Journal of Theoretical and Applied Information Technology
30th November 2017. Vol.95. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CO-DEPENDENCE RELATIONSHIP BETWEEN MASTER DATA MANAGEMENT AND DATA QUALITY: A REVIEW

[1]FAIZURA HANEEM, [2]AZRI AZMI, [3]NAZRI KAMA

[1,2,3,4]Advanced Informatics School, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

E-mail: [1]hmafaizura2@live.utm.my, [2]azriazmi@utm.my, [3]mdnazri@utm.my

ABSTRACT

Master Data Management refers to the consolidation, integration and standardization of master data from multiple data sources into a centralized system to support data quality improvement in an organization. Nevertheless, while Master Data Management came into prominence in the information systems field of study, there is a lack of review papers for this topic have been published. Hence, this paper reports the results of a systematic literature review on the Master Data Management research topic. It aims to summarize the research progress of Master Data Management since 2000 to July 2016 and to review the association of Master Data Management and Data Quality. Search strategies with relevant keywords were used to identify literature from seven prestigious academic databases, namely 1) ACM Digital Library; 2) Emerald; 3) IEEE; 4) Science Direct; 5) Scopus; 6) Springer Link; 7) Web of Science, and one industry research database, namely Gartner. Additionally, the study made use of Google Scholar to find more related literature on the MDM research topic. From the review, 777 articles were found during the initial search and 347 relevant articles were filtered out for the analysis of MDM research progress. Then, out of the relevant articles, 49 were selected to discuss the association of MDM and Data Quality. This paper is a first academic systematic literature review on the progress of Master Data Management and its association with Data Quality. The result of the review shows that Master Data Management came into prominence from 2009 in parallel with the Big Data movement. Most researchers describe Master Data Management as a means to resolve data quality issues encountered during the management of multiple data sources. It ensures better data quality in the organization by combining a set of processes, data governance, and technology implementations.

Keywords: *MDM, Data Quality, Systematic Literature Review*

1. INTRODUCTION

Master data represent the most relevant business entities in the organization such as customer, products or suppliers [1]. Master data are the references for transactional data which relatively unchanged where there would not be a single transactional data without master data. In most organizations, the problem in managing master data is that the master data are scattered across various business units, applications and database systems. Master data which have similar information (i.e. individual profile, agency's corporate information) are redundant within organization since they have been stored and managed in silo by each business units.

Master Data Management (MDM) has been used to enable an organization or enterprise to associates all of its master data to a single reference repository [2]. This repository provides a standardized center of definitions that can be leveraged across different business units in an organization [3]. By having this MDM reference, it is expected that the data redundancies and inconsistencies can be further reduced hence data quality of the organization would be improved [4] [5] [6].

To the best of our knowledge, there has yet to be presented a systematic literature review report on the association between MDM and Data Quality in the research domain. While both concepts are obviously closely related to each other, major streams of research have explored them in a rather isolated fashion. There is a strong school around data quality research, and there are contributions focusing on data management in general and MDM in particular. This observation can also be made in the practitioners' realm where

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

master data management typically is a responsible of operations and supply chain management while data quality often can be found in financials and sales. So, an academic investigation of the relationship between the two fields of research is a merit to both for practitioners and researchers.

The structure of this paper starts with the background of MDM and data quality definition. We then present the review methodology and subsequently discuss the review results based on the research questions. Finally, we conclude the paper and recommend the future works.

2. LITERATURE REVIEW

2.1 Master Data Management

In 2006, MDM was described as a process of creating and maintaining the values and master data and the relationship between them [7]. Master data are defined as critical business data in an organization, shared across several different systems or organizational units, serve as reference for transactional data, and rarely changed [4], [8], [9]. In addition, [10] described MDM is an application-independent process which explains, owns and manages core business data entities. It ensures the consistency and accuracy of this data by providing a single set of guidelines for their management and thereby creates a common view of key company data, which may or may not be held in a common data source.

The MDM term was further explained by researchers based on the researcher's perspective and research context. In 2011, [11] defined MDM as bringing master data together to enable the employment of master data management services such as data governance and stewardship, data quality , metadata, hierarchy and overall the data lifecycle management. Despite various MDM terms defined by earlier researchers, there are similarities that exist between them. It can be summarized that MDM is not just about the technology. It is also a management of shared core data to reduce redundancy and ensure better data quality through standardized reference with a combination of process, governance and technology. It aims to serve data as a 'single reference of truth' to the consumers by consolidating and integrating the master data from multiple data sources into a central system.

2.2 Data Quality

Data Quality is "the measure of the agreement between the data views represented by an information systems and that same data in the real world" [12]. There are six core dimensions in measuring data quality which are: 1) completeness; 2) uniqueness; 3) timeliness; 4) validity; 5) accuracy; and 6) consistency [13]. The issue of data quality is increasingly important in information systems as well as organizations are depending on multiple data sources of data to make decisions [14]. Poor data quality in information systems particularly in managing multiple data sources has affected every application domain [15], [16]. It has led to a problem of duplication, inaccuracy and inconsistency of information [17], [18]. A study by [19] stated that in average, organizations are disbursing hundreds of thousands of dollars direct cost in data cleansing and other activities to improve the quality of information they use to conduct business. Moreover, the hidden cost of data quality issues such as lost opportunities, low productivity, waste, and myriads of other consequences, is believed to be much higher than these direct costs. MDM tends to resolve data quality issues that have been encountered during the management of multiple data sources in an organization by introducing a set of processes, data governance, and technology implementations.

3. RESEARCH METHODOLOGY

MDM is related to the Information Systems (IS), Information Technology (IT) and Data Management fields of study [20]–[22]**.** Thus, this research adapted the systematic review methodology that is designed particularly for the IS research [23]. This study simplified the review protocol suggested by [23] to the following four stages enumerated as follows: 1) research questions; 2) search strategy design; 3) study selection; and 4) analyses of findings (see Fig.1).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

First, a set of research questions were identified based on the research's aims. Second, the search strategies were designed by determining the sources of databases, search keywords and searching criteria. Third, the study selection was conducted by filtering duplicate articles. Fourth, the quality assessment was performed to select the relevant articles and finally the analyses of the findings were performed against the selected relevant articles.

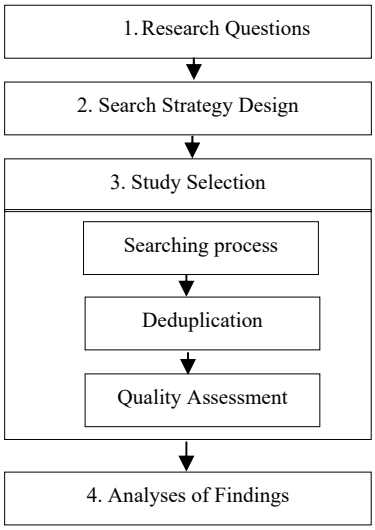


Fig. 1. Four stages in review protocol

3.1 Research Questions

This paper aims to review the progress of Master Data Management research topic and its association with data quality. To achieve this aim, four research questions were formulated as shown in Table I.

Table I. Research Questions

ID	Research Questions
RQ1	How did the numbers of articles vary by year?
RQ2	Who is leading the MDM research among the selected source of databases?
RQ3	How do the MDM articles vary in different publication types?
RQ4	How does MDM associate with data quality?

3.2 Search Strategy Design

The description of the search strategy designed in this research consists of databases, search keywords and search criteria.

3.2.1 Databases

Nine electronic repositories were used for this review study. Seven from academic repositories namely: 1) ACM Digital Library; 2) Emerald; 3) IEEE; 4) Science Direct; 5) Scopus; 6) Springer Link; 7) Web of Science and one (1) industry research repository, namely Gartner. Additionally, the study also includes Google Scholar to find more related articles on the MDM. Title, abstract and index terms were used to conduct searches for journals, proceedings, books, book chapters and industry research.

3.2.2 Search Keywords

There are three steps involved in constructing the search keywords of this review [24].

i. Identification of alternative spellings and synonyms for major terms.
ii. Identification of keywords in relevant papers or books.
iii. Usage of the Boolean OR to incorporate alternative spellings and synonyms.

The initial search strings are ("master data"), ("management"), ("Master Data Management"), and ("MDM"). Then, the search strings were joint using "AND" and "OR" Boolean. The search strings were inputted to each electronic repository to retrieve the articles based on the titles, abstracts, contents and keywords, depending on the advanced search facility provided by the database.

3.2.3 Search Criteria

There are two search criteria used during the searching process which are:

i. The language used in the paper is English.
ii. The paper is categorized as journal, proceeding, book, book chapter and industry research.

3.3 Study Selection

Initially, 767 articles were identified by using the search keywords from all eight academic and industry research repositories. Additionally, the searching process continued with a manual search of articles from Google Scholar database, in cases where the articles were not indexed in the selected electronic repositories. From the manual search, ten

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

additional articles were found. During the searching process, the metadata of all 777 identified articles were gathered and tabulated in a list using Microsoft Excel. There are six columns in the list which are: 1) Database; 2) Article title; 3) Abstract; 4) Year; 5) Publication Type; and 6) DOI/ISBN/ISSN Number.

Then, de-duplication process was performed against the list to eliminate the duplicated copies of the identified articles that exist across databases [25]. During the de-duplication process, 42 duplicate articles were found and removed. This process compressed the list from 777 articles, to only 735 identical articles.

Next, quality assessment was conducted by performing the practical screening against the 735 identical articles. Practical screening is the activity of screening the title and abstract of the articles based on quality assessment criteria to check the relevancy of the articles [23]. The quality assessment criteria are listed in Table II.

Table II. Quality Assessment Criteria

Phase	Aim	ID	Criteria
First Phase	To answer RQ1-3	QA1.1	The main context of the article is MDM
		QA1.2	The objective of the article is clearly stated
		QA1.3	The papers are primary study or original research
Second Phase	To answer RQ4	QA2.1	The article contains 'data quality' phrase in its title, abstract and keywords.
		QA2.2	The article that stated the association of MDM and data quality

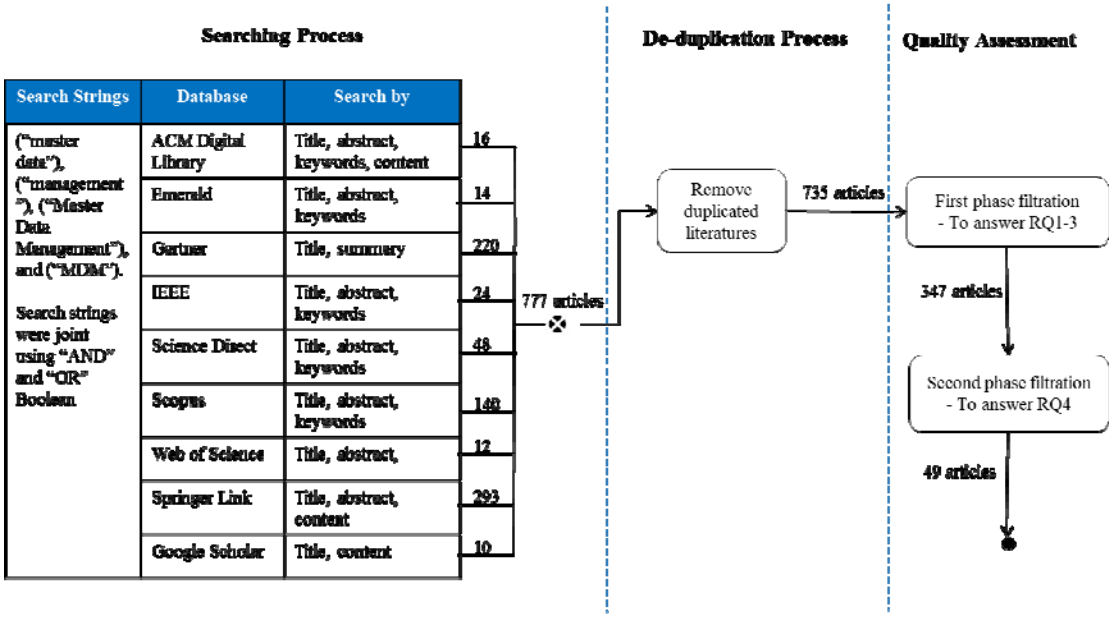


Fig. 2. Study Selection Process

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table III. Discussion Topic Of Data Quality In 49 Related Articles

Data quality issues on multiple data sources management			MDM resolve data quality issues			Data quality is the key success of MDM		
Duplication	**Inaccuracy**	**Inconsistency**	**Process**	**Data Governance**	**Technology**	**Assessment**	**Integration**	**Assurance**
[26] [27] [28] [29] [8] [30] [31] [32] [18] [33] [17] [34] [35] [36] [37] [38]	[39] [26] [40] [41] [42] [27] [29] [30] [31] [32] [18] [20] [43] [17] [44] [35] [45] [38]	[39] [26] [41] [46] [42] [27] [29] [8] [30] [18] [20] [43] [17] [35] [38]	[27] [31] [17] [34] [47] [48] [38] [49]	[50] [51] [8] [31] [52] [20] [43] [33] [53] [54] [37] [55] [56] [48] [57] [38]	[41] [58] [59] [42] [60] [28] [61] [32] [62] [18] [63] [64] [36] [65] [66] [67] [49]	[31] [68]	[60] [61] [31] [63]	[26] [31] [69]

Two phases of filtration were involved in this process. The first filtration listed 347 relevant articles that are based on quality assessment criteria in answering research question RQ1-3. Then, second filtration was performed in answering research question RQ4. Out of 347 relevant articles, there are 49 articles have been selected as listed in Table III for second filtration phase. Overall, Fig. 2. illustrates the selection process of this review which consists of the searching stage, de-duplication process and quality assessment stage.

3.4 Analyses of Findings

This stage analyses 347 relevant articles in answering RQ 1-3. Then it discusses on the 49 further selected articles in answering RQ 4. The analyses of findings are presented in the following Section 4.

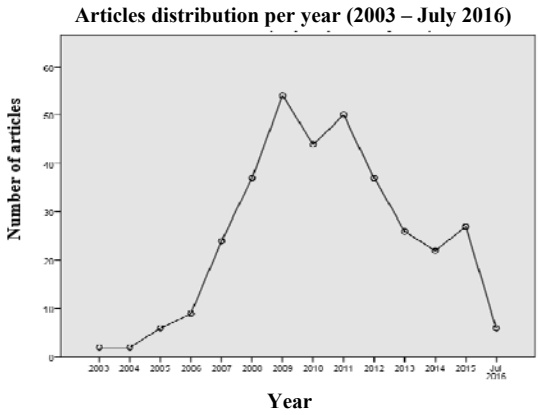


Fig. 3. Distribution Of Articles By Publication Year

4. FINDINGS AND DISCUSSION

The analyses of findings are reported based on the formulated research questions.

4.1 RQ1 - How did the numbers of articles vary by year?

Fig. 3. illustrates the distribution of the 347 related articles by year, regardless of the publication type. In 2003, "Master Data Management" was firstly described by Gartner in the analysis of SAP MDM solution to manage and maintain the distributed master data within organization [70]. Since then, the research interest of this topic increased dramatically until 2009, dropped slightly in 2010 and increased again in 2011. At this point, the interest constantly decreased until 2014 but rose again in 2015. However, the score for the year 2016 could not being concluded since it only shows the number of the articles from January to June 2017.

4.2 RQ2 - Who is leading the MDM research among the selected source of databases?

Knowing which databases that devoted on this research topic would lead a new researcher of the MDM to the right sources.

Fig. 4. illustrates the related articles distribution by sources. The chart only displays eight databases excluding Web of Science (WOS) database since all the related articles from WOS database are duplicated with the articles from Scopus database.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

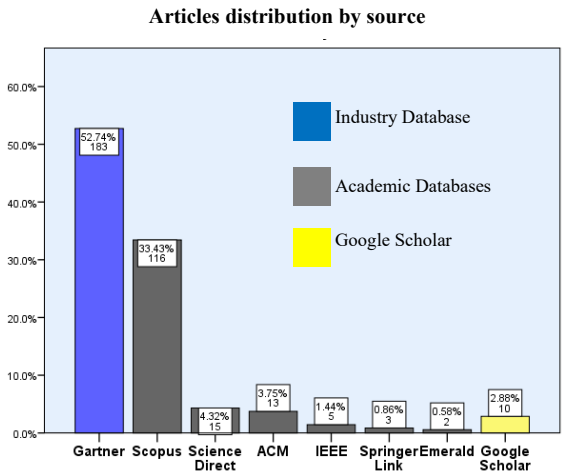


Fig. 4. Distribution Of Articles By Sources

The chart testifies that a total of 183 (52.7%) articles are come from the industry research by Gartner and followed by academic research from Scopus: 116 (33.4), Science Direct: 15 (4.3%); ACM: 13 (3.8%); Google Scholar: 10 (2.9%); IEEE: 5 (1.4%); Springer Link: 3 (0.9%); and Emerald: 2 (0.6%). %). Google scholar recorded 10 articles which equal to 2.88%. This shows that industry research by Gartner has led the MDM research as compared to academic research.

4.3 RQ3 - How do the MDM articles vary in different publication types?

Fig. 5. describes the distribution of related articles by the publication types.

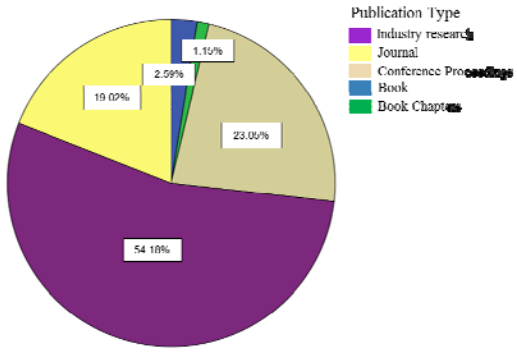


Fig. 5. Distribution of Articles by Publication Types

It is worthy to note that a total of 54.18% of the articles have been published in industry research. This is followed by publication type of conference proceedings of 23.05% and journals of 19.02%. The rest are books and book chapters of 2.59% and 1.15% respectively. This significantly shows that this topic needs an improvement of evidence in an academic research in filling in the gaps from the industry research.

4.4 RQ4 – How does MDM associate with Data Quality?

Data quality is an important role in the success in implementing MDM where it supports the trustworthiness of master data [71], [72]. To answer the RQ 4, the second filtration phase has been performed by refining the articles that contain 'data quality' phrase in its title, abstract and keywords. Using our interpretations as a basis, the association between MDM and data quality in existing articles can be categorized into three higher-order comparative topics: 1) data quality issues on multiple data sources management, 2) how MDM resolve the data quality issues of multiple data sources management, and 3) data quality is the key success of MDM. The 49 refined articles and the comparative topics are shown in Table III.

With duplicated, inconsistent, and inaccurate data across multiple data sources, there is a high demand to have an integrated data management in an organisation [36]. There is a lot of traditional enterprise data integration, data warehouse and data mart technologies have been utilized to meet the demand [35]. Nevertheless, when there is a conflict between two similar data from different data sources, these technologies cannot support a real time validation of these data [29]. They only capable to resolve batch mode of data. Thus the MDM is a better real-time solution to address the weaknesses of the traditional enterprise data integration, data warehouse and data mart technologies for the real time validation [44].

4.4.2 MDM resolve Data Quality issues of multiple data sources management

MDM would resolve data quality issues of duplication, inaccurate, inconsistency data across multiple data sources management [20]. With MDM, the common critical data that give a value across business unit and departments will be consolidated into central system and these datasets will be referred as highly accurate and authorized data by data consumers for a common good [56], [67]. According to [4], [11], MDM is not just about

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

a technology, it is an approach to ensure data quality through a combination of processes, data governance, and technology implementations. These three elements play a critical role in resolving data quality issues of multiple data sources management.

4.4.2.1 Processes

MDM consist of two main processes which are: 1) Entity Resolution (ER) Process, and 2) Entity Identity Information Management (EIIM) [17], [27], [31], [34], [48], [49]. Entity Resolution (ER) Process or also known as record linking or de-duplication is the essential process that have been recognized as a main data cleansing to remove the duplicate records and to promote data quality in database systems [38], [73], [74]. ER determines the accurate data when there are multiple entities that have been identified from several sources which referring to the same set of entities [34]. ER process consists of two main processes which are 1) determination of two records that referring to the same entity, and 2) selection of the best which called survivor record. The first step compares the identity information of two records using a set of matching rules to determine that records are duplicates for the same entity. Next process selects one survivor record between those two records and this survivor record will be passed to the next process. In regards to the record de-duplication, ER is primarily a data cleansing process where it addresses the data quality issues of redundant and data duplication prior to data integration process [75].

Even though ER is an essential process for effective MDM [47], but itself independently is insufficient enough to manage the life cycle of identity information. Identity information is a collection of attribute-value pairs that describe the characteristics of the entity that serve to distinguish one entity from another. For example, a student name attribute with a value such as 'James Smith' would be identity information. However, because there may be other students with the same name, additional identity information such as birthdate or address may be required to fully disambiguate one student from another [48]. The goal of EIIM is to sustain the identity integrity over time. Entity Identity Information Management (EIIM) is a basic requirement in MDM as it is a process that associates ER and data structures that represent the identity of an entity into specific operational configurations [21]. These operational configurations are all executed together to maintain the entity identity integrity of master data over time [48]. With regards to data quality , EIIM is not limited to MDM but it can be applied to other types of systems such as RDM systems, referent tracking systems and social media [76]. Overall, the MDM implementation would promote data quality of an organisation by incrementally reducing the amount of duplicated data and providing authoritative master data to the data consumers throughout an enterprise [11], [77].

4.4.2.2 Data Governance

MDM builds quality into data management processes through clearly documented roles and responsibilities under data governance [50], [52]–[56]. The three important aspects of data governance for MDM are managing key data entities and critical data elements, ensuring the observance of information policies, and documenting and ensuring accountability for maintaining high-quality master data [20], [48], [78]. To accomplish these tasks, an effective team of people which have a clean-cut mission statement and well-defined roles and responsibilities are very vital to be established. This team should be an association between business people and IT staff [33], [37], [57].

According to [51], on one hand, the business people would play the role of MDM Champion, Information Steward and MDM Process Manager. The MDM Champion not only builds the business case for MDM but also elicits buy-in from other business participants and assures that the organization is fully aware of the project and its impact. This person either provides or must secure executive sponsorship and works directly with the IT Architect. The Information Steward defines the objectives for data quality and evaluates the results of the technology solution; he or she collaborates with the IT Data Steward. The Process Manager not only defines the processes for master data management but also helps manage them. The Process Manager should work in parallel with the IT System Manager.

On the other hand, IT staff would consist of Architect, Data Steward and System Manager. The Architect designs the enterprise-level strategy for master data management applications and assures that executives are able to follow the MDM process. The Data Steward identifies the requirements for fixing existing problems in data quality, makes sure that the level of quality remains high over time and ensures that the technology

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

solution supports any pre-existing data governance requirements. The System Manager ensures that the MDM technology supports existing technology platforms. Table IV summarizes the responsibilities for each role.

4.4.2.3 Technology Implementations

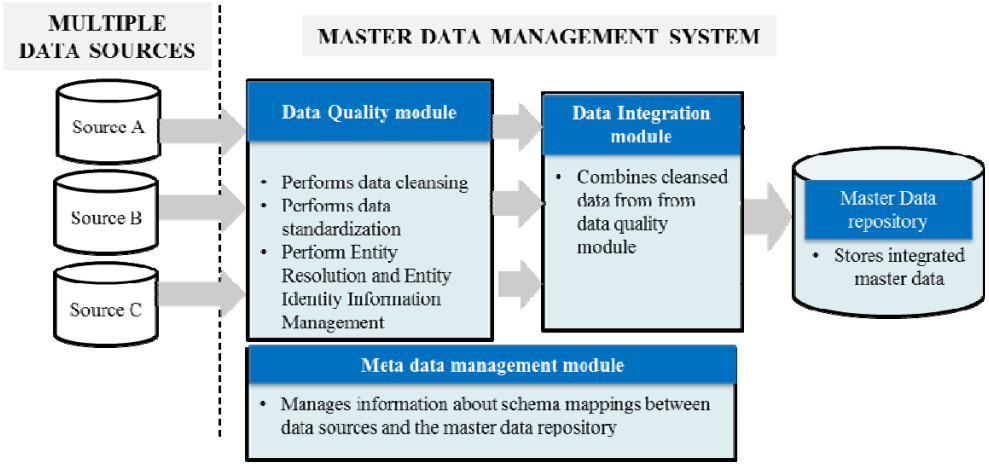
In solving existing data quality problems, the MDM technology solutions are designed to be a system that integrates the multiple data sources into a single unified view [36]. Even though there are various of architectures for MDM implementation [67], typically, MDM system consist of four main modules namely: 1) data integration module; 2) master data repository; 3) metadata repository; and 4) data quality module [18]. Fig. 7 illustrates the interrelation among the modules.

The data quality module pre-processes the input data from the sources to the MDM system through data cleaning and data standardization techniques. This module is also performing Entity Resolution (ER) Process and Entity Identity Information Management (EIIM). The data integration module then combines pre-processed data from data quality module to provide a unified view of them by using schema mappings. After that, the master data repository stores integrated master data processed by data integration module. The metadata repository manages information about schema mappings between data sources and the master data repository[41], [42].

Table IV. Roles And Responsibilities In MDM [51]

Team	Roles	Responsibilities
Business	MDM Champion	Identifies MDM risk and creates a mitigation plan Secures executive sponsorship Builds the business case for MDM Assures business buy-in and project visibility
	Information Steward	Defines information objects for MDM Defines objectives for data quality Evaluates results of pilot project Samples ongoing data quality
	MDM Process Manager	Defines and manages MDM processes Performs impact analysis for process changes Ensures successful rollout of solution in various departments Facilitates change management
IT	Architect	Designs enterprise-level strategy for MDM applications Identifies solutions that have strategic fit and portfolio compatibility Provides executive-level visibility into the MDM process
	Data Steward	Defines data governance requirements for maintaining master data Identifies requirements for both existing and on-going data quality Defines system integration with existing data flow for master data objects Ensures that technology solution supports governance requirements
	System Manager.	Ensures that MDM technology supports existing platforms Ensures that master data component integrates with source data solutions Ensures ease of integration with existing business applications

Fig. 7: Four Main Typical Modules In MDM System



ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In recent years, the development of MDM systems has been fostered by large software vendors [15], [37]. To name a few, Oracle Master Data Management, IBM InfoSphere Master Data Management, SAP NetWeaver MDM and Informatica MDM [30], [35], [43]. While most of the MDM vendors tend to develop a comprehensive solution for MDM, yet there is still a lack of strength in other aspects of data quality beyond matching [80]. For example, data profiling is needed to assess the state of master data quality in the initial stages of an MDM effort, but some of the MDM system does not have that capabilities. For this reason, a growing number of integration and partnerships are arising between MDM vendors and data quality tools vendors [58], [60], [61], [63]–[66].

4.4.3 Data quality is the key success of MDM

Master data are defined as critical business data in an organization, shared across several different systems or organizational units, serve as reference for transactional data, and rarely changed [8], [77]. Storing a high-quality of master data in the master data repository is one of the critical success factors of the MDM implementation [27], [39], [40], [46]. To achieve that, data quality assurance must be in place throughout the master data lifecycle [81]. Master data lifecycle phases consist of three main stages which are: 1) Assessment; 2) Integration, and 3) Assurance [27]. Fig. 8. illustrates the master data lifecycle.

The assessment is an initial process where core data entities from multiple data sources are identified to be stored in master data repository. It is important to stress the fact that master data selection are always identified from the business requirements [68], [82]. With the assistance of tools and technologies, the assessment process identifies and analyses candidate master data sets, primary keys, foreign keys, entity relationship, and pre-defined business rules before any data integration can begin [31].

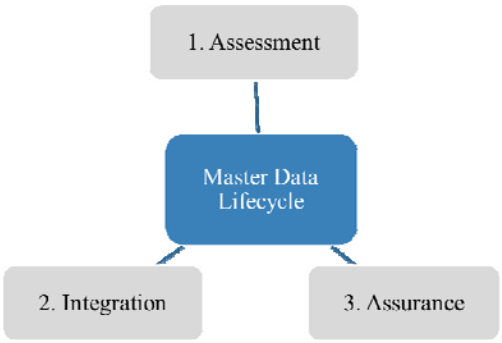


Fig. 8. Master Data Lifecycle [27]

The integration stage is a process of linking up the identified core data together prior to store them into master data repository [27]. In ensuring the data quality, the identified core data are gone through the cleansing, standardization, matching, and linkage process. After completing the integration processes, the integrated data are ready to be stored in master data repository and published to data consumers. Besides using MDM solution during this stage, it also may require assistance from other data quality tools to perform these processes [60], [61], [63].

The final stage of master data lifecycle is an assurance stage. MDM is not going to be a case of "build once and they will come." Changes in the master data requirements due to new business requirements or data user complaints may trigger a refinement of master data object design. Auditing and monitoring compliance with defined data quality expectations coupled with effective issue response and tracking, along with strong data stewardship within a consensus-based governance model, will ensure ongoing compliance with application quality objectives. According to [43], the key challenges in managing master data is poor data quality . Hence, it is inevitable to instill data quality assurance in MDM implementation with the focus of master data [69]. Without a focus on managing the quality of master data, the organization runs a risk of repeating from an enterprise information management program to just another unsynchronized data silo [26].

5. CONCLUSION

Although Master Data Management came into prominence in the Information Systems field of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

study and there have been many contributions on the topic for almost two decades, but there is a lack of review papers for this topic have been published. This paper reports the results of a systematic literature review on the MDM research progress and how it associates with data quality. The review was conducted by assessing articles from nine databases from academic and industry research which are: 1) ACM Digital Library; 2) Emerald; 3) Gartner; 4) IEEE; 5) Science Direct; 6) Scopus; 7) Springer Link; 8) Web of Science; and 9) Google Scholar. From the review, it shows that the topic has received increasing attention especially from the industry as compared to the academic community. As far as data quality concern, it can be concluded that data quality has codependence relationship with Master Data Management. On one hand, MDM could resolve data quality issues encountered during the multiple data sources management. This is by implementing a set of processes, predefined data governance, and technology implementations. On the other hand, the key success of MDM implementation is depending on the high-quality of master data stored in the master data repository of MDM. MDM implementation success can be achieved by ensuring data quality throughout the master data lifecycle. For future works, it is highly recommended to explore the association of the MDM with other topics such as big data, data modelling, or business intelligence.

6. ACKNOWLEDGEMENT

The study is financially supported by a Fundamental Research Grant Scheme (Vote No. R.K130000.7838.4F861) awarded by Ministry of Higher Education of Malaysia and University Teknologi Malaysia.

REFERENCES:

[1] B. Otto, K. M. Hüner, and H. Österle, "Toward a functional reference model for master data quality management," *Inf. Syst. E-bus. Manag.*, vol. 10, no. 3, pp. 395–425, 2012.

[2] H. Scheidl, "Master Data Managment Maturity and Technology Assessment," University of Turku, 2011.

[3] DAMA, *The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK Guide), 1st Edition*. New Jersey: Technics Publications, 2009.

[4] A. Dreibelbis, E. Hechler, I. Milman, M. Oberhofer, P. van Run, and D. Wolfson, *Enterprise Master Data Management: An SOA Approach to Managing Core Information*. Pearson Education, 2008.

[5] S. Lerche, "Achieving customer data integration through master data management in enterprise information management," 2014.

[6] R. F. Smallwood, *Information governance: concepts, strategies, and best practices*. John Wiley & Sons, 2014.

[7] M. Shin, "Master Data Management : A Framework and Approach for Implementation," in *11th International Conference on Information Quality (ICIQ)*, 2006.

[8] B. Otto and K. Hüner, "Functional reference architecture for corporate master data management," 2009.

[9] Y. Cao, T. Deng, W. Fan, and F. Geerts, "On the data complexity of relative information completeness," *Inf. Syst.*, vol. 45, pp. 18–34, 2014.

[10] H. A. Smith and J. D. McKeen, "Developments in Practice XXX : Master Data Management : Salvation Or Snake Oil ?," *Commun. Assoc. Inf. Syst.*, vol. 23, no. 1, p. 4, 2008.

[11] D. Cervo and M. Allen, *Master Data Management in Practice: Achieving True Customer MDM*. New Jersey: John Wiley & Sons, 2011.

[12] K. Orr, "Data quality and systems theory," *Commun. ACM*, vol. 41, no. 2, pp. 66–71, 1998.

[13] DAMA UK Working Group, "The six primary dimensions for data quality assessment," 2013.

[14] N. K. Yeganeh, S. Sadiq, and M. A. Sharaf, "A framework for data quality aware query systems," *Inf. Syst.*, vol. 46, pp. 24–44, 2014.

[15] M. Ouzzani, P. Papotti, and E. Rahm, "Introduction to the special issue on data quality," *Inf. Syst.*, vol. 38, no. 6, pp. 885–886, 2013.

[16] F. Naumann and C. Rolker, "Assessment methods for information quality criteria," 2000.

[17] S. Ben Hassine-Guetari and B. Laboisse, "Managing Multisource Database: Between

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Theory and Practice," in *16th International Conference on Information Quality (ICIQ)*, 2011, pp. 507–521.

[18] H. Galhardas, L. Torres, and J. Damásio, "Master data management: a proof of concept," in *15th International Conference on Information Quality (ICIQ)*, 2010.

[19] S. Lee, "Challenges and Opportunities in Information Quality," in *E-Commerce Technology and the 4th IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services*, 2007, p. 481.

[20] A. Cleven and F. Wortmann, "Uncovering four strategies to approach master data management," in *System Sciences (HICSS), 2010 43rd Hawaii International Conference on. IEEE*, 2010, pp. 1–10.

[21] B. Otto, "How to design the master data architecture: Findings from a case study at Bosch," *Int. J. Inf. Manage.*, vol. 32, no. 4, pp. 337–346, 2012.

[22] Silvola, Risto, O. Jaaskelainen, H. Kropsu-Vehkapera, and H. Haapasalo, "Managing one master data – challenges and preconditions," *Ind. Manag. Data Syst.*, vol. 111, no. 1, pp. 146–162, 2011.

[23] C. Okoli and K. Schabram, "Working Papers on Information Systems A Guide to Conducting a Systematic Literature Review of Information Systems Research," *Sprout Work. Pap. Inf. Syst.*, vol. 10, no. 26, pp. 1–51, 2010.

[24] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," *Tech. Rep. EBSE-2007-01*, 2007.

[25] Q. He, Z. Li, and X. Zhang, "Data deduplication techniques," in *Future Information Technology and Management Engineering (FITME), 2010 International Conference*, 2010, vol. 1, pp. 430–433.

[26] G. F. Knolmayer and M. Röthlin, "Quality of Material Master Data and Its Effect on the Usefulness of Distributed ERP Systems," in *International Conference on Conceptual Modeling*, 2006, pp. 362–371.

[27] D. Loshin, "Chapter 5 - Data Quality and MDM BT - Master Data Management," in *The MK/OMG Press*, Boston: Morgan Kaufmann, 2009, pp. 87–103.

[28] J. Kokemuller and A. Weisbecker, "Master data management: Products and Research," in *14th International Conference on Information Quality (ICIQ)*, 2009, vol. 8–18.

[29] K. K. A. Looi, *MDM for Customer Data*. MC Press Online, LP, 2009.

[30] D. Loshin, *Master data management*. Morgan Kaufmann, 2010.

[31] D. Loshin, "19 - Master Data Management and Data Quality BT - The Practitioner's Guide to Data Quality Improvement," in *MK Series on Business Intelligence*, Boston: Morgan Kaufmann, 2011, pp. 327–350.

[32] M. Krizevnik and M. B. Juric, "Improved SOA Persistence Architectural Model," *ACM SIGSOFT Softw. Eng. Notes*, vol. 35, no. 3, 2010.

[33] Roxane Edjlali and T. Friedman, "Overcome Barriers to Improving Data Quality : Lessons From the EMEA MDM Summit Workshop What You Need to Know," *Gartner, March (ID G00211687)*, 2011.

[34] K. Murthy and P. Deshpande, "Exploiting evidence from unstructured data to enhance master data management," in *Proceedings of the VLDB Endowment*, 2012, vol. 5, no. 12, pp. 1862–1873.

[35] E. Gomede and R. M. Barros, "Master Data Management and Data Warehouse: An architectural approach for improved decision-making," in *Information Systems and Technologies (CISTI), 2013 8th Iberian Conference on. IEEE*, 2013, pp. 1–4.

[36] F. Versini, P. Gruau, P. André, N. Bouvard, L. Gache, A. Givaudan, G. Machet, V. Marchand, M. Raschas, A. Vabois, G. Vincent, and E. Zarnitsky, "Master data management in the health industry: What models are required to meet tomorrow's ethical, regulatory and economic requirements," *S.T.P. Pharma Prat.*, vol. 23, no. 3, pp. 123–187, 2013.

[37] E. Buffenoir and I. Bourdon, "Reconciling complex organizations and data management: the Panopticon paradigm," *arXiv Prepr. arXiv1210.6800*, 2012.

[38] M. Allen and Delton Cervo, *Multi-Domain Master Data Management: Advanced MDM and Data Governance in Practice*. Morgan Kaufmann, 2015.

[39] Andrew White, D. Newman, D. Logan, and J. Radcliffe, "Mastering Master Data

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Management," *Gartner, January (ID G00136958)*, 2006.

[40] T. Friedman, "Trust and Compliance Issues Spur Data Quality Investments," *Gartner, Sept. (ID G00151255)*, 2007.

[41] S. J. Lowry, P. B. Warner, and E. Deaubl, "NOAO imaging metadata quality improvement : A case study of the evolution of a service oriented system," in *Proceedings of the Conference on Object-Oriented Programming Systems, Languages, and Applications, OOPSLA*, 2008, pp. 675–683.

[42] Y. Luh, C. Pan, and C. Wei, "An innovative design methodology for the metadata in master data management system," in *International Journal of Innovative Computing, Information and Control*, 2008, vol. 4, no. 3, pp. 627–637.

[43] R. Silvola, O. Jaaskelainen, H. Kropsu-Vehkapera, and H. Haapasalo, "Managing one master data-challenges and preconditions," *Ind. Manag. Data Syst.*, vol. 111, no. 1, pp. 146–162, 2011.

[44] D. Sammon, F. Adam, T. Nagle, and S. Carlsson, "Making Sense of the Master Data Management (MDM) Concept: Old Wine in New Bottles or New Wine in Old Bottles?," *DSS*, pp. 175–186, 2010.

[45] D. E. O'Leary, "Embedding AI and Crowdsourcing in the Big Data Lake," *IEEE Intell. Syst.*, vol. 29, no. 5, pp. 70–73, Sep. 2014.

[46] A. Winkelman, D. Beverungen, and C. Janiesch, "Improving the Quality of Article Master Data : Specification of an Integrated Master Data Platform for Promotions in Retail," in *European Conference on Information Systems (ECIS)*, 2008, pp. 1859–1870.

[47] J. Kobielus, "Big Identity's double-edged sword: Wielding it responsibly," *IBM Data Management Magazine*, 2013.

[48] J. R. Talburt and Y. Zhou, *Entity Information Life Cycle for Big Data: Master Data Management and Information Integration*. Morgan Kaufmann, 2015.

[49] S. Kikuchi, "Structural analysis of issues in exchanging qualified master data under cloud computing," *Int. J. Comput. Sci. Eng.*, vol. 11, no. 3, pp. 297–311, 2015.

[50] A. Duff, "Master data management roles-Their part in data quality implementation," in *10th International Conference on Information Quality (ICIQ)*, 2005.

[51] Ventana Research, "Building Successful Master Data Management Teams," 2007. [Online]. Available: ftp://ftp.software.ibm.com/software/emea/de/db2/WP_MDM-by-Ventana-Research.pdf.

[52] T. Friedman, "Case Study : Smith & Nephew Focuses on Data Governance as First Step in MDM Program What You Need to Know," *Gartner, April (ID G00175102)*, 2010.

[53] P. W. Sigmon, "Shattering the data governance myth," *IBM Data Management Magazine*, 2013.

[54] E. Buffenoir and I. Bourdon, "Managing Extended Organizations and Data Governance," in *Digital Enterprise Design and Management 2013*, Springer, 2013, pp. 135–145.

[55] W. McKnight, "Chapter Seven - Master Data Management: One Chapter Here, but Ramifications Everywhere BT - Information Management," Boston: Morgan Kaufmann, 2014, pp. 67–77.

[56] Anand, R. and Hobson, S. F. and Lee, J. and Liu, X. and Wang, Y. and Yang, and Jeaha, "Federation of multi-level master data management systems," U.S. Patent No. 8,635,249, 2014.

[57] Oracle, "All the Ingredients for Success : Data Governance , Data Quality and Master Data Management," 2011.

[58] S. Judah and T. Friedman, "Magic Quadrant for Data Quality Tools," *Gartner, Novemb. (ID G00272508)*, 2015.

[59] S. Orlando, A. White, and J. Radcliffe, "SAP Explains the Next Phase of Master Data Management at What You Need to Know," *Gartner, May (ID G00158070)*, 2008.

[60] T. Friedman and A. Bitterer, "Magic Quadrant for Data Quality Tools What You Need to Know," *Gartner, June (ID G00167657)*, 2009.

[61] T. Friedman and A. Bitterer, "Magic Quadrant for Data Quality Tools What You Need to Know," *Gartner, June (ID G00200603)*, 2010.

[62] J. Radcliffe, T. Friedman, and A. Bitterer, "Informatica Moves Into the MDM Market by Acquiring Siperian," *Gartner, Febr. (ID*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

G00174235), 2010.

[63] T. Friedman and A. Bitterer, "Magic Quadrant for Data Quality Tools What You Need to Know," *Gartner, July (ID G00214013)*, 2011.

[64] T. Friedman, "Magic Quadrant for Data Quality Tools," *Gartner, August (ID G00214013)*, 2012.

[65] T. Friedman, "Magic Quadrant for Data Quality Tools," *Gartner, Oct. (ID G00252509)*, 2013.

[66] S. Judah and T. Friedman, "Magic Quadrant for Data Quality Tools," *Gartner, Novemb. (ID G00261683)*, 2014.

[67] E. Baghi, S. Schlosser, V. Ebner, B. Otto, and H. Oesterle, "Toward a Decision Model for Master Data Application Architecture," in *System Sciences (HICSS), 2014 47th Hawaii International Conference*, 2014, pp. 3827–3836.

[68] T. Frisendal, "Opportunity: Reliable Business Information and MDM," in *Design Thinking Business Analysis*, Springer, 2012, pp. 87–90.

[69] P. Hosangadi, "Instill confidence through solid data quality," *IBM Data Management Magazine*, 2014.

[70] A. White and D. Hope-Ross, "SAP's MDM Shows Potential, but Is Rated 'Caution,'" *Gartner, Sept. (ID G00117327)*, 2003.

[71] P. Bonnet, *Enterprise Data Governance: Reference and Master Data Management Semantic Modeling*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013.

[72] D. Loshin, "Data Governance for Master Data Management and Beyond," 2013.

[73] F. Naumann and M. Herschel, *An Introduction to Duplicate Detection*. Morgan and Claypool Publishers, 2010.

[74] J. R. Talburt, *Entity Resolution and Information Quality*, 1st ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2010.

[75] T. N. Herzog, F. J. Scheuren, and W. E. D. Winkler, *Data Quality and Record Linkage Techniques*. Springer Science & Business Media, 2007.

[76] D. Mahata and J. R. Talburt, "A Framework for Collecting and Managing Entity Identity Information from Social Media," in *19th International Conference on Information Quality (ICIQ)*, 2014.

[77] A. Dreibelbis, E. Hechler, I. Milman, M. Oberhofer, P. van Run, and D. Wolfson, *Enterprise Master Data Management: An SOA Approach to Managing Core Information*. Boston: Pearson Education, 2008.

[78] D. Loshin, "Chapter 4 - Data Governance for Master Data Management," in *Master Data Management*, D. Loshin, Ed. Boston: Morgan Kaufmann, 2009, pp. 67–86.

[79] A. White, "Oracle Improves Its MDM Product Data Offering With Oracle Product Data Quality Cleansing and Matching Server What You Need to Know," *Gartner, April (ID G00167061)*, 2009.

[80] T. Friedman, "Q & A : Common Questions on Data Integration and Data Quality From Gartner â€™ s MDM Summit," *Gartner, Febr. (ID G00164983)*, 2009.

[81] Ofner, M. Hubert, K. Straub, B. Otto, and H. Oesterle, "Management of the master data lifecycle: a framework for analysis," *J. Enterp. Inf. Manag.*, vol. 26, no. 4, pp. 472–491, 2013.

[82] M. Delbaere and R. Ferreira, "Addressing the data aspects of compliance with industry models," *IBM Syst. J.*, vol. 46, no. 2, pp. 319–334, 2007.