

EVENT LOG STANDARDS IN BPM DOMAIN

BEATA GONTAR, ZBIGNIEW GONTAR

Department of Computer Science, Management Faculty, University of Łódź

Automated Business Process Discovery (alias Process Intelligence, Process Analysis, Workflow Mining or Process Mining) includes methods, standards and tools to support the discovery and analysis of operational business processes. Paper aims to review ongoing standards for storing and managing event log data, which is starting point for process mining, and verify in what range chosen standards meet the needs of event log analysis. We discuss CWAD, BPAF, PROV-DM, PROV-O, MXML, SA-MXML, and XES formats. Findings of the paper are based on the literature review and analysis of selected software.

Keywords: automated business process discovery, process intelligence, process mining, event logs standards

1. Introduction

Integrating Business Process Management (BPM), Data Warehouses and Business Intelligence (BI) is one of the emerging trends in Business Informatics [1]. BI refers to the traditional systems that assist managers in making decisions based on the facts of the business data stored in Data Warehouses. In the context of BPM, Data Warehouse covers event logs, and analytical systems perform event log analysis, in the similar way, in which traditional BI systems perform analysis of data from databases. All these circumstances cause the need for standardization in event log in BPM domain. Recent evolution of event logging could be one of manifestations of the trend, which was mentioned above. Let us notice, that event logs

refer not only to the BPM domain. *Application domain, as well as the purpose of an event log, the addressed the targeted business layer and the logging type decide about what kind of events are logged, which event data is considered, and how events are technically represented depends* [2]. J. Becker et al. classified event formats into four domains, i.e. Business Process Management (BPM), Complex Event Processing (CEP), System Interoperability (SI), and IT Security. Generally speaking, there are two main resources for standardization efforts in this area, i.e. business represented by standardization community like W3C, Workflow Management Coalition (WfMC), etc., and academic community represented by IEEE Task Force on Process Mining. Apart from that, there are many proprietary process log formats.

In principle, for each domain mentioned above, data mining/process mining algorithms could be applied (see: references to selected conference papers [5, 11]). In this paper, we focus on a BPM domain, in which the main role in academic community seems to have IEEE Task Force on Process Mining. The aim of this article is to present the selected data log formats, and verify in what range chosen standards meet the needs of event log analysis.

2. A brief history

Process logs files are available in the BPM systems since the late 1980s, and early 1990s [14]. Most BPMS used proprietary formats for log representation, which causes difficulties in attempts to use event logs in Business Activity Monitoring (BAM) and process mining. The first attempts to standardize the process logs are the effect of an effort of Workflow Management Coalition (WfMC). In 1996-1998, WfMC proposed Common Workflow Audit Data (CWAD). In [13], zur Muhelen put some examples of audit trails on the base of CWAD model, including tables from IBM MQSeries Workflow, Staffware 2000, and Carnot Process Engine. In 2008 – 2009, WfMC proposed XML audit data format, inspired by CWAD, called Business Process Analytics Format (BPAF). According to Paul Buhler [4], in 2011, *JBoss RiftSaw BPEL engine became the first substantial BPAF implementation in the world, mapping Apache ODE (Orchestration Director Engine) events onto BPAF*. In 2012, RiftSaw BPEL engine was implemented in JBoss SwitchYard service delivery framework, and JBoss Overlord, which is a SOA Governance platform that provides a new Business Activity Monitoring (BAM) capability with a Gadget-based interface. In 2012, Provenance Working Group in World Wide Web Consortium (W3C) works on development of Provenance Data Model (PROV-DM), and Provenance Ontology (PROV-O) standards. Both refer to the provenance concept, and aimed in providing provenance information with the data in a semantic framework.

Academic community proposed its own way to standardize process logs. In 2005, a group of researchers from Technische Universiteit Eindhoven (TU/e),

proposed Mining XML Format (MXML). In 2007, on the base of EU-funded SUPER project, TU/e community proposed Semantically Annotated Mining eXtensible Markup Language (SA-MXML). In 2009, Christian Günther et al. from TU/e, and Fluxicon Inc. proposed Extensible Event Stream (XES) format, adopted in 2010 as the Task Force on Process Mining standard. In parallel with the work of standardization in the field of event logs, TU/e community worked on software for process mining, resulted in developing ProM framework.

According to Michael zur Muehlen et al., the first mention of the use of the logs for the purposes of analysis comes from 1996 [13]. In 2012, IEEE Task Force on Process Mining released a Process Mining Manifesto, which describes three basic types of process mining, referring to the efforts of TU/e community, i.e. *discovery*, *conformance checking*, and *enhancement*. Process model discovery relates to constructing complete and compact process models able to reproduce the observed behaviour of the process instances. Conformance checking means diagnosis on the base of observed reality in the sense of checking whether the modelled behaviour matches the observed behaviour. Process model enhancement is the projection of information extracted from the logs onto the model, to make the tacit knowledge explicit and facilitate better understanding of the process model. This understanding of process mining dates back to the time of developing in 2003 α -algorithm by van der Aalst, Weijter and Mărușter from TU/e [12].

3. Event log standards

Among the most important event log standards specified in [2] (see: Table 1.), we chose BPAF, CWAD, PROV-DM, PROV-O, MXML, SA-MXML, and XES for detailed exploration.

Common Workflow Audit Data (CWAD) was released for standardisation process audit events by WfMC in 1996 (first published version of CWAD 1.0 is dated of 1 November 1996). In 1998, WfMC released CWAD Specification 1.1 (22 September 1998). According to the WfMC reference model (Fig. 1), *workflow engine captures events occurred in processes, activities, and manual work, and persist them in a form that complies with the abstract CWAD model, which defined not how the data should be stored, but what information was to be gathered and made available for analysis*. Motivation for event logs auditing was recovery, proof of process execution, visualization of the flow of expired process instances, compensation, Business Process Reengineering (BPR). So as to implement the CWAD model, corresponding database table was needed.

The Common Workflow Audit Data (CWAD) structure consists of a prefix, a body and a suffix (see: Fig. 2).

Table 1. Classification of event formats

Domain	Event formats	Standardized by
BPM	Business Process Analytics Format (BPAF)	Workflow Management Coalition (WfMC)
	Provenance Data Model (PROV-DM)	World Wide Web Consortium (W3C)
	Provenance Ontology (PROV-O)	World Wide Web Consortium (W3C)
	Common Workflow Audit Data (CWAD)	Workflow Management Coalition (WfMC)
	Mining eXtensible Markup Language (MXML)	Academic community / van der Aalst et al.
	Semantically Annotated Mining eXtensible Markup Language (SA-MXML)	Academic community / Werner Janush, Anna Karla Alves de Medeiros, Wil van der Aalst, Peter van der Brand, Ton Weijters (TU/e)
	Extensible Event Stream (XES)	IEEE Task Force on Process Mining / Christian Günther (TU/e- Fluxicon)
CEP	RuleCore Event Format (RCEF)	ruleCore
	WebSphere Business Events XML Format (WBEF)	IBM Corporation
SI	Common Base Event (CBE)	IBM Corporation initiative. Standardized by OASIS under WSDM Event Format (WEF)
	Common Event Expression (CEE)	The MITRE Corporation
	Web Services Distributed Management Event Format (WSDM WEF)	OASIS
IT Security	Common Event Format (CEF)	ArcSight
	Intrusion Detection Message Exchange Format (IDEMF)	Internet Engineering Task Force (IETF)
	Incident Object Description Exchange Format (IODEF)	Internet Engineering Task Force (IETF)
	Distributed Audit Service (XDAS) Common Audit Event Record Format (XCAERF)	The Open Group
	Event Metamodel and Profile (EMP)	Object Management Group (OMG)

Source: own preparation on the base of [2]

In this sense, database transaction concepts, on the base of CWAD model, could be extended for workflow purposes, and data structure could be enhanced to accommodate proprietary attributes (see: [13]). *The body part of the data is variable depending on the event type that caused the audit trail to be written. Recorded events related to the state changes: instantiation, start, and completion, and could be caused by a change to the process instance, activity instance or a change to a work item. The audit data consisted of three kinds of information: Basic, Discretionary and Private.* A process instance can have a parent process instance called initial pro-

cess instance in the data model. An activity has an application associated with it. A user has a corresponding domain and a corresponding node (network location). A process instance and an activity instance can have attributes. The attributes are of a certain data type and have a value and a length. A process instance, an activity instance and a work item have a state at a certain point in time.

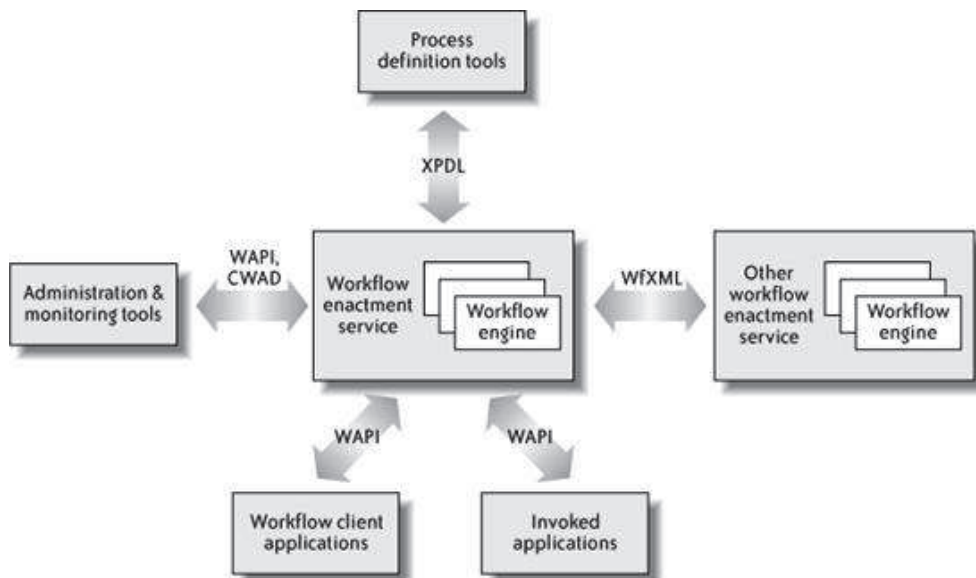


Figure 1. WfMC reference model. Source: WfMC

Zur Muehlen et al. proposed audit trail table structure and data mart structure for the CWAD model [13].

Business Process Analytics Format (BPAF) history dates back to the 2000ties. Version 2.0 of BPAF, released in 2008 by WfMC, was extension of CWAD version 1.1 in the sense of XML Schema for Audit Data Events. Version 2.0 R1 introduced revised states for Audit Data Events, alignment with Process State Model, and WS-Human Task (23 April 2008). In Version 2.0 R2, released 8 December 2008, State Model Graphics and Legend, Information of Event Data Format, and Information on BPAF State Model were added. The structure of events in the BPAF notation includes the event ID, a reference to the process model, the process instance and its state, the name of the process and activities, information related to the activities, information about the previous state action or process, and additional business. This extended information could be applied to social data analysis, in the sense of identification the structure of relations between actors in the process, and traditional data analysis derived from statistics, machine learning and artificial intelligence, in the sense of their exploration of the processes.

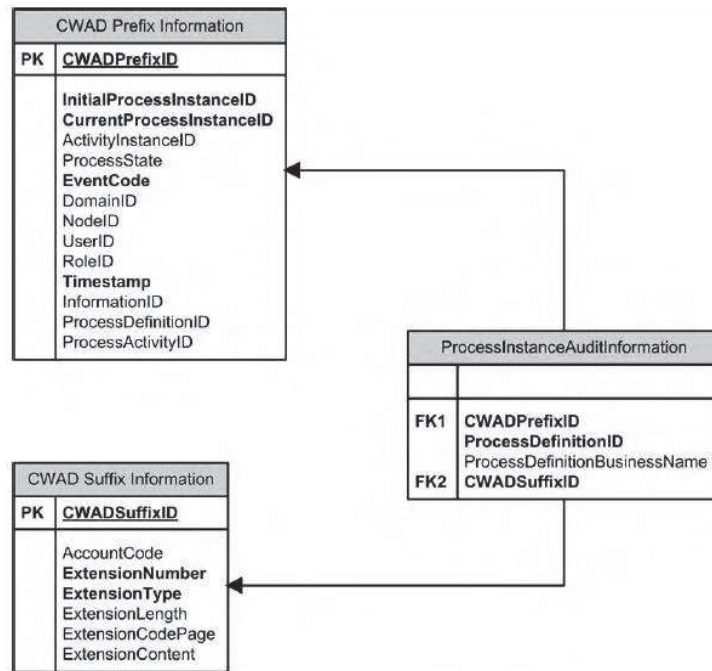


Figure 2. CWAD structure. Source: [13], and WfMC

Modelling the process dynamics on the base of BPAF data is based on the concept of state process model. Similar solutions have been implemented in BPEL4People and WS-HumanTask. In BPAF, state process model is defined in terms of activities and processes. We can distinguish the initial state (running instance) and final state (completed instance). The process (or action) may be ready to run, suspended, executed, cancelled or terminated. The event log can save the status of the process in terms of running / completed at a general level or at the detailed level, and in this way enabling the analysis of the dynamics of the process or activity.

Provenance Data Model (PROV-DM) and Provenance Ontology PROV-O, introduced in 2012, deal with the problem of *from where an entity originated, how it was produced and what has happened to it since creation* [9]. Workflow could be described as an entity, and in this sense its provenance could be expressed in PROV-DM. Having workflow described in PROV-DM, it is possible to perform provenance analysis, e.g. detecting duplicate records in workflow results [3]. PROV-DM specification, referring to the Web resources, includes a schema (Figure 3), consistency constraints and interface rules to the schema, and a language for recording provenance facts.

Application of PROV-O enable performing the following analysis: business process compliance [4] (*determine that execution was compliant with a higher level BPMN model of understanding, verification, that participants correctly followed a choreography protocol*), regulatory compliance (*with mobile process participants executing distributed lightweight workflow applications we can aggregate the events into a coherent history – the merging of the semantic graphs does this naturally*), process mining (*use of advanced pattern matching against semantic business process events to accomplish process mining and/or identify workflow patterns*), and context aware process execution (*enable greater use of sense and response techniques for process agility*).

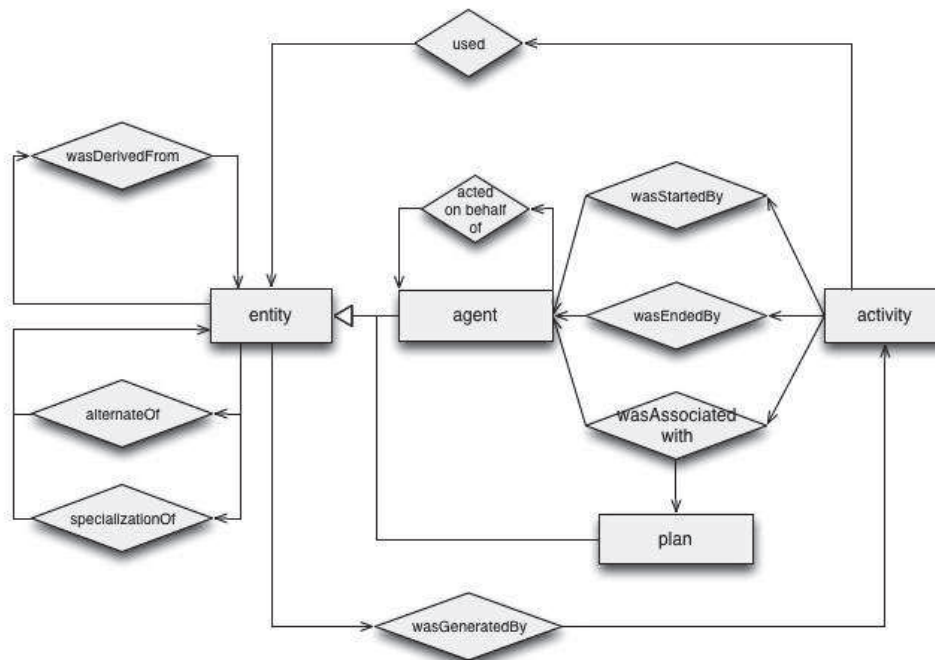


Figure 3. PROV schema. Source: [9]

Mining XML (MXML) started in 2003 as an academic initiative to share a common input format among different mining tools [12]. Defining MXML was the first step towards the creation of a repository on which process mining researchers can test their algorithms [16]. MXML was defined in [17].

According to the schema for the MXML format (Figure 4), Workflow Log contains one or more Processes with additional Data, and information about the Source of the log.

Every Process has zero or more Process Instances, and every Process Instance has zero or more Audit Trail Entry, which in turn have Workflow Model Element Id, Event Type, and optionally Time Stamp and Originator information. The event type is related to the state of the task, i.e.: schedule, assign, reassign, start, resume, suspend, auto skip, manual skip, withdraw, complete, ate abort, pi abort and unknown. It is possible also to log additional information.

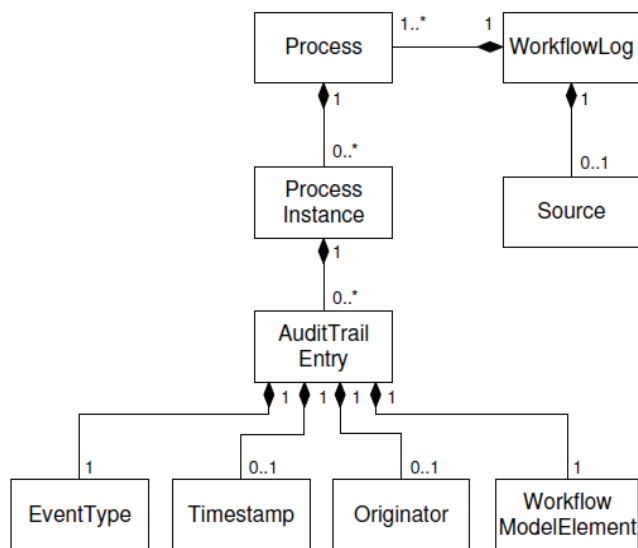


Figure 4. MXML meta model. Source: [17]

Semantically Annotated Mining eXtensible Markup Language (SA-MXML) is a semantic version of the MXML format, introduced in 2008 by TU/e community. Elements of SA-MXML (apart from Audit Trail Entry and Time Stamp) have an attribute (model Reference), which refers to ontologies with the defined process concepts, and represents semantic description of process. SA-MXML format is backwards compatible with MXML format.

eXtensible Event Stream (XES) was introduced in 2009 by Christian W. Günther from Fluxicon Process Laboratories – a spin-off of the process mining research done at TU/e. According to the schema for the XES format (Figure 5), process (Log) and track (Trace) of the process are defined by their elements (Event). Classifier, defined through a set of attributes, refer to the events. The extensions are defined in the sense of concept, life cycle, organizational, time, and semantic, and are used to define a set of attributes on any levels of the XES log hierarchy.

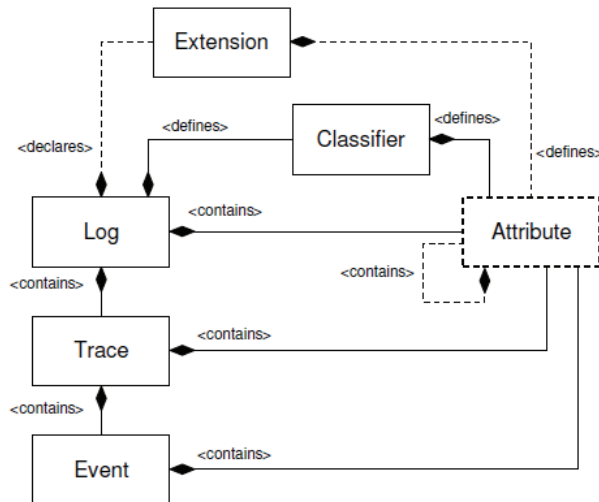


Figure 5. XES meta model. Source: [18]

4. Application comparison

Software tools for process mining are still in its infancy. The flagship product of academic community is Process Mining (ProM) framework. List of currently available process mining algorithms, implemented in ProM, is impressive (*there are more than 100 packages containing more than 400 plug-ins available, and each plug-in refers to control-flow mining techniques, the organizational perspective, mining less-structured, flexible processes, the verification of process models, verification of Linear Temporal Logic formulas on a log, performance analysis, log filters, etc.*: <http://www.promtools.org/prom6/>). Most of them are experimental and with mainly scientific importance.

There are two spin-offs recruited from TU/e community, i.e.: Fluxicon (<http://fluxicon.com/>) (Nitro, Disco) and Futura Process Intelligence – acquired by Perceptive Software. The tools originated in TU/e environment, explore MXML, SA-MXML, and XES standards, while the other vendors use proprietary format.

The most common formats used in software packages are MXML and XES. In fact, the domain specific requirements are not fullfield. For instance, two promising approaches - BPAF and XES, can represent specific even type in BPM. BPAF is used for state models and process audit events. XES has a flexible structure which allows for modeling business events. It is one of the reasons the vendors supports own solutions [19].

Table 2. Comparison of tools delivered by vendors with used standards

Software	CWAD	BPAF	PROV-DM	PROV-O	MXML	SA-MXML	XES
ARIS Process Performance Manager (Software AG)							
Perceptive Software (Perceptive Process Mining)							
Disco (Fluxicon)					Y		Y
Interstage Automated Process Discovery (Fujitsu)							
Nitro (Fluxicon)							Y
ProM (Academic)					Y	Y	Y
ProM Import (Academic)					Y	Y	Y
QPR Process Analyser/Analysis (QPR Software)							
XESame (Academic)							Y

5. Summary and conclusion

This article focuses on the Common Workflow Audit Data (CWAD), Business Process Analytics Format (BPAF), Provenance Data Model (PROV-DM), Provenance Ontology (PROV-O), Mining XML (MXML) and Extensible Event Stream (XES) as the standards that are currently the most expressive and the easiest for integration with the monitoring and diagnosis level of BPM life cycle.

Complexity of data: *distributed, heterogeneous data from software and hardware components, several types of logs: networks, databases, servers, applications; and heterogeneity of data log formats: proprietary data formats, standardization problems, manpower costs to configure, maintain and tune collector/adapters, specific code needed for each format, difficult to correlate for end-to-end problem diagnostics; instrumentation: cost for ownership of all formats analysis tools, complexity to interoperate several analysis tools, standard compliance* – all these obstacles course the problems in developing process mining systems.

Currently, process mining algorithms are limited to discovery, conformance checking, and enhancement (see: Process Mining Manifesto, <http://www.win.tue.nl/ieeetfpm/>). In the future, process mining algorithms could be extended for emotional process analysis [12], provenance analysis [9] and semantic analysis.

REFERENCES

- [1] Ballard C., White C., McDonald S., Myllymaki J., McDowell S., Goerlich O., Neroda A., (2005) *Business Performance Management... Meets Business Intelligence*. International Business Machines Corporation.
- [2] Becker J., Matzner M., Müller O., Walter M., (2011) *A Review of Event Formats as Enablers of event-driven BPM*. In: Proceedings of the 5th International Workshop on event-driven Business Process Management (edBPM 2011), Lecture Notes in Business Information Processing. Clermont-Ferrand, France.
- [3] Belhajjame K., Missier P., Goble C.A., (2012) *Detecting Duplicate Records in Workflow Results*. In: Proceedings IPAW'12. Springer-Verlag.
- [4] Buhler P., (2012) *Toward the Semantic Processing of Business Process Events*. In: Proceedings of the SOA Cloud Service Technology Symposium. London, UK.
- [5] Dasgupta D., Rodríguez J., Balachandran S., (2006) *Mining security events in a distributed agent society*. In: Proceedings of Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security Conference. SPIE.
- [6] de Medeiros A., Karla A., van der Aalst W., Pedrinaci C., (2008) *Semantic Process Mining Tools: Core Building Blocks*. ECIS 2008 Proceedings. Paper 96. <http://aisel.aisnet.org/ecis2008/96>.
- [7] Folino F., Greco G., Guzzo A., Pontieri L., (2009) *Discovering Multi-perspective Process Models: The Case of Loosely-Structured Processes*. in: Filipe J., Cordeiro J. (eds.), Enterprise Information Systems 10th International Conference, ICEIS 2008, Barcelona, Spain, June 12-16, 2008, Revised Selected Papers. 2009, XVII, 361 p., Springer.
- [8] McLellan M., (1996) *Workflow Metrics. One of the great benefits of workflow management*. In: Oesterle, H., Vogler, P. (eds.) Praxis des Workflow-Management, pp. 301-318, Vieweg, Braunschweig.
- [9] Moreau L., Missier P. (eds.), (2012) *PROV-DM: The PROV Data Model*. W3C.
- [10] Schroder M. (ed.), (2012) *Emotion Markup Language (EmotionML) 1.0*. W3C.
- [11] Shukla V., (2011) *Extending Oracle CEP with Predictive Analytics*. Retrieved from: https://blogs.oracle.com/CEP/entry/extending_oracle_cep_with_pred. 19.10.2012.
- [12] van der Aalst W.M.P., Weijter A.J.M.M., Maruster L., (2003) *Workflow Mining: Discovering process models from event logs*. In: IEEE Transactions on Knowledge and Data Engineering, vol 16.
- [13] zur Muehlen M., (2002) *Workflow-based Process Controlling. Foundation, Design, and Application of Workflow-driven Process Information Systems*. Logos Verlag.
- [14] zur Muehlen M., Swenson K.D., (2010) *BPAF: A Standard for the Interchange of Process Analytics Data (September 3, 2010)*. Business Process Management Workshops: BPM 2010 International Workshops and Education Track, Hoboken, NJ, USA, September 13-15, 2010. Available at SSRN: <http://ssrn.com/abstract=2030857>

- [15] The Workflow Management Coalition, (1998) Workflow Management Coalition Audit Data Specification. The Workflow Management Coalition Specification.
- [16] Saravanan M.S, Rama Sree R.J, (2011) *Application of Mining Algorithms using ProM and Weka Tools*. IJCST Vol. 2, Issue 3, September 2011.
- [17] van Dongen B.F., van der Aalst W.M.P., (2005) *A Meta Model for Process Mining Data*. In Proceedings of the CAiSE WORKSHOPS, volume 2. FEUP.
- [18] Gunther C. W., (2009) *XES Standard Definition*. 1.0. <http://www.xes-standard.org/>
- [19] Becker J., Matzner M., Muller O., Walter M., *A Review of Event Formats as Enablers of event-driven BPM*. <http://www.springerlink.com/content/jjr464703773r1w3/> (access: 12.X.2012)