be assessed and the results might be combined for stating the global data management maturity level.

Next, data quality maturity levels are going to be described, and the data quality goals associated to each one are going to be highlighted.

MATURITY LEVELS AND QUALITY GOALS.

We define five data quality maturity levels: Initial, Definition, Integration, Quantitative Management and Optimising. For each level, we address specific activities, which aim to obtain specific data quality goals. These specific goals are achieved by getting some typical products. Next, maturity levels for DMP are described, identifying the most important objectives and the product that must be generated.

Initial

A *DMP* is said to be at **Initial Level** when no managed and coordinated efforts are made in order to assure data quality. Success in these organizations depends on some people who are supposed to be qualified enough to develop these mentioned efforts.

Definition

A *DMP* is said to be at **Definition Level** or **Defined** when efforts are made in order to draw the entire process, identifying and defining all components (both active and passive), their relationships and the way in which these ones are developed according to a previous project. Therefore, a DMP is said to be defined when a DMP Project (DMPP) is managed. For managing a DMPP it is necessary to develop the next management activities:

- **DMP Project Management**. The main goal of this activity is to create a plan for coordinating efforts and elaborate a document, which clearly describes an agenda of activities and a budget for developing the DMP. This document can be done by following [28] Within some planned activities, the next process must be undertaken: a data requirements management, an analysis of these requirements, a design of a solution for satisfying them, an implementation of the process based on previous design, and testing for implemented process. Any technique or tools, like PERT or CPM, used on project development may be used here.
- **Data requirements management**. Wang et al (1994) identified three types of user requirements: Product Requirements, Quality Product Requirements and Data Quality Requirements. All of them must be collected and documented. IEEE 830.1 ([30]) could be used for guidance when elaborating on these documents. By assimilating to data quality, two products must be done: a DMP User Requirements Specification (DMPURS), which will have a Quality Process Requirements Specification (QPURS) and a Data Quality User Requirements Specification (DQURS). These requirements are the starting point for modeling the DMP, the database or data warehouse where data and data quality issues are going to be saved. Some graphical representation might be used for each model For example, for DMP, the IP-MAP, which was given by [50]; and for database or data warehouse and data quality issues, the extended entity-relationship model proposed by [61]. From these products, an analysis might be achieved for giving a first approach to a solution, which will consists of dimensions and metrics, data sources and targets, and a place where data and other components are going to be saved ([40])
- **Data quality dimensions and metrics management**. Having metrics for measuring DMP efficiency may help to improve it. This evolves several matters, which can be found out from DQURS. So data quality issues must be controlled ([24]), a quite comprehensive set of data quality dimensions (or parameters) and data quality indicators (or metrics) must be defined ([61]) Many authors like [47] or [12] have explained how data quality dimensions can be identified and how to measure certain data quality issues in certain application and environment. The ideal situation would be to have a universal data quality dimensions set with appropriate metrics, but, unfortunately, this is not possible ([46]) due to the fact that data quality depends directly on data problems. However there is a classification that begins to be broadly used. It is the one proposed by [52] and referred to in section 2. This classification is one which groups data quality dimensions into four categories: Intrinsic to data (Accuracy, Objectivity, Believability, Reputation), Accessibility (Accessibility, Access security), Contextual (Relevancy, Value-Added, Timeliness, Completeness, Amount of data) and Representational

(Interpretability, Ease of understanding, Concise representation, Consistent representation) Data quality managers are encouraged to find out the one which best suits their problem, having into account several factors, which are briefly described in fifth section. For determining these metrics, generic and usable methodologies might be used like [29] or GQM, given by [54] Even authors like [40] have proposed a more specific data quality measurement framework with several and concrete data quality issues to measure: data quality of data models, data quality of data values, data quality of data representation or data quality of information policy. On the other hand, some authors like [2, 6, 7, 57] have proposed metrics for measuring specific issues of specific components of DMP; in sixth section, and as example, some metrics are developed, by using GQM ([54]) An important aspect of measurements is the need of automating measurements, as required by [22]. All these metrics will help to improve the DMP.

- **Data sources and data targets management**. Due to particular intrinsic characteristics of data, it is necessary to identify and to document data sources as well as data targets. [2] and [22] discuss these issues and give several ways for treating information from multiple sources. Such sources like targets must be identified, defined and limited from DMPURS. In a data warehouse environment, tools and techniques like ETLs ones must be used in order to unify semantic and formats of incoming data ([13])
- **Database or data warehouse development or acquisition project management**. Raw data must be collected and stored in an appropriate warehouse. Typically, this place has been a database or a data warehouse, which can be seen as a passive component of DMP. Some organizations develop their own data products or perhaps can acquire it from third companies. So that data quality can be assured, it is highly recommended that an acquisition project or a development project is managed having taken into account both DMPURS and DQURS. This activity may also include other minor sub activities like **Configuration Management**, **Maintenance Management** or **Commer-cial solution election management** for instance.

For climbing from the Initial Level to the Definition Level, a plan for developing the DMP must be drawn and followed.

Integration

A *DMP* is said to be at **Integration Level** or **Integrated** when level 2 is raised and many efforts are made in order to develop and execute this according to organizational data quality polices. This therefore implies having several standardized data quality issues. And this level is focused on the way in which organizations capture all knowledge from their experience, and make it reusable by adding it to an organizational culture in order to avoid last minute errors or to enable it to produce a good work from the beginning. For this, standards or guidelines must be redacted and referred to for all issues, and paying particular attention to these standards tries to promote consistent use of better tools and methods ([25]). Therefore, several activities must be achieved in order to define certain data quality goals through organizational standards. As these activities are performed, organizations are getting these standards and guidelines that may help to guidance for DMP development. As the more used and refined these standards are, the more quality the data product has, and thus the more satisfied the stakeholders are. The next activities are defined as the key ones for getting an organizational culture of data quality:

- **Data Quality Team Management**. Data quality initiatives need people to support all of the activities that must be performed. These people must work according to the organization's ideas and trends. Among their many abilities must be data quality and to have managerial skills. [48] points out the need for high managers to lead data quality initiatives. This implies the need to select people who take care of data quality through DMP, and support activities related to it, like standardization and measurements.
- **Data quality product verification and validation**. All information products must be verified and validated, in order to avoid defects or disagreements with user requirements. A technique that can be used, might be software inspections ([15, 21]), but adapted to data quality issues. A more specific methodology, which can be used, is *data testing* proposed by [39], expanding to DMP, because the Data Testing Model is limited to data stored in an information system. In order to coordinate efforts a plan for testing could be designed and drawn up by following for instance, [27]
- **Risk and poor data quality impact Management**. Authors like [12] states that it is necessary to delimit risk for determining the impact of poor data quality to DMP. [20] proposes a methodology, which can be adapted to data quality issues in order to collect and document all risk.
- **Data quality standardization Management**. All lessons learned through specific experiences might be recollected and documented. An example of standardization could be the sixth process of TDQM ([13]), which seeks to accomplish these issues "by integrating quality management beliefs, principles, and methods into the culture of the enterprise". Only by incorporating last data quality experiences to new DMP, is it possible to develop better ones.
- **Organizational Processes Management**. A way of coming to truth all above-mentioned efforts consist of defining data quality policies, which affect not only a concrete DMP, but an entire organization. The Data Quality Team must work on data policies, which reflect organizational culture. [40] presents the elements that are the subject of data policy design. For organizations have a real data quality culture, because all of their processeswether or not

they are related to data, must take into account data quality issues in order to improve it.

A DMP reaches Integration Level when it is developed under the organizational support. Obviously, this fact implies, first, that the DMP is managed through a project (therefore, is defined); secondly that this organizational support for data quality issues exists.

Quantitative management

A *DMP* is said to be at **Quantitative Management Level** or **quantitative managed** when it is integrated into an organizational data quality culture and many efforts are made in order to take several measures related to DMP and its components. Therefore, the main goal of this level is to obtain the quantitative compliance that the DMP performance over some reasonable time period is as consistent as required in terms of variation and stability ([16]). In order to satisfy this proposal, a Plan for Measurements Management must be drawn and followed. So, the next activity must be executed:

- **DMP Measurements Management**. As Meredith (2002) states, a plan for software quality measurements starts with the decision of taking measures, which can be seen as tools for improving software quality. This implies the need to choose what, when and how to measure, how to represent these measures and to whom. Since metrics about DMP have been drawn at Definition Level (*what* question), the plan must focus on the remainder of these questions. With respect to the *when* question, the answer is when measurements do not alter DMP. With respect to the *how* question, some algorithm might be outlined in order to make measurements repeatable, and even more, automatizable. Finally, as important as metrics is in the way of representing results, many authors like [13, 40] propose the use SPC as a way of representing data about DMP. On the other hand, another complementary way of representation is the one proposed by [25, 26], in which Kiviat's diagrams are used to relate several aspects of the DPM or the data quality components. The way depends on the work, which will be developed later.

A DMP reaches Quantitative Management Level when, once integrated, some metrics are defined in order to control its variation and stability. All these metrics can be defined into the organizational culture.

Optimizing

A *DMP* is said to be at **Optimizing Level** when quantitative measurements taken at previous levels are used in efforts in order to detect defect sources or identify ways to optimise the entire processes. Two kinds of activities might be executed:

- **Causal Analysis for Defect Prevention.** [51] offer a framework for defect prevention, which could be used in a data quality environment in order to identify, and analyze root causes of data process defect, and modify for avoiding it. For an easier identification of data quality elements, which can be root of problems, Ishikawa's diagram proposed by [13] in the fifth process of TDQM may be used.
- **Organizational Development and Innovation**. This clause is basic in the main idea of a continuous improvement. Learned lessons in DMP must be basis not only for defect prevention, but also for continuous improvements.

IMPORTANT FACTORS TO TAKE INTO ACCOUNT WHEN DRAWING DATA QUALITY METRICS.

As showed, some metrics for measuring certain issues on data quality are required. Many of these metrics are developed on ad hoc way to solve specific problems ([46]) so, each organization uses their own criteria over their own data problems for choosing metrics and there is no a set of universal metrics applicable to any case. Due to this fact, authors agree that being relatively easy to choose data quality dimensions which best describe a data quality problem, however, developers might take into account several factors when drawing metrics for specific problems which impact on validity of achieved results. In this section, many factors are going to be shown. Some of them have been observed from authors' experiences related on data quality literature found in not academic publications. These factors are:

- **Reasons to measure.** [57] highlights the reasons for which is intended to measure as one of the most important factors when choosing metrics. Metrics can be defined for controlling a project evolution, for evaluating the degree of customer satisfaction, for optimizing data product quality, … All these mentioned reasons can impact on the definition of distinct metrics for the same data quality dimensions, since they are related directly to the rol of stakeholder.
- **Dependence of the Environment of Operation**. All the factors might be fitted here as technological (as the operating system, the proper database management system, network, performance computer, failure tolerance of system...) as human (skills with computers of the people who works with data –[38]-, previous knowledge of quality of information...), which can affect to results of measurements processes. For example, let be supposed somebody is interested in measuring timeliness of a data set. The metric entitled "*Time in seconds that a transaction takes to be completed from the moment in which user finish data entry process until data has been processed*" can be chosen: nobody can assure that the same transaction with the same data load should take the same time in finishing, since it will depend on the typing speed featured by person who introduces data, of the degree of congestion of the net, of the transactions engine of the database....
- **Capacity of judgment data quality.** It is said quality is a subjective concept. To determine if data is or not good for an application it is necessary to emit a judgment from a data quality dimension point

of view. This judgment must be endorsed by a serie of tests, which might allow to obtain the same result in case of the judgment had to repeat in any other circumstances. To emit a judgment consists of giving a qualification or valuation for data according to the quality criterion expressed by the data quality dimension. This capacity of prosecution can relapse on a sufficiently qualified user, on the proper database or on some another external mechanism:

- **Dependence on the user**. To be able to measure it is necessary to ask to users their opinion (qualification) that deserves an information in a certain dimension of quality. Their response will be subjective. And a subjective measurement can influence *"very negatively*", "*negatively*", "*normally*", "*positively*" or "*very positively*" in the final valuation of this data: they can present a series of problems or connotations (created interests, lack or mistakes of formation, disinterestedness...) that will do the measurement not quite correct as that would be wished. These problems are not an object of study of this report.
- **Dependence of the base or store of information**. [17,18] proposes capacity of evaluation to determine a valid value for a certain metrics should relapse into the proper database across the detailed definition of metadata (for instance, as the ranges of values for the domain of the attributes) or across the existence of business rules ([40]) which decide the above mentioned ranges of values: for example a person who has less than eighteen years old should have a null value in the number of the driving license in a database, or certain lendings might not be allowed in a library to whom they do not verify a series of conditions...
- **Spatial and temporal location of the measurement.** The time to realize a measurement must be perfectly defined respect to the procedure that is doing the valuation of the quality of information. This implies to know what, when, how, and where to measure and who must realize the measurement. For example, if the reliability is measuring up (*what*) of the information in data update operation is necessary to know which person or process (*who*) will measure after the update (*when*) the affected tables (*where*) applying some mechanism of valuation (*how*).

It is necessary to take into account these four groups of factors at the moment of defining metric for data qualities dimensions not only for the validity of the measurements but for the way in which the proper procedures will be implemented to measure: it does not have the same computational complexity to ask to a user a value for certain metrics of data quality than to have to do a query to a knowledge base to obtain this value. Therefore, the chosen metrics for a data quality dimension should be representative, ideal and not redundant.

DEFINITION OF DATA QUALITY METRICS.

It is possible to apply the method "*Goal - Question - Metric*" (GQM) by [54] for which a goal is defined and progressively refined by making questions for which metrics are defined for giving enough quantitative information as answers to these questions.

To help in the task of finding specific metrics for data quality characteristics, suggestions made in ISO 9126 ([32]) can be consulted.

Next, two data quality metrics are proposed and defined for data quality dimensions. First of them depends directly on metadata (e.g. definition of the ranges of the values for domains). The second one is a generic example of how a metric, which manages values created as a result of a judgment or evaluation, might be defined. As before explained, this judgment might have been done by a user, or according to any defined business rules.

Metric based on metadata: metric for correction.

The next metric is defined to measure the data quality dimension of *correction* of a database. The most interesting way of showing these values is as the rate of attributes at table and database level which is incorrect, understanding for incorrect values those which do not belong to the range of valuies defined for the domain of this attribute. Here it is:

Goal: To determine the degree of correction of a database.

Questions:

1. When an attribute is incorrect?
2. How is it possible to determine the Rate of Incorrect Values for an attribute of a table in a database?
3. How is it possible to determine the Rate of Incorrect Values for a database?

Metrics:

- **NUMBER OF INCORRECT VALUES FOR AN ATTRIBUTE (NVI (A)).**
    - An attribute is said to be incorrect when saved value does not belong to the range of values for the defined domain for this attribute. In this range of values it might be or not be the null one.
    - NVI(A) is defined as the number of rows in which stored values for a certain attribute A does not belong to the range of values for the defined domain for this value A in the metadata.
- **NUMBER OF ROWS OF A TABLE (NRT (T)).**
    - NRT(T) is defined as the number of rows that has a certain table T.
- **RATIO OF INCORRECT VALUES FOR AN ATTRIBUTE A (RVI (A, T)).**
    - It is defined as the relation between the number of incorrect values for an attribute A and the number of rows that the table has where the attribute is defined.

$$RVI(A,T) = \frac{NVI(A)}{NRT(T)}$$

- **RATIO OF INCORRECT VALUES IN A TABLE (RVIT (T)).**

    - It is defined as the arithmetical average of incorrect values for all the attributes of the table**:**

$$RVIT(T) = \frac{\sum_{i=1}^{NRT(T)} NVI(A_i)}{NRT(T)}$$

- **NUMBER OF TABLES OF A DATABASE (NT (BD)):**

    - It is defined as the number of tables a database has**.**

- **RATIO OF INCORRECT VALUES FOR A DATABASE (RIVBD (BD)):**

    - It is defined as the geometric average of the incorrect values for the tables of a database**:**

$$RIVBD(BD) = \frac{\sum_{i=1}^{NT(BD)} RVIT(T_i)}{NT(BD)}$$

All of these metrics can be generalized for any kind of database as object-oriented ones, by adapting these formulas to the used technology.

A generic metric based on values created as a result of an evaluation.

The next metric is a generic one based on the fact that after having evaluated the stored values in the attribute A of a table T with respect to a certain dimension of quality D, a value V is generated and is stored in the attribute A' of the same table or of another one. To obtain this value V, mechanisms M had been to be implemented for producing it by asking direct questions to a qualified user ("*please, specify a value* V_I *for attribute* A *according to the dimension* D "), or by evaluating business rules ("*if the information is X and it has been produced by source* S, *then its degree of* D_I *is* V_I"). But it is necessary to highlight metrics must be independent from the mechanism M that has generated this value: mixed mechanisms might be defined and metrics might continue being valid.

Goal*:* To determine data quality level of an attribute A of a database BD respect to the dimension D.

Questions:

1. Which is the value V for a quality attribute A' where there is stored the result of evaluating the attribute A of the table T respect to dimension D?
2. How can the average value of the values V for an attribute A of a table in a base or store of information be determined?

Metrics:

- **VALUE OF QUALITY INDICATOR OF THE ATTRIBUTE A OF THE TABLE T WITH RESPECT TO THE DIMENSION D (VQI (A, T, D))**

    - That is defined as**:**

        - If A ∉ T then VQI(A,T,D)= ∅ = NULL

        - IF A ∈ T then VQI(A,T,D)= V = EVALUATION (M, T, A, D)

            Being M the mechanism of evaluation. But it is insisted on that the metrics must be independent from evaluation mechanism.

- **AVERAGE OF QUALITY INDICATOR OF A OF A TABLE WITH RESPECT TO THE DIMENSION D (AQI (A, T, D))**

    - It is defined as the arithmetical average of the values of the quality indicator of the attribute A with respect to the dimension D:

$$AQI(A,T,D) = \frac{\sum_{i=1}^{NRT(T)} VQI(A_i,T.D)}{NRT(T)}$$

Set of metrics for a problem.

In some occasions, it is necessary to define a set of metrics to describe in a suitable way a concrete situation of data quality. Once upon chosen the best fitting dimensions and metrics, it proposes to draw in a table as the following one all the information related to the decisions on the chosen metric. Table 1 resumes this issue.

CONCLUSSIONS AND FUTURE WORK

We have just presented a data quality model based on maturity levels. This work addresses main issues on drawing a DMP, identifying all components and established relationships, and for highlighting quality aspects for processes and for data governed by data quality policies. The idea for making a DMP stemmed from the sensation that data quality measurements represented by third companies' works was insufficient because some present areas lacking information are needed to be filled. The idea of drawing maturity levels came from quality ideas and were formalized through Humphrey's ([25]) concepts. Data quality aspects are formalised mainly trough TDQM's and other principles. Technical aspects of data warehouse management were first observed from DWQ.

Organizations must learn and formally model their data quality management so that major data problems sources can be identified. Once identified, initiatives for avoiding them or for improving efficiency can be arranged.

We are currently working on several issues, such as a data quality management questionnaire for each level,

and a major classification of both active and passive components, and their relationships. This is done by studying literature as well as from our own experiences. For an empirical validation of our advances, we are working with several enterprises, which have accepted to implement this work through their DMPs. This will allow refining maturity levels and quality goals.
The main contribution of this paper is to provide a framework where all these researchers can work to unify all data quality concepts and related trends.

Table 1: Summary of Metrics

Metrics	Measured Dimension	Allowed Values	Dependence of the Environment	Resource - Mechanism
RIVD (BD)	Correction	[0,1]	No	Meta Data
AQI(T, A,D)	D	[a, b]	No	M

Where:

- *a* and *b* the minimal and maximum values respectively for *V* and
- M is the mechanism of evaluation.

ACKNOWLEDGEMENTS

This research work has been developed under CALDEA project, financed by CICYT (TIC 2000-0024-P4-02)

REFERENCES AND CITATIONS

[1] Ballou D., Wang R., Pazer, H., Tayi, G.K. (1998) *Modeling Information Manufacturing Systems to Determine Information Product Quality*. Management Science 44(4), 1998, Pp 462-484

[2] Ballou, D. and Tayi, G.K. (1999). *Enhancing data quality in Data Warehouse Environments*. Communications of the ACM, January 1999/ Vol 42, No I

[3] Basili, V.R. y Weiss, D. (1984). *A methodology for Collecting Valid Software Engineering Data*. IEEE Transactions on Software Engineering. SE-10. No. 6. pp.728-738.

[4] Bicego, A., and Kuvaja, D. (1993) *Bootstrap, Europe's Assesment Method*, IEEE Software, pp93-95.

[5] Bobrowski, M., Marré, M., Yankelevich, D. (1998) *A software Engineering View of Data Quality*. Proceedings of Second International Software Quality in Europe. Belgium. November 1998.

[6] Bouzeghoub, M. and Kedad, Z. (2000) *A quality-based framework for Physical Data Warehouse Design*. Proceedings of the International Workshop on Design and Management of Data Warehouse (DMDW'2000) Stockholm, Sweden, June 5-6, 2000.

[7] Calero, C. and Piattini, M. (2002) Metrics for databases: a way to assure the quality. In: Information and Database Quality. Kluwer Academic Publishers

[8] Celko, J. (1995): *Don´t Warehouse Dirty Data. Datamation*, 15 October, pp. 42-52.

[9] Coallier, F. (1994) *How ISO 9001 fits into the software world*, IEEE Software. January 1994, pp 98 -100

[10] Crosby, P.B. (1979) Quality is free. New York, New York: McGraw Hill

[11] Deming, W.E. (1986) *Out of the crisis*. Cambridge, MA. MIT Center for advanced engineering study

[12] English, L.P. (1999) *Improving Data Warehouse and Business Information Quality: Methods for reducing costs and increasing Profits*. Willey & Sons.

[13] English, L.P. (2002) *Total Quality data Manaegment (TQdM), Methodology for Information Quality Improvement*. In: Information and Database Quality. Kluwer Academic Publishers

[14] Eppler, M.J. (2001) *Increasing Information Quality through Knowledge Management Systems Services*. Proceedings of the 2001 International Symposium on Information System and Engineering (ISE'01) June 25-28, 2001. Las Vegas, Nevada, USA

[15] Fagan, M. (1976) *Design and Code Inspections to Reduce Errors in Program* Development.IBM Systems Journal 15, *3* Pp 182-211.

[16] Florac, W, A.and Carleton, A.D. (2002) *Using Statistical Process Control to Measure Software Process. In Fundamental Concepts for the Software Quality Engineer*. Taz Daughtrey Editor. American Society for Quality.

[17] Forino, R., (2000). *Data e.Quality: The Data Quality Assessment, Part I.* DM Review Online in August 2000. www.dmreview.com/editorial/dmreview/print_action.cfm?EdID=2600.(December, 2002)

[18] Forino, R., (2000b). *Data e.Quality: The Data Quality Assessment, Part II.* DM Review Online in September 2000. www.dmreview.com/editorial/dmreview/print_action.cfm?EdID=2755 .(December, 2002)

[19] Genero, M. and Piattini, M.(2002) *Quality in conceptual Modelling.* In Information and Database Quality. Kluwer Academic Publishers

[20] Getto G. (2002) *Risk Management Supporting Quality Management of Software Acquisition Projects. In Fundamental Concepts for the Software Quality Engineer*. Taz Daughtrey Editor. American Society for Quality.

[21] Gilb, T. and Graham D. (1993) *Software Inspection*. London: Addison-Wesley Longman.

[22] Hinrinchs, H. and Aden, T. (2001) *An ISO 9001:2000 compliant quality management System for Data Integration in Data Warehouse System*. Proceedings of the International Workshop on Design and Management of Data Warehouse (DMDW'2001) Interlaken, Switzerland, June 4, 2001.

[23] Hoxmaier, J.A. (2001), *Dimensions of Database Quality. In Developing Quality Complex Database*

Systems: Practices, Techniques, and Technologies. Editor Shirley Becker. Idea Group Publishing.

[24] Huang, K.T., Lee, Y., Wang, R. (1999) *Quality Information and Knowledge*. Prentice Hall, Upper Saddle River

[25] Humphrey, W. (1989) *Managing the software process*, Addison – Wesley, Reading Mass., 1989

[26] Humprhey, W. (2002) *The Software Standard Profile. In Fundamental Concepts for the Software Quality Engineer*. Taz Daughtrey Editor. American Society for Quality.

[27] IEEE (1986) *IEEE STD 1012-1986 IEEE Standard for Software Verification and Validation Plans.*

[28] IEEE (1987) *IEEE STD 1058.1-1987 IEEE Standard for Software Project Management Plans*.

[29] IEEE (1992) *IEEE STD 1061-1992. IEEE Standard for a Software Quality Metrics Methodology*.

[30] IEEE (1998) *IEEE guide to Software Requirements Specification. IEEE Std. 830-1998*. New York (USA) IEEE Computer Society.

[31] ISO (2001) ISO/IEC 9126: *Information technology – Software product evaluation – Quality characteristics and guidelines for their use*, ISO/IEC Standard 9126, Geneva , http://www.iso.ch

[32] ISO 9126 (1999). *Software Product Evaluation-Quality Characteristics and Guidelines for their Use*. ISO/IEC Standard 9126. Geneva

[33] Jarke M., Jeusfeld, M.A., Quix, C., Vassiliadis, P. (1998) *Architecture and Quality in Data Warehouses*. Proceedings of Tenth International Conference CAiSE'98 Pisa. Italy, Springer-Verlag, pp 93-113

[34] Jarke, M., and Vassiliou, Y. (1997). *Data Warehouse Quality: A review of the DWQ Project.* Proceedings of Second Conference on Information Quality. Massachusetts Institute of Technology, Cambridge

[35] Juran, J.M. (1988) *Juran on Planning for Quality*. New York, New York: McMillan.

[36] Kahn, B., Strong, D., Wang, R. (2002) *Information Quality Benchmarks: Product and Service Performance*. Communications of the ACM April 2002/Vol. 45, No. 4

[37] Kimball, R., (2000). *Indicators of Quality*. Inteligent Enterprise 3(6). Pp 20-23

[38] Kimball, R., (2000b) *Is your data correct?* Intelligent Enterprise. www.intelligententerprise.com/001205/webhouse1_1.shtml (December, 2002)

[39] Kiszkurno, E., Yankelevich, D. (2001) *Data Testing*. ASSE 2001 Proceedings of SADIO.

[40] Loshin D. (2001) *Enterprises Knowledge Management: The Data Quality Approach*. Morgan Kauffman, San Francisco (California)

[41] Madnick, S., and Wang, R. (1992) *Introduction to TDQM Program Research*. Available on http://web.mit.edu/tdqm/papers/92/92-01.html. Last access on December, 2002.

[42] Meredith, D.C. (2002) *Managing with Metrics: Theory into Practice*. In Fundamental Concepts for the Software Quality Engineer. Taz Daughtrey Editor. American Society for Quality.

[43] Motha, W. M. and Viktor H.L. (2000). *Expanding Organizational Excellence: The Interplay between Data Quality and Organizational Performance*, International Conference on Systems, Cybernetics and Informatics (SCI'2001), Orlando: USA, July 22-25, Volume XI, pp.60-65.

[44] Paulk, M., and Konrad, M.(1994) *An overview of ISO's SPICE project* American Programmer, pp16-20

[45] Paulk, M., C. Weber, B. Curtis, and Chrissis, M. (1995) *The Capability Maturity Model Guideline for Improving the software Process*, Addison-Wesley, Reading, Mass.

[46] Pipino, L., Lee, Y., Wang, R. (2002) *Data Quality Assessment* Communications of the ACM April 2002/Vol. 45, No. 4ve

[47] Redman, T.C. (1996) *Data Quality for the Information Age.* Artech House Publishers, Boston.

[48] Redman, T.C. (2001) *Data Quality: The Field Guide*. Digital Press

[49] SEI. (2002) *Capability Maturity Model® Integration (CMMI^{SM}), Version 1.1 CMMI^{SM} (CMMI-SE/SW/IPPD/SS, V1.1) Staged Representation CMU/SEI-2002-TR-012 ESC-TR-2002-012*. en http://www.sei.cmu.edu/publications/documents/02.reports/02tr002.html (December 2002)

[50] Shankaranarayanan, G. Wang, R., Ziad, M. (2000) *IP-MAP:Representing the Manufacture of an Information Product*. Proceedings of the 2000 Conference on Information Quality.

[51] Smith, C., and Heights, A. (2002) *Defect Prevention: The road less Traveled*. In Fundamental Concepts for the Software Quality Engineer. Taz Daughtrey Editor. American Society for Quality.

[52] Strong, D.M., Lee, Y. W., Wang R.Y. (1997) *Data Quality in context*. Communications of the ACM. May 1997, pp 103-110

[53] Strong, D.M., Lee, Y. W., Wang R.Y. (1997b) *Ten potho-les in the road to information quality*. IEEE Computer August 1997, pp 38-46

[54] van Soligen, R. and Berghout, E. (1999) *The Goal/ Question/Metric Metodology: a practical guide for quality improvement of software development*. Ed McGraw-Hill

[55] Vassiliadis, P. (2000) *Data Warehouse Modelling and Data Quality Issues. Knowledge and Data Base Systems Laboratory Computer Science Division* Department of Electrical and Computer EngineeringNational Technical University of Athens Zographou 157 73, Athens, Greece

[56] Wand, Y. and Wang, R. (1996) Anchoring Data Quality Dimensions in Ontological Foundations. Communica-tions of the ACM (CACM), 39, (11), pp 86-95

[57] Wang, R., (1998) A product perspective on data quality management. Communications of the ACM. February 1998 Vol 41(2) pp58-65

[58] Wang, R., Kon, H., y Madnick S. (1993) *Data Quality Requirements Analisys and Modeling*. Published in the Ninth International Conference of Data Engineering Vienna, Austria. Pp 670 – 677

[59] Wang, R., Storey V.C., Firth, C. F. (1995) *A frame-work for analysis of Data Quality Research*. IEEE Transactions on Knowledge and Data Engineering, Vol 7(4). Pp 623-640.

[60] Wang, R., Strong., D., Guarascio, L.M. (1994) *Beyond Accuracy: What Data Quality Means to Data Consumers*. Journal of Management Information Systems, Spring 1996, Volume 12, No. 4. pp. 5-33.

[61] Wang, R., Ziad, M., Lee, Y.W. (2001*) Data quality*. Kluwer Academic Publishers. Massachusets (USA)

---

[i] Data Warehousing Institute has noticed on February of 2002, annual losses of six hundreds millions of dollars in American organizations due to data quality problems (Intelligent Enterprise 5(6)- March 2002, pp 12)