AMCIS 2022 Proceedings | SIG EntSys - Enterprise System

Aug 10th, 12:00 AM

Towards Augmented MDM: Overview of Design and Function Areas – A Literature Review

Hendrik Roth
Heilbronn University of Applied Sciences, hroth@stud.hs-heilbronn.de

Simon Paul Mönch
Heilbronn University of Applied Sciences, simon-moe@web.de

Thomas Schäffer
Hochschule Heilbronn, thomas.schaeffer@hs-heilbronn.de

Towards Augmented MDM: Overview of Design and Function Areas – A Literature Review

Completed Research

Hendrik Roth
Heilbronn University of Applied Sciences
hroth@stud.hs-heilbronn.de

Simon Mönch
Heilbronn University of Applied Sciences
smoench@stud.hs-heilbronn.de

Thomas Schäffer
Heilbronn University of Applied Sciences
thomas.schaeffer@hs-heilbronn.de

Abstract

Nowadays, the handling of data is of great importance for companies due to the increasing amount of data by digitalization. Time-consuming tasks in master data management (MDM) must be automated to provide data-driven business models with adequate data quality in real time and thus achieve higher data value. To increase the level of automation in companies, technologies as artificial intelligence are used and applied in information systems, including systems for MDM. The corresponding tasks can be summarized under the term augmented MDM. However, it is not entirely clear which of these processes can fall under the scope of augmented MDM. This paper presents a systematic literature review of 20 examined research articles published in four literature and conference databases to determine design areas and functions of augmented MDM. The findings are one design element "systems" with eleven functions and a proposed definition of terms related to augmented MDM.

Keywords

Artificial Intelligence, Augmented MDM, Data Quality, Master Data Management.

Introduction

The digital transformation represents new business models and the optimization of existing business (Hentschel et al. 2019). As the amount of data increases, it becomes more and more important to ensure that the data contained in information systems has adequate data quality and is available in real time (Leyh et al. 2017; AlSuwaidan 2019). Therefore, companies need a corporate master data management (MDM) (Ren et al. 2020; Otto et al. 2007). Master data is an essential business asset as it represents the fundamental data of an enterprise (Otto et al. 2007). The management of this data is a time-consuming process (Han et al. 2017; Otto et al. 2007). For that reason, the level of automation in companies needs to be increased to make the variety and vast amount of master data manageable today and in the future (Walter et al. 2022; Pontello et al. 2021). The impact of data changes in the business context needs to be well understood, and changes to master data require special precautions (Labadie et al. 2020; Singh and Singh 2022). To automate these tasks, technologies like artificial intelligence (AI), machine learning (ML) and natural language processing (NLP) are used. For this, the company needs a high quality of data to ensure the outcome of the process automation is viable (Walter et al. 2022; Pontello et al. 2021). High data quality is achieved as soon as the data is fit for use referring to the satisfaction of the requirements "fitness for use" (Walter et al. 2022, Wang and Strong 1996).

Augmented Master Data Management (augmented MDM) aims to assist in general tasks that help to maintain master data and keep a high data quality (Judah and White 2020). With the use of diverse techniques different tasks within the information systems of a company are supported, facilitated, improved and automated. Currently, there is no consistent definition of the term "augmented MDM" other than that of Judah and White (2020), who have presented this term as a generic term for discussion. Thus, it is not yet completely clear which processes augmented MDM can ultimately encompass or which new functions can be covered by augmented MDM.

This paper attempts to categorize the term augmented MDM within the scope of MDM. Furthermore, it has the goal to introduce the term augmented MDM to help scientists in the future comprehending the differences between MDM in general and the most recent developments, summarized under the concept augmented MDM. For that reason, we set up the hypothesis that augmented MDM comprises the automation of master data management. To obtain a clear vision, we aim to answer the following research question:

RQ: What are the design areas and associated functions of augmented MDM?

This paper is structured as follows. First, we explain the relevant terms of augmented MDM and the design areas of MDM. In the next step, we present the methodological procedure for the systematic literature analysis. This is followed by the findings where we used a concept matrix to visualize the information. Finally, we summarize the results and derive implications for further research.

Theoretical Background

Design and function areas of MDM

Master data management can be designed based on three areas: strategy level, organization level and system level (Otto and Österle 2015; Otto and Hüner 2009). The strategy level builds the top level of the design areas and contains the master data strategy. The second level is the organizational level and includes master data controlling systems, master data processes as well as methods and the master data organization. The third level, the system level, represents the whole master data architecture and the master data application systems. This level describes which application systems are to be used to implement master data management. That is why there must be clarified which associated functional areas this application systems have to support. There are six functional areas: master data lifecycle management (LCM), metadata management and master data modelling (MEM), master data quality management (DQM), master data integration (INT), cross functions (CRF), and administration (Pontello et al. 2021; Otto and Hüner 2009). Later, we will leave out the administration area as well the subfunctions *search* and *workflow management* of cross functions in this paper because these functions are just necessary for administrative usage of MDM applications and hence not relevant for this research. LCM covers *data creation, data maintenance, data deactivation* and *data archiving* functions. MEM consists of d*ata modelling, model analysis* and m*etadata management*, while DQM includes *data analysis, data enrichment* and *data cleansing* functions. *Data import, data transformation* and *data export* functions are part of master data integration. Cross functions contain a*utomation* and r*eport* functions (Pontello et al. 2021; Otto and Hüner 2009). Figure 1 shows the three design areas and five function areas which were used to fully cover MDM in this research.

Augmented MDM

According to Judah and White (2020) the definition of augmented MDM is as follows: „Augmented MDM is the application of graph, *AI/ML*, *NLP* and similar technologies to master data management. Augmented MDM extends traditional MDM capabilities to reduce some manual data stewardship and discovery tasks and enables the creation of contextual application data by exposing previously unknown relationships between master and application data attributes." Deduced the goal of augmented MDM is to minimize and, if possible, completely or partly automate the manual tasks of master data management. In consequence Augmented MDM solutions enable higher levels of automation in data governance and stewardship tasks and can indirectly lead to improved business outcomes for example by increasing the accuracy and consistency of governance processes by using AI/ML, such as entity resolution (Judah and White 2020). By this definition we can comply that augmented MDM is supporting the digital transformations requirements in between master data objects and transaction data. Judah and White (2020) also mention that it is currently not possible to integrate these functions into every MDM use case that relates to governance processes and operational MDM.

Research Methodology

To identify the design areas and associated functions of augmented MDM we conducted a literature review based on the context of augmented MDM. Therefore, we used the methods of vom Brocke et al. (2009) and Webster and Watson (2002). For the literature search the AIS Electronic Library, IEEE Xplore Digital Library, Science Direct and Springer Link were used.

Since there is no literature that contains the term „augmented MDM” except the publication from Judah and White (2020), we extended our search term with the topics Judah and White (2020) called in context of augmented MDM. Consequently, we used the following search term to perform electronic searches: [((“master data” OR “master data management”) AND (“nlp” OR “graph” OR “machine learning” OR “automation”)) OR „augmented MDM"]. There was no specific time selection. The searches resulted in a total of 777 publications. After excluding duplicates and irrelevant papers by screening each abstract and keywords or, if necessary, the full article for checking thematic relevance, 18 relevant publications for our research focus were left. Based on this literature we performed a forward and backward search as Webster and Watson (2002) recommend, hence the number of relevant publications increased to 29. We left out 9 papers that described only technologies such as NLP or ML from a technical or mathematical perspective, but it was unclear whether this could be applied to master data, since master data was not referenced. Finally, there were 20 relevant papers left.

The 20 relevant publications were read in their entirety and analyzed using Mayring's (2000) qualitative content analysis method. As categories the references of Otto and Hüner (2009) and Pontello et al. (2021) for the design areas and function architecture of MDM were used. In the first iteration the papers were assigned to design areas by using a four-eye-principle as proposed by Peffers et al. (2012). After this we observed that all the papers could be categorized into the system level, mostly in application systems of master data. This is the reason why there was a need to assign the functional categories and areas of MDM based on Pontello et al. (2021) and Otto and Hüner (2009) as a second iteration with a four-eye-principle by Peffers et al. (2012) as well. Figure 1 visualizes the research methodology for comprehension.

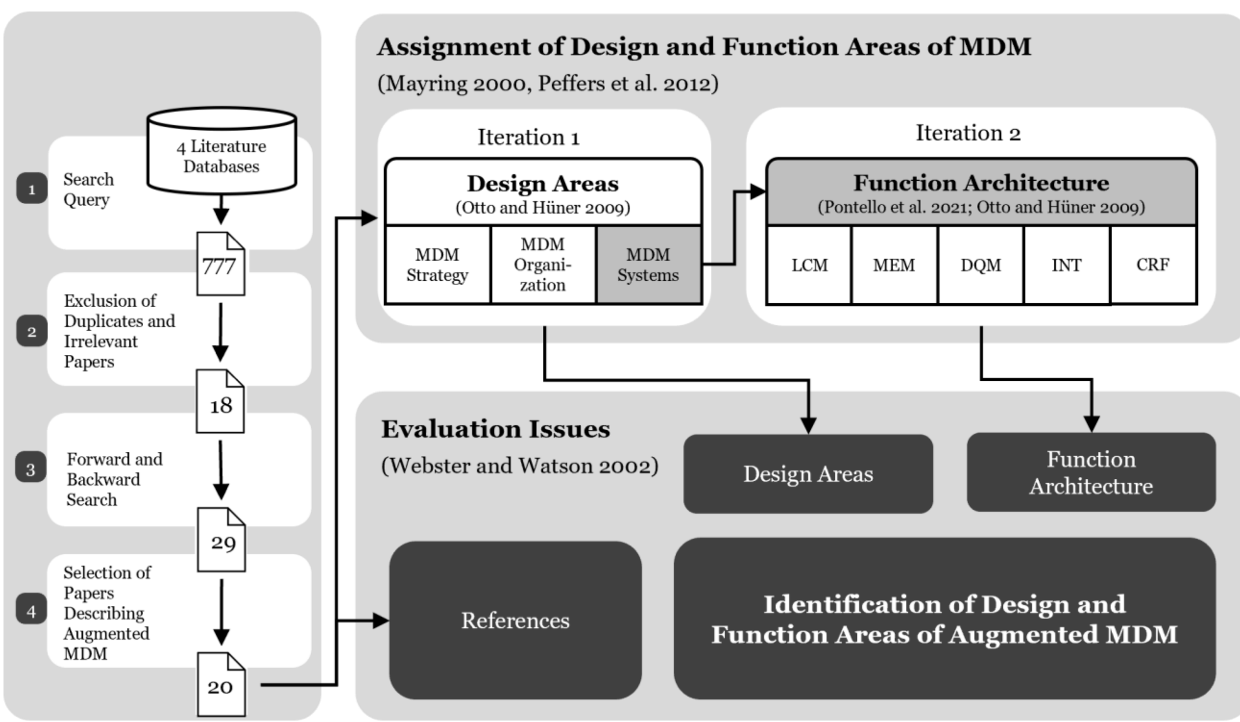

Figure 1. Overview of the research methodology

Findings

In the following section we will describe the findings regarding the design and associated functional areas covered by the literature review in the context of augmented MDM.

To accomplish this, table 1 gives an overview of the design and function areas as identified in the literature. All papers are at the design area system level, because each of them describes the technical realization of an architecture or specific functions of an application system. For this reason, it can be stated that augmented MDM involves system-level implementation. The associated functions are described in detail later.

	Design Areas			**Function Architecture**														
				LCM				**MEM**			**DQM**			**INT**			**CRF**	
Reference	MDM Strategy	MDM Organization	MDM Systems	Data Creation	Data Maintenance	Data Deactivation	Data Archiving	Data Modelling	Model Analysis	Metadata Management	Data Analysis	Data Enrichment	Data Cleansing	Data Import	Data Transformation	Data Export	Reporting	Automation
AlSuwaidan 2019			X								X		X	X	X	X		X
Altendeitering 2021			X						X		X		X	X	X	X		X
Altendeitering and Tomczyk 2022			X								X	X	X	X	X	X		X
Bodganov et al. 2020			X						X		X		X	X		X	X	X
Dong et al. 2018			X								X		X	X		X	X	X
Ehrlinger et al. 2019			X						X		X			X		X	X	X
Han et al. 2017			X						X									X
Kern et al. 2021			X						X	X			X	X				X
Kulkov 2021			X								X		X					X
Lettner et al. 2021			X						X		X		X	X		X	X	X
Lohmer et al. 2021			X							X		X	X	X		X		X
Mahanta and Mohamed 2020			X						X		X			X			X	X
Pontello et al. 2021			X						X				X					X
Ren et al. 2020			X		X			X	X				X	X	X	X		X
Salem and Abdo 2017			X						X		X		X		X			X
Sawarkar and Kodati 2021			X						X	X	X		X					X
Vasiliev et al. 2021			X					X	X					X	X	X	X	X
Walter et al. 2022			X								X						X	X
Yi and Ye 2020			X										X					X
Zhao et al. 2020			X					X	X		X		X	X				X

Table 1. Overview of Design Areas and Function Architecture for augmented MDM

Lifecycle Management (LCM)

The functional area lifecycle management contains the functions of d*ata creation, data maintenance, data deactivation* and *data archiving*. Some functions which are in the scope of lifecycle management are handled as well by other functions (e.g., data quality management). The mentioned functions in lifecycle management do refer to general and security aspects (e.g., false data format cannot be implemented) of MDM ensuring that the data management is reasonable with functions like conditional prices (e.g., buy more, pay less) (Otto and Hüner 2009). The only authors who describe *data maintenance* in the context of lifecycle management are Ren et al. (2020). Data is pulled from multiple companies and then stored separately in the appropriate layers for further processing, while maintaining the data in every layer, until it is added into the ontology layer.

Metadata Management (MEM)

This section describes the function architecture within the tasks of the metadata management which include d*ata modelling, model analysis* and m*etadata management* itself. Most papers refer to the *model analysis* in metadata management.

Data Modelling

Zhao et al. (2020) simulate a master data network and model the data entities in order to visualize data sharing and linkage among supply chains. With this network model the authors are able to describe and quantify multi-type relationships among the entities (Zhao et al. 2020). Hence, *data modelling* enables the users to classify master data (Zhao et al. 2020; Otto and Hüner 2009). Visualizing data models can be supported by using *visual notation for OWL ontologies* (*VOWL)* on *GraphDB* (Vasiliev et al. 2021). *VOWL* automatically creates a visualized version of the *GraphDB* which then can be stored as a versioning of the recent database schema including the database schema itself (Vasiliev et al. 2021).

Model Analysis

Due to the rising amount of data, domain experts cannot handle the manual identification of dependencies or relations of master data sets because it is time-consuming (Han et al. 2017). Using *association rule mining*, large datasets can be automatically correlated based on variables' characteristics (Han et al. 2017). For this task, the *support vector machine* algorithm can be used in combination with *association rule learning* (Altendeitering 2021). It is necessary to recognize relationships between data types for a better understanding of master data. Therefore, *unsupervised learning* like *k-means* or *principal component analysis* (PCA) can be used (Pontello et al. 2021; Kern et al. 2021). By using a *graph database*, relations between data can be better found automatically as well as ontology analyses can be better performed with *machine learning* algorithms (Ren et al. 2020). Functional dependencies can be identified with *miner algorithms* (Salem and Abdo 2017). The use of *entity resolution* and *contextual embedding methods* is for capturing relationships among data columns (Sawarkar and Kodati 2021). *Set pair analysis* makes it possible to find indirect relationships between table-to-table master data in the supply chain (Zhao et al. 2020).

Metadata Management

One of the subfunctions of metadata management is the identification of metadata. Here, there is automation potential with *support vector machine* algorithm for identifying metadata (Kern et al. 2021). The *support vector machine* algorithm can be also used for *data cleansing* and detection of duplicates as well as supporting in glossary creation (Kern et al. 2021). For supporting *metadata management*, *distributed ledger technology* (DLT) can help for creating reliable metadata due to digital signatures (Lohmer et al. 2021). Visualization of metadata is also part of augmented MDM because *unsupervised machine learning* algorithms can support automated visualization (Kern et al. 2021).

Data Quality Management (DQM)

The data quality management section describes *data analysis, data enrichment* and *data cleansing*. In this section, special attention is given to d*ata analysis* and *data cleansing*. *Data analysis* and d*ata cleansing* are frequently occurring tasks in the field of data quality management, in contrast to *data enrichment*. *Data enrichment* has two entries, whereas *data cleansing* and *data analysis* have 28 entries in sum.

Data Analysis

Data analysis refers to different functions like rule-based plausibility checks, data quality measurements, compliance checks, graphical analyses, real-time analyses, and statistical evaluations (Bodganov et al. 2020; Altendeitering 2021; Mahanta and Mohamed 2020; Otto and Hüner 2009). It is possible to evaluate incoming data against automatically defined data quality rules based on *quality rule mining* (Altendeitering 2021). This reduces errors and time-consumption. Moreover, data quality tools are available that assist in detecting inconsistencies or automatically execute data quality analysis tasks when necessary (Altendeitering and Tomczyk 2022). In addition, data sets can be analyzed and hence measured for data quality by *support vector machine* algorithm (Bogdanov et al. 2020). Automatically finding anomalies with *ML* for improving data quality can also be done (Bogdanov et al. 2020). Prediction analysis is also supported by *AI* techniques (Bogdanov et al. 2020). Data detection, detection and processing of inconsistent data can be automated with algorithms like *k-nearest neighbor*, too (Dong et al. 2018). With *entity models*, automated rule-based data quality checks can be performed (Ehrlinger et al. 2019; Lettner et al. 2021). Automated rule-based data quality controls through multiple data quality dimensions by analyzing data sets can be done with *ML* (Walter et al. 2022). Real-time analytics are also part of augmented MDM by a simulated quality profile without human intervention (Mahanta and Mohamed 2020).

Data Enrichment

Data enrichment combines or compares internal data with external data in order to identify possible errors in the company's own data stock (Otto and Hüner 2009; Altendeitering and Tomczyk 2022). *Distributed ledger technologies* such as *blockchain* ensures that the completeness of the data is guaranteed and signatures received from network wide actors make changes easily traceable and thus increase the trustworthiness of the system (Lohmer et al. 2021). In addition, *data enrichment* has the function to add value to the existing data and improve the understanding for business users (Otto and Hüner 2009; Altendeitering and Tomczyk 2022).

Data Cleansing

An automated *data cleansing* process is unavoidable for enhancing reliable data (Kulkov 2021). Moreover, it is essential for MDM because otherwise an organization could make bad decisions due to bad data quality which would lead to more costs (AlSuwaidan 2019; Salem and Abdo 2017). There are multiple automated *data cleansing* methods for different data types (AlSuwaidan 2019). The automated data cleaning process includes tasks such as filling in missing values and removing whitespaces (Altendeitering and Tomczyk 2022). Real-time *data cleansing* is possible with *DLT* (Bogdanov et al. 2020). With *DLT* like *blockchain technology*, duplications can also be reduced, and the accuracy of the data can be ensured (Lohmer et al. 2021). Another automated *data cleansing* task is clearing text data using *NLP* (Bogdanov et al. 2020). With *random forest* algorithm it is possible to identify duplicate records (Bogdanov et al. 2020). Methods based on trees or *support vector machine* algorithm can help prevent invalid data from entering an enterprise system (Pontello et al. 2021). Additionally, an automated *data cleansing* system identifies incorrect data and generates possible data repairing proposing (Ren et al. 2020). Furthermore, there is a *t-repair* algorithm for data repairing based on miner technique (Salem and Abdo 2017). Pattern recognition is part of augmented MDM either. There are a few simple *AI* algorithms possible for pattern recognition like *k-nearest neighbors* or *linear regression* (Pontello et al. 2021) as well as *deep neural networks* (Kulkov 2021). The comparison and identification of duplicates can be done with *entity resolution* (Bogdanov et al. 2020). An *entity recognition* algorithm can be used to parse and standardize data values (Dong et al. 2018). However, there is a practical workflow for automated master data cleansing too. Firstly, the data will be corrected and cleaned using *NLP* like *trie* and *wordnet*. Secondly, the data attributes and their values can be automatic split using e.g., *lattice-LSTM (long short-term memory*) (Yi and Ye 2020). A *knowledge-based data cleaning system* by crowdsourcing can be used for automatic data cleansing as well (Zhao et al. 2020). Data can be distinguished as valid and invalid and *top-k possible repairs* can be generated by the system (Zhao et al 2020).

Integration (INT)

This chapter describes the identified functions in the context of data integration. Especially *data import* and *export* are described, while the *transformation* is mentioned mostly in the context of extract, transform, load- (ETL) processes.

Data Import and Export

Data import and export is necessary to process data from one system into another (Otto and Hüner 2009; Vasiliev et al. 2021). Therefore, many automated or semi-automated *ETL*-processes are required to reduce the time exposure and connect the enterprise systems of a company (Salem and Abdo 2017). MDM systems often offer *ETL*-functions (or pipelines) and *API*s to process the data according to the system or user needs (Vasiliev et al. 2021; Altendeitering et al. 2021). These functions are essential to ensure that an appropriate data format is used without the need to manually transform the data (Vasiliev et al. 2021).

Data Transformation

Overall, the transformation is frequently mentioned within *ETL*-processes (Vasiliev et al. 2021; Altendeitering and Tomczyk 2022). The purpose of *data transformation* is to store data in a proper format or structure for querying and analysis purposes (Salem and Abdo 2017). Thus, *data transformation* is crucial when data is processed from one system into another (Ren et al. 2020). Salem and Abdo (2017) mention that *data mining* provides methods and technologies to reduce errors and minimize manual work steps in transformation processes, as well as providing useful information for decision making purposes.

Cross functions (CRF)

In this subsection, functions of the functional area *reporting* and *automation* are described. Especially the functional area *automation* is relevant for this research, because it is the only functional area to which every paper was assigned.

Reporting

Due to better data analysis opportunities by using *AI*, it is possible to have real-time quantification of data quality. This makes real-time data quality reports in an aligned data quality process possible (Bogdanov et al. 2020). Furthermore, there is a real-time data quality control system with automated alarm monitoring and information if there is problem (poor quality) data detected (Dong et al. 2018). Other data quality measurement tools like *DaQL* or *DaQL 2.0* support automatically visual analytics generation with results of the automated measured data quality (Ehrlinger et al. 2019; Lettner et al. 2021). To sum this up, automatic data measurement/control techniques with e.g., *ML*, can be integrated into a data quality dashboard for sending the results of the quality controls directly to a dashboard and hence generate automatically reports (Walter et al. 2022).

Automation

Table 1 shows, that every literature we analyzed describes automation processes of diverse tasks without human involvement. In this context, automation is distinguished from other functions because only the other functions mentioned are described as automated. In this respect, automation cannot be regarded as a function on its own but is rather a state. Thus, the mentioned functions are part of Cross function *automation* (Otto and Hüner 2009). Especially data quality management and metadata management tasks are currently discussed and automated in the scope of augmented MDM as table 1 reveals. As previously described, the automation through in this context identified technologies leads to reducing time-consuming tasks, makes things more efficient and gives new opportunities in enterprise systems such like real-time data quality checks due to advanced data analysis.

Table 2 provides an overview of the technologies used to assist MDM and assigns them to the function architecture of MDM as identified above. For example: *Association Rule Learning* was mentioned in the subfunction *Model Analysis* of *Metadata Management* (MEM) → it is marked with an 'x'.

	Function Architecture														
	LCM				**MEM**			**DQM**			**INT**			**CRF**	
Technologies	Data Creation	Data Maintenance	Data Deactivation	Data Archiving	Data Modelling	Model Analysis	Metadata Management	Data Analysis	Data Enrichment	Data Cleansing	Data Import	Data Transformation	Data Export	Reporting	Automation
Association Rule Learning						x									x
Association Rule Mining						x									x
Blockchain									x						x
Contextual Embedding						x									x
DaQL/DaQL 2.0														x	x
Deep Neural Networks										x					x
DLT							x			x					x
Entity Recognition										x					x
Entity Resolution						x				x					x
ETL-Pipelines		x									x	x	x		x
Graph DB						x									x
K-Means						x									x
K-Nearest Neighbor								x		x					x
Lattice-LSTM										x					x
Linear Regression										x					x
Miner Algorithms						x				x					x
NLP										x					x
PCA						x									x
Quality Rule Mining								x							x
Random Forest										x					x
Set Pair Analysis						x									x
Support-Vector-Machine						x	x	x		x					x
Top-K Possible										x					x
T-Repair										x					x
Trie										x					x
VOWL					x										x
Wordnet										x					x

Table 2. Assignment of identified Technologies to Function Architecture

Discussion

As far as we know, this is the first research besides Judah and White (2020) that identifies areas of augmented MDM and clarifies the term augmented MDM in research context. Our findings are all part of system level design area. This does not mean that augmented MDM must be in this design area of master data management. Augmented MDM is more considered to describe automation and extension of MDM as all papers analyzed are about automating functions of MDM. Hence, the thesis that augmented MDM includes the automation of MDM, can be confirmed. If this can be adopted to strategic or organizational design and function areas, it is augmented MDM too. While identifying the automatized functions by the literature, it was also possible to detect technologies that came along with these functions (see table 2). After bringing all previous findings together, we can consolidate the definition of augmented MDM of Judah and White (2020) by including the automation aspect. Thus, we present the deduced definition:

„Augmented MDM is the application of recent technologies to master data management. Augmented MDM extends and automates traditional capabilities to reduce some manual data stewardship, discovery and data quality management tasks and enables the creation of contextual application data by exposing previously unknown relationships between master and application data attributes."

Augmented MDM is used to make master data management more efficient due to automation and gains new opportunities that are not possible without used technology. Thereby augmented MDM includes associated functions based on system level which are *metadata management*, *master data modeling*, *master data quality management*, *lifecycle management, master data integration*, and *cross functions*.

Conclusion

This paper gives an overview of design areas and associated functions which are covered by augmented MDM. Our research shows that augmented MDM currently includes functions at the system level design area. This level is decisive for the quality of the data available in the company's information systems and is therefore crucial for the company's business viability especially for data driven business. The focus is set on automating tasks of metadata management, master data quality management, master data integration, and cross functions such as reporting. In this context, there were some technologies identified that fall under augmented MDM. Since the paper did not focus on the technologies mainly, but rather they have come along through the functions, further research can systematically analyze which technologies augmented MDM may cover. Perhaps by simply looking at technologies, additional functions or design areas can be identified that fall under augmented MDM. This can be an approach for function areas regarding the white spots in table 1, e.g., acquire and creation processes like automating *data creation*, *data enrichment* and *data transformation*, as well as *data maintenance*. Moreover, there may be ways for applying *ML* at strategy level for data governance functions, e.g., automatically choosing roles and responsibilities by automated analysis who does what in the past. Furthermore, the term augmented MDM was demystified and can be used from now on in future research whenever it comes to automizing tasks in the context of master data management with tools like *AI/ML*, *NLP*, *DLT*, *data mining* and *graph* for gaining benefits. Future work may consider looking at new application areas in augmented MDM such as organizational and strategic level. Likewise, new concepts can be created that enable partial automation at the very beginning e.g., an information system based on *AI*. This implies that the information system is able to provide data in real time with appropriate data quality. Furthermore, it would be interesting to create an indicator for the evaluation of the degree of automation possible with the help of augmented MDM. This could determine how many functions, processes and tasks can be automated at all with the help of augmented MDM as well as answering the question at which point it is helpful to consider the usage of augmented MDM.

References

AlSuwaidan, L. 2019. "Data Management Model for Internet of Everything," in *Mobile Web and Intelligent Information Systems*, I. Awan, M. Younas, P. Ünal and M. Aleksy (eds.), Cham: Springer International Publishing, pp. 331-341.

Altendeitering, M. 2021. "Mining Data Quality Rules for Data Migrations: A Case Study on Material Master Data," in *Leveraging Applications of Formal Methods, Verification and Validation*, T. Margaria and B. Steffen (eds.), Cham: Springer International Publishing, pp. 178-191.

Altendeitering, M., and Tomczyk, M. 2022. "A Functional Taxonomy of Data Quality Tools: Insights from Science and Practice," *in Proceedings of Wirtschaftsinformatik 2022 (Vol. 4)*.

Bogdanov, A., Degtyarev, A., Shchegoleva, N., and Khvatov, V. 2020. "Data Quality in a Decentralized Environment," in *Proceedings of ICCSA 2020*, pp. 58-71.

Dong, X., He, H., Li, C., Liu, Y., and Xiong, H. 2018. "Scene-Based Big Data Quality Management Framework," in *Proceedings of ICPCSEE 2018,* pp. 122-139.

Ehrlinger, L., Haunschmid, V., Palazzini, D., and Lettner, C. 2019. "A DaQL to Monitor Data Quality in Machine Learning Applications," in *Proceedings of DEXA 2019*, pp. 227-237.

Han, W., Borges, J., Neumayer, P., Ding, Y., Riedel, T., and Beigl, M. 2017. "Interestingness Classification of Association Rules for Master Data," in *Proceedings of ICDM 2017,* pp. 237-245.

Hentschel, R., Leyh, C., and Baumhauer, T. 2019. “Critical Success Factors for the Implementation and Adoption of Cloud Services in SMEs,” in *Proceedings of HICSS,* 2019.
Judah, S., and White, A. 2020. “Hype Cycle for Data and Analytics Governance and Master Data Management,” available at https://www.gartner.com/en/documents/3987607, *accessed on Apr 24 2022*.
Kern, C. J., Schäffer, T., and Stelzer, D. 2021. “Towards Augmenting Metadata Management by Machine Learning,” in *Proceedings of INFORMATIK 2021,* pp. 1467–1476.
Kulkov, I. 2021. “The role of artificial intelligence in business transformation: A case of pharmaceutical companies,” *Technology in Society* (66).
Labadie, C., Eurich, M., and Legner, C. 2020. “Empowering Data Consumers to Work with Data: Data Documentation for the Enterprise Context,” in *Proceedings of Wirtschaftsinformatik 2020*.
Lettner, C., Stumptner, R., Fragner, R., Rauchenzauner, F., Ehrlinger,. 2021. “DaQL 2.0: Measure Data Quality based on Entity Models,” *Procedia Computer Science* (180), pp. 772-777.
Leyh, C., Gebhardt, A., and Berton, P. 2017. “Implementing ERP Systems in Higher Education Institutes: Critical Success Factors Revisited,” in *Proceedings of FedCSIS,* pp. 913-917.
Lohmer, J., Bohlen, L., and Lasch, R. 2021. “Blockchain-Based Master Data Management in Supply Chains: A Design Science Study,” in *Proceedings of APMS,* pp. 51-61.
Mahanta, P., and Mohamed, A.-G. 2020. “A Hybrid Approach to Insightful Business Impacts,” in *Proceedings of OTM 2019,* pp. 155-160.
Mayring, P. 2000. “Qualitative Content Analysis,” *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research,* Art. 20, http://nbn-resolving.de/urn:nbn:de:0114-fqs0002204, *accessed on Apr 24 2022*.
Otto, B., and Hüner, K. 2009. “Functional Reference Architecture for Corporate Master Data Management,” 2009, BE HSG / CC CDQ / 21.
Otto, B., and Oesterle, H. 2015. “Corporate Data Quality: Prerequisite for Successful Business Models”.
Otto, B., and Reichert, A. 2010. “Organizing Master Data Management: Findings from an Expert Survey,” in *Proceedings of ACM Symposium on Applied Computing 2010*, pp. 106-110.
Otto, B., Weber, K., Schmidt, A., and Osl, P. 2007. “Towards a Framework for Corporate Data Quality Management,” in *Proceedings of ACIS 2007*.
Peffers, K., Rothenberger, M., Tuunanen, T., and Vaezi, R. 2012. “Design Science Research Evaluation,” in *Proceedings of DESRIST 2012,* pp. 398-410.
Pontello, V., Beckmann, H., and Lanquillon, C. 2021. “Meta-learning approach for implementation of AI methods in the context of CRISP-DM with case studies from master data management,” in *Proceedings of ICE/ITMC 2021*, pp. 1-9.
Ren, L., Zhang, Z., Zhao, C., and Zhang, G. 2020. “Cloud-Based Master Data Platform for Smart Manufacturing Process,” in *Proceedings* of *SmartGIFT 2020*, pp. 163-170.
Sawarkar, K., and Kodati, M. 2021. “Automated Metadata Harmonization Using Entity Resolution and Contextual Embedding,” in *Proceedings of Intelligent Computing 2021*, pp. 129-138.
Singh, S., and Singh, J. 2022. “A Survey on Master Data Management Techniques for Business Perspective,” in *Proceedings of CIIR 2022*, pp. 609-617.
Vasiliev, D. A., Ghiran, A., Buchmann, R. A. 2021. “Evaluation of Data Integration Plans based on Graph Data,” *Procedia Computer Science* (192), pp. 1041-1050.
vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., and Cleven, A. 2009. “Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process,” in *Proceedings of ECIS 2009*.
Walter, V., Gyoery, A., and and Legner, C. (2022). “Deploying machine learning based data quality controls – Design principles and insights from the field,*” in Proceedings of Wirtschaftsinformatik 2022*.
Wang, R.Y., and Strong, D.M., 1996. “Beyond Accuracy: What Data Quality Means to Data Consumers,” J. Manag. Inf. Syst. (12), pp. 5-33.
Webster, J., and Watson, R. T. “J. Webster and R. T. Watson. 2002. Analyzing the Past to Prepare the Future: Writing a Literature Review,” MIS Quarterly (26:2), pp. 13–23.
Yi, S., and Ye, H. 2020. “A Practical Workflow in Cleaning Master Data,” in *Proceedings* of *ISPA/BDCloud/SocialCom/SustainCom 2020*, pp. 1311-1314.
Zhao, C., Ren, L., Zhang, Z., and Meng, Z. 2020. “Master data management for manufacturing big data: a method of evaluation for data network,” in *Proceedings of World Wide Web* (23:2), pp. 1407-1421.