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

doi:10.3233/SHTI230034

Multi-Perspective Process Mining Interfaces for HL7 AuditEvent Repositories: XES and OCEL

Andreas POINTNER[a,1], Oliver KRAUSS[a], Andreas ERHARD[a], Andreas SCHULER[a, b] and Emmanuel HELM[a, b]
[a] *Department of Medical and Bioinformatics, University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria*
[b] *Research Department Advanced Information Systems and Technology, University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria*

Abstract. Background: Medical information systems frequently use event logging, but these logs are not suitable for process mining as they are not logged in a standardized format. Objectives: Our goal is to enrich medical event logs for use in process mining. Method: We present an approach to convert events from standards-based repositories into the XES and OCEL formats commonly used in process mining. Results: We tested this approach using simulated data from the Austrian breast cancer screening program. Conclusion: We aim to apply it to analyze care guidelines and improve hospital processes in the future.

Keywords. Data Mining; Process Assessment, Health Care; Standard of Care; Electronic Health Records

1. Introduction

Process-oriented data analysis techniques such as process mining have become increasingly popular in the healthcare domain in recent years, as it allows for the identification and optimization of care delivery processes, leading to improved patient outcomes, cost savings and enhanced care efficiency.

In their recent work, Munoz-Gama et al. [1] highlight the potential, the distinguishing characteristics, and the challenges of process mining for healthcare (PM4H). The ability of process mining techniques to answer process-related questions in the healthcare domain has been proven in numerous case studies [2]. However, there is no evidence of systematic adoption as of early 2022 [1]. To increase the adoption, healthcare specific methods and techniques should be developed [1].

We want to contribute to this by showing how existing healthcare data standards can be utilized to provide ready-to-use process mining data, i.e., standardized event logs. Our previous work [3] has already had an impact on the development of HL7 FHIR[2] and we

---

[1] Corresponding Author: Andreas Pointner, University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria, E-Mail: andreas.pointner@fh-hagenberg.at

[2] HL7®, FHIR® and the FHIR® logo are the registered trademarks of Health Level Seven International and their use does not constitute endorsement by HL7®.

continue to strive to incorporate the process-oriented data science perspective in the further development of the standard.

The process mining community defined a set of standard formats, that are used within various tools to discover processes, like XES and OCEL [4]. However, healthcare information systems do not provide data in these formats.

Munoz-Gama et al. [1] identified ten challenges for the further development of process mining in healthcare. This paper aims to address two of those challenges: "C1: Design Dedicated/Tailored Methodologies and Frameworks", and "C9: Complement HISs with the Process Perspective". C1 defines that a framework developed in the area of process mining for healthcare should guide researchers and practitioners through the various phases of PM4H analysis [1]. C9 addresses the source of the data. Existing health information systems (HISs), e.g., electronic health records (EHRs), should be complemented with a process perspective [1].

The question we aim to answer here is: how can HISs provide their data to a process mining tool, so that: (1) different perspectives on the processes are supported, and (2) the data is provided in process mining suitable formats?

To overcome this challenge of data integration, some approaches focus on existing HISs' audit logs. In 2013, Cruz-Correia et al. [5] were the first to look at standardized audit logs for process mining. Helm et al. [3], [6] and Murillas et al. [7] followed with approaches to make these data sources more easily accessible for PM4H.

Recently, Jones et al. [8], Chen et al. [9], and Noshad et al. [10] also utilized audit log data for process mining. All three applied unsupervised machine learning techniques to analyze practitioners' interactions with their HISs and to visualize complex care pathways. They also highlight the usefulness of audit log data for process mining.

In 2020, Adler-Milstein et al. [11] concluded: "As researchers become aware of this new source of data ... the tools for working with audit log data will improve. In parallel, there will be pressure to develop common terminologies and data models for audit log data, which will make it easier to work with these data across multiple institutions and with different EHRs at scale.".

2. Methods

The international standards development organization Health Level Seven (HL7) produces the world's most used standards for healthcare interoperability [12]. Their latest addition to the HL7 standards family is called Fast Healthcare Interoperability Resources (FHIR). FHIR-based APIs and implementations of EHRs become increasingly common.

Following up on the work of Helm et al. [3], who developed a simple transformation approach, we focus on a solution for the mapping of audit log data to process mining suitable formats, XES and OCEL. These two data formats enable the usage of process mining suitable formats. In addition, we provide the means to query different perspectives of the process.

2.1. HL7 FHIR AuditEvent

HL7 FHIR AuditEvent[3] provides an information model that can be used for logging events. The documentation specifies it as "a record of an event relevant for purposes such

[3] https://build.fhir.org/auditevent.html

as operations, … and performance analysis". It allows storing relevant information about an event, like the timestamp when it occurred, and the action that was performed during the event. In addition, it can store relations to other "objects" that are involved in the event, like the patient, medical devices, or practitioners.

While in [3], the authors had to extend this resource to be suitable for process mining, recent additions allow us to use the standard as-is. This enables storing the type of the event that occurred, or a workflow context for single events.

2.2. XES and OCEL

The 1849-2016 - IEEE Standard for eXtensible Event Stream (XES) [13] defines an information model for "Achieving Interoperability in Event Logs and Event Streams". An XES event log has a hierarchical structure, comprising log, trace, and event. Log (the process) contains a collection of traces (execution instances) and a trace contains a collection of events [13]. XES is supported by most process mining tool vendors.

OCEL: A Standard for Object-Centric Event Logs [4] is a more recent standard that aims to overcome the main shortcoming of XES. Modern information systems that support processes cannot be reduced to a single case notion [4] (cf. trace in XES). Thus, OCEL does not only store the identifier, activity, and timestamp to a certain event, but also a list of relevant objects, that are associated with it. In contrast to XES, OCEL does not group its events in traces, and thus has no hierarchical structure.

2.3. Transformation

Each model transformation starts with a given input model that is transformed into a resulting output model. The input for our transformation is always a set of AuditEvent elements called Bundle – a grouping resource in HL7 FHIR. They are transformed into a single log element on the output model, independently of the used standard.

In contrast to a set of HL7 FHIR AuditEvent elements, the XES log consists of a hierarchical data structure. Thus, we must perform grouping operations for the Bundle of AuditEvents. Figure 1 depicts the mapping of an AuditEvent resource (left) to an XES event log (right).

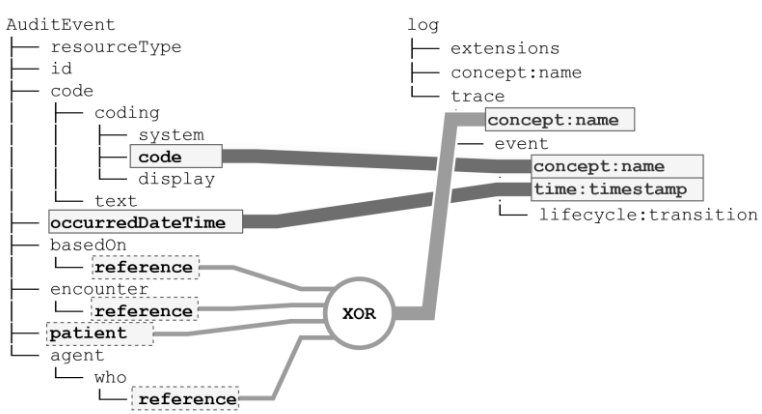

Figure 1: Transformation between the HL7 FHIR AuditEvent (left) and the XES log (right). The lines indicate which concepts are mapped to which. The XOR node indicates which alternatives can be used to map log.concept:name in contrast to the default (AuditEvent.encounter.reference → log.trace.concept:name).

In contrast to the XES transformation, the mapping between AuditEvents and OCEL logs is simple (cf. Figure 2). Both information models provide a similar approach to structure an event log.

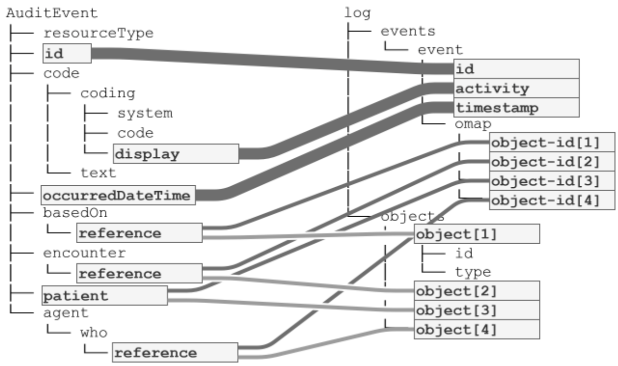

Figure 2: Transformation between the HL7 FHIR AuditEvent (left) and the OCEL log (right). The lines indicate which concepts are mapped to which.

3. Results

In this work, we simulate the mammography double diagnosis process and generate events out of it. This process is part of the Austrian breast cancer screening program “früh erkennen” (detect early) [14]. We use this example as an evaluation for our approach and will show the transformation results in this chapter.

For each transformation concept, we provide a reference implementation written in Java and openly available on GitHub, cf. [15], [16]. The implementation of the transformation uses the Graph Transformation Framework presented in [17]. It is based on autogenerated models for XES and OCEL [18], [19], that were also made publicly available. The XML Schema specifications of the respective standards were used to generate Java domain class models. For AuditEvents, we rely on the Hapi FHIR Structures R5 [20].

```
<log xes.version="1.0">
 ...
 <trace>
  <string value="Enc/868"
     key="concept:name"/>
  <event>
   <string value="ATS"
     key="concept:name"/>
   <date
     value="2022-01-31"
     key="timestamp"/>
   <string
     value="complete"
     key="transition"/>
  </event>
  ...
 </trace>
 ...
</log>
```

```
<log>
 <events>
  <event>
   <string value="ATS" key="activity"/>
   <date value="2022-01-31" key="ts"/>
   <list key="omap">
    <string value="CP2" key="obj-id"/>
    ...
   </list>
  </event>
  ...
 </events>
 <objects>
  <object>
   <string value="CP2" key="id"/>
  </object>
  ...
 </objects>
</log>
```

Figure 3: Simplified excerpt from the automatically transformed XES (left) and OCEL (right) logs.

Figure 3 shows a part of the resulting XES and OCEL documents, after applying the transformation on the AuditEvents.

In addition to our actual goal, we show the reconstructed process using the process mining tool Disco [21] to create a Directly-Follows Graphs in Figure 4. It shows an application for our approach, which is reconstructing the process from the AuditEvents using the XES audit logs.

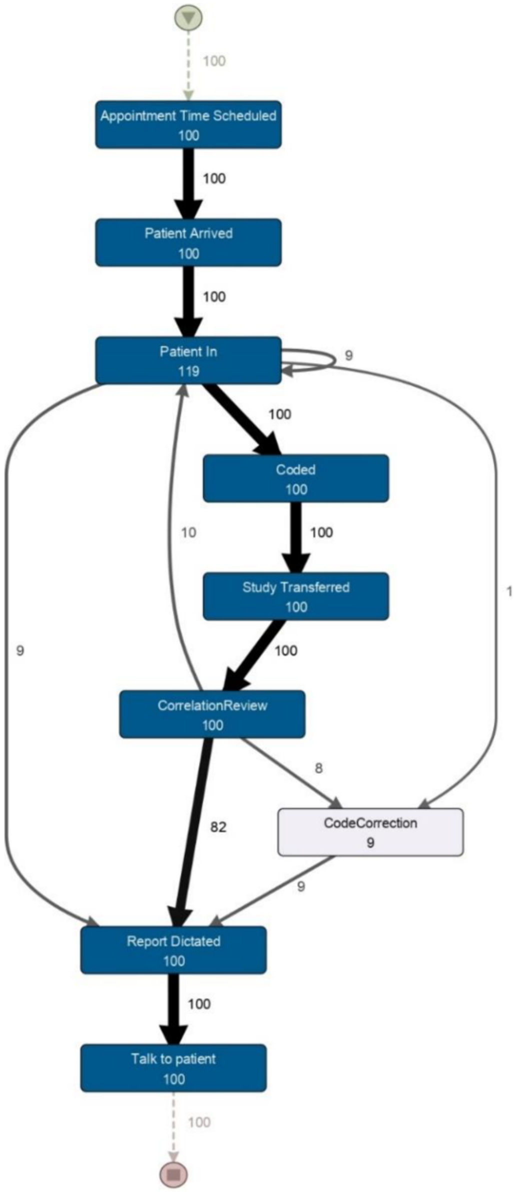

Figure 4: Reconstructed process using the transformed XES log and the process mining tool Disco.

4. Discussion

In this work, we showed a concept to utilize an event log for process mining. We made several assumptions to create a test setup and validated the approach based on parts of the Austrian breast cancer screening process. However, the approach, and especially our assumptions, come with some strengths and weaknesses.

The approach aims at a relevant research direction. Adler-Milstein et al. [11] and Rule et al. [22] highlight the high potential in the analysis of EHR audit logs. Even the simple event log of our simulated process instantly provides an overview and valuable insights when visualized. The approach is based on widely adopted, open standards. HL7 FHIR is adopted by all major EHR vendors. Microsoft, Google, and Apple also provide implementation frameworks and services for FHIR. XES is the de-facto standard for process mining data, at least in the scientific community.

We assume the event log to be a “flat” collection of AuditEvent resources, e.g., in a text file or in a NOSQL database. In practice, this might not be the case. The data could be stored, e.g., in a relational database. However, the concept of the mapping, i.e., which elements should be mapped and how, remains relevant. The approach was only tested on simulated data. While this suffices to validate our transformation, real-world data comes with additional challenges. For example, Cruz-Correia et al. [5] already identified several data quality issues when trying process mining on real-world audit logs.

Funding

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

References

[1] J. Munoz-Gama *et al.*, 'Process mining for healthcare: Characteristics and challenges', *Journal of Biomedical Informatics*, vol. 127, p. 103994, 2022.
[2] E. De Roock and N. Martin, 'Process mining in healthcare–An updated perspective on the state of the art', *Journal of biomedical informatics*, p. 103995, 2022.
[3] E. Helm, O. Krauss, A. Lin, A. Pointner, A. Schuler, and J. Küng, 'Process Mining on FHIR - An Open Standards-Based Process Analytics Approach for Healthcare', in *Process Mining Workshops*, vol. 406, S. Leemans and H. Leopold, Eds. Cham: Springer International Publishing, 2021, pp. 343–355. doi: 10.1007/978-3-030-72693-5_26.
[4] A. F. Ghahfarokhi, G. Park, A. Berti, and W. M. P. van der Aalst, 'OCEL: A Standard for Object-Centric Event Logs', in *New Trends in Database and Information Systems*, Cham, 2021, pp. 169–175.
[5] R. Cruz-Correia *et al.*, 'Analysis of the quality of hospital information systems audit trails', *BMC medical informatics and decision making*, vol. 13, no. 1, p. 84, 2013.
[6] E. Helm and F. Paster, 'First steps towards process mining in distributed health information systems', *IJET*, vol. 61, no. 2, pp. 137–142, 2015.
[7] E. G. L. de Murillas, E. Helm, H. A. Reijers, and J. Küng, 'Audit trails in OpenSLEX: paving the road for process mining in healthcare', in *Int. Conference on Information Technology in Bio- and Medical Informatics*, 2017, pp. 82–91.
[8] B. Jones, X. Zhang, B. A. Malin, and Y. Chen, 'Learning tasks of pediatric providers from Electronic Health Record audit logs', in *AMIA Annual Symposium Proceedings*, 2020, vol. 2020, p. 612.
[9] B. Chen *et al.*, 'Mining tasks and task characteristics from electronic health record audit logs with unsupervised machine learning', *Journal of the American Medical Informatics Association*, vol. 28, no. 6, pp. 1168–1177, 2021.
[10] M. Noshad, C. C. Rose, and J. H. Chen, 'Signal from the noise: A mixed graphical and quantitative process mining approach to evaluate care pathways applied to emergency stroke care', *Journal of Biomedical Informatics*, vol. 127, p. 104004, 2022.
[11] J. Adler-Milstein, J. S. Adelman, M. Tai-Seale, V. L. Patel, and C. Dymek, 'EHR audit logs: a new goldmine for health services research?', *Journal of biomedical informatics*, vol. 101, p. 103343, 2020.
[12] T. Benson and G. Grieve, *Principles of Health Interoperability*. Springer, 2021.
[13] IEEE Task Force on Process Mining, 'IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams', *IEEE Std.*, vol. 1849–2016, pp. 53–57, 2016, doi: 10.1109/IEEESTD.2016.7740858.
[14] K. Jeitler, T. Semlitsch, N. Posch, A. Siebenhofer, and K. Horvath, 'Breast cancer screening in Austria: Key figures, age limits, screening intervals and evidence', *Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen*, vol. 109, no. 4–5, pp. 363–370, 2015.
[15] A. Pointner, O. Krauss, A. Schuler, and E. Helm, 'Transformation from HL7® FHIR® AuditEvent to OCEL'. Jul. 2022. doi: 10.5281/zenodo.6901919.
[16] A. Pointner, O. Krauss, A. Schuler, and E. Helm, 'Transformation from HL7® FHIR® AuditEvent to XES'. Jul. 2022. doi: 10.5281/zenodo.6901825.
[17] A. Schuler, A. Pointner, and C. Praschl, 'Graph Transformation Framework'. Jul. 2022. doi: 10.5281/zenodo.4314721.
[18] A. Pointner, 'Xes-model: 1.0.0'. Zenodo, Jun. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6760266
[19] A. Pointner, 'Ocel-model: 1.0.0'. Zenodo, Jul. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6861509
[20] 'AuditEvent - hapi-fhir-structures-r5 6.2.2 javadoc'. Feb. 2023. [Online]. Available: https://javadoc.io/doc/ca.uhn.hapi.fhir/hapi-fhir-structures-r5/6.2.2/org/hl7/fhir/r5/model/AuditEvent.html
[21] C. W. Günther and A. Rozinat, 'Disco: Discover Your Processes.', *BPM (Demos)*, vol. 940, no. 1, pp. 40–44, 2012.
[22] A. Rule, M. F. Chiang, and M. R. Hribar, 'Using electronic health record audit logs to study clinical activity: a systematic review of aims, measures, and methods', *Journal of the American Medical Informatics Association*, vol. 27, no. 3, pp. 480–490, 2020.