Computer Standards & Interfaces (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-6X and ISO/IEC 33000

Ana G. Carretero*, Fernando Gualo, Ismael Caballero, Mario Piattini

Alarcos Research Group. Institute of Software and System Technologies. Escuela Superior de Informática Paseo de la Universidad 4, 13071 Ciudad Real, Spain

ARTICLE INFO

Keywords:
Data quality
Data governance
Data quality management
Data management
Data improvement
Maturity model

ABSTRACT

Organizations are increasingly becoming realized that the higher the levels of quality of data they use in their business processes, the larger the organizational performance can be. Therefore, it is highly recommended to pay special attention to data quality by institutionalizing a set of the best practices related to the disciplines of data-related disciplines, namely, data management, data quality management and data governance.

After investigated on the field, and because of our research, we developed MAMD – *Modelo Alarcos de Mejora de Datos*-, a framework for assessing and improving the levels of data quality in organizations, in which we aligned and established the relationships between the three disciplines. Our aim was to provide organizations with sound artefacts, which can help them to efficiently implement data-related strategies to achieve adequate levels of data quality, and consequently, better organizational performance. Grounding our proposal on a process-oriented approach, we initially developed two components for MAMD: (1) a process reference model addressing the best practices of the previously mentioned data-related disciplines, and (2) an assessment and improvement model of the level of implantation of these practices. The process reference model is based on the principles of ISO 8000-61, which we complemented by adding specifics on data governance processes, and specifics on data management processes. The evaluation model is grounded on ISO 8000-62, and therefore aligned to ISO/IEC 33000. After having tested the usability of MAMD in several case studies, and after having analysed the conclusions raised from the learnt lessons, this paper describes the changes we introduced to the first version of MAMD to make the framework easier to apply (more easily auditable, and more easily implementable by consultancy). The paper also describes the application of the new version of MAMD to a new case study to check the efficiency of the changes. So, the main contribution of this paper is the presentation of the new version of MAMD.

1. Introduction

The potential of the organizations to develop their mission and to find new paths to innovate on an increasingly competitive market is mainly grounded on data. Due to this fact, organizations are becoming more and more conscious that the higher level of quality of the data, the higher the benefits they can obtain. So, they understand that enough resources in deploying solutions to achieve adequate levels of data quality shall be invested to meet intended and future uses of data in their business processes to ensure adequate levels of data quality.

This cannot be done in an isolated way, but it should involve the whole organization over the time. Hence, ensuring data quality is a task which must: (1) be planned well enough in advance; (2) consider clear objectives aligned with organizational strategy; (3) assign adequately qualified human, and sufficient materials and economic resources. Only then, commensurate results with organizational potential can be guaranteed. It can be said that the assurance of data quality levels must be achieved by implementing running in an integrated way data management, data quality management and data governance programs. In this paper, we will discuss how organizations can achieve this goal by means of the continuous and integrated implantation of best practices of these three disciplines.

To face up with a solution, we thought about an approach based on the idea of assessing and improving the “data processes” of the organization, similarly to what is done with software process in software developing companies. This enabled us to ground our proposal in the domain of existing software process solutions. To facilitate software processes improvement to organizations, there are

* Corresponding author.
E-mail addresses: AnaIsabel.Gomez@uclm.es (A.G. Carretero), Fernando.Gualo@uclm.es (F. Gualo), Ismael.Caballero@uclm.es (I. Caballero), Mario.Piattini@uclm.es (M. Piattini).

http://dx.doi.org/10.1016/j.csi.2016.11.008
Received 23 September 2016; Received in revised form 21 November 2016; Accepted 24 November 2016

alternatives based on *de iure* and *de facto* standards like COBIT [1], CMMI [2], ISO/IEC 15504 [3], ISO/IEC 33000 [4]... unfortunately, they do not specifically address low levels of data quality concerns, and it is not easy to use them directly with data management, data quality management and data governance disciplines.

However, in recent times, new process-oriented initiatives (DMM [5] or ISO 8000-60 [6]) emerged to cope with these disciplines. After a detailed study, we concluded that DMM had two important concerns: its application is not as easier as desirable (too deeply detailed, and too many requirements to meet), and it is focused primarily on financial domain. On the other hand, we realized that because of its general purpose, ISO 8000-6X is easier to apply and use than DMM, although it does not explicitly cover neither data government processes aspects and it nor does fully address data management processes.

To fill this gap, and because of our research we developed the Alarcos' Model for Data Improvement (MAMD stands for "*Modelo Alarcos de Mejora de Datos*" in Spanish). Our main goal was to create a framework that allows organizations to plan and execute continuous improvement projects based on PDCA cycle to progressively set up the best practices of the data-related disciplines, namely, data management, data quality management, and data governance. These three disciplines are strongly interrelated, and all of them working together are aimed to assure that the organization is using data with adequate levels of quality. This means that the used data has quality enough as to make successful the execution of the organizational business processes.

According to other software process approaches, MAMD consists of two main components:

A process reference model that extends ISO 8000-61 [7] with the foundations and best practices of data governance processes and some data management processes, not already included. More details about the process reference model will be introduced in Section 3.2.

An assessment and improvement model based on ISO/IEC 33000 [4]. We decided to ground our proposal on ISO/IEC 33000 because it is a generic framework for the assessment and improvement of some process characteristics (like capability and maturity); in this sense, ISO/IEC 33000 supersedes ISO/IEC 15504, mainly focused on software processes. As part of the assessment and improvement model, an organizational data maturity model was also included. This data maturity model specifies in an ordered way of achievement, the set of processes that companies should implement to warranty that they are using data with the adequate levels of quality for their business processes. In Section 3.3, the assessment and improvement model will be described.

After conducted several case studies [8] in order to check the usability of the framework, we realized that it was not as much easier to apply as we intend, and consequently, it would not help organizations to assure adequate levels of data quality. In our investigation, "easier to apply" is meant to be both "easily auditable" (it is easy to conduct an assessment of the organizational maturity level of the organization with regard to the three data-related disciplines), and "easily implementable" (it easily guides organizations in achieving changes to their structure to set up those best practices which contribute to assure adequate levels of data quality, that is to say, to achieve a more mature organization with regard to the three data-related disciplines.) In this sense, we analyzed the learnt lesson after the case studies, and we decided to make changes to the framework, to make it easier to apply. These changes affect to both components of MAMD, which has evolved to MAMD v2.0.

The main contribution of this paper is the presentation and description of the structure and components of MAMD framework v2.0, to better enable organizations to assure better levels of quality for the data used in their business processes.

This paper is structured as follows: Section 2 presents works related to the existing maturity models (both for software process and for "data process"). Section 3 presents the rationale of the changes we made to evolve MAMD from v1.0 to v2.0. Section 4 describes a new case study we conducted to check the usability of MAMD v.2.0. Section 5 introduces some conclusions raised and introduce some future lines of work that we consider necessary to continuously improve MAMD. Lastly, we include some acknowledge and references.

2. Related work

This section brings to the arena some related work corresponding to the core knowledge necessary to understand the scope and content of our proposal.

2.1. Data management, data quality management, and data governance concepts

The MAMD framework is based on the three aforementioned data-related disciplines: data management [9], data quality management [10,11] and data governance [12]. They are strongly dependant one from the others, and it does not make sense to implement them independently, because all working together better contributes to assure adequate levels of quality. This dependence is observed by [13]- where is revealed that the actual investigation in data quality involve the obvious need of adding certain governance, management, and technical aspects. The description of the three disciplines is showed below:

1. **Data governance** is aimed to design and implement data management and data quality strategies, which allows the alignment of data strategies to business organizational strategies. Such strategies are implemented as organizational policies. This will give support to the business needs by providing the necessary resources to both areas and monitoring the use of the resources regarding the strategic objectives of the organization.
2. From our perspective and for the sake of simplicity, we consider that **data management** implements and maintains a technological data infrastructure that must support business requirements. The requirements will be expressed through the data management policies. Likewise, the specific data quality requirements and their management shall be supported by the technological infrastructure.
3. **Data quality management** implements and maintains a data quality organizational culture that shall produce, maintain, perform, and communicate data quality management good practices that must be applied by data management. The actions previously mentioned shall satisfy the data quality specific requirements that ensure the organization processes success.

2.2. Scope of the existing data maturity models

A maturity model can be understood as a tool used to organize a set of elements ordered according to a given criterion [14]. In the domain of this work, the criterion is related to organizational maturity with respect of guarantee the success of business processes by means of the implementation of best practices of data quality management, data management and data governance, what enables adequate levels of data quality for the tasks at hands.

The first researcher to apply the concept of maturity model in the field of computer science was probably Humphrey in 1987 [15]. He used it to explain organizations; how to have more capable processes to produce high quality software. Specifically, in data quality domain, English was the first one to apply the maturity concept to data management at the same time as he included the notion of "data quality" in [14]. Since then, there has been many works related to data management that try to address this issue. Following subsections will go deep into such data quality management maturity models.

Regarding the scope of "*data management practices*" [16], it is easy to see how the evolution of the field has found data quality management and data governance. By the end of the twentieth century,