ARTICLE IN PRESS

Computers in Human Behavior xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Computers in Human Behavior

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

MD3M: The master data management maturity model

Marco Spruit *, Katharina Pietzka

Department of Information and Computing Sciences, Utrecht University, The Netherlands

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Master data management
Maturity assessment
Information management

ABSTRACT

This research aims to assess the master data maturity of an organization. It is based on thorough literature study to derive the main concepts and best practices in master data maturity assessment. A maturity matrix relating 13 focus areas and 65 capabilities was designed and validated. Furthermore, an assessment questionnaire was developed which can be used to assess the master data management maturity. Emphasis is laid on the academic validity of the model development process. Our extensive case study provides an example of iterative human learning, behavior and collaboration resulting from technological needs in a large-scale infrastructural network. Concludingly, this research uncovers reasons and incentives for prudent master data management and provides a benchmarking tool with which different organizations can compare their levels.

1. Introduction

Information and data became increasingly important and a crucial competitive factor over time. Additional to the three sector theory, a quarternary sector has been defined. That additional sector contains information-based activities (Beniger, 1986; Kenessey, 1987). This shows the development towards an information-centered economy. Research in this area therefore has a significant economic and societal relevance.

Master data are the data describing the most relevant business entities, on which the activities of an organization are based, e.g. counterparties, products or employees. In contrast to transactional data (invoices, orders, etc.) and inventory data, master data are oriented towards the attributes. They describe the main characteristics of objects in the real world. Single master data entities are rarely being changed, for instance the properties of some kind of material. Instances of master data classes are relatively constant, especially if they are compared with transactional data. Master data is the reference for transactional data. There would not be a single order or delivery without master data (Otto & Hüner, 2009).

Starting from a particular size, every organization has to deal with the question of how to integrate master data from different units or areas. Furthermore, the organizational setup of the firm plays a big role. Does the firm operate in a data-intense business? In areas with strict regulations on the traceability of events and accountability (like pharmaceutical industries, finance, trading), with data as a main source for added value (like the finance industry), or with the urgent need for efficiency and agility (production, technical industries), a systematic integration of data is crucial for the business (Wegener, 2008). The recent development shows that organizations have to cope with short innovation cycles and market launch times. Furthermore, the complexity is increasing due to globally harmonized business processes and global customer services. This results in shorter decision cycles basing on more information (Kumar, 2010; Otto & Hüner, 2009).

As stated before, there are many reasons why a company might consider Master Data Management (MDM), which we define as *"the management of the consistent and uniform subset of business entities that describe the core activities of an enterprise"*. It is general consensus and common sense that correct, available and timely data are of great importance and can be a competitive advantage (Borghoff & Pareschi, 1997; Kahn, Strong, & Wang, 2002; Otto & Hüner, 2009). However, many companies have insufficient data management strategies. Especially bigger companies struggle with the huge amount of data and have no sufficient strategy to exploit the data (Davenport & Prusak, 2000; Otto & Hüner, 2009).

The objectives of this research paper are of both a practical and an academic nature. From a corporate point of view, the objective is to give organizations the possibility to assess their own MDM maturity and benchmark against other organizations. This situation leads to the following research question: *How can a company's current state in Master Data Management be measured to identify potential improvement areas?*

The remainder of this paper is structured as follows. The following chapter introduces the research approach that was followed and on which the whole research is based on. Then the maturity model is presented. Afterwards, the validation is presented. The research is discussed and conclusions are drawn as well as fields for further research presented.

* Corresponding author.
E-mail address: m.r.spruit@uu.nl (M. Spruit).

http://dx.doi.org/10.1016/j.chb.2014.09.030
0747-5632/

2. Research approach

To assess the maturity of the master data management of an enterprise, we propose the MDM maturity model. The MDM maturity model is a means of assessing the whole process of master data management including the data point of view and also focusing on the whole operational process.

In order to keep this research consistent according to academic requirements, the development will be based on guidelines and frameworks from academia. As already mentioned in earlier sections, this research is based on the design science approach described by Hevner, March, and Park (2004). It is also in accordance with the research of Becker, Knackstedt, and Pöppelbuß (2009) who published a paper on the development of maturity models for IT management and proposed a procedure model for the development. Also they base their research on the design science approach to achieve a reasonable catalog of requirements for the design of maturity models.

According to Becker et al. (2009) a maturity model is an artifact that aims at solving the problem of defining an organization's current status regarding their capabilities and deriving means for improvements. They developed a procedure consisting of eight steps for developing scientifically valid maturity models in accordance with Hevner et al. (2004).

The first step is the comparison with existing maturity models. Researchers should have an overview about what already exists, so that their development can be based on already existing models or improve an already existing one. This instruction takes into account Hevner's 'Guideline 6: Design as an artifact' and 'Guideline 4: Research Contributions'. Hevner's sixth guideline: 'Design as a Search Process', which implies the need of an iterative solution development, refining, evaluating and possible enhancement, is considered in the second and third step. The second advises an iterative procedure saying that models must be developed step by step and the third gives advice about the evaluation. All principles and preconditions for developing a maturity model as well as usefulness, quality and effectiveness must be evaluated iteratively. This one is also in accordance with Hevner's third guideline about evaluation, stating that results must be evaluated with appropriate scientific grounding. The next instruction deals with the topic of multi-methodological procedure. The development of maturity models employs a variety of research methods, where the application needs to be well-founded and finely tuned. Here, the guideline of 'Research Rigor' stating that selected methods have to be rigorously attuned is applied. Hevner's second guideline about 'Problem relevance' recommends that the problem solving artifacts should be innovative but also relevant to researchers and/or practitioners. This requires thus a precise definition of the problem and therefore is in line with the fifth and sixth procedural step; identification of the problem relevance and definition of the problem. These imply that the solution's relevance must be demonstrated and that the future application domain, the conditions and the benefits must be defined before designing the model. Hevner's seventh guideline (Communication of Research) finds place in the last two principles. The results should be targeted at specific user groups. The rule 'Targeted Presentation of Results' indicates that the presentation of the model must be targeted at an audience regarding the conditions of the applications. The last one (Scientific Documentation) recommends the detailed documentation of the design process, taking into consideration every relevant step in the design process (Becker et al., 2009; Hevner et al., 2004).

The development of the master data management maturity model followed the guidelines presented above to ensure the validity for an academic design science research.

3. The master data management maturity model: MD3M

3.1. Maturity levels

The following Table 1 gives an overview of the amount of maturity levels and their meaning based on IT Governance Institute (2000), Butler (2011), Kumar (2010) and Loshin (2010), respectively.

From this information, the decision was taken to exclude a level with a non-existing maturity. The ignorance of existing issues within the organization concerning master data management will be considered as no maturity at all and therefore not be considered in the matrix as a unique level. Hence, the first level will be – like it is consensus among the different models – an initial one. On this level, the attention has just been raised for the topic and initial plans are developed to investigate and tackle the problem. Consequentially, there are isolated measures initiated by single units or persons in the organization without relation to others. They are meant to solve internal problems of the particular unit that are due to the insufficient MDM in place. On the third level, the initiatives start to be aligned among each other and awareness is in place for the initiatives and problems of other units. This level is called organized. Different units start to collaborate for certain projects. On the following level, the initiatives tend to be best practices. The organization adopted common frameworks and implemented them. Processes have been defined and are adhered to. The fifth level is called optimized. On this level, all processes concerning MDM are optimized for the purposes of the organization beyond the best practices. The maximum of benefits can be drawn from the MDM initiatives and it is constantly reviewed whether the circumstances change.

The following Table 2 depicts the chosen maturity levels for the MDM maturity model and a short description. They are mainly based on the COBIT model earlier described, except that here the level zero is left out as an explicit maturity level. The COBIT framework is based on the same perspective like this research and it provides the most widely known maturity level stages. Furthermore, it is not too detailed for this top-down view and since this is an initial prototype, the overall view seems to be the best approach. Achieving a maturity level means all capabilities of the stage are fulfilled.

3.2. Key topics and focus areas

The key topics and the focus areas were developed with a bottom up approach. They were deployed after a thorough analysis of the literature study of existing MDM models and other literature and studies on the topic. They were chosen to cover all aspects of master data management that are relevant for an organization. An appropriate granularity was developed. All factors from the presented models are somehow covered in the new MDM maturity model, even though the order is different. This is due to the finer granularity and the structuring with the key topics and the focus areas within. Furthermore, there was practical input from the company regarding topic-unrelated practices. So the only input for this model was general practice that a company follows, but no company-specific factors. The model is designed to fit in general to all – especially bigger – companies dealing with master data. For small enterprises, the effort to implement elaborate MDM would be exaggerated. The model was deployed and validated in a loop approach. The initial draft mainly based on literature was presented to experts within the company and evaluated with them. Iteratively, the final matrix was developed, ensuring that it covers all important aspects of the topic.

Concerning the key topics, the goal is to find mutually exclusive and logical chunks in which the focus area can be grouped

Table 1
A comparison of existing maturity levels in master data management.

Level	COBIT	Oracle	IMN	DataFlux
0	Non-existent	–	–	–
1	Initial	Marginal	Initial	Initial
2	Repeatable		Isolated	Reactive
3	Defined process	Stable	Organized	Managed
4	Managed and measurable	Best practice	Unified	Practice
5	Optimized	Transformational	Optimized	Strategic performance

Table 2
Description of the derived MD3M maturity levels.

Level	Description
1: Initial	A first awareness for issues regarding the topic of MDM has been raised on an operational level. Initial steps are initialized
2: Repeatable	Measures from individuals are conducted to solve individual problems. No connection to other units or projects. Still operational
3:Defined process	First collaborations take place on a tactical level. Awareness was created for the existence of other initiatives
4: Managed and measurable	Best practices are in place for handling of MDM. There are defined processes on a tactical level
5: Optimized	Optimized handling of MDM. The organization's efficiency has been improved. Tactical approach on the topic

logically. Resulting from the previous research and the focus on MDM, the choice is to have a topic-oriented approach. This approach is chosen because processes might look different across different companies. If the MD3M was too much focused on processes, it would not be generic anymore and could not be applied by all companies. The other option would be to keep the process descriptions very general that it would still work for every company. Therefore, the model is organized around topics that can be included in any process depending on the organizational setup.

There are five key topics identified, with thirteen focus areas in total:

Data model. This key topic deals with the data and the infrastructural and organizational view on it. It contains topics like what data is considered as master data, how the data is structured, which systems use what data, and where the data is stored.

Davenport and Prusak mention the aspects of 'Contextualization', meaning that it must be clear for which purpose data is collected, 'Categorization' meaning that key components are known and 'Condensation', meaning a summary of the data is essential for overview purposes (Davenport & Prusak, 2000). It is crucial to have a shared understanding of the data in the organization. Shared definitions help align the understanding of different stakeholders with e.g. technical and non-technical perspectives (EDUCAUSE, 2009; IBM, 2007). It is important to have one single view on the data and on the systems that use the data in order to consolidate different sources into one for the whole enterprise (IBM, 2007; Loshin, 2010). Experience has proven that the understanding of elements and terms differ throughout the business units if they are not explicitly agreed upon and communicated.

The following are the three focus areas belonging under this category:

- Definition of master data.
- Master data model.
- Data landscape.

Data quality. The key topic DATA QUALITY is dedicated at data quality in all regards. It includes finding ways to assess the data quality and assess it, finding ways to improve data quality and investigating the reasons for and impact of quality issues. Furthermore, the organization can assess what the most frequent and critical sources for problems are. Data quality is very important for an organization. Therefore, the organization must find out which quality the data has, how this can impact the business, where the sources for poor quality lie and how to improve the data. Data is an organizational asset, so the quality must be good in order to be a competitive advantage and in order not to cause problems. Therefore, data quality must be known and improved (Pipino, Lee, & Wang, 2002; Redman, 1995). In order to do this, objective criteria must be defined that match the needs of the organization (Kahn et al., 2002). Awareness must be raised that poor data quality impacts the business and therefore necessary measures can be taken in an appropriate seriousness (Spruit and Linden, in preparation).

The focus areas are the following:

- Assessment of data quality.
- Impact on business.
- Awareness of quality gaps.
- Improvement.

Usage & Ownership. This key topic is dedicated at defining who uses the data in which systems. Which employee has read/write access and is it clear why people are granted or denied access to certain data? The organization can find out if there are data ownership concepts implemented and see whether the historical grown way still displays the needs. It is generally consensus throughout academia and practitioners that divided responsibilities shared with different people are not effective. Therefore, ownership concepts have proven to be an adequate concept (EDUCAUSE, 2009; IBM, 2007; Loshin, 2010). Furthermore, due to data privacy and data protection reasons, data has to be distributed to appropriate users and not be made available for users without access rights. Of course, data availability must be ensured at all times (Anderson & Moore, 2007; Casassa Mont & Beato, 2007; EDUCAUSE, 2009; IBM, 2007; Loshin, 2010).

Under the umbrella of usage and ownership fall the following focus areas:

- Data usage.
- Data ownership.
- Data access.

Data protection. This section is about the technical security of data; whether and how it is secured against possible incidents. Incidents can be of different kinds; either failure of components, software bugs or steered by people on purpose, like sabotage, hacking, fraud or theft. To ensure confidentiality and the running business, data must be protected (IBM, 2007; Loshin, 2010;

Shaw, Chen, Harris, & Huang, 2009). This protection must be conducted via physical measures and software precautions (Bernard, 2007; Borgman, 2000).

This key topic contains one focus area:

- Data protection.

Maintenance. Here, the focus is on physical storage and the data lifecycle. The main points to investigate are how the data is stored and how is the data treated during the lifecycle. Organizations rely on software solutions. These solutions regularly become faster and more performant quickly. In order to make use of the technical innovations, systems should be up to date and be maintained properly. Furthermore, the inserted data should be kept clean as well, so outdated data is to be removed according to the data lifecycle (Bowker, Baker, Millerand, & Ribes, 2010; Youssef, Butrico & Da Silva, 2008).

This area contains the following sub-topics:

- Storage.
- Data lifecycle.

3.3. Maturity model

The above topics and focus areas regarding MDM lead to the following Master Data Maturity Model (MD3M). The overview is shown in Table 3. The complete MD3M is listed in the Appendix A. The accompanying MD3M questionnaire is also freely available online (Spruit & Pietzka, 2014).

4. Case study: NRGCORP

To validate the model and provide practical information, the whole process was executed at a large case company. This section describes the practical implementation of the before mentioned MDM Maturity Model. We will first introduce the case company, then describe the assessment process and, finally, elaborate on the MD3M assessment results as shown in Table 3.

The case company where our research took place, will now be described anonymously where possible. We will refer to the company as NRGCORP, a trading company in the energy sector. It is part of a stock organization in North-Western Europe which is, in turn, an entity within a global player in the Energy sector which currently employs about 90 K people and generates revenues of about € 100B. NRGCORP trades on many platforms with a broad portfolio of both standard and exotic products. The regional unit under investigation employs over 1000 people from 40 countries whom are active on more than 20 exchanges and in over 40 countries, conducting roughly 600,000 trades per year.

NRGCORP has several reasons for investigating the internal master data landscape and the improvement potential. The business the company is operating in, is by nature very data-intense and is strongly dependent on high quality data which needs to be highly available. If the data is incorrect, the company faces a high risk; the company might even not be able to operate normally and lose profits. In addition, side-effects on third parties might also lead to legal problems. Finally, the company was founded as a merger of different organizations, and faces another merger in the next few years which will further complicate the already complicated master data setups.

Having outlined the case company's profile, we will now continue to report on our MD3M application to assess NRGCORP's current master data management maturity. This includes filling in the questionnaire and assessing the maturity level of the organization (Spruit & Pietzka, 2014. It is obviously important that responsible and knowledgeable persons have to fill in the questionnaire. At NRGCORP we found experts who work with the master data in their daily work. These answers were used to calculate the company's master data maturity.

After conducting several interviews with NRGCORP employees from different departments with a different view on the data and a different focus on the importance of data, a data model was developed. Since the data landscape in a global energy trading organization is very complex, the model was not developed with a data entity level granularity. It is meant to communicate the logical structure on a detailed level without losing focus. Furthermore, a more detailed granularity would increase the model's complexity while decreasing understandability. To also improve understandability, the data model was grouped and colorcoded. The model was revised iteratively in collaboration with key employees.

The results of the questionnaire's application at NRGCORP and the final maturity matrix is shown in Table 3. Summarizing the results, the single maturity levels are as follows. For the business functions 'Data Model', 'Usage and Ownership' and 'Maintenance', the maturity is 1. The business function 'Security' even reaches maximum maturity of 5. However, 'Data Quality' remains at a maturity of 0. Therefore, the overall master data management

Table 3
Overview of the master data management Maturity Model (MD3M) with the results of our case study at NRGCORP.

	Initial	Repeatable	Defined process	Managed & measurable	Optimized
Data model					
Definition of master data	Implemented	Implemented	Implemented	Missing	Missing
Master data model	Implemented	Implemented	Implemented	Missing	Missing
Data Landscape	Implemented	Missing	Implemented	Implemented	Missing
Data quality					
Assessment of data quality	Missing	Implemented	Implemented	Missing	Missing
Impact on business	Implemented	Implemented	Implemented	Implemented	Implemented
Reasons/sources for poor quality	Implemented	Implemented	Missing	Implemented	Implemented
Improvement	Missing	Implemented	Missing	Implemented	Missing
Usage & ownership					
Data usage	Implemented	Implemented	Missing	Missing	Missing
Data ownership	Implemented	Missing	Missing	Missing	Missing
Data access	Implemented	Implemented	Implemented	Missing	Implemented
Data protection					
Data Protection	Implemented	Implemented	Implemented	Implemented	Implemented
Maintenance					
Storage	Implemented	Missing	Missing	Missing	Missing
Data lifecycle	Implemented	Implemented	Missing	Missing	Implemented

Please cite this article in press as: Spruit, M., & Pietzka, K. MD3M: The master data management maturity model. *Computers in Human Behavior* (2014), http://dx.doi.org/10.1016/j.chb.2014.09.030

maturity level of NRGCORP is 0 as well. However, this also implies that implementing this particular capability will lead to a direct increase in overall maturity.

The 'Data Quality' value is so low because two capabilities on the first level are not implemented yet at NRGCORP. The company does not yet have a feeling about the quality of data and in which regards the quality adds value for the company. Furthermore, there are no processes in place to figure out the sufficiency of data quality. Two other capabilities are already implemented, however. The employees at EET are aware of the fact that bad data quality affects the business' reputation. Furthermore, they know about different reasons causing bad master data in the firm (see Fig. 1).

To conclude, even though the overall maturity is still 0, the total percentage of implemented capabilities versus missing ones is approximately 60% already: From the 65 defined capabilities, NRGCORP implements 38 (58.5%). Finally, Table 4 shows an overall tendency to be seen from more implemented capabilities towards less implemented ones throughout the rise of maturity levels. This can be derived from the previous table. The first maturity level is with 84.6% implemented and the last two ones have a less than half as high implementation quote.

5. Validation

To ensure validity of this research, the research was evaluated along Yin's criteria (Yin, 2003). Yin proposes for single case study research to 'Construct Validity' through the usage of multiple evidence sources and through establishing a chain of evidence (Yin, 2003). This takes place when collecting data. The MD3M is based on several sources of academia and models from practice. These were investigated and compared to serve as a basis for the developed model. Additionally, it is suggested to have the key informants review the draft study. This was conducted via evaluation loops with the experts that were consulted. For ensuring 'External Validity', it is advised to make use of theory in single case studies (Yin, 2003). The research design is based on a thorough literature study and a comparison of existing models. For ensuring 'Reliability', it is recommended to demonstrate that the study can be reconducted and achieve the same results. This study is thoroughly documented, so a researcher would be able to conduct this research at another case company and is likely to achieve the same results. The results are extensively documented, so can serve as input for a case study database. Since this is the first study in this particular field, there cannot be a database filled with equivalent studies. The adherence to the described criteria shows the validity and soundness of this research.

Table 4
Implemented versus missing MD3M capabilities per maturity level at NRGCORP.

Maturity per level	Total	Implemented	Missing
1	13 (100%)	11 (84.6)	2 (15.4)
2	13 (100%)	10 (76.9)	3 (23.1)
3	13 (100%)	7 (53.8)	6 (46.2)
4	13 (100%)	5 (38.5)	8 (61.5)
5	13 (100%)	5 (38.5)	8 (61.5)

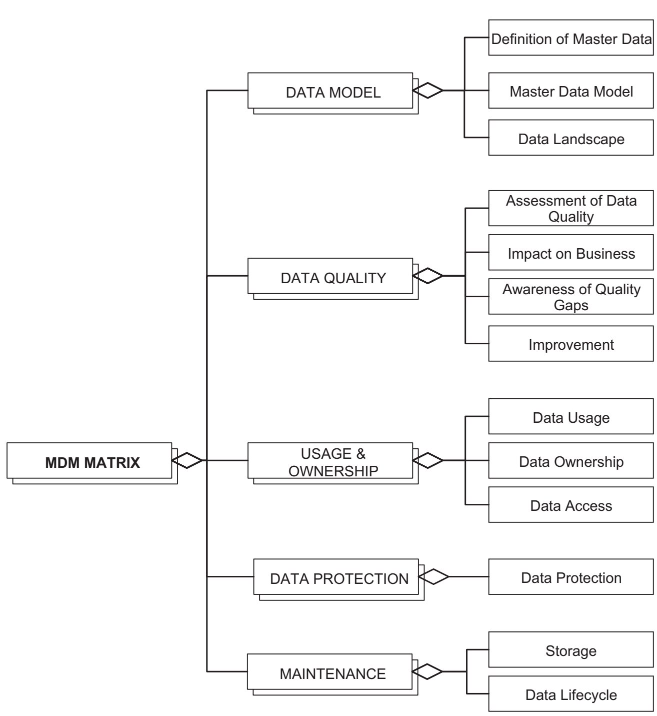

Fig. 1. Key topics and focus areas in master data management.

6. Discussion

The MD3M consists of five levels of maturity. The levels' descriptions are kept rather broad because they are used to describe the maturity level of very distinct capabilities situated in different topics. The level aims at being able to describe all capabilities properly.

The first level is called 'initial'. It describes a first awareness on the topic of MDM for the different focus areas. The second one displays a 'repeatable' state, meaning that insular measures have been initialized in different departments, but they have no connection to other projects. The third 'defined process' level indicates first collaborations and awareness on an inter-divisional level. The fourth level contains 'managed and measurable' processes that are in place in the organization. Processes are defined. The fifth one is 'optimized'. On this level, the efficiency has been improved due to MDM measures. The attention is on a tactical level.

The levels proved to be of an appropriate granularity. The levels cover enough information to be helpful in assessing the maturity and providing a status update on the current situation in an organization. Additionally, they are not too specialized that the interviewees would get lost in details that they could not answer without consulting specialists for every capability. Experts with a good overview about master data activities can answer the questions without too much effort. This is also an important factor. If answering consumed too much time, the organizations would be reluctant to participate because the benefits would only show after a lot of input. For this, the resources might not be available.

So, concluding, it appears that the granularity is on a good level and helps both practitioners and academics in assessing and comparing master data maturity levels in different organizations.

The capabilities were developed based on literature and best practices and then validated by experts. So, it can be assumed that these are reliable. In the case study, there is a development to be seen that indicates a tendency towards less implemented capabilities with rising maturity. This justifies the impression that the ordering is reasonable.

Whilst developing the maturity model and applying at the case company, it appeared that there are three big improvement areas, namely knowledge management, process management and data landscape management. Knowledge management is about managing what the people in the organization know and making sure that nothing gets lost. In general, it is important to have a functioning knowledge management policy in place and create an atmosphere in which sharing is more appreciated than employees gathering knowledge for themselves. Process management is all about creating feasible, lean processes that guide the employees and do

Please cite this article in press as: Spruit, M., & Pietzka, K. MD3M: The master data management maturity model. *Computers in Human Behavior* (2014), http://dx.doi.org/10.1016/j.chb.2014.09.030

not create any unnecessary overload, but still considering all necessary steps. The last field data landscape management is about the set up of the data. This needs to be in form and reflect the reality in a sensible manner.

The key topic 'Data Model' is about the streamlining definitions. Different business units have different understandings of the same key terms due to different perspectives on the same topic and possible ignorance of others' requirements, which is a state to be terminated. Actions in this field clearly belong to the field of knowledge management which is about sharing knowledge across an organization. Furthermore, it belongs in the category of process management because clearly defined processes could help with this problem. A structured data model delivers much precious information about the data landscape in an organization which is important to keep a good overview in order to see improvement potential.

The key topic 'Data Quality' needs processes and knowledge management in order to become mature. In order to figure out the actual quality of data elements, there must be processes in place to reach all business units and to have a reliable path to walk along. Knowledge management is definitely needed when trying to figure out the quality of data. Since quality here is defined as fitness for use, the fitness can only be evaluated if every business unit shares their requirements. It is also relevant that business units share information in order to define the impact that quality can have on the whole business.

Within 'Usage & Ownership', all three aspects are covered. There must be processes in place for granting or denying access to systems or data. Everyone must know how he can get access and the ones deciding must have processes to know what to do. Knowledge sharing is important because this helps defining which user groups need access to which data. Additionally, a data landscape overview is beneficiary for users to having an overview about existing sources of information, so nothing is set up or saved redundantly.

In the field of 'Data Protection', process management is an important factor. To keep data safe and secure, there must be processes in place that must be adhered to. Only when this is given, data is secured against unauthorized access or other violations.

Concerning 'Maintenance', Data landscape has an important role. This key topic is about how to store the data and maintaining the data lifecycle. Therefore, information on storage or data lifecycle, which are by definition close to the data layer, belong fully to the field of data landscape.

These three improvement areas can be approached either reactively or proactively. The reactions of the MD3M's results are clearly focused on improvement of the organizational setup towards the management of master data. Actions can be classified into two categories from the point of time, at which the snapshot is taken. Both approaches are important because the old mistakes and inefficiencies need to be resolved but the roots must be extinguished as well in order to avoid upcoming mistakes. First, there can be reactions to the situation. Beyond that, there are proactive improvement possibilities. An example for reactive would be correcting errors in data that are already in the systems. Proactive measures are changing the way the data are entered; e.g. implement a system where elements to be inserted are not written down manually, but chosen from drop-down lists in order to avoid spelling mistakes.

7. Conclusion and future work

Our extensive case study provides an example of iterative human learning, behavior and collaboration resulting from technological needs in a large-scale infrastructural network. However, one possible limitation of the research at hand is the small amount of case companies. This might lead to restricted generalizability. Therefore, a higher amount of case studies at different companies would uncover possible shortcomings. Thus, applying the MD3M at different companies would give comparable results. The experts could be asked for feedback and show if there are any aspects left out. It would be particularly interesting whether there are differences for companies of a different size. Additionally, the more frequent application could help uncover possible inconsistencies regarding the capabilities.

Another improvement opportunity could be the feedback by experts. More experts could have been consulted to comment on the model and give possible improvement ideas. Alternatively, there could have been another approach of developing the matrix criteria. Experts could be asked to develop their own criteria in which maturity is to be achieved. The downside of this approach that it is very time consuming. Developing a model that covers all or at least most possible factors will take time. Then, cross-checking would have been needed. There, the experts could have discussed the developed approaches and derive one optimized one together. Again, this would be very time consuming because it must be based on research before.

Another approach could have been to derive the maturity model first from practical opinions and then cross-check it with academia. For this, experts from different companies could present their ideas which would be compared and then brought in relation to existing literature on the topic.

Furthermore, an interesting aspect would be if the different capabilities have different weights. If some are more important than others, it could lead to another matrix and other parts to emphasize. The aspect of importance was not considered in this research.

Finally, the results could have been discussed with experts who could have been asked to draw own conclusions in order to find possible hidden conclusions that the author might not have found.

Information and their efficient use is a big competitive factor. Organizational assets do not anymore only consist of physical products to process or sell. The information that organizations obtain are valuable. Therefore, it is really important for organizations to integrate the internal data in a sensible way. The businesses' pace in increasing steadily in terms of innovation cycles which leads to more complexity. The companies need to have short decision cycles based on thorough data. If the organizational landscape changes, e.g. due to mergers, it results in much effort to integrate the different counterparties' data. Synergies may arise from good data management, if different business units benefit from other units' work.

This research's goal was to provide an overview about existing master data maturity models and then derive a model which can be used to assess an organization's master data management maturity. This will help the organization to position them and see if they are underperforming and where they could improve. In order to do so, there was an extensive literature study conducted covering relevant terms and aspects in the field. Then different maturity models were compared with each other. From this, criteria were developed in which a company should achieve maturity when willing to perform efficiently. These criteria were grouped into the five key topics Data Model, Data Quality, Usage & Ownership, Data Protection and Maintenance. These five groups are subdivided into one to four focus areas, in which sub-maturity can be achieved. Additionally, dependencies between single capabilities were identified and influential factors which do not apply to all groups of organizations. The goal was to have a maturity model that is applicable for all sizes of companies and all different industries. Therefore these were necessary, so the model can be applied even though some things apply for one kind of organization, but do not for another one. This maturity can be assessed by answering a questionnaire which was developed in accordance with the matrix. Both were validated with experts.

Summarizing, this model has been proven to be a good possibility to analyze an organization's master data maturity and benchmark it against other organizations. This helps an organization to see whether it has a maturity that is adequate to its company size and whether there are weak spots to work on.

Appendix A

MD3M: The master data management maturity model

Initial	Repeatable	Defined process	Managed & measurable	Optimized
Data model				
Initial	*Repeatable*	*Defined process*	*Managed & measurable*	*Optimized*
A basic understanding of master data exists within some units or within individuals	First cooperative definitions have been made between single units. Discussions are held about the topic	The definition bases on more information from different departments and is a cooperative result. Fewer units have their individual understanding, but thriving towards a shared definition	There is an official definition of master data for the organization with regard to the special circumstances of the organization. This definition is known by all parties involved and can easily be found on a centrally accessible space	There are interfaces for data of different organizations that need to exchange data on a regular basis. Standard formats are established
There are initial attempts to design a model. Probably, there are already some models focusing on data for a particular topic	The different departments can give an overview about master data and how it is interrelated relevant in their scope. There is no knowledge about the data model for the other departments	The different departments can give an overview about master data and how it is interrelated relevant in their scope. Some knowledge already exists about master data objects in other key topics	An enterprise wide master data model was constructed and agreed upon throughout the different units which are concerned with master data	The enterprise wide master data model is maintained regularly. A clear plan with the intervals and the responsibilities concerning the maintenance exists and is communicated throughout the relevant roles
There is an overview about systems that use or access master data	There is a full overview about which systems have reading or writing access to data	It is pointed out if data is stored and accessed redundantly	There is a consistent inventory of all data sources and by which systems they are used. Redundancies are pointed out and concepts are developed to resolve them	There is a consistent inventory of all data sources and by which systems they are used. Redundancies are solved. The data logic is scalable. Superfluous systems are substituted
Data quality				
Initial	*Repeatable*	*Defined process*	*Managed & measurable*	*Optimized*
There is a feeling about data being of good or bad quality	It is clearly stated which aspects are part of data quality and need to be measured in terms of assessing data quality	Data quality is defined regarding the requirements of different stakeholders	Data quality is measured objectively and for each piece of master data it is known which quality it has	The data quality assessment is conducted regularly for every group of data
The organization knows that quality issues in certain data will impact the business from a reputational point of view	The organization knows that quality issues in certain data will impact specific parts of the business as direct monetary loss	The organization knows how bad master data impacts the business from a monetary perspective	The organization knows how bad master data impacts the business from a non-monetary perspective, i.e. loss in reputation, lacking customer retention etc	The organization can state how insufficient master data influences the business in monetary and non-monetary terms and can classify this in financial arguments
The competence team is aware of the fact that there are different reasons for poor data quality	The organization can state which reasons for poor data quality occur in the organization	There are patterns investigated about poor data quality	The employees are aware of the reasons and sources of bad master data quality in their daily work and the consequences thereof	The organization is aware of different reasons for poor data and where they are existent in house. The company knows where the weak spots are and what the reason for that weakness is
The organization figures out areas in which the data quality is not sufficient	There is awareness of the importance of high quality data in terms of efficiency and effectiveness	The organization has a benchmarking system in place to assess whether the data quality is sufficient or not	Improvement measures are installed to improve the data quality	The organization regularly assesses the data quality along the benchmarking system and ensures that the data quality stays within the defined quality

(*continued on next page*)

Appendix A (*continued*)

MD3M: The master data management maturity model

Usage & ownership				
Initial	*Repeatable*	*Defined process*	*Managed & measurable*	*Optimized*
The organization knows for the area of master data who is using which data	It is known if every employee uses the data he has. The employee knows where to get the needed data	Every source of data that an employee might need it communicated to him and he is given access to	Data repositories are maintained regularly and do not get outdated, ergo unusable	The employees use the possibilities they have and are not reluctant to use certain systems to obtain data from
Data elements are owned by individuals/ departments	Data elements are owned by logically consistent roles/departments. The owner defines usage, purpose and content of data	Responsible persons for data are openly communicated and known throughout the company. The data owner has defined responsibilities for treatment of the data	Data stewards are established for chunks of data	Data stewardship is promoted and fixed in the role description of the job. Data quality standards are defined and adhered to
There is a defined process how to get access to data	Access to data is denied to unauthorized personnel	Every employee has access to the data he needs to fulfill his job	Every employee has access to the data he needs to fulfill his work and only this data. He does not have access to data that he either does not need or should not be seeing	Every employee knows which sources he has access to and what he can find there for his purposes
Data protection				
Initial	*Repeatable*	*Defined process*	*Managed & measurable*	*Optimized*
The technical requirements for data protection are fulfilled	Access to data must be activated on request	There are rules for which roles data access can be granted	Passwords exist for systems with data access which have to adhere to common security standards and have to be changed regularly	Awareness for data protection must be raised among the employees
Maintenance				
Initial	*Repeatable*	*Defined process*	*Managed & measurable*	*Optimized*
The data is stored in a persistent, performing way	The data logic is regularly checked for up-to-datedness	Automatic tools regularly check for redundancies and duplicates	The data base logic is regularly checked for persistence, performance and efficiency	The data is stored in an innovative way with possibilities of forecasting and analysis
The organization is aware of the fact that data has a lifecycle and that data structure will change over time	Data is considered as an organizational asset	Guidelines must be established for treating data over the lifecycle	For every data item, a single source of truth is established	The entering, updating and deleting of data is automatically logged by the systems to decrease documentation effort and facilitate auditing

References

Anderson, R., & Moore, T. (2007). The economics of information security: A survey and open questions. *Technology*, 1–27.

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing maturity models for IT management. *Business & Information Systems Engineering, 1*(3), 213–222.

Beniger, J. (1986). *The control revolution: Technological and economic origins of the information society* (1st ed.). Cambridge, MA: Harvard University Press. 493p.

Bernard, R. (2007). Information lifecycle security risk assessment: A tool for closing security gaps. *Computers & Security, 26*(1), 26–30. http://dx.doi.org/10.1016/j.cose.2006.12.005.

Borghoff, U., & Pareschi, R. (1997). Information technology for knowledge management. *Journal of Universal Computer Science, 3*(8), 835–842.

Borgman, C. (2000). *The premise and the promise of a global information infrastructure. From gutenberg to the global information infrastructure: Access to information in the networked world* (1st ed.). Boston, Massachusetts: MIT Press, pp. 1–34.

Bowker, G., Baker, K., Millerand, F., & Ribes, D. (2010). Toward information infrastructure studies: Ways of knowing in a networked environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International handbook of internet research* (1st ed., pp. 97–117). Dordrecht: Springer Science + Business Media.

Butler, D. (2011). Oracle MDM Maturity Model. <http://blogs.oracle.com/mdm/entry/oracle_mdm_maturity_model> Retrieved 12.12.11.

Casassa Mont, M., & Beato, F. (2007). On parametric obligation policies: Enabling privacy-aware information lifecycle management in enterprises. In: Eighth IEEE international workshop on policies for distributed systems and networks POLICY07 (2007) (pp. 51–55).

Davenport, T., & Prusak, L. (2000). *Working knowledge – How organizations manage what they know*. Boston, Massachusetts: Harvard Business School Press.

EDUCAUSE Center For Applied Research. (2009). Data stewardship, security, and policies. ECAR Research Study 8 (pp. 75–91).

Hevner, A., March, S., & Park, J. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105.

IBM (2007). The IBM data governance council maturity model: Building a roadmap for effective data governance.

IT Governance Institute (2000). Cobit 4.1 – Executive summary framework. Reproduction.

Please cite this article in press as: Spruit, M., & Pietzka, K. MD3M: The master data management maturity model. *Computers in Human Behavior* (2014), http://dx.doi.org/10.1016/j.chb.2014.09.030

Kahn, B., Strong, D., & Wang, R. (2002). Information quality benchmarks: Product and service performance. *Communications of the ACM, 45*(4), 184–193.

Kenessey, Z. (1987). The primary, secondary, tertiary and quaternary sectors of the economy. *Review of Income and Wealth*.

Kumar, S. (2010). MDM maturity model. *Information Management Newsletters, March 15, 2010*. <http://www.information-management.com/newsletters/master_data_management_mdm_maturity_model-10017331-1.html> Retrieved 13.12.11.

Loshin, D. (2010). MDM components and the maturity model. A Dataflux white paper.

Otto, B., & Hüner, K. M. (2009). Funktionsarchitektur für unternehmensweites Stammdatenmanagement Inhaltsverzeichnis. *Data Management*.

Pipino, L., Lee, Y., & Wang, R. (2002). Data quality assessment. *Communications of the ACM, 45*(4), 211–218.

Redman, T. (1995). Improve data quality for competitive advantage. *Sloan Management Review, 36*(2), 99–107.

Shaw, R., Chen, C., Harris, A., & Huang, H. (2009). The impact of information richness on information security awareness training effectiveness. *Computers & Education, 52*, 92–100.

Spruit, M., & Linden, V. van der (submitted for publication). BIDQI: The business impacts of data quality interdependencies model. *Applied Computing and Informatics*.

Spruit, M., Pietzka, K. (2014). The MD3M questionnaire: Assessing master data management maturity. Technical report UU-CS-2014-022, Department of information and computing sciences, Utrecht University.

Wegener, H. (2008). Metadaten, Referenzdaten, Stammdaten. In B. Dinter & R. Winter (Eds.), *Integrierte Informationslogistik* (pp. 189–210). Berlin: Springer.

Yin, R. (2003). *Case study research – Design and methods*. Thousand Oaks: SAGE Publications.

Youssef, L., Butrico, M., & Da Silva, D. (2008). *Toward a Unified Ontology of Cloud Computing*. Grid Computing Environments Workshop (GCE) (pp. 1-10). Austin, Texas.

Please cite this article in press as: Spruit, M., & Pietzka, K. MD3M: The master data management maturity model. *Computers in Human Behavior* (2014), http://dx.doi.org/10.1016/j.chb.2014.09.030