A case study on assessing the organizational maturity of data management, data quality management and data governance by means of MAMD.

ANA G. CARRETERO, University of Castilla-La Mancha
ALBERTO FREITAS, University of Porto
RICARDO CORREIA, University of Porto
MARIO PIATTINI, University of Castilla-La Mancha

Abstract. Organizations are increasingly becoming aware about the higher the levels of quality of the data they use in their business process are, the larger the performance of the business process will be. To maximize this performance, it is necessary that the organizations institutionalize some good practices related to data management, data quality management and data governance.

As a result of our investigations, we have developed the MAMD Framework, which provides a process reference model and an assessment and improvement methodology. The main contribution of this paper is to briefly introduce MAMD, and to show how it has been applied in a case study to assess the organizational maturity level of a Computer Science and IT Department of a University Hospital.

Key Words. Data Quality, Data Management, Data Quality Management, Data Governance.

1. INTRODUCTION

Organizations need data to feed their business processes. So, the potential for organizations to carry out their mission rely on data. To be more competitive in this global market implies improving the existing ways - and if it is possible to find new ones - to better exploit the owned data. In this line, organizations are becoming increasingly aware that the larger amount of data, the larger yield can obtain from them. Therefore, it seems reasonable to think that organizations should invest sufficient resources to deploy solutions in order to assure owned data have a level of quality commensurate to their potential.

Assuring the quality of the data is therefore a task that should be planned with sufficient time, should be done with clear objectives that are aligned to the organizational strategy, and involving adequate human, material and economic resources. Only proportional results may thus be guaranteed to the potential of the organization. This assurance of the quality of the data should be done through the implementation of integrated programs of data management, data governance, and data quality management.

To facilitate the improvement of the processes organizations software, there are alternatives based on standards such as COBIT [1], CMMI [11], ISO/IEC 15504 [9],... but unfortunately they do not specifically address the problem of the data, nor it is easy to customize them to face up in a joint way the disciplines of data management, the data governance, and data quality management. However, in recent times, some process-oriented initiatives are emerging like DMM [10] or ISO 8000-60 [2], which are trying to deal with the way in which organizations can face up with the implementation of these plans. After a detailed study, we conclude that DMM had two major problems: it was not easy to apply and it was mainly focused on the financial sector. On the other hand, in spite of ISO 8000-60, is a general-purpose framework more easily applicable, it does not explicitly include processes related to data governance, neither fully address all data management processes.

To mitigate these shortcomings, and as main result of our research, we have developed the Alarcos' Data Improvement Model (called MAMD from the Spanish acronym of *anonymized for blind review*). Our goal was to develop a framework that would allow organizations to plan and execute projects for continuous data improvement projects based on PDCA, in order to achieve partial increments of improvements. MAMD consists of two main components:

- A Process Reference Model (PRM) that extends ISO 8000-61 [3] with the data governance processes , and some other data management processes,
- An assessment and improvement model based on ISO/IEC 33000 [4].

The main contribution of this paper is two-fold: on a hand it briefly introduces the components of MAMD, and on the other, it describes a case study that we conducted to empirically validate MAMD.

The rest of the paper is structured as follows: section 2 briefly introduce the MAMD components. Section 3 describes how MAMD has been used to evaluate the level of organizational maturity of data governance, data management and data quality management in a Computer Science and IT Department of a University Hospital, which will be called CSD for confidentiality reasons. Finally section 4 contains some conclusions that have been obtained as a result of the evaluation and introduces some future works that we consider would help to improve MAMD.

2. ALARCOS' DATA IMPROVEMENT MODEL

As stated in the introductory section, MAMD consists of (1) a process reference model that brings together the principles of data quality management introduced in ISO 8000-61 extended with the specific data governance processes as well as some data management processes, and (2) an Assessment and Improvement Model based on the various parts of ISO/IEC 33000 family. The following sections describe both components.

2.1 Process Reference Model

According to the clause 5.3.1 of ISO/IEC 33004 [6], a process reference model is defined as a set of processes that collectively can better support the organizational process model. MAMD Process Reference Model consists of 21 process grouped into the areas of data management, data governance, and data quality management quality. They have been identified by mapping the processes in ISO 8000-61, DMM, COBIT, and DAMA, and describes the purpose of each of them. Due to paper length restrictions, we cannot include the whole documentation for all of the processes (this includes Process Results, and Activities). In the following, the description of the identified processes is shown.

Data Management Processes.

- **DM.1. Data requirement management.** This process aims at collecting and validate requirements referral to necessary data to manage the organization successfully.
- **DM.2. Technological infrastructure management.** The goal of this process is to specify and maintain the necessary technological infrastructure to support data meaning shared between applications.

- **DM.3. Historical data management.** The process addresses how to maintain and perform necessary policies to organizational historical data management.
- **DM.4. Data security management.** This process is aimed to define and enable mechanisms to make possible confidentiality, integrity, accessibility or availability, authenticity, non-repudiation, consistency, isolation, and data audit.
- **DM.5. Configuration management.** The process addresses how to define the processes by which an organization demand, determines, approves, and implements the reachable plans and evaluates the changes of data lifecycle.
- **DM.6. Master Data management**. This process is aimed to identify the relevant concepts to organization business domain and the organizational data strategy alignment around these master data.
- **DM.7. Data design.** The goal of this process is to develop a consistent data model, complete, comprehensive and extensible that covers the data requirements of all organizational units. In addition, the data model shall be aligned to the organizational data strategy**.**
- **DM.8. Data sources and data targets establishment.** The process addresses how to identify and characterize each data sources and destinations used in original business processes, as well as the agreements and interactions with providers and customers.
- **DM.9. Data integration**. The goal of this process is to ensure data integrity through flow control and relationships with transferred data to application systems or data bases.

Processes related to Data Quality Management.

- **DQM.1. Data Quality Planning.** This process is aimed at introducing the corresponding data quality actions and mechanisms into the various organizational resource dealing with data.
- **DQM.2. Data Quality Control and Monitoring.** The main aim of this process is to establish the corresponding actions to monitor the levels of quality on the data used through the organization as well as to fix the found non-conformities in data.
- **DQM.3. Data Quality Assurance.** This process is aimed at implementing the necessary changes to organizational resources so that it can be assured that levels of data quality will be enough to successfully run the organizational business processes.
- **DQM.4. Data Quality Improvement.** The goal of this process is to implement a continuous improvement cycle based on PDCA model to data in organizational repositories and business processes.

Processes related to Data Governance.

- **DG.1. Data strategy establishment.** The process addresses how to identify and prioritize data management objectives, and work according to these prioritization to give support to the corporate strategic objectives.
- **DG.2. Data lifecycle management.** The goal of this process is to identify the importance degree of data have to different business processes in corresponding stages.
- **DG.3. Data Value Management.** This process is aimed at determining the organizational value of data according to the data strategy.

- **DG.4. Standards, policies and procedures definition.** This process is aimed to establish those standards, policies, good practices and procedures to data management, data quality management and data governance to support as better as possible the data quality strategy.
- **DG.5. Human resources management**. The process address how to manage needs adequately to required specific formation to the human resources specifically destined to data management, data quality management and data governance.
- **DG.6. Financial resources management**. The goal of this process is to develop plans for financial resources provisioning and maintaining that can give support to organizational data strategy.
- **DG.7. Data organization strategies monitoring**. This process is aimed to develop and measure key indicators for monitoring the achievement of data management strategy and check that it is being actually aligned with the organizational data strategy.
- **DG.8. Change management in data strategy**. The goal of this process is to maintain coherently organizational data strategy according to the evolution of corporate strategic objectives.

2.2 Assessment Model

The assessment model is based on ISO/IEC 33020 [8], and includes an assessment methodology that consists of five steps and a maturity model.

The assessment of the level of organizational maturity corresponds to an assessment of type 1 (see ISO/IEC 33002 clause 4.6.1.1 [5]), and therefore, at least four instances of four organizational processes in the scope of the evaluation must be inspected. As a result of the assessment, a certain level of maturity will be provided, and as a specification of the scope of the improvement, a maturity level will be also be provided.

The maturity levels proposed in MAMD, along with their meaning and the processes that are included, are detailed below (see table I):

- **Maturity level 0 or Immature:** the organization cannot provide evidence about the effective implementation of good practices addressed by the process reference model. Therefore, there are no guaranties that their data is being used adequately.
- **Maturity level 1 or Basic:** the organization can evidence that it uses a set of basic good practices oriented to provide the minimum support necessary to the data management required to successfully support their business processes. Nevertheless, no special attention is given to data governance and data quality.
- **Maturity level 2 or Managed:** the organization, in addition to the data management activities, is becoming aware of the need of quality for the data they used in their business process. Some control and monitoring over the quality of data is to be done and evidence that the organization works on the development of policies and standards for aligning data and data quality to their organizational strategy.
- **Maturity level 3 or Established:** the organization can evidence that they can address complex data management activities, they already are worried about the data quality by conducting planning as they become aware of the value of data to establish sound data strategies.

- **Maturity level 4 or Predictable:** the organization can evidence that they monitor the data strategies in order to assure that data is being used with adequate levels of quality that are aligned to the organizational strategies.
- **Maturity level 5 or Innovating:** the organization can evidence that it uses a set of good practices oriented to guarantee that organizational data strategies are evolving and several actions are executed to assure that data quality is being continuously.

Table I. Processes of MAMD per levels of Maturity Model

Maturity Level	Processes
1	DM.2. Technological infrastructure management DM.4. Data security management
2	DM.1. Data requirement management DM.3. Historical data management DG.2. Data lifecycle management DG.4. Standards, policies and procedures definition DQM.2. Data Quality Control and Monitoring DM.5. Configuration management
3	DM.9. Data integration DQM.1. Data Quality Planning DM.8. Data sources and data targets establishment DM.6. Master Data management DM.7. Data design DG.1. Data strategy management DG.5. Human resources management
4	DG.7. Data organization strategies monitoring DQM.3. Data Quality Assurance DG.6. Financial resources management DG.3. Data Value Management
5	DG.8. Change management in data strategy DQM.4. Data Quality Improvement

The level of maturity is computed according to the level of capacity that the processes of the processes involved in the evaluation reference model. The level of capacity is computed taking into account the degree of institutionalization of the best practices and the processes attributes described in ISO/IEC 33020.

To compute the level of capability of the processes, it is necessary to inspect the different types of evidence (direct, indirect, interviews and documentation) collected for each one of the instances of the business processes chosen to perform the assessment. This is to be done according to the nine Process Attributes as described in ISO/IEC 33004 (see table 2).

Table II. Processes Attributes described in ISO/IEC 33004

Capability Level	Goal of the Process Attribute
1. Performed	PA.1.1. Process performance
2. Managed	PA.2.1. Performance management PA.2.2. Work product management
3. Established	PA.3.1. Process definition PA.3.2. Process deployment
4. Predictable	PA.4.1. Quantitative analysis PA.4.2. Quantitative control
5. Innovating	PA.5.1. Process innovation PA.5.2. Process innovation implementation

As a result of the assessment of the capability level, a score of {*Not Achieved (N), Partially Achieved (P), Achieved Full (F), Largely Achieved (F)*} for each of the attributes of the process as specified in ISO/IEC 33020 is obtained.

CASE STUDY : APPLYING MAMD AT THE COMPUTER SCIENCE DEPARTMENT OF A MEDICAL FACULTY

3. In this section, the results obtained in the case study we conducted Computer Science and IT Department of a Faculty of Medicine (hereafter CSD - for confidentiality reasons) are shown. These results have allowed to refine and validate the model. The evaluation has been carried out following the principles of assessment set out in ISO 33010 [7]. Names and the corresponding filiations of the people who participated in the case study are anonymized for confidentiality reasons.

Description of the Organization

3.1 The CSD is in charge to provide with computer science and IT services for the reminder of the six academic departments in the university hospital. The number of employees of the department is twenty two people, who support most of the business processes of the reminder departments. The team is multidisciplinary and there are several mathematicians, several computer scientific and several IT specialists who are in charge to maintain the specific IT infrastructure that support the academic business processes of the CSD.

3.2 Scope of the assessment

As it was a pilot project, along with a representative of the head of CSD, we decided that the scope of the assessment was to evaluate only up to level 2. It was a class 1 assessment. This involves inspection of evidence of the institutionalization of best practices contained in the reference model of processes in at least four instances of processes of the Organization: one main and three auxiliary. But in order to simplify the effort that CSD staff should made, it was decided to choose only three processes, which were identified during the planning phase (see section 3.3).

3.3 Planning the Assessment

It is important to highlight that the planning has two stages: (1) a first initial stage in which we met representatives of CSD, in which it was chosen the processes that were going to be discussed during the evaluation, and (2) a second one in which we planned when and to whom we needed to interview in order to document the evidence about the institutionalization of the good practices of data management, data quality management and data governance.

With regard to the identification of the business processes to be considered as part of the assessment, and taking into account the interest of the CSD on optimizing the most critical processes of the department the following ones were chosen:

— Main process (MP): **Pharmacology Data Repositories Maintenance**. This business process is aimed at recovering, documenting and unifying the results obtained in different Pharmacology testbeds made in different experiments. These data should become available for different academic and medical studies.

— Auxiliary Process 1 (AP1): **Biostatistics Report Generation**. This process is in charge to support the generation of various report required by other CSD teams,

mainly from outside the department. These reports are used for making decisions on different scenarios.

– Auxiliary Process 2 (AP2): **Clinical software maintenance**. This process is run by the IT team. It is important to highlight that we are interested in the data flow required to execute the corresponding jobs, not in the software maintenance (it can make some misunderstandings)

The assessment was planned to be executed during a third week of June 2015. We interviewed people in charge of the Pharmacology, Biostatistics and IT teams. It is quite interesting to mention that the Pharmacology team follows ISO 9001 guidelines, just for convenience, even when the whole department is not ISO 9001 certified

3.4 Collecting evidence

We use a work script including a checklist of the expected results, and a questionnaire containing specific questions aimed at collecting evidence from the execution of data management, data quality management and data governance processes.

The questionnaire, which covered the processes involved in level 1 and level 2 of MAMD (see section 2.2) and the attributes of processes AP.1.1, AP.2.1, AP.2, included questions for persons identified during the planning (see section 3.3). Apart from the answers, the CSD workers provided some other evidence, which were catalogued as well as the relevant documentation. The full catalogue of the gathered evidence cannot be included due to confidentiality reasons. In the following, for each process within the two first maturity levels in Table 1, we include some conclusions raised after inspecting the gathered evidence. Please, be advised that we first inspected evidence in the Main Process, and in the case that there not exists, we inspected AP1, and if not, we inspected AP2.

- **DM.2. Technological infrastructure management**. The PP is conveniently supported by the CSD internal technological infrastructure. In this sense, the PA.1.1 is satisfied (even when it can be considered as externalized from the Pharma team). The management of the technological infrastructure itself is done by member of the IT team, who is responsible for collecting specific requirements, validating and developing the technological solutions that supports the works of the reminder teams. In this sense, both PA.2.1, and PA.2.2 are fully satisfied. Even when we obtained sufficient evidence to meet the assessment requirements and produce a conclusion, we also inspected the two other processes. We decided to include the results about the assessment only when necessary.
- **DM.4. Data security management.** The Pharma Team specified the data security requirements, which also include the legislation regarding to privacy which rules in the country of the CSD. This satisfies PA.1.1. In order to make operative the data security restrictions, the IT team developed the corresponding mechanisms to support the MP, fully satisfying the PA.2.1 and PA.2.2.
- **DM.1. Data requirement management**. The MP is fully described in the internal documentation (in fact, for convenience, they follow ISO 9001) and consequently, the Pharma team have already properly identified data requirements during the description of the process. With respect to the management of the process, we discovered that the objectives in managing the requirements are identified, and the process is planned, even when it is not monitored, and the Pharma team provided sufficient resources and human

resources ad hoc. With respect to the management of work products, we could see that although there are templates predefined for the collection of data requirements and various controls and revisions of the generated documents are made, these tasks are carried out also in a way ad hoc. So PA.2.1 and PA.2.2 can be considered as largely satisfied.

- **DM.3. Historical data management**. We gather sufficient evidence to state that the Pharma team maintains strict policies about the historical data management because it is important for the success of the consequent reports. And they are supported by the IT team. This makes PA.1.1 be considered as fully satisfied. As this process is very important to the Pharma team they pay special attention to the performance management as well as the work product management. Therefore, both PA.2.1 and PA.2.2 can be considered as fully satisfied.

- **DG.2. Data lifecycle management.** Due to the awareness of the importance of the data by the Pharma team, they have successfully stablished the life cycle of the data they use in the MP. They provided evidence that they meet of the results of this process, and consequently we can state that the PA.1.1 is fully satisfied. As they worked in this process in an ad hoc way to manage the data lifecycle, unfortunately, they were not able to provide any evidence that allows is to declare the PA.2.1and PA.2.2 are fully satisfied..

- **DG.4. Standards, policies and procedures definition**. As part of the management of the Pharma team, they have adapted and properly implemented a series of policies, practices and procedures that are based on the general policies of the CSD. Many of these policies are strongly oriented to properly manage the data and estimate its quality. During the conduction of the questionnaires they provided us with answers and some other proofs that allows us to state that the results of this process is achieved (and consequently the PA.1.1 is fully satisfied). The work of defining a catalog of standards, policies and procedures is not executed in a predefined way, no responsibilities are assigned and resources are allocated only when necessary. So, we could not obtain evidence enough to state that PA.2.1 and PA.2.2 are at least satisfied. Similar conclusions were raised when we analyzed the evidence we obtained for the AP.1 and for the AP.2.

- **DQM.2. Data Quality Control and Monitoring**. As they have been working over the years, the Pharma team has collected some data quality rules that they used to check the validity and consistency of the data they received. From the answers to the questions they provided during the questionnaires and from the part of the works they showed, we could state that they execute some kind of non-formal data quality control and monitoring. Consequently, we concluded that PA.1.1 is partially satisfied. Even when they demonstrated knowledge enough to perform the data quality monitoring tasks, they did not follow a predefined process, and they could not provide any evidence of that. Therefore, PA.2.1 and PA.2.2 were considered as not satisfied. This led us to consider the inspection of the AP.1 and, if required the inspection of AP.2. When inspecting AP.1, we discovered that the Biostatistics team follows a predefined process, with the allocation of the necessary resources and responsibilities. In addition, they documented all of the concerns related to the results of the data quality control and monitoring processes. During the interview, the Biostatistics team provided answers and several evidence that

allows us to state that not only PA.1.1 is fully satisfied, but also PA.2.1 and PA.2.2 are largely satisfied.

- **DM.5. Configuration management**. Changes made to the data definition along the years when executing the MP, were not adequately managed. Therefore, the Pharma team could not provide us with any evidence that allowed us to raise a conclusion about the satisfaction of PA.1.1. For the AP.1, as they did not need to run any configuration management process, the same conclusions were obtained. This led us to inspect the AP.2. Even when they often some configuration management for the software maintenance task, they did not apply these tasks on the data they used. Therefore, we raised the conclusion that none of the PA.1.1, PA.2.1 or PA.2.2 were at least partially satisfied.

Validating data

3.5 The assessment team confirmed the objectivity of the collected evidence and we validated them with respect to other evidence provided for the same processes. It was considered that the set of evidence provided was sufficient and quite representative for the purpose and scope of the evaluation.

Deriving results

3.6 Based on the conclusions obtained from the collected evidence and following the guidelines provided in clause 4.2 of ISO 33010 for the scoring of the process attributes, we derived the results for each one of the process attributes – shown as {N, P, L, F} in Table III.

Table III. Summary of the scoring of the Process Attributes when assessing the CSD

AP.2.2	F	F	L	F	F	N	L	N
AP.2.1	F	F	L	F	F	N	L	N
AP.1.1	F	F	F	F	F	F	F	N
	DM.2	DM.4	DM.1	DM.3	DG.2	DG.4	DQM.2.	DM.5
	Maturity Level 1		**Maturity Level 2**					

To consider a maturity level as "consolidated" is necessary that all the processes of the process reference model involved in the previous levels are in the "F" state, and for the level being assessed, the process would be at least at in a "L" state.

3.7 In this case study, as it can be sawn in table III, all processes in the maturity level 1, have been scored as "F" for the AP.1.1. . This enables that the organization can be said to be in ML1 (it has the ML1 consolidated). But some of the other processes have been scored different than "L" or "F", what makes that the maturity level 2 cannot be considered as "consolidated". In this situation, the higher maturity level consolidated is the 1, and consequently we can state that the CSD is at ML1.

Report results

Finally, after the assessment, we held a meeting with representatives of the CSD to inform them of the results of the assessment. In this meeting, we did not only informed about the obtained level of maturity (maturity 1 level or Basic), but also about the

strong and the weak points (non-conformities) of the organization with regard to data management, data quality management and data governance.

Finally, the assessment team provide to the CSD some suggestions about how to improve the organization until to consolidate the maturity level 2.

4. CONCLUSIONS

We have found that application of MAMD can really provide benefits for organizations, benefits that result from working with data that have adequate levels of quality. In any case, it is necessary to apply MAMD in more case studies, and take advantage of the lessons learned to better refine the model.

Also we want to quantitatively establish up to what extent the improvement of the level of maturity of data management, data quality management, and data governance can offer a clear advantage for the organizations.

ACKNOWLEDGMENTS

This work has been funded by VILMA project (Consejería de Educación, Ciencia y Cultura de la Junta de Comunidades de Castilla La Mancha, y Fondo Europeo de Desarrollo Regional FEDER, PEII-2014-048-P) and SEQUOIA project (Ministerio de Economía y Competitividad and Fondo Europeo de Desarrollo Regional FEDER, TIN2015-63502-C3-1-R).

REFERENCES

[1] ISACA, 2013. *COBIT 5: Enabling Information*.
[2] ISO, 2015. ISO 8000-60: Data Quality Management : The Overview of Process Assessment ISO.
[3] ISO, 2015. ISO 8000-61: Data quality: Information & data quality management process reference model ISO.
[4] ISO, 2015. ISO/IEC 33000: Information technology: Process assessment ISO.
[5] ISO, 2015. ISO/IEC 33002: Information technology: Process assessment: Requirements for performing process assessment ISO.
[6] ISO, 2015. ISO/IEC 33004: Information technology: Process assessment: Requirements for process reference, process assessment and maturity models ISO.
[7] ISO, 2015. ISO/IEC 33010: Information technology: Process assessment: Guide to performing assessment ISO.
[8] ISO, 2015. ISO/IEC 33020: Information technology: Process assessment: Process measurement framework for assessment of process capability ISO.
[9] ISO/IEC-JTC1/SC7, 2004. *ISO/IEC 15504-1:2004, Information Technology - Process Assessment - Part 1: Concepts and Vocabulary*. International Organization for Standarization, Geneva, Switzerland.
[10] SEI, 2014. DMM: Data Management Maturity Model SEI, Pittsburgh.
[11] TEAM, C.P., 2002. Capability Maturity Model® Integration (CMMI SM), Version 1.1. *CMMI for Systems Engineering, Software Engineering, Integrated Product and Process Development, and Supplier Sourcing (CMMI-SE/SW/IPPD/SS, V1. 1).*