Society Briefs

Moe T. Wynn
*Queensland University of Technology, AUSTRALIA*

Wil van der Aalst
*Technical University of Aachen, GERMANY*

Eric Verbeek
*Eindhoven University of Technology, THE NETHERLANDS*

Bruno Di Stefano
*Nuptek Systems Ltd., CANADA*

# The IEEE XES Standard for Process Mining: Experiences, Adoption, and Revision

## Introduction

The IEEE Standards Association (SA) officially published the XES Standard as IEEE Std 1849-2016: IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams on 11 November 2016. This standard has been sponsored by the IEEE Computational Intelligence Society (CIS) Standards Committee. Through the XES Standard, event data can be transported from the system where it was generated to the system in which it can be stored and analyzed, without losing semantics. Next to providing a standardized syntax and semantics, the XES Standard also supports the introduction of new extensions to define additional concepts in a flexible manner. The standard allows for the exchange of event data between different process mining tools.

Since then, the IEEE Task Force on Process Mining has been driving various initiatives to increase the adoption of the standard format for event data in academia and practice. The XES standard has made a significant impact in advancing the field of process mining by enabling the release of real-life event logs in a standard format.

Recently, a revision of this standard that was approved by the IEEE SA, as a result of which the IEEE 1849-2016 Standard was superseded by the IEEE 1849-2023 Standard. The revised standard (IEEE 1849-2023) which was published on 9 September 2023 incorporates the learnings from the applications of the standard in the community and incorporates several new extensions.

This paper describes the experiences of the working group in (re-) defining a standard exchange format for event data, highlights its impact to date, and calls for wider adoption of the XES standard in the community.

## What is XES?

The eXtensible Event Stream (XES) Standard permits data to be transferred from the location where it was generated to the location where it can be stored and analyzed, that is, from the system that generated the data to the data scientist that analyzes the data. The XES Standard enforces that this transfer and storage is done in a standardized way, that is, in a way that is clear and well-understood. For this reason, the XES Standard fixes the syntax of the data (how to store the data?), the semantics of the data (what does the stored data actually mean?), and ways to extend the latter.

Figure 1 shows the XML serialization for the XES Standard as a state machine flow diagram. The main part of the diagram is the part containing the log, traces, events, and attributes. As the diagram shows, all these elements may have any number of attributes, and an attribute can be of seven different types (six simple types and one list type).

A *classifier* assigns an identity to each event, which makes it comparable to others (via their assigned identity). Examples of such identities include the descriptive name of the event, the descriptive name of the case the event relates to, and the descriptive name of the cause of the event.

An *extension* defines for every type of element a (possibly empty) set of attributes. The extension provides points of reference for interpreting these attributes, and, thus, for their containing elements. Extensions therefore are primarily a vehicle for attaching semantics to a set of defined attributes per element. Extensions have many possible uses.

*Digital Object Identifier 10.1109/MCI.2023.3333141*
*Date of current version: 8 January 2024*

Corresponding author: Moe T. Wynn
(e-mail: m.wynn@qut.edu.au).

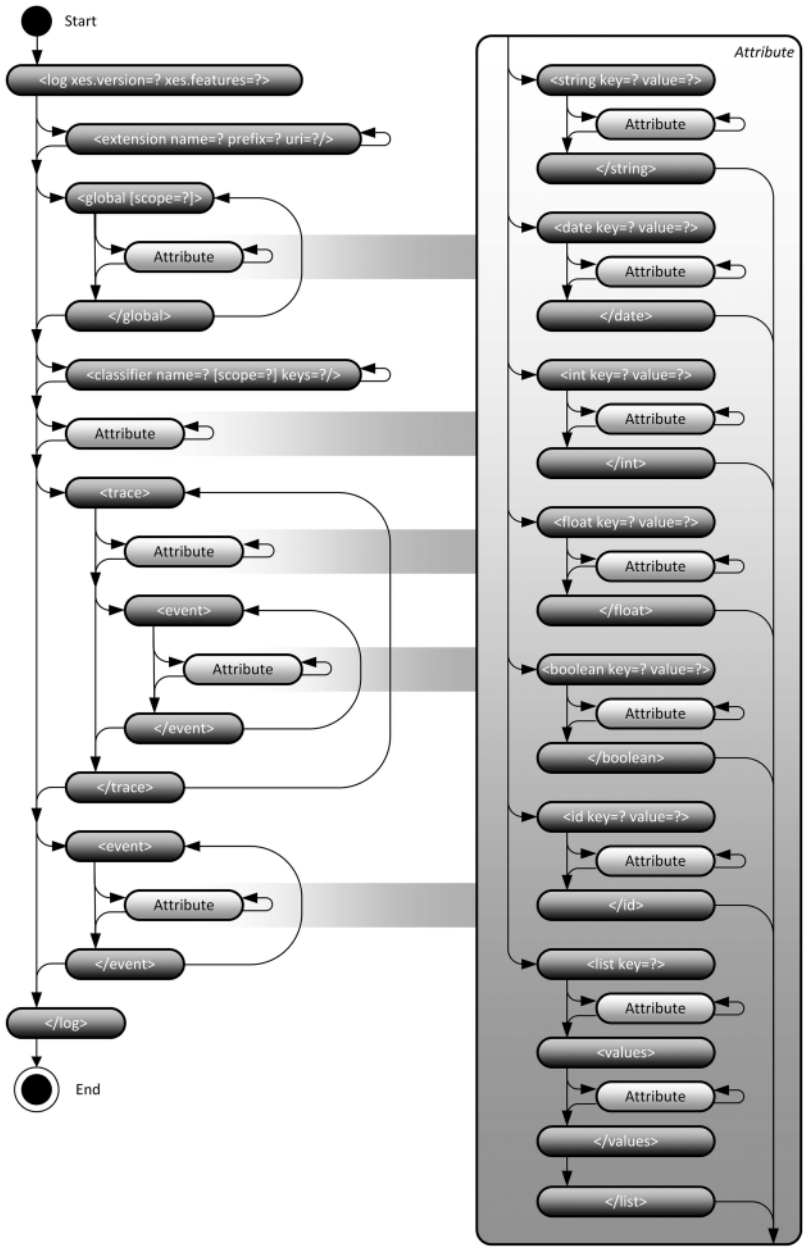


**FIGURE 1** State machine flow diagram for the XES XML serialization.

One important use is to introduce a set of commonly understood attributes which are vital for a specific perspective or dimension of event log analysis (and which may even not have been foreseen at the time of developing this Standard). As an example, the *Concept* extension stores a generally understood name for any element. For logs, the name attribute may store the name of the process having been executed. For traces, the name attribute usually stores the case ID. For events, the name attribute represents the name of the event, e.g., the name of the executed activity represented by the event. Other uses of extensions include the definition of generally-understood attributes for a specific application domain (for example, medical attributes for hospital processes), or for supporting special features or requirements of a specific application.

XES recognizes and treats all extensions as equal, independent from their source. This allows users of the format to extend it, in order to fit any purpose or domain setting. However, there are recurring requirements for information stored in event logs, which demand a fixed and universally understood semantics. For this purpose, a number of extensions have been *standardized*. When creating logs for a specific domain, or also when designing log-analyzing techniques, one should consider using these standardized extensions[1], since they allow for a wider level of understanding of the contents of event logs.

## Impact on the Field

The first process mining algorithms were developed in the late 1990-ties [1], [2]. This was a response to the limited adoption of workflow management technology. Organizations were struggling to model their processes in such a way that a workflow management system could be configured based on such models and support the corresponding processes. This is the reason that in the beginning, process mining was also referred to as "workflow mining". This turned out to be somewhat naive. Real-world processes often exhibit a lot of variability, and this is precisely the reason why process mining is so valuable. Instead of automating observed processes, the primary use case is to improve processes by providing transparency and uncovering unknown performance and compliance problems.

The first process mining algorithms focused on the task of *process discovery*, i.e., creating a process model based on event data. It is relatively easy to discover a process model without concurrency, e.g., a Directly-Follows Graph (DFG), but it is challenging to discover a model that incorporates concurrency (e.g., a Petri net of BPMN model) [1], [2]. After the first wave of process discovery algorithms, the scope was expanded to include *conformance checking*, *performance analysis*, *comparative process mining*, *predictive process mining*, and *action-oriented process mining*. Conformance checking relates events in the event log to activities in the process model and compares both. The goal is

[1]The full list of standardized extensions can be found at https://www.tf-pm.org/resources/xes-standard/about-xes/standard-extensions.

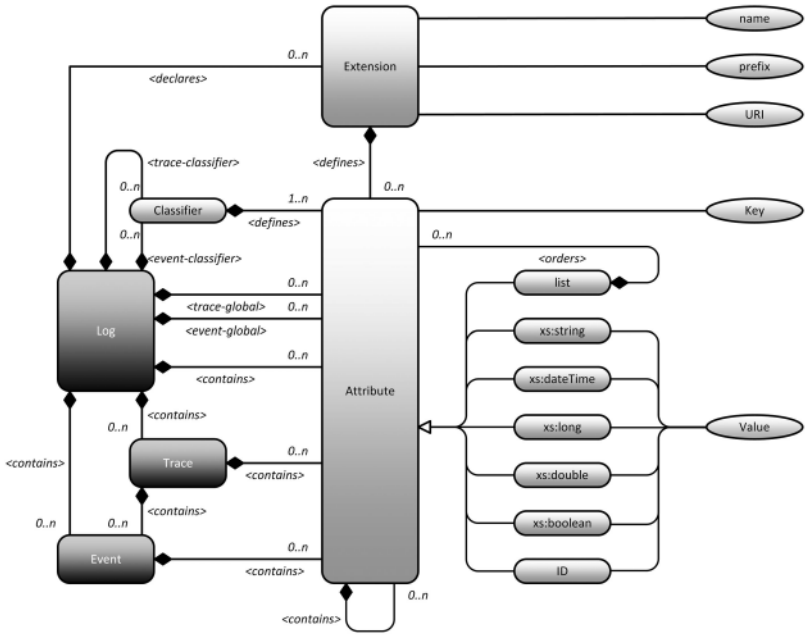


**FIGURE 2** XES meta-model.

to find commonalities and discrepancies between the modeled behavior and the observed behavior. The *process-mining manifesto*, written in 2011 [9], already provided a pretty comprehensive overview of the process-mining discipline. The manifesto was written shortly after the first version of XES was created. In 2010, the IEEE Task Force on Process Mining released the first informal version of XES (www.xes-standard.org). However, it would take another six years to become an official IEEE standard [6].

The process-mining manifesto and the XES standard helped to unify terms and provide a common understanding of process mining. Before 2010, process mining was mainly a research topic with very little industry adoption. In 2007, Futura Process Intelligence was the first dedicated process-mining provider. Later, many more process mining companies followed. For example, Celonis started in 2011. Now, there are over 40 companies providing process mining software (see processmining.org). Recently, Gartner released a "Magic Quadrant" for process mining tools [7], illustrating that process mining is now considered to be a separate software category.

Many process mining tools support XES (e.g., Apromore, Celonis, Disco, Mehrwerk, Minit, myInvenio, PM4Py, ProM, QPR, and Signavio). However, the most significant contribution of XES was not the XML exchange format; it was the XES meta-model shown in Figure 2 standardizing the core concepts. It is fair to say that most process mining tools agree on the core concepts as defined in the XES standard.

Process mining has been adopted in many industries. Examples include [2]: consulting (Deloitte, Ernst & Young, KPMG, and PwC), finance and insurance (Rabobank, Wells Fargo, Hypovereinsbank, Caixa General, ADAC, APG, Suncorp, VTB, etc.), logistics and transport (Uber, Deutsche Bahn, Lufthansa, Airbus, Schukat, Vanderlande, etc.), production (ABB, Siemens, BMW, Fiat, Bosch, AkzoNobel, Bayer, Neste, etc.), healthcare, biomedicine, and pharmacy (Uniklinik RWTH Aachen, Charite University Hospital, GE Healthcare, Philips, Medtronic, Pfizer, Bayer, AstraZeneca, etc.), telecom (Deutsche Telekom, Vodafone, A1 Telekom Austria, Telekom Italia, etc.), food and retail (Edeka, MediaMarkt, Globus, Zalando, AB InBev, etc.), energy (Uniper, Chevron, Shell, BP, E.ON, etc.), and IT services (Dell, Xerox, IBM, Nokia, ServiceNow, etc.). Especially in Europe, process mining has been widely adopted. Here, it has become fairly standard to apply process mining to standard processes such as Purchase-to-Pay (P2P) and Order-to-Cash (O2C). However, process mining is a generic technology and is expected to expand vertically (different processes) and horizontally (other organizations and geographic regions).

## Usage in Scientific Literature

Searching for "process mining" and "XES" as mandatory terms using Elsevier's Scopus, we found 621 publications (including 380 conference papers, 203 journal papers, and 22 book chapters) in the period from 2010 until 2023 (search conducted on 18-9-2023). The number of citations has been steadily growing to around 75 per year.

Many process mining papers use the so-called *BPI Challenges* (BPIC) logs and the *Process Discovery Contest* (PDC) logs, which are all provided in XES format. For example, the BPI Challenge 2017 [5] has been downloaded 6500 times, viewed 14.843 times, and cited 73 times (statistics from data.4tu.nl on 18-9-2023). The BPI Challenge 2012 [4] has been downloaded 5063 times, viewed 14.296 times, and cited 151 times (statistics from data.4tu.nl on 18-9-2023). Two other widely used XES datasets are: the Sepsis log [8], downloaded 4217 times, and cited 61 times) and the Road traffic log [3] downloaded 3961 times, and cited 95 times).

Using XES allows for an unambiguous interpretation of the event data. This makes it easy to compare analysis results. This is vital for the development of the process-mining discipline. These download and citation statistics show that the impact of XES has been significant.

## Standardization Process and Next Steps

The formal revision process of the XES standard started in early 2022 with the submission of the Project Approval Request (PAR) to the IEEE SA, which accepted the PAR on 24 March 2022. An expression of interest to join the

working group was then sent out to the community. The kick-off meeting of the working group was held on 8 June 2022.

The P1849-2023 working group is made up of 14 members with Moe Wynn as chair, Wil van der Aalst as vice-chair and Eric Vereek as the secretary. The other working group members are: Arif Selcuk Ogrenci, Claudio Di Ciccio, Estefania Serral Asensio, Julian Lebherz, Marcus Brenscheidt, Nancy Landreville, Rajesh Murthy, Sergio Correia, Andreea Ungureanu, Peter Blank, and Philipp Herrmann.

Several meetings were held in 2022 to revise the standard. The first external version for the XES Standard was sent to the IEEE SA for Mandatory Editorial Coordination on 26 Oct 2022 and changes were made based on the comments received. The balloting of the standard started in Jan 2023 and was completed on 21 March 2023 with 100% approval from 13 balloting members and no comments. The XES Standard was submitted on 22 March 2023 to the IEEE SA Revision Committee (RevCom). During its meeting on 15 May 2023, RevCom recommended this version of the XES Standard be approved. During its meeting on 4 June 2023, IEEE SA accepted this recommendation and approved the revised XES Standard. After another editorial process, the final version of the XES Standard was published by IEEE SA on 8 Sep 2023 [12].

The published standard can be found in the IEEE Digital Library (through the URL https://standards.ieee.org/ieee/1849/10907/), and can be referred to using the DOI 10.1109/IEEESTD.2023.10267858.

As a next step, we would like to put a call to the process mining tool vendors to certify their tools' compliance with the revised XES standard. Certification can provide several benefits. First, there is a guarantee of accuracy for import and export formats between certified tools. Second, the certification can act as an effective marketing tool for testing, calibration and measurement, and a passport to submit tenders to contractors that require verified event-data creation tools. Finally, the certification process can provide a performance benchmark of a tool against other process mining tools in the market as it provides an independent technical evaluation as a measure of their treatment of the standardized event data format.

## Conclusion

In 2016, the XES standard was approved by the IEEE SA as the IEEE 1849-2016 - IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams [11]. Since then, the IEEE Task Force on Process Mining has been undertaking a range of activities, including the revision of the standard, publishing data sets using the XES format, promoting the use of the standard format in academia and industry, creating a certification mechanism for tools that have adopted it. A revised standard (IEEE 1849-2023) was published on 9 September 2023, which now supersedes the earlier 2016 standard.

The steering committee of the IEEE Task Force on Process Mining also conducted an online survey with 289 participants spanning the roles of practitioners, researchers, software vendors, and end-users to collate the experiences of the international process mining community for collating and extracting event data input for process mining [10]. This survey unearthed several challenges related to the extraction of event data input for process mining. One such challenge discussed was the need for a richer notion of event data. An object-centric event data allows for the analysis of data in situations where there are potential multiple case identifiers. Although the XES standard supports the introduction of extensions, it may come at the cost of complexity. Thus, the Task Force has undertaken work to revisit the core concepts in an event log and proposed a draft conceptual data model for an object-centric event log that can support complex data structures (including many-to-many relationships between multiple objects, cases and events). The standardized Artifact Lifecycle extension of the XES standard can also act as a stepping stone for such developments.

## Acknowledgment

This work was supported in part by the IEEE Computational Intelligence Society (CIS) Standards Committee and IEEE staff members – in particular, Vanessa Lalitte, and Julie Alessi during the formal standard revision process, and in part by the Steering Committee of the IEEE Task Force on Process Mining.

## References

[1] W. M. P. van der Aalst, *Process Mining: Data Science in Action*. Berlin, Germany: Springer, Berlin, 2016.

[2] W. M. P. van der Aalst and J. Carmona Eds., *Process Mining Handbook, (Lecture Notes in Business Information Processing Series)*, vol. 448. Berlin, Germany: Springer, 2022.

[3] M. de Leoni and F. Mannhardt, "Road traffic fine management process," Eindhoven Univ. Technol., Feb. 10, 2015, doi: 10.4121/UUID:270FD440-1057-4FB9-89A9-B699B47990F5.

[4] B. van Dongen, "BPI Challenge 2021," Eindhoven Univ. Technol., Eindhoven, The Netherlands, Apr. 2012, doi: 10.4121/UUID:3926DB30-F712-4394-AEBC-75976070E91F.

[5] B. van Dongen, "BPI Challenge 2017," Eindhoven Univ. Technol., Eindhoven, The Netherlands, Feb. 2017, doi: 10.4121/UUID:5F3067DF-F10B-45DA-B98B-86AE4C7A310B.

[6] G. Acampora, A. Vitiello, B. D. Stefano, W. van der Aalst, C. Günther, and E. Verbeek, "IEEE 1849: The XES standard - the second IEEE standard sponsored by IEEE computational intelligence society," *IEEE Comput. Intell. Mag.*, vol. 12, no. 2, pp. 4–8, 2017.

[7] M. Kerremans, K. Iijima, A. Sachelarescu, N. Duffy, and D. Sugden, "Magic quadrant for process mining tools, gartner research note GG00774746," 2023. [Online]. Available: www.gartner.com

[8] F. Mannhardt, "Sepsis cases - event log," Eindhoven Univ. Technol., Dec. 2016, doi: 10.4121/UUID:915D2BFB-7E84-49AD-A286-DC35F063A460.

[9] W. van der Aalst et al., "Process mining manifesto," in *Proc. Int. Conf. Bus. Process Manage.*, 2012, pp. 169–194.

[10] M. T. Wynn et al., "Rethinking the input for process mining: Insights from the XES survey and workshop," in *Proc. Int. Conf. Process Mining*, 2021, pp. 3–16.

[11] *IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams*, IEEE Standard 1849-2016, 2016, doi: 10.1109/IEEESTD.2016.7740858.

[12] *IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams*, IEEE Standard 1849-2023, 2023, doi: 10.1109/IEEESTD.2023.10267858.