# IMPROVING INFORMATION QUALITY MANAGEMENT USING CALDEA AND EVAMECAL

(Practice-Oriented Paper)

**Ismael Caballero, Angelica Caro, Mario Piattini, Coral Calero**
Alarcos Research Group, ESI (UCLM), Spain
{Ismael.Caballero, Mario.Piattini, Coral.Calero}@uclm.es, ancaro@inf-cr.uclm.es}

**Abstract**: The main aim of this paper is to show how a framework to assess and improve data and information quality management has been applied to an organization. The framework consists of two elements: a model for data and information quality management (CALDEA) and an assessment and improvement methodology (EVAMECAL) which uses CALDEA as a reference. An underlying concept to apply the framework is the Information Management Process (IMP), which allows us to model how data (raw material) is transformed into information (data products). The idea is to identify the IMP, and then, using CALDEA as the reference point, assess and improve the process by applying EVAMECAL.

**Key Words**: Data and Information Quality, Data and Information Management, Information Management Process Assessment and Evaluation of Information Management Process.

## 1. INTRODUCTION & BACKGROUND.

The problem which we posed as requiring a solution is the need of organizations to have an integrative framework that allows the assessing and improving of data and information quality. The goal is for better decisions to be taken [31 and 35]. These decisions directly affect both the success of the business [13, 17 and 30] and/or even the overall efficiency of organizations [4, 26 and 31] since data and information are to be seen as among the most important assets for organizations [21].

Although many researching lines covering different specific issues on data and information quality have been appearing over the last few years [13 and 38], there is still a lack of well-founded and practical approaches to assess or even guarantee in an integrative way a required degree of data and information quality [17] because information quality goes beyond the definition of data and information quality dimensions and metrics [40]. Our idea of "*integrative*" includes not only technical data and information quality goals, but also managerial ones which can guide the organization towards achieving the best data and information quality management.

The starting point was to consider that an Information System can be seen as a set of production processes, which are controlled by several quality management processes. These quality management processes can affect one or several production processes or even some of the elements that take part in the different processes. Moreover, these management processes can implement several quality and quality management policies. This initial consideration agrees with several researching lines which propose the treating of information as a product [36], where data is considered as the raw input material in the process of production ([1, 8, 21 and 27]), and information (data products) as the output of the production processes.

This analogy drawn between information and typical products enables organizations to apply the classic principles and foundations of product quality management to quality management of data and information products [14] from an engineering point of view. This is implicitly or explicitly suggested by several authors (such as [3]). It is precisely this analogy that leads us to think about a production process as part of the Information System. As said previously, this transformation / production process can be affected by one or several (quality) management processes.

For a data product to be able to become sound information, data (raw materials of this process) needs to be gathered, stored, consulted and presented/displayed according to certain rules of data and information quality, so that the production processes can generate data products that satisfy user requirements and needs. They have to be products where the workers handling those data have the sufficient training to do it consistently [19]. All these quality rules are part of the quality management processes. The quality management process could be defined as the set of activities and processes that an organization can develop in order to (1) design and establish several quality policies, (2) identify techniques and procedures to assure that the organizational resources for data and information are the right ones for the present and potential uses of the data [2], that they fit the use of a product or service [24], or look for conformity with the requirements established previously [7] so as to guarantee the customer requirements.

We need to draw up some reasoning to include and to connect this set of concepts and foundations. If the set mentioned is compared to the definition of Software Process given by [16], and considering that the quality of any product cannot simply be assured by inspecting the product or by making mere statistical controls, it might make us think that a good approach to data and information management in organizations could be close through the definition of the concept of the "*Information Management Software Process (IMP)*" which is intended to model how data and information and data and information quality is managed for a specific application. In this way, data and information quality is going to be managed by assessing and improving a specific IMP.

This hypothesis is supported by [29], who explains that all software processes are formed by two sub-processes: one of production itself and another one of management. This latter affects the production sub-process, controlling it so that the execution or of some of the activities of the sub-process of management does not directly affect the quality of the developed data and information product.
With this reasoning, and bearing in mind that there exist several assessment and improvement frameworks for software processes like CMMI [32] or ISO 15504 [22], our aim was to develop a framework adapted to data and information quality issues.
Taking into account this IMP concept as a software process, an Information system (IS) could be considered as a set of IMPs sharing several organizational resources. So, if an organization wants to improve its data and information quality level in an overall sense, an identification of IMPs and subsequent assessment and improvement must be done. To guide this activity two important elements should be provided.

- A data and information quality management model, known as **CALDEA**, along the lines of CMMI and SPICE, where the required and structured Key Processes Areas (KPA) to satisfy for each level are described. These KPAs focus on management and technical issues. For each KPA, some tools, techniques, standards, and practices and metrics as required are proposed, although each organization can use the ones that suits its needs best.
- An assessment and improvement methodology, known as **EVAMECAL**, along the lines of CBA-IPI [10], SCAMPI [34] or SPICE [23] which consists of a set of steps that provides a basis for data/information quality measurement and achieves improvement through proactive management

Looking at the variety of researching lines on different data and information quality issues, our aim was to develop a framework that is as universal as possible, understanding "*universal*" as meaning that it is valid for both any situation and any organization.

The remainder of the paper is structured as follows: Section two summarizes CALDEA; the third section gives a brief explanation of EVAMECAL. Our experience using the framework at CENPRI, a federal organization which uses data and information on a large scale for "*Tribute Management* from its contributors*"* is shown in the fourth section. Finally, in the fifth section, several conclusions are set out.

# 2. CALDEA: THE DATA AND INFORMATION QUALITY MANAGEMENT MODEL.

As said above, CALDEA is the reference model. The assessments are going to be done in terms of CALDEA and the improvements are going to be planned in those same terms.
For this reason, CALDEA was given a structure which can be easily comprehensible, usable and affordable for organizations. The need to develop elements which can be used alongside other well-known and widely-used software standards such as CMMI and ISO/IEC 15504 was also borne in mind. The thought structure which best fits these requirements was that of the CMMI maturity-staged level standard. So, 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 data and information quality goals and their relative importance, providing a systematic and concise set of criteria according to which information can be assessed [13]. For each level, CALDEA addresses specific KPAs, which meet specific information quality goals. These KPAs focus on not only on technical but also on managerial issues, providing the basis for information quality measurement and management [13]. 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, as well as their activities and tasks, are based on both CMMI's KPAs [32] and lessons learnt as a result of our experiences in industrial and scientific initiatives to do with data and information quality.
The levels and their KPAs are the follows:

- **Initial Level:** An IMP is said to be at Initial Level when no efforts are made to achieve any information quality goals.
- **Definition Level:** An IMP is said to be at Definition Level or Defined when it has been defined and planned. This implies identifying all its components and their inter-relationship. To achieve this goal, the following KPAs need to be satisfied:
    - ***(IQATM) Information Quality Assurance Team Management.*** It contains activities addressing the need to choose a team which can lead data and information quality management initiatives according to the organization's ideas and trends.
    - ***(IPM) IMP Project Management***. The main goal of this KPA is to create a plan for coordinating efforts and to draw up a document which clearly describes an agenda of activities and a budget for optimizing the IMP.
    - ***(URM) User Requirements Management***. The main aim of this KPA is to collate and to document the three kinds of User Requirements identified by [37]: 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)-.
    - ***(DSTM) Data Sources and Data Targets Management***. Due to particular intrinsic characteristics of data, we should identify and 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 [28].

- ***(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 assure information quality of a higher level, 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.
- ***(IQM) Information Quality Management in IMP Components***. The main goal of this KPA is to identify from the URS-IQ the data and information quality dimensions for each information quality component that must be controlled [20 and 21], as well as the metrics adapted for each one of those dimensions [6, 12, 25 and 30].

- **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 lessons learnt about information quality. This knowledge, derived from information quality standards and policies is used with the aim of avoiding previous errors and to allow us to do better work in the future. The following KPAs must be satisfied:
  - ***(VV) Information Products and IMP Components Validation and Verification***. The aim of this KPA is to get a validation and a verification of both information products and IMP components, to correct defects and/or discordances with the USR-IMP, USR-IQ and the organizational information quality policies.
  - ***(RM) Risk and Poor Information Quality Impact Management***. It is necessary to determine the impact of risks which originate from poor quality of information in the IMP so as to limit them at organizational level [11]. This KPA proposes several activities aimed at reaching this goal.
  - ***(IQSM) Information Quality Standardization Management***. All lessons learnt through specific experiences should be properly gathered, documented and transmitted to the organizational knowledge base. This KPA contains several activities whose objective is to guide organizations towards achieving this goal.
  - ***(OIQPM) Organizational Information Quality Policies Management***. A way to implement all the efforts mentioned above consists of defining information quality policies based on standards defined previously, 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.
- **Quantitative Management:** An IMP is said to be at a Quantitative Management Level or Quantitatively Managed when integration is complete (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 which demonstrates that IMP performance over a reasonable time period remains as consistent as is required. That is defined in terms of variation and stability through a reliable set of measurements [15] 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 [18 and 28]. Since metrics about the IMP have been drawn up at Definition Level, a plan must be made, focusing on how to get values for these metrics. This KPA contains several activities for achieving this goal.
  - ***(AMP) IMP Measurement Plan Automation Management***. To increase the reliability and repeatability of measures, measurement procedures and algorithms (defined at MM

KPA) must be automated as required by [19]. This KPA aims to study all the issues in relation to the automation of these management procedures.

- **Optimizing Level:** An IMP is said to be at *Optimizing* level when, when quantitatively managed, the measurements obtained are used to develop a continuous improvement process, eliminating defects or 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 such as 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 whose job it is to remove defects detected in affected resources.
    - ***(IODM) Innovation and Organizational Development Management.*** This is the basis for the concept of continuous improvement. As with previous KPA, in this case, the results can be used to improve the IMP in terms of higher performance, more efficient time-planning or a lower budget .Lessons Learnt in IMP must provide a basis not only for prevention of defects, but also for continuous improvements

Table 1 summarizes the KPAs defined for each level and the acronyms used. For a further discussion on the contents of each KPA, see [5].

| | ACRONYM | STANDS FOR | CD | | ACRONYM | STANDS FOR | CD | | ACRONYM | STANDS FOR | CD |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Definition Level (2) | **IQMTM** | *Information Quality Management Team Management.* | 10 % | Integration Level (3) | **VV** | *Information Products and IMP Components Validation and Verification.* | 25% | Quantitative Management Level (4) | **MM** | *IMP Measurement Management* | 70 % |
| | **URM** | *User Requirements Management* | 15 % | | **RM** | *Risk and Poor Information Quality Impact Management* | 25% | | **AMP** | *IMP Measurement Plan Automation Management.* | 30 % |
| | **IPM** | *IMP Project Management.* | 25 % | | **OIQPM** | *Organizational Information Quality Policies Management* | 25% | Optimizing Level (5) | **CADPM** | *Causal Analysis for Defects Prevention Management* | 50 % |
| | **DIQM** | *Data and Information Quality Management in IMP Components* | 10 % | | **IQSM** | *Information Quality Standarization Management* | 25% | | **IODM** | *Innovation and Organizational Development Management* | 50 % |
| | **DSTM** | *Data Sources and Data Targets Management* | 25 % | | | | | | | | |
| | **AIMPM** | *Database or Data Warehouse Acquisition, Development and Maintenance Project* | 25% | | | | | | | | |

***Table 1.*** *KPAs, their acronyms and their Criticalness Degree (CD) for each maturity level.*

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

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. As well as this adaptation, a new step has been added. This is to standardize the lessons learnt through the most recent experiences.

The definition of goals, the measurement processes, the analysis criteria and the improvement plans are carried out in terms of information quality maturity levels given by CALDEA. As each maturity level groups several KPAs, each KPA groups several activities, and each activity take into account several components (any of the hardware or software elements or even persons belonging or not to the Information System which takes a place in the IMP), the assessment process has the task of generating a state for each one. Table 2 gathers the mentioned states.

| ITEMS | POSSIBLE STATES |
|---|---|
| **Maturity Level** | {"*Achieved*", "*Not Achieved*"} |
| **KPA** | {"*Fully Satisfied*", "*Satisfied*", "*Partially Satisfied*", "*Not Satisfied*"} |
| **Activity in each KPA** | {"*Fully Executed*", "*Executed*", "*Partially Executed*", "*Not Executed*"} |
| **Component** | {"*Fully Optimized*", "*Optimized*", "*Partially Optimized*", "*Not Optimized*"} |

***Table 2.*** *Possible states for each item.*

These states for Maturity Levels are calculated by using several questionnaires and several Information Quality Values (IQVs) for each item; the IQVs are calculated by making a weighted averaged of the values of several metrics and/or responses to questionnaires. The weight for each IQV depends on the relative importance of the studied item in relation to a major item. So, for a Maturity Level having several KPAs, each KPA has a relative importance (criticalness degree). Let us call *ML-IQV* to the IQV for a Maturity Level, *KPA-IQV* to the IQV for each KPA and *CD* to the Criticalness Degree for each KPA. To calculate the ML-IQV for a given Maturity Level, the formula in figure 1 must be applied (it is similar to the activities for each KPA and for components in each activity):

$$ML - IQV = \sum_{i=1}^{n} KPA - IQV_i * CD_i$$

***Figure 1.*** *Formula to calculate ML-IQV.*

Proposed values for the Criticalness Degree for each KPA are shown in table 2, in the column tabulated as "*CD*".

There is equivalence between the values of IQV (which are compatible to [35]) and the corresponding states defined for each item (see table 2). This equivalence is summarized in table 4:

| VALUE RANGE FOR IQV | | ITEM STATE |
|---|---|---|
| **IQV FOR COMPONENTS, ACTIVITIES, AND KPAS** | 0 ≤ IQV ≤ 20 | "*Not Optimized/Executed/Satisfied*" |
| | 21 ≤ IQV ≤ 60 | "*Partially Optimized/Executed/Satisfied*" |
| | 61 ≤ IQV ≤ 85 | "*Optimized/Executed/Satisfied*" |
| | 86 ≤ IQV ≤ 100 | "*Fully Optimized/Executed/Satisfied*" |
| **IQV FOR MATURITY LEVELS** | 0 ≤ ML-IQV ≤ 90 | "*Not Achieved*" |
| | 91 ≤ ML-IQV ≤ 100 | "*Achieved*" *if all inferior levels are at "Achieved"* |

***Table 3.*** *Equivalence of Value Range for IQVs and corresponding states.*

Combining the activities required for a methodology of assessment and improvement applied to data and information quality as proposed by [2] and with Deming's PDCA cycle in mind, we can give a brief summary of the definition of EVAMECAL as follows:

- **EVAMECAL – PLAN**:
  - ***EMC-P.1. Assessment of the current state of data and information quality of the IMP***. The main goal of this step is to determine what the current state of the IMP is, in terms of information quality maturity levels and IQVs. This implies the next sub-steps:
    - *EMC-P.1.1. Assessment of Data and Information Management Maturity Level of*

*the IMP*, using a set of questionnaires that has been developed.
        - *EMC-P.1.2. Calculation of IQV for IMP Components*, by measuring the chosen information quality dimensions and metrics for each component, in order to calculate the corresponding A-IQV.
    - ***EMC-P.2. Definitions of improvement goals in terms of data and information quality maturity level***. Any information quality initiative must set out a plan covering the goals to be improved. For the purpose of defining a plan for improving the most critical items first, EVAMECAL proposes to use the IQV as guidance for the identification of items mentioned.
- **EVAMECAL-DO**:
    - ***EMC-D.1. Analysis of potential causes and development of an improvement plan***. The main goal of this activity is to determine the reasons why an IMP is not working as expected. To achieve this goal, the two following sub-steps are proposed:
        - *EMC-D.1.1. Analysis and Comprehension of the problem*. To check whether any of the analyzed components have registered any kind of problems, several tests that allow the identification of these troubles and their possible sources need to be designed and executed.
        - *EMC-D.1.2. Detailed Analysis of real causes of problems*. The main aim of this activity is to discover the real causes of problems in data and information quality.
        - *EMC-D.1.3. Development of a plan for improvements*, taking into account the required goals. A plan establishing the actions that must be performed to obtain a set of improvements for the IMP, it also defines when and how these actions must be executed, by whom, and what resources are implied.
    - ***EMC-D.2. Execution of the improvement plan***. Once an improvement plan has been drawn up, and its viability has been approved, all actions for correction described in the plan must be carried out if the resources necessary are available.
- **EVAMECAL –CHECK**:
    - ***EMC-C.1. Check the efficiency of the improvement plan***. To validate the success of the plan empirically, once more a set of tests must be performed. This implies measuring the current state of the maturity level again and checking whether information quality goals have been achieved.
- **EVAMECAL-ACT**:
    - ***EMC-A.1. Obtain Conclusions***. After checking the efficiency of the improvement plan, some conclusions are drawn with respect to the problems identified and their origins. These conclusions may be the basis for avoiding future problems or for solving similar ones.
    - ***EMC-A.2. Standardize the learned lesson in order to avoid future problems***. The Information Management Team must standardize the knowledge acquired through experience, the aim being to avoid future problems and to get better results.

# 4. A REAL EXPERIENCE USING THE PROPOSED FRAMEWORK.

The framework was applied to different organizations. In the following sub-section, the results of the application to an information system consulting company are summarized.

## *4.1. Description of the organization.*

The organization in which the framework has been applied is the Centro Provincial de Informática (**CENPRI**) of the Excelentísima Diputación Provincial of Ciudad Real. The main aim of this organization is to give the necessary hardware and software support to the Diputación Provincial (Provincial Council)

itself and to the town councils of the province of Ciudad Real. It also gives support to the training of nearly fifty per cent (50%) of the eight hundred workers who use a computer in their daily job. For software support, the CENPRI develops its own applications using DBMS Ingress. They have wide and solid experience in the use of this DBMS, experience demonstrated at an international level. All applications have been developed under internal software quality rules, policies and standards. These standards have been developed internally by adapting several other software engineering standards to the work carried out there every day. .The experience accumulated over many years in the data and information quality field is taken into account, along with that gained from solving issues arising in the area of the quality of data and information being used.
Work sessions were conducted from September 2004 to February 2005 with the Director of the CENPRI.

## *4.2. Description of the studied IMP.*

Amongst all those IMPs managed by the CENPRI, "*Tribute Management*" was decided on as an object of study, since it was considered to be one of the most important IMPs of the organization.
Basically, the phases of this process are the following ones:

- Collection of data from the different town councils of the province in any possible format (CD, diskette, electronic mail, …).
- Cleaning of the data before it is stored into the databases. Cleaning them consists, fundamentally, of assuring that the format is the right one and that data is sound. In order to achieve this goal, data is going to be compared to supposedly trustworthy stored data. With unsuccessful data, a report with errors must be generated. All of these errors must be solved by the administrative personnel using other traditional tools and techniques.
- Once data is clean, it must be stored in corresponding databases. If new errors arise, then they must be gathered in a report so as to be corrected as soon as possible.
- The processes of generation and receipt emission are executed. Then receipts are sent to the contributors.
- During the period of payment, several associate banks receive the money from the contributors and when this time is over, they send the corresponding data about what has been paid in to CENPRI. This data is validated before being stored at the database.
- Once the data is validated, the information compiled is sent back to the different municipal governments (town councils).

All these steps are drawn using IP-MAP [33] in figure 2.
In this process, information coming from municipal government (data sources and format in which incoming data is sent), and output information (emission and sending out of receipts to the contributors), are of great importance (data product targets) since they are the main goal of this process. The organization has invested a major part of its efforts since 1991 in improving data and information quality for these two elements. These efforts have been directed towards the improvement of data quality by the production of several filters that allow the identification and correction of data which have erroneous values and formats. During this time, these filters have allowed them to improve, the proportion of taxes collected from 75% in 1992, to 95%, at present. This has meant a greater amount of income for the Diputación Provincial of Ciudad Real (Provincial Government). Regarding data and information quality dimensions, it should be pointed out that although they do not have an explicit definition. Suitable treatment of this aspect has been sought implicitly, in most cases. As regards the metrics related to the corresponding quality dimensions; some of these have been produced to measure the degree of efficiency of the software developed.

## *4.3. Description of the application of the framework.*

EVAMECAL was applied to the "*Tribute Management*" IMP to assess its maturity level and to make a list of all possible improvements that it could undergo. It was agreed that the focus would be on KPAs for the level of Definition only. This study took place as a pilot scheme so as not to disturb the rest of the organization's work.

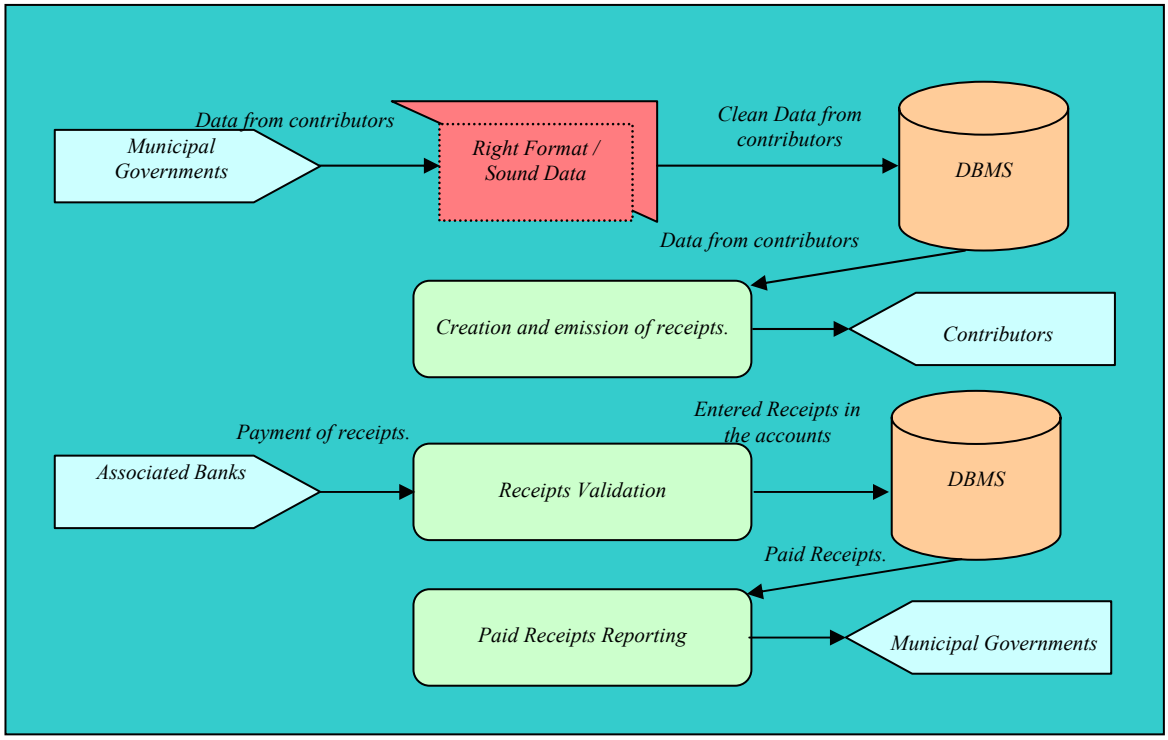


***Figure 2:*** *Modelling the "Tribute Management" IMP using IP-MAP*

The application of the framework was planned for the period from September 2004 to February 2005. The assessment work (EMC–P.1.) took about three and a half weeks. After the assessment period, a list of improvement proposals (EMC-P.2) was suggested to the organization. Once the root causes for disparities in relation to the reference model were identified, an improvement plan was drawn up (EMC-D.1.3): an Improvement Implementation Time (hereafter IIT) of five months for EMC-D.2 was agreed on, before beginning the EMC-C.1 step. By the first of February 2005, the reassessment work (EMC-C.1.) began again and this took about two weeks. Finally, and although earlier dates for completion of this were part of the agreement, the EVAMECAL-ACT phase is still being carried out, since members of CENPRI are still working hard on conclusions and the way to standardize it. Table 4 shows the agenda that is proposed for the application of the framework.

Conclusions about assessment and improvement are presented and corresponding graphs and diagrams about quantitative results in the evaluation are drawn in the following sections.

### 4.3.1. Execution of EVAMECAL-PLAN

The two main goals of EVAMECAL-PLAN are, on one hand, to determine the actual state of the IMP, and on the other hand to define the improvement goals.

In order to determine the present state of the IMP, the questionnaires were sent to the Director of the CENPRI, and IQV were calculated when possible. The main conclusions of the assessment were the following:

- (IQMTM) **Data and Information Quality Management Team (IQMT) Management**. The organization does not yet possess any team specialized in managing data and information quality.

It had no roles or responsibilities regarding this, and no standards, techniques and tools were observable. Nevertheless, all related tasks and activities are carried out by the Director of the CENPRI, although there is no documented catalogue of activities and tasks. The KPA-IQV was calculated as 18 (from results of questionnaires), thus the KPA was in a "*Not Satisfied*" level. As a measure to improve this KPA, the creation of a Data and Information Quality Management Team and the identification of roles and responsibilities were proposed.

- (IPM) **IMP Project Management**. Something similar to the drawing up of a project for this IMP had already been achieved, since the organization had managed this process for years using its experience gained over that time, This has never been seen as a data transformation process but in spite of this, much of the work required by CALDEA in relation to this issue had already been achieved. From the replies in the questionnaires, this KPA has obtained a value of 42 *("Partially Satisfied*"). As an improvement proposal, the formalization of a project for the IMP that was assessed here was suggested. This suggestion involved making several estimations whose aims were to define a life cycle for data and a plan for its correct implantation.

| PHASE | STEPS | ACTIVITIES | AGREED STARTING DATE |
|---|---|---|---|
| **EVAMECAL – PLAN** | EMC-P.1. Assessment of the current state of data and information quality of the IMP. | EMC-P.1.1. Assessment of Data and Information Management Maturity Level of the IMP. | First Week of September 2004 |
| | | EMC-P.1.2. Calculation of IQV for IMP Components. | Second Week of September 2004 |
| | EMC-P.2. Definitions of improvement goals in terms of information quality maturity level. | | Third Week of September 2004 |
| **EVAMECAL-DO** | EMC-D.1. Analysis of potential causes and development of an improvement plan. | EMC-D.1.1. Analysis and Comprehension of the problem. | Fourth Week of September 2004 |
| | | EMC-D.1.2. Detailed Analysis of real causes of problems. | Fourth Week of September 2004 |
| | | EMC-D.1.3. Development of a plan for improvements taking into account the required goals. | First Week of October 2004 |
| | EMC-D.2. Execution of the improvement plan. | | Second Week of October 2004 |
| **EVAMECAL –CHECK** | EMC-C.1. Check the efficiency of the improvement plan. | | First Week of February 2005 |
| **EVAMECAL-ACT** | EMC-A.1. Obtain Conclusions. | | First Week of March 2005 |
| | EMC-A.2. Standardize the learned lesson in order to avoid future problems. | | First Week of April 2005 |

***Table 4.*** *Agenda for application of the framework at CENPRI***.**

- (URM) **User Requirement Management**. User and user quality requirements for procedures were managed properly, but data and information quality requirements were dealt with in an implicit way, without any formalization or documentation. As the organization has enough infrastructure to support this KPA, and considering the answers of the interlocutor, this KPA was assessed with a value of 45 ("*Partially Satisfied*"). With the aim of improving the data and information quality requirement management, a formalization and documentation of this KPA was proposed.
- (DSTM) **Data Sources and Data Target Management**. Data sources and targets were

completely identified, and several filters and mechanisms have been developed to give uniformity in the formats of the data. This situation caused this KPA to fall within the level of "Fully Satisfied" (87). No improvement proposals were suggested at this KPA since there were other more important goals to achieve.

- (AIMPM) **Database or Data Warehouse Acquisition, Development or Maintenance Project Management**. The database, where data is stored, belongs to the organization and has been suitably designed and developed to support "*Tribute Management"*, following internal standards and development methodologies. The procedure and data models were designed and developed with the database. They used the DBMS Ingress; nevertheless, no part of the aspect of data and information quality had been monitored. The qualification obtained for this KPA-IQV was 59 ("*Partially Satisfied*") as recommended by the replies to questionnaires. Since data and information management quality user requirements have not been observed, and although several mechanisms to avoid respective data and information quality problems have been implemented, the main improvement proposal was to implement those kinds of requirements when possible. Another suggestion was to get proper documentation of the database linked to the IMP, and to put a configuration control for changes into operation.
- (IQM) **Information Quality Management in IMP components**. Although many data and information quality initiatives have been implemented, the organization does not yet have a sense of the dimension of data and information quality or the metrics to deal with several data and information quality trouble points. All they have done is to identify several problems, and they have hardly worked at all towards solving these or in avoiding them.For this reason, the qualification obtained in this KPA was only a 12 ("*Not Satisfied*"). So, as an improvement proposal, the choice of the most appropriate data and information quality dimensions and metrics for each one of the components of the IMP was forcefully suggested and encouraged.

**4.3.2. Execution of EVAMECAL-DO and EVAMECAL-CHECK.**

Having carried out the assessment as well as proposing a plan for the implementation of the improvements and just as the agenda agreed upon had set out, the work to achieve the improvements were carried out during the Improvement Implementation Time (IIT). When this period ended the reassessment work to check the efficiency of the plan begun, obtaining the following results:

- (IQMTM) **Data and Information Quality Management Team Management**. To satisfy the improvement proposal for this KPA, the need to create a IQMT was demonstrated to the organization. Once this need was accepted, the organization set aside a series of human and material resources to having an IMQT which has put into operation all the activities and tasks related to data and information quality. After the reassessment work, the qualification for this KPA was a value of 62, thereby achieving a state of "*Satisfied*".
- (IPM) **IMP Project Management**. This KPA was not satisfied, because the data and information management area had not been monitored, and because several basic project rules had not been used. The IQMT worked on this KPA, trying to implement all suggested improvement proposals during the IIT. In the end, several goals were achieved and a qualification of 70 ("*Satisfied*") for the KPA-IQV was obtained.
- (URM) **User Requirement Management**. To satisfy the main goals of this KPA, all data and information quality user requirements must be gathered, documented and observed in an appropriate way. As the organization has worked with other types of different user requirements, an observation of this kind of requirements can be done easily. After the IIT, all types of requirements were gathered and documented, but they were not completely implemented. The new evaluation of this KPA got a "*Satisfied*" (75) qualification
- (DSTM) **Data Sources and Data Target Management**. As no improvement proposals were suggested, no activities were planned or executed. This KPA kept the same qualification of 87 ("*Fully Satisfied*").

- (AIMPM) **Database or Data Warehouse Acquisition, Development or Maintenance Project Management**. The main improvement proposal depended on the execution of an improvement plan for the URM KPA. As this plan was successfully implemented, the introduction of data and information management requirements would, logically, also be done. But after IIT, not all requirements were implemented, due to limits on time. For this reason, the reassessment work gave a qualification of 71 ("*Satisfied*") for this KPA.
- (IQM) **Information Quality Management in IMP components**. The problem when achieving this KPA was that the workers of the IQMT did not have a sense of data and information quality dimensions and metrics. So, in order to achieve the main goal of the improvement proposal, the concept itself and other foundations related to data and information quality dimensions and metrics were explained to the members of IQMT. After the IIT, some dimensions for data value quality and for the relational schema of the database were identified and some metrics from these dimensions were derived and applied. The qualification of this KPA after the improvements went up to "*Satisfied*" (68)

Figure 3 shows the results of the application of the framework to the IMP. Making the corresponding calculations, the "*Tribute Management*" IMP has "*Not Achieved*" the level of maturity of "*Definition*" either before (ML-IQV=45,8) or after (ML-IQV=78,4) the improvements

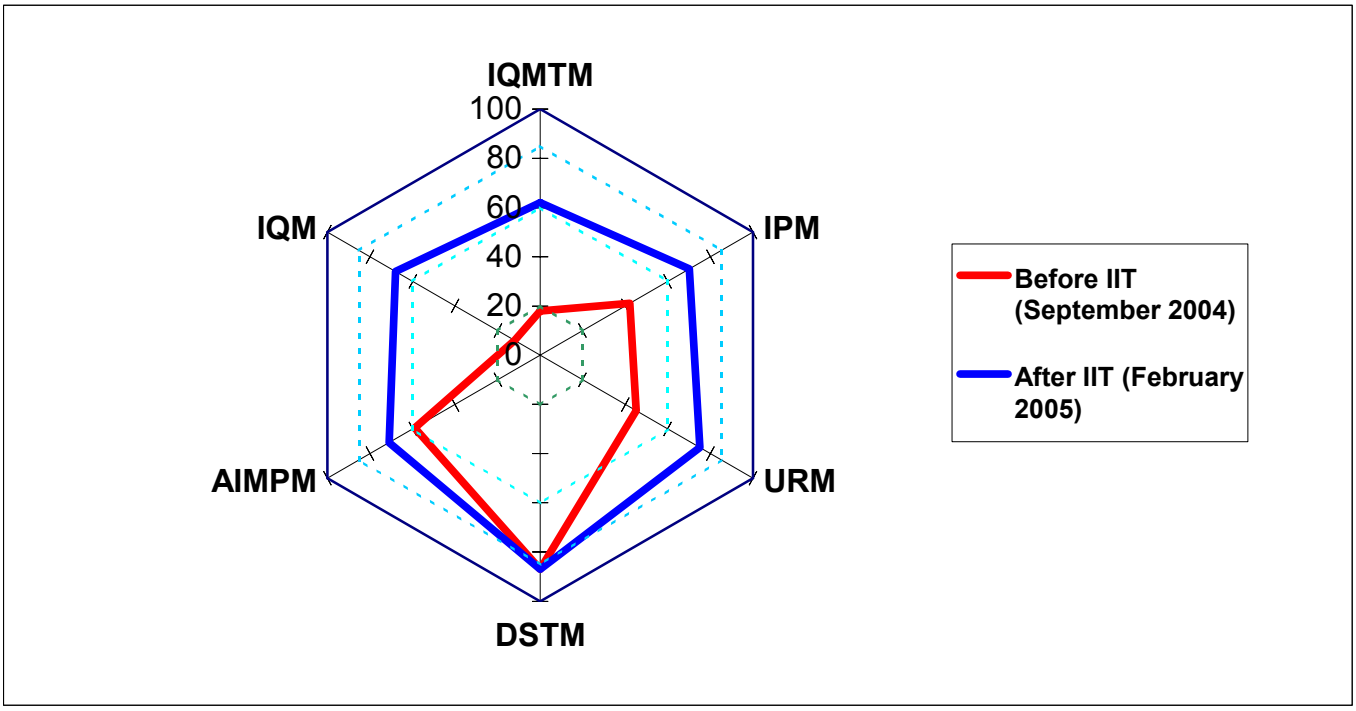


Figure 3. States of the KPAs for Level of Definition before and after IIT.

# 5. CONCLUSIONS.

In this paper, the results of the application of a framework for assessing and improving information quality management have been presented.

It is important to realize that the framework has only focused on KPAs of Level of Definition, and the application of the framework has been done in a limited time framework that had been agreed on beforehand. After the experience, several conclusions were reached:

- Related to the framework:
  - CALDEA successfully supports all issues related to data and information quality management. Although one of the most important goals for the framework was to develop a universal and independent one, a set of tools and techniques might be gathered in order to give a complete catalogue to the organizations which are not yet familiar with any.
  - EVAMECAL can be applied again and again until IMP gets the qualification of "*Achieved*" for this level. Since the reference model has KPA in different maturity levels which can be interdependent, it would be not possible to apply EVAMECAL to the same IPM at the same time for the purpose of assessing and improving several KPAs addressed at different levels.
  - As an IMP could be extended to the collaborative system, the framework could be used with this kind of systems.
  - It would be interesting to set the criticalness degree finely for each element (KPA, activities and components) in order to get a universal set of these,, valid for any situation and for any organization. It is important to get this if we are to establish a benchmark which allows the organization to make comparisons between them.
  - It would be also interesting to develop an economic model that allows the organization to estimate how much the application of the framework may cost. Regarding the time required for the application, this is highly dependent on how great a quantity of material resources are given over to it, and on the infrastructure of the organization.
  - Tools for automating the process of the application of the framework are needed to make this work easier.
  - As demonstrated by the experience presented here, the framework is very usable, and the most important issues are addressed properly.
- Related to the organization:
  - Organizations who realize what their data and information quality problems are, need to have some commitment towards solving them.. This commitment implies dedicating a quantity of human and material resources to initiatives in data and information quality.
  - Once the organization has become aware of the reference model, and with the experience obtained in applying the framework, it can design new IMPs in line with CALDEA so as to get the best data and information management quality from the early stages of developments.

## 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).

## 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 I. January 1999.

[3] 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.

[4] Burgess, M.S.E., Gray, W.A., and 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.

[5] Caballero, I., Gómez, O., and Piattini, M. (2004) "Getting Better Information Quality By Assessing And Improving Information Quality Management". *Proceedings of the Ninth International Conference on Information Quality (ICIQ'2004)*, 44-58

[6] 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.

[7] Crosby, P.B. *Quality is free*. New York, New York: McGraw Hill, 1979

[8] Dedeke, A. (2003). "Managing Processes for Information Quality: A Framework for Process Metadata Development". *Proceedings of the Eigth International Conference on Information Quality (ICIQ'2003)*, 224-233

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

[10] 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.

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

[12] Eppler, M. J. *Managing Information Quality*. Springer. Germany. 2003.

[13] 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

[14] Firth, C., and Wang, R. (1993). Closing the Data Quality Gap: Using ISO 9000 to study Data Quality. En http://web.mit.edu/tdqm/papers/93/pass1/93-03.html (last access on June 2005)

[15] 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.

[16] Fuggeta, A. "Software Process: A roadmap. The future of Software Engineering", Ed. A. Finkelstein *ACM, Press*. 2000. Pp.27-34.

[17] Gertz, M., Tamer, M., Saake, G., and Sattler, K., "Report on the Dagstuhl Seminar "Data Quality on the web"". *Proceedings of SIGMOD record.* Vol.33, No.1. March 2004.

[18] 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

[19] Hinrichs, H. "CLIQ- Inteligent Data Quality Management". *Proceedings of the fourth IEEE international Baltic Workshop on databases and Information System*. 2000

[20] Hoxmaier, J.A. "Dimensions of Database Quality". In *Developing Quality Complex Database Systems: Practices, Techniques, and Technologies*. Editor Shirley Becker. Idea Group Publishing. 2001

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

[22] ISO IEC 15504 *TR2: 1998, part 2: A reference model for processes and process capability*, ISO/IEC JTC1/SC7, 1998.

[23] ISO IEC 15504 TR2*: 1998, Software Process Assessment – Part 7: Guide for use in process improvement*, ISO/IEC JTC1/SC7, 1998.
[24] Juran, J.M. *Juran on Planning for Quality*. New York, New York: McMillan. 1988
[25] Kahn, B., Strong, D., and Wang, R. "Information Quality Benchmarks: Product and Service Performance". *Communications of the ACM* April 2002/Vol. 45, No. 4
[26] Kim, W., and Choi, B., "Towards Quantifying Data Quality Costs" in *Journal of Object Technology*, Vol. 2, no.4,. Pp. 69-76. July-August 2003
[27] Lee, Y.W., Strong, D., Kahn, B., and Wang, R. "AIMQ: a methodology for information quality assessment". *Information & Management*, 2001
[28] Loshin D. *Enterprises Knowledge Management: The Data Quality Approach*. Morgan Kauffman, San Francisco (California), 2001
[29] McLeod, R. (1990). "Management Information Systems". New York, NY: McMillan Publishing
[30] Pipino, L., Lee, Y., and Wang, R. "Data Quality Assessment". *Communications of the ACM* April 2002/Vol. 45, No. 4
[31] Redman, T.C. *Data Quality for the Information Age*. Artech House Publishers, Boston. 1996
[32] SEI. 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. In http://www.sei.cmu.edu/publications/ documents/02.reports/02tr002.html (last access on June 2004)
[33] Shankaranarayanan, G. Wang, R., and Ziad, M. "IP-MAP: Representing the Manufacture of an Information Product". *Proceedings of the 2000 Conference on Information Quality*. 2000
[34] Standard CMMI *SM Appraisal Method for Process Improvement (SCAMPI SM)*, Version1.1: Method Definition Document. CMU/SEI-2001-HB-001.
[35] Strong, D.M., Lee, Y. W., and Wang R.Y. "Data Quality in context". *Communications of the ACM*., pp 103-110. May 1997
[36] Wang, R. "A product perspective on data quality management". *Communications of the ACM*. Vol 41(2) pp58-65. February 1998
[37] Wang, R., Kon, H., and Madnick S. "Data Quality Requirements Analysis and Modeling". Published in the *Ninth International Conference of Data Engineering* Vienna, Austria., Pp 670 – 677. 1993
[38] Wang, R., Storey V.C., Firth, C. F. "A framework for analysis of Data Quality Research". *IEEE Transactions on Knowledge and Data Engineerin*g, Vol 7(4). Pp 623-640. 1995
[39] Wang, R., Ziad, M., and Lee, Y.W. *Data quality. Kluwer Academic Publishers*. Massachussets (USA). 2001
[40] Xu, H., "Would organization size matter for data quality?". *Proceedings of the Eighth Conference on Information Quality ICIQ'2003*. Pp. 365-379. 2003.