

GETTING BETTER INFORMATION QUALITY BY ASSESSING AND IMPROVING INFORMATION QUALITY MANAGEMENT

(Research-in-Progress)

Ismael Caballero, Óscar Gómez, Mario Piattini

Alarcos Research Group, ESI (UCLM), Spain

{[Ismael.Caballero, Mario.Piattini](mailto:Ismael.Caballero@uclm.es)}@uclm.es, ogomezg@mixmail.com

ABSTRACT: Information quality has already become a decisive factor in information - dependent business. Much has been said about the growing importance of data and information quality, and many researching lines over the last decade have looked at specific data and information quality issues from different standpoints. However data and information quality goes beyond the definition of data quality dimensions, there is still lack of an integrative framework, which can guide organizations in the assessment and improvement of data and information quality in a coordinated and global way. In this paper a framework to fill this gap is proposed. This framework is based on the Information Management Process (IMP) concept and it consists of two main components: an Information Quality Management Model structured in Maturity Levels (CALDEA) and an Assessment and Improvement Methodology (EVAMECAL). The methodology allows the assessment of an IMP in terms of maturity levels given by CALDEA, which can also be used as guidance for improvements. In the paper, a tool for automating the assessment and improvement is also briefly described.

KEY WORDS: Data and Information Quality, Data and Information Management, Information Management Process Assessment and Evaluation of Information Management Process.

1. INTRODUCTION

Currently, in this IT era, it would be true to state that information has already become a decisive factor of the information economy [16, 20, 47] since it is the basis for tactical, strategic or operational decisions [48, 54]. Many researchers and organizations have recognized this fact as the need to consider data and information as one of the most important assets [27]. Poor data and information quality will have a negative impact on the global efficiency of organizations [7, 40, 48]. It is a widely known fact that dealing with data and information problems is not a trivial issue, it is not free and many resources are required [61] because quality assurance is a complex process, in which the difference between costs and required quality is closely linked to the context of the application and organization requirements [6]. Much has been said about the growing importance of information quality, and many researching lines have appeared over the last decade dealing with different and specific data and information quality issues from different standpoints [16, 58], although it is still missing of well-founded and practical approaches to assess or even guarantee a required degree of information quality in an integrative way [20] because

information quality goes beyond the definition of data quality dimensions [61]. Thus, organizations need an integrative framework, which allows the assessment and improvement of information quality by coordinating and organizing all resources (including humans, software, hardware, economics and data) in such a way, that, knowledge about the company and the “*information manufacturing processes*” [56] can guide them through information quality efforts. Many of these efforts begin with an assessment of the current of data/information quality and carry on with optimization procedures. In spite of the fact that some researches have provided several information quality measurements and/or assessment methodologies [13, 15, 27, 42, 45, 48, 59, 60], none of them are focused on how to optimize information quality by coordinating group efforts or commitments extended to the entire organization in both analytic and pragmatic ways [16]. The reason is that data and information quality have not normally been understood as organizational issues, which can cover several departments, each one with its human resources and software applications.

In order to draw up the mentioned integrative framework, we looked for several characteristics that it must possess. In [16], the following four conditions for a good information quality framework are established:

1. It should provide a systematic and concise set of criteria according to which information can be evaluated.
2. It should provide a scheme to analyze and solve information quality problems.
3. It should provide the basis for information quality measurement and proactive management.
4. It should provide the research community with a conceptual map that can be used to structure a variety of approaches, theories, and information quality related phenomena.

Our aim was to develop a universal (valid for any kind of organization) and integrative (covering the entire company and letting researchers address their future works) framework satisfying these conditions. The proposed framework is intended to aid organizations to achieve their information quality goals through management. We stated the basis of our proposal by following two principles: first, considering information as a product [56] (which allows us to look at information from an engineering point of view [4, 39] as in other typical quality manufacturing foundations [10, 11, 38]) and secondly, taking into account the Software Process definition given by [19] (which allows us to identify who is doing what, when, using which resources and how, for managing the “non-ideal” characteristics of data and information). These two principles let us define the concept of **Information Management Process (IMP)** as a view of both Information Manufacturing and Information Quality Management processes. Therefore, an IMP is intended to model how information and information quality is managed for a specific application. In this way, information quality is going to be managed by assessing and improving a specific IMP. Although several frameworks for assessing and improving software processes already exists such as CMM [28, 46], CMMI [50], ISO 9001 [9], BootStrap [3] and SPICE [35], none of them have focused on information quality nor even take it into account as an important issue. The next two steps were to establish what conditions an IMP must possess to assure several data and information quality goals in different contexts and how to achieve these goals. On one hand, we found out that by separating information goals in several levels (as in CMMI's) it made it possible and easier in an incremental way to achieve partial data and information quality objectives through management. On the other hand, it was necessary to establish the relationship between assessment and improvement. As a result, our proposal defines two main components:

- An information quality management model based on maturity staged levels, known as **CALDEA**, in the style of CMMI and SPICE, where required and structured Key Processes Areas (KPA) to satisfy for each level are described. These KPAs are focused on management and technical issues. For each KPA, some tools, techniques, standards, and practices and metrics as required, are proposed, but not imposed trying to make a universal framework. Thus, each organization can use the ones that best fit to its needs.

- An assessment and improvement methodology, known as **EVAMECAL**, in the style of CBA-IPI [12], SCAMPI [53] or SPICE [36] which consists of a set of steps that provides a basis for data/information quality measurement and achieves improvement through proactive management.

This structure presents an important benefit for organizations, which know the previously mentioned standard: they can use their experience when optimizing their IMP, since the framework is aligned to them.

The remainder of this paper is structured as follows: In section 2, the Information Quality Model based on Maturity Levels is briefly summarized. In section 3, the Assessment and improvement information Quality Methodology is described. In section 4, a tool for automating part of the methodology is presented. In section 5, some conclusions are outlined and several future researching lines are detailed.

2. CALDEA: A INFORMATION QUALITY MANAGEMENT MODEL

CALDEA defines five information quality management maturity levels for an IMP: Initial, Definition, Integration, Quantitative Management and Optimizing. The levels are ordered by taking into account several information quality goals and their relative importance, providing a systematic and concise set of criteria according to which information can be assessed [16]. Thus, at higher levels where more information quality issues are assured, it possible to state that more organizational requirements are satisfied. It also possible to affirm that the higher the information quality maturity levels an organization has reached for its most important IMPs, the more competitive this organization can be due to the lack of information quality problems. The levels are established in the same way as the staged ones of CMMI, because it appears to be easier to work with a well-defined sequence of improvements (which cover from basic management project fundamentals to complex data quality management issues). As previously mentioned, for each level, CALDEA addresses specific KPAs, which meet specific information quality goals. These KPAs are focused on not only technical but also managerial issues, providing the basis for information quality measurement and management and integrating both aspects in order to compensate the lack of integrative frameworks mentioned in the introduction. Each KPA has been broken down into activities and tasks, which can be satisfied by using several techniques, practices and tools. We should point out that the chosen KPAs, and their activities and tasks are based on both CMMI's KPAs [50] and learned lesson as a result of our experiences in industrial and scientific initiatives regarding information quality. Although, the contents of this research paper are in continuous revision in order to achieve a theoretical validation, due to paper length restrictions, neither of the proposed techniques nor tools can be detailed.

2.1 Initial Level

An IMP is said to be at Initial Level when no efforts are made in order to achieve any information quality goals.

2.2 Definition Level

An IMP is said to be at Definition Level or Defined when it has been defined and planned. This implies to identifying all its components and their relationship. To achieve this goal, the following KPAs need to be satisfied:

- **(IQATM) Information Quality Assurance Team Management.** Data and information management quality initiatives required people having direct responsibility for them and for their integrity to support all of the activities that must be performed. These people must work in accordance with the organization's ideas and trends and must encourage the entire organization to show commitment to

information quality policies [2], by making corresponding efforts in support of the activities of this maturity model. Among their abilities must be data and information quality and managerial ones. [49] points out the need for high managers to lead data and information quality initiatives. This implies selecting people to take care of data quality through IMP, and support activities related to it, like standardization and measurements. Any techniques or tools related to human resources management may be used.

- **(IPM) IMP Project Management.** The main goal of this KPA is to create a plan for coordinating efforts and draw up a document, which clearly describes an agenda of activities and a budget for optimizing the IMP. This document can be made by following [31] for instance. Amongst other planned activities, the following must be carried out: a data and information requirements management, an analysis of these requirements, the design of a solution for satisfying them, the implementation of the process based on previous design, and testing of the implemented process. Any techniques or tools used on project development may be used here, for instance, PERT or CPM.
- **(URM) User Requirements Management.** User Requirements must be collected and documented. Three kinds of requirements might be identified [57]: those related to the final product (URS), those related to IMP – which must be gathered in the User Requirement Specification for IMP document (URS-IMP) - and those related to Information Quality –which must be gathered in the Information Quality User Requirements Specification (URS-IQ)-. These requirements are the starting point for modelling the IMP, the database or data warehouse and other procedures. Some tools and techniques can help developers to get each document [33]. There are some known graphical techniques for these issues. For example, for IMP, IP-MAP, given by [51]; for database or data warehouse and data quality issues, the extended entity-relationship model proposed by [60].
- **(DSTM) 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 from URS-IMP in order to avoid problems like uncontrolled data redundancy or problems with data format interchange [43]. [1, 25] discuss these issues and suggest several ways of treating information from multiple sources. In a data warehouse environment, tools and techniques like ETLs ones must be used in order to unify the semantics and formats of incoming data [13].
- **(AIMPM) Database or Data Warehouse Acquisition, Development or Maintenance Project Management.** Raw data must be collected and stored in an appropriate database or a data warehouse. In order to better assure information quality, it is highly recommendable to draw up a project for acquisition, development or maintenance of a database or a data warehouse management system, supporting both URS-IQ and URS-IMP. This KPA may also include other minor sub activities like Data Quality Assurance [37], Configuration Management, Maintenance Management or Commercial Solution Election Management.
- **(IQM) Information Quality Management in IMP Components.** The use of metrics for measuring IMP efficiency may help to improve it. It is therefore necessary to identify from the URS-IQ the dimensions of quality of information (in the same way as [34] proposes for software) for each information quality component that must be controlled [26, 27], as well as the metrics adapted for each one of those dimensions [15, 39, 47]. A description of data quality dimension and a discussion about which are the most important can be found in the works of the majority of the consulted authors. In order to make our proposal as universal as possible no dimensions are mandatory, as this is not possible [47] due to the fact that data and information quality depend directly on data/information problems. However, as guidance, the adopted as standard data quality dimensions set proposed by [54] is recommended to information quality managers who are encouraged to find out the one that best fit to their problems. For helping them, generic methodologies might be used like [32] or GQM, given by [55]. Even authors like [43] have proposed a more specific data and information quality measurement framework with several and specific 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, 5, 8] have proposed metrics for

measuring specific issues of specific components of IMP. An important aspect of measurements is the need to automate measurements, as required by [24].

2.3 Integration Level

An IMP is said to be at Integration Level or Integrated when besides being Defined (Definition Level has been achieved), many efforts are made in order to assure that the IMP complies with organizational information quality requirements, standards and policies. This implies standardizing different information quality learned lessons through information quality standards and policies in order to avoid previous errors and to allow to do better work in the future. The following KPAs must be satisfied:

- **(VV) Information Products and IMP Components Validation and Verification.** Both information products and IMP components must be verified and validated to correct defects and/or discordances with the USR-IMP, USR-IQ and the organizational information quality policies. A technique that can be used could be software inspections [17, 22] but adapted to data/information quality issues. A more specific methodology, which can be used, is *data testing* proposed by [41], expanded to IMP, 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, [30].
- **(RM) Risk and Poor Information Quality Impact Management.** It is necessary to determine the impact of risks due to the poor quality of information in the IMP in order to limit them at organizational level [13]. [21] proposes a methodology, which can be adapted to information quality issues in order to collect and document all risks.
- **(IQSM) Information Quality Standardization Management.** All lessons learned through specific experiences should be properly gathered, documented and transmitted to the organizational knowledge base. An example of standardization could be the sixth process of TDQM [14], which seeks to accomplish these issues “*by integrating quality management beliefs, principles, and methods into the culture of the enterprise*”. Only by incorporating the latest data and information quality management experiences, the IMP performance will be higher than it would be otherwise.
- **(OIQPM) Organizational Information Quality Policies Management.** A way to implement all the efforts previously mentioned, consists of defining information quality policies based on the previously defined standards affecting not only single IMPs, but also the whole organization. The Information Quality Management Team must work on data and information quality policies, which reflect organizational culture. [43] presents the elements that are the subject of data/information policy design. Organizations can be said to have an information quality culture when all their processes, related or not to information and information quality management, take into account data quality issues in order to improve it.

2.4 Quantitative Management

An IMP is said to be at a Quantitative Management Level or Quantitatively Managed when being integrated (Integration level has been achieved) several Measurement Plans have been developed and implemented and measurement procedures have been automated. Therefore, the main information quality goal of this level is to obtain an automated quantitative compliance that IMP performance over a reasonable time period, remains as consistent as required in terms of variation and stability through a reliable set of measurements [18] of information quality dimensions of IMP. This level is composed of the following KPA:

- **(MM) IMP Measurement Management.** The aim of this KPA is to get some metrics that must be used to check conformity to specifications [23, 43]. As [44] states, a plan for software quality measurements starts with the decision to take measures. This implies choosing “*what*”, “*when*” and “*how*” to measure and how to represent these measures and to “*whom*”. Since metrics about the IMP have been drawn up at Definition Level (“*what*” question), the plan must focus on the remaining

questions. As regards the “*when*” question, the answer is when measurements do not alter the IMP results. As far as the “*how*” question is concerned, some algorithms might be outlined in order to make measurements repeatable. Finally, as important as the metrics is the way in which the results are represented and to “*whom*”. Many authors such as [13, 43] propose the use of several control diagrams as a way of representing data about IMP. Another complementary way of representing the results is that proposed by [29], in which Kiviat’s diagrams are used to relate several aspects of the IMP or the data/information quality components.

- **(AMP) IMP Measurement Plan Automation Management.** In order to increase the reliability and repeatability of measures, measurement procedures and algorithms (defined at MM KPA) must be automated as required by [24]. This KPA aims to study all the issues in relation to the automation of these management procedures.

2.5. Optimizing Level

An IMP is said to be at Optimizing Level when being quantitatively managed the obtained measurements are used to develop a continuous improvement process by eliminating defects or by proposing and implementing several improvements. The following two KPAs must be satisfied:

- **(CADPM) Causal Analysis for Defects Prevention Management.** From the study of the measurements results, some typical quality techniques and tools like Statistical Control Process (SPC) can be applied to detect defects of information quality and identify their root causes. The conclusions obtained must provide a basis for a corresponding maintenance process for removing detected defects in affected resources. [52] offers a framework for defect prevention.
- **(IODM) Innovation and Organizational Development Management.** This is the basis for the concept of continuous improvement. Similar to previous KPA, here, the results can be used to improve the IMP in terms of higher performance, more efficient planned time or lower budget. Learned lessons in IMP must provide a basis not only for defect prevention, but also for continuous improvements

3. EVAMECAL: A METHODOLOGY FOR ASSESSING AND IMPROVING INFORMATION QUALITY MANAGEMENT.

3.1. Previous Considerations.

Before beginning the presentation of the methodology it is useful to explain the terms in which the assessment is going to be carried out and the improvement which we aim to plan and achieve. Different states for items (Maturity Levels, KPAs, Activities and IMP’s components) are going to be classified:

- Each Maturity Level can be in one of these states: {“*Achieved*”, “*Not Achieved*”}
- Each KPA can be in one of these following states: {“*Fully Satisfied*”, “*Satisfied*”, “*Partially Satisfied*”, “*Not Satisfied*”}.
- Each Activity in each KPA can be in one of these states: {“*Fully Executed*”, “*Executed*”, “*Partially Executed*”, “*Not Executed*”}
- Each component can be in one of these states: {“*Fully Optimized*”, “*Optimized*”, “*Partially Optimized*”, “*Not Optimized*”}.

Now the question is to define how an item can achieve a certain state. The first thought was that each organization should define when a certain state is achieved: so, for instance, an organization can define that a KPA is “*Fully Satisfied*” when all activities contained in the KPA are at last “*Executed*”; or it is “*Satisfied*” when, for example, at least eight out of ten activities are “*Executed*”; and so on. However this

solution has two main problems: firstly, it may not reflect the degree of importance that each items has for the information quality; and secondly there will be no way comparing organizations and by so doing establish an Information Quality Benchmark.

In order to solve the first problem and with the possibility of automation of assessment and improvement in mind, we decided to introduce the criticalness degree concept for each KPA in the Maturity Level, for each Activity in the KPA and for each Component in the Activity. It is important to highlight another advantage: they can also be used as guidance when improvement is required: the most critical item must

CRITICALNESS DEGREE	
Definition Level	
(IQATM) Information Quality Assurance Team Management	10 %
(IPM) IMP Project Management	15 %
(URM) User Requirements Management.	25 %
(DSTM) Data Sources and Data Targets Management.	10 %
(ADMPPM) Database or data warehouse Acquisition, development or maintenance Project Management	25 %
(IQM) Information Quality Management in IMP Components.	25%
Integration Level	
(VV) Information Products and IMP Components Validation and Verification.	25%
(RM) Risk and Poor Information Quality Impact Management	25%
(IQSM) Information Quality Standardization Management	25%
(OIQPM) Organizational Information Quality Policies Management	25%
Quantitative Management Level	
(MM) IMP Measurement Management	70 %
(AMP) IMP Measurement Plan Automation Management.	30 %
Optimizing Level	
(CADPM) Causal Analysis for Defects Prevention Management	50%
(IODM) Innovation and Organizational Development Management.	50%

Table 1. Criticalness Degrees

This represents part of our current researching line: our objective is to get a theoretical and empirically validation for these thresholds for criticalness degrees based on the demands of the different organizations and to set finely the ranges for quantification.

3.2. The Methodology.

As stated, the main aim of EVAMECAL is to assess and improve a specific IMP of a given organization. It was originally based on Deming's PDCA, but it has been restructured by following the Define-Measure-Analyze-Improve, and added a new step of Standardize the learned lessons through the most recent experiences.

The definition of goals, the measurement processes, the analysis criteria and the improvement plans are made in terms of information quality maturity levels given by CALDEA. Very briefly summarized, EVAMECAL could be enunciated as follows:

1. **Measure the current state of maturity level.** The main goal of this step is to determine which is the

be optimized the first. The next step was to define the Information Quality Value (IQV) for an item as the weighted average of all components of this item taking into account the criticalness degree. Thus, an Activity – Information Quality Value (A-IQV) can be defined as the weighted average of the values of data/information quality dimension of the IMP component, taking into account the criticalness degree of each component; An KPA-IQV can be defined as the weighted average of all A-IQVs for this KPA; And a Maturity Level –IQV (ML-IQV) can be defined as the weighted average of all KPA-IQVs for this level. This IQV can also allow to define ranges for states.

We proposed, taking into account our real experiences, several values for each criticalness degree and limits for the ranges of states, which are a hypothesis according to the supposed degree of importance. Table 1 shows the proposed values for each KPA's criticalness degree. On the other hand, and having four states for components, activities and KPAs, we have established (as at [36]'s style), that if $0 \leq IQV \leq 20$, the state is “*Not Optimized/Executed/Satisfied*” if $21 \leq IQV \leq 60$, the state is “*Partially Optimized/Executed/Satisfied*”; if $61 \leq IQV \leq 90$, the state is “*Optimized/Executed/Satisfied*”, and finally if $91 \leq IQV \leq 100$ “*Fully Optimized/Executed/Satisfied*”; for Maturity Level if $0 \leq ML-IQV \leq 90$ the state is “*Not Achieved*”, and if $91 \leq ML-IQV \leq 100$ and all KPAs are at least at a “*Satisfied*” state, the level is said to be “*Achieved*”.

current state of the IMP in terms of information quality maturity levels and IQVs. To do this, a set of surveys must be conducted, and if some metrics have been defined, some measures must be taken in relation to the specific IMP components. This implies the next substeps:

- a. **Conduct the proposed surveys**, in order to get some information about the IMP, a set of questionnaires have been proposed. In the next subsection a description of these questionnaires can be found.
- b. **Measure the chosen information quality dimensions for each component**, in order to calculate the corresponding A-IQV.
2. **Define goals in terms of information quality maturity level for improvement**. Any information quality initiative must define a plan covering the goals to be improved. In order to define a plan, EVAMECAL defines a rules system, which allows the activities and KPAs that must first be improved to be identified. This rule system is not going to be detailed here due to paper length restriction.
3. **Develop a plan for improvements taking into account the required goals**. A plan establishes what actions must be performed to obtain a set of improvements for the IMP. It also defines when and how must execute these actions, by whom, and what resources are implied. A plan can be defined only to obtain improvements to a simple activity or to reach higher levels.
4. **Check the efficiency of the plan**. In order to empirically validate the success of the plan, a set of tests must be executed again. This implies measuring the current state of the maturity level again, and check if information quality goals have been achieved.
5. **Standardize the learned lesson in order to avoid future problems**. The Information Management Team must standardize the knowledge acquired through experiences, in order to avoid future problems and get better results.

3.3. The surveys.

For the assessment process a set of surveys has been drawn up. This set consists of four different classes of questionnaires, with different goals:

- Delimit and characterize the organization, a total of fifteen questions.
- Delimit and characterize the IMP to be assessed, a total of six questions.
- Assess the degree of achievement of each maturity level: several questions organized in different and selective blocks have been developed. The questions are focused on the KPAs, activities, tasks, proposed techniques, tools and practices and required developed products. The idea of organizing the questionnaire in several depth levels is so that questions will be asked from top going down only if

DEPTH LEVEL	1	2	3
Level Maturity 2	12	18	82
Level Maturity 3	8	7	24
Level Maturity 4	3	2	11
Level Maturity 5	4	2	11
Total Questions per depth level	28	31	131
Total Questions Questionnaire	190		

Table 2. Number of questions by maturity and depth levels.

necessary. Thus, the block of questions of the first depth level serves to evaluate if KPAs are satisfied or not, avoiding at this depth level other questions which are not important for establishing specific aspects about the accomplishment of the more specific issues, which are dealt with in lower depth levels. Thus, if all the answers to the questions of the first depth level

differ from “*Not Satisfied*”, then questions in the level immediately below should be answered, and so on. Altogether one hundred and ninety questions would be answered in the case of all the answers to the questions of the first and the second depth level differing from “*Not Satisfied*”. The answers to these questions must be a number between 0 and 100, in order to calculate the corresponding IQV for each component, activity, KPA and Maturity Level. We should point out that we are working on checking the validity and efficiency of each block and question inside the blocks for each depth level. Due to length restrictions none of the questions are included in this paper.

- Finally, collate and compare the answers with previous questionnaires, there is a last block of questions with descriptive and textual language.

4. EVATOOL: A TOOL TO AUTOMATE EVAMECAL.

In order to facilitate the process of assessment and improvement, a tool is being developed. It consists of the following five main components:

- EVATOOL-QC (*EVATOOL – Questionnaire Conductor*) whose main aim is to automate the making of questionnaires,
- EVATOOL-M, (*EVATOOL- Measurer*) whose main aim is to automate the measurements,
- EVATOOL-D (*EVATOOL-Drawer*), whose main aim is to draw the entire IMP by following IP-MAP
- EVATOOL-IA (*EVATOOL-Improvement Advisor*), whose main aim is to give advice in relation to the improvements: it implements the rule systems by taking into account the present state of the IMP and with regards to criticalness degree, it recommends the points to be improved.
- EVATOOL-P (*EVATOOL-Planner*) whose main aim is to develop a plan for improvement when a set of them is required.

Next, a brief summary of EVATOOL foundations is going to be presented.

4.1. Design Principles of EVATOOL

EVATOOL is designed to support the user in applying the EVAMECAL methodology throughout the whole proposed process, from the initial stages to the last ones, and is aimed at the users not concerned with the assessment (but perhaps the most important source of information) to the top-level managers involved in taking the enterprise to the highest levels of quality.

EVATOOL has been built on the following foundations:

- Flexibility: In the widest sense, EVATOOL's design enables the manager to change as much as possible because of
 - DBMS independence: EVATOOL only needs a classical SQL engine to work due to the exclusive use of standard SQL features. Only relations, primary/foreign keys and simple data types (integers and strings) are needed in a SQL system to adapt EVATOOL to run. Currently, EVATOOL has been extensively tested on MySQL and runs without problems on MS-Access.
 - DBMS abstraction layer: If an assessment/company/user needs to store data in whatever way, there is a DBMS abstraction layer that isolates the core of the application.
 - Easy to install/deploy. In the simplest case, only a web browser is needed to use EVATOOL, without loss of either aesthetical or practical features.
- Easy to use: Almost no knowledge is required to operate it. It works for two types of users and in both cases the tool remains easy to use. For managers, there is a single administration tool that enables them to create new surveys, delete or modify them. This tool is graphically operated and resembles to the classical DB administration tools. For simple users there are questions grouped in stages, which the user must respond to for the manager to gain knowledge about quality in the Information Management Process developed in the firm.
- Easy to maintain/update. New surveys/features/stages can be added to questionnaires. This feature has been considered as one of the most important, due to pretended universality of the proposed framework.
- Complete. EVATOOL must provide a complete support to each one of the steps of the process recommended by EVAMECAL either allowing data storing or data processing (the most common activities).

4.2. EVATOOL Architecture.

The architecture of EVATOOL is reflected in figure 1.

The tool adopts a classical three-tier architecture, proven to be a useful, practical, change-proof and flexible architecture pattern. In the lowest level, the DBMS Independence Module works together with a DBMS providing the rest of the application the operations needed to store and process data. As explained above, the objective is to keep EVATOOL as flexible as possible, and this module provides the required independence. To keep the program easy to adapt, the relational model implemented in the DBMS is created with the most well-known features of SQL, and is retrieved by the DBMS Independence Module through ODBC, the current “*de-facto*” standard to communicate SQL engines with applications.

The DBMS independence module operates with internal data structures representing data, tables and relationships to keep the program easy to adapt and maintain. The structures represent the classical SQL features already commented: Data, tables and relationships. This module is the proxy between the core application and the final data stored in disk.

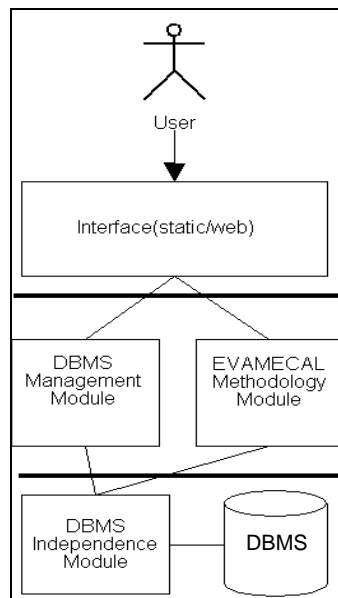


Figure 1. EVATOOL Architecture

The intermediate layer must accommodate two types of operations. User-related or manager-related actions. Thus, the application has been divided to isolate unrelated implementation of actions. The user-related module, in brief, allows EVATOOL to guide the user along EVAMECAL steps, process data provided, and give feedback. The manager-related module enables directors to examine data provided by users, retrieve calculations performed by EVATOOL and update/modify/delete surveys.

These two intermediate modules are separate and must remain so, due to implementation and logical aspects. Between implementation aspects, separation allows low coupling and thus maintainability. Between logical aspects we can say that user- and manager-operations are not related and therefore, these modules have no reason to live in the same abstraction.

The highest level provides the interface for the final user. It has been designed to be identical in both (stand-on and web) versions to keep EVATOOL easy to use. It is also used in the same way in both versions but early on, the differences between the static and web version, like performance (static version) vs. adaptability (web version), mark certain trade-offs for managers.

4.3. Using EVATOOL

Ease of use, as said above, is one of the design principles on which the tool has been designed. EVATOOL guides the user through the different surveys asking him to respond, and offering help by means of pop-up messages that explain the questions and help him to offer better answers.

For managers, EVATOOL has different possibilities:

- When conducting assessment, managers can vary levels of acceptance in the different KPA's explained in EVAMECAL; vary questions and order of questions (sometimes certain questions will only be asked if former questions have been answered positively or negatively). In Figure 2, a window allowing directors to modify thresholds of acceptance for different KPA's (the tool doesn't allow errors such as values under 0% or over 100%)
- When updating the tool, new questionnaires can be added or deleted, as well as adding, modifying or deleting questions from questionnaires. Figure 4 shows the usability of the program, which, as said above, is one of the foundation principles for EVATOOL. As can be seen the dialogue is easy to follow and it is clear what the user has to do.

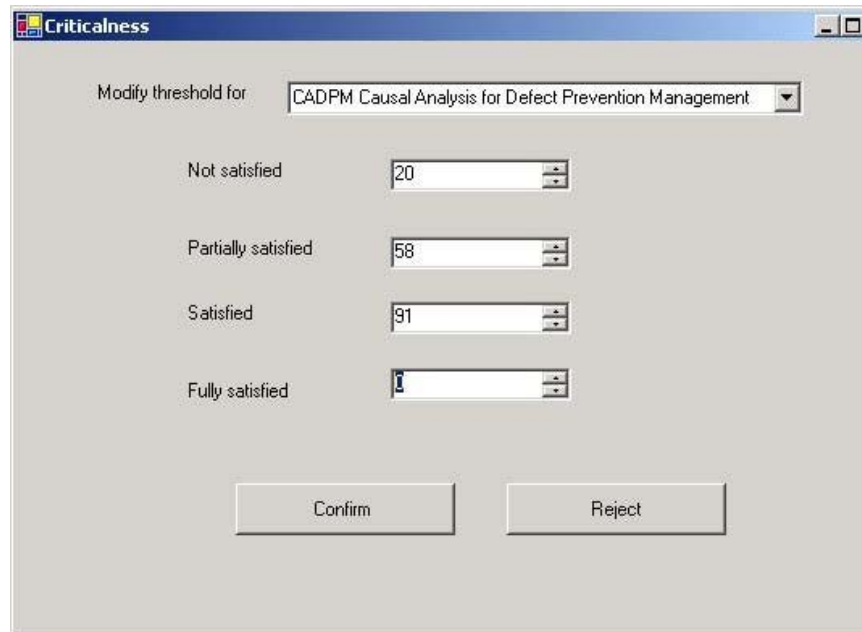


Figure 2: Modifying thresholds

- When processing collected data, managers can update/modify criticalness degrees for the different KPA's to support the required universality of the framework (see Figure 3). This allows managers to focus their efforts on evaluation without having to recalculate. This feature means that EVATOOL clearly resembles a spreadsheet, making it easier to use and to understand.

4.4. Future developments

- Port the stand-on version to C++ with WxWidgets (formerly WxWindows) to allow use in not-connected computers not using Microsoft Windows. Currently, EVATOOL has been implemented using C# on Windows platforms, which makes it a not-so-easy tool to port. (However, in computers connected to a network, this problems doesn't exist, as EVATOOL can also be run too as a Web Service, requiring only a standard web browser to use it)
- Make EVATOOL visually attractive by following EVATOOL appearance. All the previous efforts have focused on functionality rather than aesthetics.

- Test results thoroughly and use them to improve EVAMECAL.
- Add on-line help.
- Incorporate security mechanisms to allow different levels in hierarchy to see none, some or all of the data calculated, processed or stored.
- Full translation of questionnaires into English (currently, only user interfaces and the tool messages have been translated from Spanish)

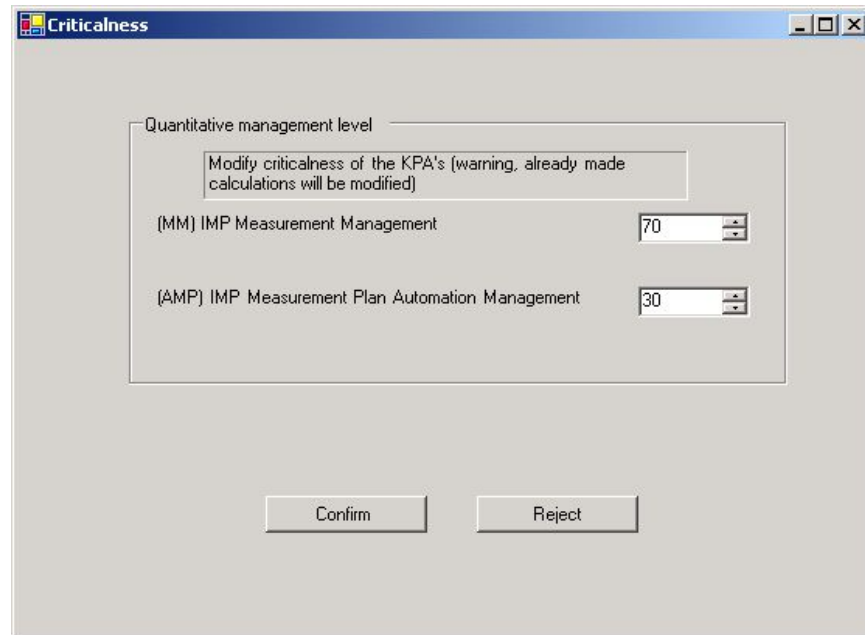


Figure 3: Modifying criticalness in QM level

5. CONCLUSIONS AND FUTURE WORK.

In this paper a framework for assessing and improving information quality management has been presented. It consists of two main components: an Information Quality Model based on Maturity Level (CALDEA) and a methodology for assessment and improvement (EVAMECAL). This structure satisfies the four conditions previously mentioned [16]: As CALDEA is structured in maturity levels with KPAs, a systematic and concise set of criteria for information quality assessment is provided, thus satisfying the first condition. By defining KPAs for each level, some of them focused on management issues, the basis for proactive management and measurements is provided, thus satisfying the third condition. And finally, by being structured in such staged levels describing KPAs and proposing (not imposing) tools, techniques, standards, ... a conceptual map has been provided for the research community in order to address a variety of approaches, theories and information quality related phenomena, thus satisfying the fourth condition. On the other hand, EVAMECAL provides a schema to analyze and solve information quality problems, thereby satisfying the second condition and allowing researchers to address their future works. To sum up we can say that the main idea of the framework is to use EVAMECAL for assessing and improving an IMP using CALDEA's levels as reference in the guidance of the optimization of the information quality in organizations.

For the future, we have planned several projects: firstly, to finish our current researching lines to refine CALDEA and EVAMECAL by applying them to different kinds of organizations (we have already applied the framework to some organizations with more or less success); Secondly, we aim to test and improve the efficiency of EVATOOL; And finally, we want to study the usefulness of current techniques,

tools and practices or developing when not exists, in order to obtain a collection of recommended ones for each context.

6. ACKNOWLEDGEMENTS

This research is part of both CALIPO- supported by Dirección General de Investigación of the Ministerio de Ciencia y Tecnología (TIC2003-07804-C05-03)- and MESSENGER project - supported by Consejería de Ciencia y Tecnología de la Junta de Comunidades de Castilla-La Mancha (PCC-03-003-1).

7. REFERENCES

- [1] Ballou D., Wang R., Pazer, H., and Tayi, G.K. Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science* 44(4). 1998. Pp. 462-484,
- [2] Ballou, D. and Tayi, G.K. Enhancing data quality in Data Warehouse Environments. *Communications of the ACM*, Vol 42, No 1. January 1999.
- [3] Bicego, A., and Kuvaja, D. Bootstrap, *Europe's Assessment Method*, IEEE Software, 1993, Pp. 93-95.
- [4] Bobrowski, M., Marré, M., and Yankelevich, D. A Software Engineering View of Data Quality. *Proceedings of Second International Software Quality in Europe*. Belgium. November 1998.
- [5] Bouzeghoub, M. and Kedad, Z. 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.
- [6] Bringel, H., Caetano, A., Tribolet, J. Business Process Modeling Towards Data Quality Assurance. *Proceeding of ICEIS'2004*. Pp 565-568. Porto, Portugal. 2004
- [7] Burgess, M.S.E., Gray, W.A., Fiddian, N.J. A flexible quality framework for use within information retrieval *Proceedings of the Eighth International Conference on Information Quality (ICIQ'2003)*. 2003. Pp 297-313.
- [8] Calero, C. and Piattini, M. Metrics for databases: a way to assure the quality. In: *Information and Database Quality*. Kluwer Academic Publishers. 2002. Pp. 57-84.
- [9] Coallier, F. How ISO 9001 fits into the software world, *IEEE Software*.. January 1994. Pp 98 –100
- [10] Crosby, P.B. *Quality is free*. New York, New York: McGraw Hill, 1979
- [11] Deming, W.E. *Out of the crisis*. Cambridge, MA. MIT Center for advanced engineering study, 1986
- [12] Dunaway, D. K. *CMM SM -Based Appraisal for Internal Process Improvement. (CBA IPI) Lead Assessor's Guide (CMU/SEI-96-HB-003)*. Software Engineering Institute, Carnegie Mellon University, Pittsburgh, 1996.
- [13] English, L.P. *Improving Data Warehouse and Business Information Quality: Methods for reducing costs and increasing Profits*. Willey & Sons, 1999.
- [14] English, L.P. Total Quality data Management (TQdM), Methodology for Information Quality Improvement. In: *Information and Database Quality*. Kluwer Academic Publishers. 2002. Pp. 85-110.
- [15] Eppler, M. J. *Managing Information Quality*. Springer. Germany. 2003.
- [16] Eppler, M. J. and Wittig, D. Conceptualizing Information Quality: A review of Information Quality Frameworks from the last ten years." *Proceedings of the 2000 Conference on Information Quality*. Pp 83-96
- [17] Fagan, M. Design and Code Inspections to Reduce Errors in Program Development. *IBM Systems Journal* 15, 3. 1976. Pp 182-211.
- [18] Florac, W, A.and Carleton, A.D. Using Statistical Process Control to Measure Software Process. In *Fundamental Concepts for the Software Quality Engineer*. Taz Daughtrey Editor. American Society for Quality. 2002. Pp.133-144.
- [19] Fuggeta, A. Software Process: A roadmap. The future of Software Engineering, ed. A. Finkelstein *ACM, Press*. 2000. Pp.27-34.
- [20] Gertz, M., Tamer, M., Saake, G. Sattler, K., Report on the Dagstuhl Seminar "Data Quality on the web". *Proceedings of SIGMOD record*. Vol.33, No.1. March 2004.

- [21] 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. Pp. 25-38, 2002
- [22] Gilb, T. and Graham D. *Software Inspection*. London: Addison-Wesley Longman. 1993
- [23] Grimmer, U., and Hinrichs, H. A methodological approach to data quality management supported by data mining. *Proceedings of the Sixth International conference on Information Quality*. Pp 217-232. 2001
- [24] Hinrichs, H. CLIQ- Intelligent Data Quality Management. *Proceedings of the fourth IEEE international Baltic Workshop on databases and Information System*. 2000
- [25] Hinrichs, H. and Aden, T. 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
- [26] Hoxmaier, J.A. Dimensions of Database Quality. In *Developing Quality Complex Database Systems: Practices, Techniques, and Technologies*. Editor Shirley Becker. Idea Group Publishing. 2001
- [27] Huang, K.T., Lee, Y., Wang, R. *Quality Information and Knowledge*. Prentice Hall, Upper Saddle River, 1999
- [28] Humphrey, W. *Managing the software process*, Addison – Wesley, Reading Mass. 1989
- [29] Humphrey, W. The Software Standard Profile. In *Fundamental Concepts for the Software Quality Engineer*. Taz Daughtrey Editor. American Society for Quality. 2002. Pp. 3-16.
- [30] IEEE STD 1012-1986 IEEE Standard for Software Verification and Validation Plans.
- [31] IEEE STD 1058.1-1987 IEEE Standard for Software Project Management Plans.
- [32] IEEE STD 1061-1992. IEEE Standard for a Software Quality Metrics Methodology.
- [33] IEEE STD 830-1998. IEEE Standard guide to Software Requirements Specification. IEEE New York (USA) IEEE Computer Society.
- [34] ISO/IEC 9126: Information technology – Software product evaluation – Quality characteristics and guidelines for their use, ISO/IEC Standard 9126, Geneva, 2001.
- [35] ISO IEC 15504 TR2: 1998, part 2: A reference model for processes and process capability, ISO/IEC JTC1/SC7, 1998.
- [36] ISO IEC 15504 TR2: 1998, Software Process Assessment – Part 7: Guide for use in process improvement, ISO/IEC JTC1/SC7, 1998.
- [37] 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
- [38] Juran, J.M. *Juran on Planning for Quality*. New York, New York: McMillan. 1988
- [39] Kahn, B., Strong, D., Wang, R. “Information Quality Benchmarks: Product and Service Performance”. *Communications of the ACM* April 2002/Vol. 45, No. 4
- [40] Kim, W., Choi, B., Towards Quantifying Data Quality Costs in *Journal of Object Technology*, Vol. 2, no.4., Pp. 69-76. July-August 2003
- [41] Kiszkurno, E., Yankelevich, D., Data Testing. ASSE 2001 Proceedings of SADIO. 2001
- [42] Lee, Y.W., Strong, D., Kahn, B., Wang, R. AIMQ: a methodology for information quality assessment. *Information & Management*, 2001
- [43] Loshin D. *Enterprises Knowledge Management: The Data Quality Approach*. Morgan Kauffman, San Francisco (California), 2001
- [44] Meredith, D.C. Managing with Metrics: Theory into Practice. In *Fundamental Concepts for the Software Quality Engineer*. Taz Daughtrey Editor. American Society for Quality. 2002. Pp145-154
- [45] Olson, J. E. *Data Quality: the accuracy dimension*. Ed. Morgan Kaufmann Publishers. 2003
- [46] Paulk, M., C. Weber, B. Curtis, and Chrissis, M. *The Capability Maturity Model Guideline for Improving the software Process*, Addison-Wesley, Reading, Mass. 1995
- [47] Pipino, L., Lee, Y., and Wang, R. Data Quality Assessment. *Communications of the ACM* April 2002/Vol. 45, No. 4
- [48] Redman, T.C. *Data Quality for the Information Age*. Artech House Publishers, Boston. 1996
- [49] Redman, T.C. *Data Quality: The Field Guide*. Digital Press. 2001
- [50] SEI. Capability Maturity Model® Integration (CMMISM), Version 1.1 CMMISM (CMMI-SE/SW/IPPD/SS, V1.1) Staged Representation CMU/SEI-2002-TR-012 ESC-TR-2002-012. In <http://www.sei.cmu.edu/publications/documents/02reports/02tr002.html> (last access on June 2004)
- [51] Shankaranarayanan, G. Wang, R., Ziad, M. IP-MAP: Representing the Manufacture of an Information Product. *Proceedings of the 2000 Conference on Information Quality*. 2000

- [52] Smith, C., and Heights, A. Defect Prevention: The road less Traveled. In *Fundamental Concepts for the Software Quality Engineer*. Taz Daughtrey Editor. American Society for Quality.2002
- [53] Standard CMMI SM Appraisal Method for Process Improvement (SCAMPI SM), Version1.1: Method Definition Document. CMU/SEI-2001-HB-001.
- [54] Strong, D.M., Lee, Y. W., Wang R.Y. Data Quality in context. *Communications of the ACM.*, pp 103-110. May 1997
- [55] Van Soligen, R. and Berghout, E. *The Goal/ Question/Metric Metodology: a practical guide for quality improvement of software development*. Ed McGraw-Hill. 1999
- [56] Wang, R., A product perspective on data quality management. *Communications of the ACM*. Vol 41(2) pp58-65. February 1998
- [57] Wang, R., Kon, H., and Madnick S. Data Quality Requirements Analisis and Modeling. Published in the *Ninth International Conference of Data Engineering* Vienna, Austria., Pp 670 – 677. 1993
- [58] Wang, R., Storey V.C., Firth, C. F. A framework for analysis of Data Quality Research. *IEEE Transactions on Knowledge and Data Engineering*, Vol 7(4). Pp 623-640. 1995
- [59] Wang, R., Strong, D., Guarascio, L.M. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, , Volume 12, No. 4. pp. 5-33. Spring 1996
- [60] Wang, R., Ziad, M., Lee, Y.W. *Data quality*. Kluwer Academic Publishers. Massachussets (USA). 2001
- [61] Xu, H., Would organization size matter for data quality. *Proceedings of the Eighth Conference on Information Quality ICIQ'2003*. Pp. 365-379. 2003.