dHealth 2023
B. Pfeifer et al. (Eds.)

doi:10.3233/SHTI230038

Extending a Data Management Maturity Model for Process Mining in Healthcare

Andreas ERHARD[1,a,b], Klaus ARTHOFER[b] and Emmanuel HELM[a]
[a] *University of Applied Sciences Upper Austria School, of Informatics, Communications and Media, 4232 Hagenberg, Austria*
[b] *University of Applied Sciences Upper Austria, School of Business & Management, 4400 Steyr, Austria*

Abstract. Background: Many components must work together to continuously improve processes in healthcare organizations. Process mining has recently developed into a discipline that can make a significant contribution here. Objectives: We want to extend an existing management tool to assess and improve the capability of organizations in this area. Method: We add a dimension to the adoption readiness assessment and maturity model for sharable clinical pathways to assess and improve event data quality. Results: We present different approaches for formal and checkpoint assessments and an embedding of the improvement strategy with examples. Conclusion: The additional dimension from the process mining domain integrates with the existing model. At all levels, links can be established between the various aspects of event data quality with existing dimensions. The model has yet to be tested in a real-world use case.

Keywords. Process Mining, Data Management Maturity Model, RAMM, OMG, BPM+ Health, Event Data

1. Introduction

Rising patient numbers and costs in the healthcare sector are driving the need for more efficient use of resources. In addition, there is an increasing need for rich data, to enable secondary use of information and improve patient outcomes. Data management maturity models (D3Ms) can help healthcare organizations optimize their use of resources and effectively use their data use and management for decision-making by defining clear goals and tracking progress. They ensure that data is accurate, complete, secure and compliant with regulations. Moreover, D3Ms provide a foundation for leveraging advanced analytics techniques like process mining, which can help healthcare organizations derive valuable insights from their data to drive better outcomes.

In this work, we introduce a new dimension for the readiness assessment and maturity model (RAMM). This dimension aims to measure the ability of organizations to use process mining based on the quality of their data sources.

This article presents maturity models specific to healthcare and discusses data quality requirements. We start with an introduction to process mining in healthcare, then give an overview of D3Ms in healthcare and explain how we add a new dimension to a

[1] Corresponding Author: Andreas Erhard, University of Applied Sciences Upper Austria, School of Informatics, Communications and Media, Hagenberg, Austria, E-Mail: Andreas.Erhard@fh-hagenberg.at

model. Finally, we discuss how recent developments in healthcare data standards help to reach higher maturity levels.

1.1. Process Mining for Healthcare

Process mining is an amalgamation of techniques aimed at extracting value-added insights from data generated during process execution, e.g., event logs [1]. It serves as a link between process science (domains such as business process management) and data science (data mining, predictive analytics). The prerequisite for different process mining techniques is event data of a certain quality [1].

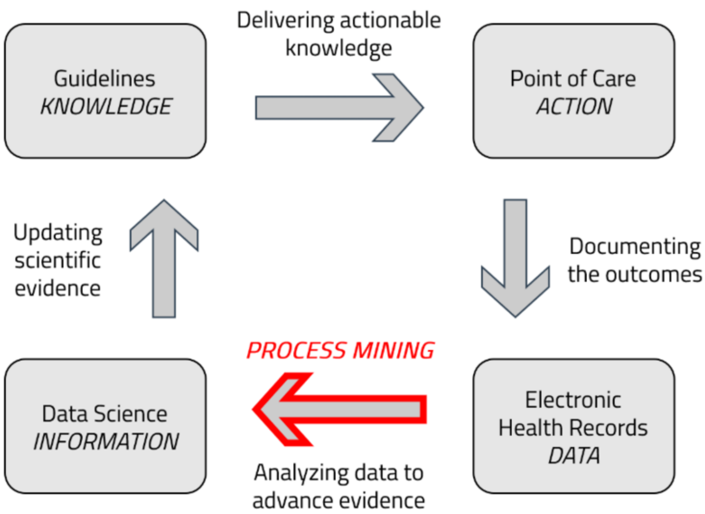

Figure 1. The data lifecycle with clinical guidelines as representation of knowledge (from [2]).

In healthcare, process mining is becoming increasingly popular [3]. Captured event data can be used for process analysis to identify improvement opportunities faster and easier [4]. In addition, process mining can be used to check compliance and control certain rules, e.g., clinical guidelines. This facilitates the monitoring of processes and control of best practices. For example, it can be used for the continuous improvement process of clinical guidelines, cf. figure 1.

1.2. Data Management Maturity Models

D3Ms provide companies with information, data, and feedback on factors that influence the current maturity level of their data management. These models typically also provide recommendations for further action to achieve a higher maturity level.

Effective data management can improve the quality of data. By utilizing the right standards and practices, it can provide transparency to the flow of data [5]. Ultimately, higher data quality promotes the ability to make better informed business decisions [6].

Published frameworks such as the Carnegie Mellon Capability Maturity Model Integration (CMMI) or the Data Management Association International's Data Management Body of Knowledge (DAMA DMBOK) recommend conducting an assessment to determine the current state of data management and related factors [7].

The starting point for our approach was basic research on which maturity models are currently available on the market and which are specifically targeted at healthcare.

In 2018, Gomes and Romáo [8] examined 26 maturity models for information systems in healthcare. Their work serves as a starting point for investigating which maturity model is suitable for our use case. In addition to these 26, we also examined the more recent (2020) Adoption Readiness Assessment and Maturity Model (RAMM) for sharable clinical pathways [9,10] by BPM+ Health of the OMG (Object Management Group). The BPM+ Health community applies business process modeling standards to clinical best practices, care pathways, and workflows directly at the point of care.

2. Methods

After reviewing the different maturity models with their different approaches, we decided to focus on the RAMM. It has a broader scope than other D3Ms, but the main reason was the unique perspective on clinical pathways, and thus, the adoption of a process-oriented approach. Next, we explain how the RAMM is structured and how a new dimension can help to complete the full data lifecycle from figure 1.

The RAMM is based on a 5-level maturity model and promotes and supports organizational change by improving readiness to adopt clinical pathways, proven workflows, and best practices [10]. The model comprises 7 dimensions:

1. Institutional Standards / Guidelines / Policies
2. Stakeholder Management (patient/caregiver(s), clinical)
3. Adoption Processes
4. Privacy, Security, Confidentiality
5. Skills and Expertise (education component)
6. Knowledge Assets, Tools and Automation
7. Goals and Measurement

Each of these 7 dimensions has 5 maturity levels. The lowest level (1) describes an ad hoc approach with an inconsistent adoption of clinical pathways. The highest level (5) implies a “Learning Health System” [9], that is based on metrics and facilitates continuous process improvement.

However, we argue that the RAMM lacks a data quality perspective, i.e., a way to quantify an organization’s ability to transform their data to information (cf. figure 1). Only with reliable, evidence-based (i.e., data-based) methods to learn from recorded activities and outcomes, metrics-based continuous process improvement is possible. This can be achieved with process mining [1,4].

2.1. Adding the Event Data Quality Dimension

The starting point for process mining is event data. To define the applicability of a data source for process mining, the *Process Mining Manifesto* [11] introduces five different maturity levels for event logs (cf. Table 1). These five levels are treated like an additional dimension to the 7 described in the RAMM. The classification for this EDQ (Event Data Quality) dimension is based on the following quality criteria derived from [11]:

- Event logs should be *trustworthy*, i.e., it should be safe to assume that the recorded events actually happened and that the attributes of the events are correct.

- Event logs should be *complete*, i.e., given a particular scope, no events may be missing.
- Any recorded event should have well-defined *semantics*.
- Moreover, the event data should be *safe*, i.e., privacy and security concerns are addressed while recording the events.

There are some overlaps with the RAMM's existing dimensions. Especially dimension 4, *Privacy, Security, Confidentiality*, already covers parts of the criteria for *safety* in the *Manifesto* [11]. However, interdependencies between different dimensions in D3Ms are not uncommon. In fact, reaching the highest level in only one dimension while staying in very low levels in all others can be a sign for incorrect assessment.

Table 1. The five maturity levels for healthcare event data quality, based on [11].

1	The first level describes event logs of poor quality, where recorded events may not correspond to reality and events may be missing. Examples for level 1 event logs are paper documents routed through the organization or paper-based medical records. While it can be possible to apply process mining techniques to these logs, it "does not make much sense" [11].
2	In the second level, events are already recorded automatically by an information system. However, there is no systematic approach to the recording of events and the information system is not part of all possible steps in the processes or it can be bypassed. This results in incomplete logs that do not always properly reflect the processes [11].
3	The third level is similar to the second level in that the information system still does not follow a systematic approach to record events. The main difference is the trustworthiness of the recorded events to match reality. Van der Aalst [11] names ERP systems as sources for level 3 logs because the recorded events, while distributed over many tables, can be assumed to be correct. Many hospital information systems are based on ERP systems like SAP and without further effort to improve their data quality, these systems are classified within maturity level 3 in most cases.
4	Level four records events not only in a complete and trustworthy, but also in a systematic way. This means that the information system is aware of the notions of process instance (e.g., a case id) and activity [11]. Workflow engines and other BPM systems can record level four logs.
5	The highest level five adds semantics on top of trustworthiness, completeness, and safety. For a log to be classified in level five, the meaning of all recorded events and their attributes must be well-defined [11]. This can be achieved by using code systems and ontologies.

3. Results

By developing the Event Data Quality (EDQ) dimension (Table 1), we aimed to create a tool for healthcare organizations to assess their ability to adopt process mining. Together with the RAMM, the EDQ dimension can be used to structure the assessment, and to guide the adoption of improvement.

3.1. EDQ Assessment

The RAMM distinguishes between formal assessments and checkpoint assessments. The former aims to provide an "externally credible" [10] measure of the maturity levels and employs more rigor in planning and conduct. Checkpoint assessments are less time-consuming and expensive. They are approached via questionnaires or interviews.

To conduct the EDQ assessment in a formal way, the work of Fox et al. [12] can be utilized. Their Care Pathway Data Quality Framework (CP-DQF) describes a thorough method to evaluate and report a system's DQ [12].

A less expensive approach that can also be adopted for checkpoint assessments, is the method developed by Kurniati et al. [13]. They conduct simple experiments to determine the suitability of the data for process mining. An experiment comprises (1) a *scope*, e.g., a specific data source like the admissions table of an EHR, (2) a *question*, e.g., if the admission table is suitable to analyze the process of cancer patient admissions, (3) a *hypothesis*, e.g., the table provides the minimum requirements for process mining (i.e., case id, activity name, timestamp), and (4) a *method*, that describes how the data will be extracted, transformed, and loaded into a process mining tool to answer the question [13]. The performance of these experiments will be documented, and any problems will be discussed. A predetermined set of repeatable experiments like these could accompany ongoing checkpoint assessments.

3.2. EDQ Improvement

EDQ improvement should be embedded in RAMM's strategy and implementation plan to develop organizational change management [10]. This transition plan comprises 10 steps closely related to the different dimensions of the maturity model. It starts with the identification of practices expected and missing (gaps), i.e., an assessment. This step should be extended with practices described in the previous section. A later step, technology planning and integration with organization technology ecosystem, is the place to embed the EDQ improvement. Based on the identified maturity level, a plan for improvement can be developed.

An example would be a level 2 assessment, where specific gaps in the electronic documentation of clinical procedures are identified. The resulting suggestions for improvement can, on the one hand, point to the completion of the process documentation and, on the other hand, initiate concrete further developments of the information systems in the direction of ERP. In addition, the use of standardized data formats is recommended, e.g., following public HL7 FHIR Implementation Guides.

A second example would be a level 4 assessment, where there is complete process documentation, but the naming of activities or entities does not follow a standardized scheme. Here, suitable code systems, e.g., SNOMED-CT, or the establishment of an organizational ontology based on a suitable domain ontology are described as suggestions for improvement.

4. Discussion and Outlook

In this work we introduced the Event Data Quality dimension that aims to measure the ability of organizations to use process mining based on the quality of their data sources. It can, for example, identify if certain reference structure data and patient master data such as case ID, patient ID, or treatment ID are available in order to sufficiently support medical treatment on the one hand and to enable process mining on the other hand.

The main weakness of this work is that our approach has not yet been tested in practice. However, since we combine two existing concepts and show their overlaps, we assume that it is functional.

Of course, there is much more to the adoption of standardized ways of working in healthcare organizations. Process management practices from manufacturing industry used to be notoriously unsuccessful when applied in healthcare [14]. Recent management tools, like the OMG's BPM+ Health RAMM [9], try to provide a more holistic approach. The RAMM helps to develop plans for organizational change management that also encompass the goal of continuous improvement [10]. Conducting assessments of the EDQ dimension together with the RAMM can help to identify the gaps to evidence-based, i.e., data-based process improvement.

Acknowledgments

This research is financed by research subsidies granted by the government of Upper Austria.

References

[1] W. van der Aalst, Process Mining: data science in action, Springer Verlag, Berlin, Heidelberg, 2016

[2] HL7 International, Clinical Decision Support Working Group, FHIR Clinical Guidelines (v1.0.0). 2021, https://hl7.org/fhir/uv/cpg/index.html, last access: 05.02.2023

[3] T. G. Erdogan and A. Tarhan, "Systematic Mapping of Process Mining Studies in Healthcare," in IEEE Access, vol. 6, pp. 24543-24567, 2018, doi: 10.1109/ACCESS.2018.2831244.

[4] J. Munoz-Gama et al., "Process mining for healthcare: Characteristics and challenges," Journal of Biomedical Informatics, vol. 127, p. 103994, 2022, doi: 10.1016/j.jbi.2022.103994.

[5] B. Otto, "Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers," *CAIS*, vol. 29, 2011, doi: 10.17705/1CAIS.02903

[6] J. Ladley, *Data governance: How to design, deploy, and sustain an effective data governance program*. London, San Diego, CA, Cambridge, MA, Oxford: Academic Press an imprint of Elsevier, 2020. [Online]. Available: https://www.sciencedirect.com/science/book/9780128158319

[7] I. Steenbek, *'Brevity is the soul of wit': DAMA-DMBOK in a nutshell.,* 2018 [Online]. Available: https://datacrossroads.nl/2018/06/25/dama-dmbok-in-a-nutshell/

[8] J. Gomes and M. Romão, "Information System Maturity Models in Healthcare," Journal of medical systems, vol. 42, no. 12, p. 235, 2018, doi: 10.1007/s10916-018-1097-0

[9] Object Management Group (OMG), "Adoption Readiness Assessment Maturity Model (RAMM) for sharable clinical pathways", https://www.bpm-plus.org/publications/index.htm, 2021, last access: 05.02.2023

[10] Object Management Group (OMG), "BPM+ Health Adoption Guide Playbook", https://www.bpm-plus.org/publications/index.htm, 2022, last access: 05.02.2023

[11] W. van der Aalst et al., "Process Mining Manifesto," in Lecture Notes in Business Information Processing, Business Process Management Workshops, F. Daniel, K. Barkaoui, and S. Dustdar, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 169–194.

[12] F. Fox, V. R. Aggarwal, H. Whelton and O. Johnson. "A data quality framework for process mining of electronic health record data," IEEE international conference on healthcare informatics (ICHI), pp. 12-21, 2018

[13] A. P. Kurniati et al. "The assessment of data quality issues for process mining in healthcare using Medical Information Mart for Intensive Care III, a freely available e-health record database," Health informatics journal, 25(4), 1878-1893, 2019

[14] A. Hellström, S. Lifvergren and J. Quist, "Process management in healthcare: investigating why it's easier said than done," Journal of Manufacturing Technology Management, Vol. 21 No. 4, pp. 499-511., 2010 https://doi.org/10.1108/17410381011046607