A Comparative Study of Workflow Mining Systems

Mahdi ABDELKAFI, Dorra BOUSABEH,
MIR@CL Laboratory/University of Sfax, Route de Tunis,
BP 242, 3021 Sakeit Ezzit, Sfax, Tunisie.
Mahdi.abdelkafi@hotmail.fr, Dorra.bousabeh@gmail.com

Lotfi BOUZGUENDA, Faiez GARGOURI
MIR@CL Laboratory/University of Sfax, Route de Tunis,
BP 242, 3021 Sakeit Ezzit, Sfax, Tunisie.
Lotfi.bouzguenda@gmail.com,Faiez.gargouri@isimsf.rnu.tn

***Abstract— this* work deals with Workflow Mining (WM) a very active and promising research area. First, in this paper we precise the main uses cases of WM, and then we present a critical and comparative study of three representative WM systems of this area: the ProM, InWolve and WorkflowMiner systems. The comparison is made according to quality criteria that we have proposed such as the capacity to filter and convert a Workflow log, the capacity to discover workflow models and the capacity to support Multi-Analysis of processes. One of the most components specific to these systems is the Workflow log which is the basis for any WM system. This paper also compares their workflow log models by considering five requirements (Usability, Refer to the three workflow models, Simplicity, EventStream and TimeStamp). Finally, we summarize this study and we explain how Multi-Agent Systems (MASs) approach to cope with organizational model discovering issue. In our work, we mean by organizational model, the organizational structures (federation, coalition, team, etc.) and the interaction protocols (contract net, auction, vote, etc.).**

Keywords-Workflow Mining; Workflow models; Multi-Agent Systems Approach; Reverse engineering of Business Processes.

I. INTRODUCTION

Workflow Mining Context. The Workflow technology aims at automating the coordination of activities composing business process [1]. A Workflow Management System (WfMS for short) permits to define, implement and perform one or several business processes. The traditional interest of a WfMS focuses on design, configuration and enactment phases [2]. The figure 1 shows the life Cycle of Workflow.

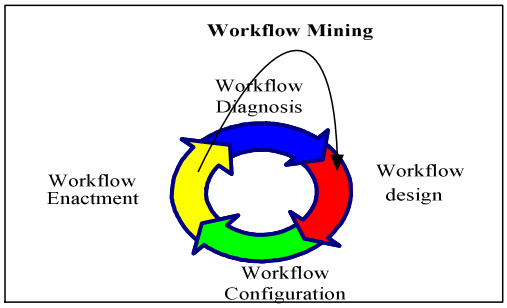

Figure 1. Life Cycle of Workflow

As a consequence, there are a few supports for diagnosis phase. Besides, the support for the design phase is limited to provide a simple editor yet that an analysis is useful for the design which is neglected. A very few WfMS propose the simulation, the checking of design step and also support to help the interpretation of data from execution traces. Although the majority of actual WfMS collect the business instances execution traces, they do no propose a support to exploit this important information. In this context, the Workflow Mining (WM) area has recently appeared and considered as a current research area. More precisely, WM aims mainly at analyzing the workflow execution traces (or Workflow log) in order to discover the Workflow models such as the Organizational Model (OM), the Informational Model (IM) and the Process Model (PM), which help the monitor to improve or propose a new workflow [3]. The WM is considered as an appropriate method to support Business Processes Reverse engineering (BPR).

According to literature, the workflow mining is used for three main cases: firstly, it is used to discover the previous models in order to design a workflow. Secondly, it is used to analyze the performances of activities composing the process and actors that perform these activities. Thirdly, it is used to check whether the deployed process corresponds to the prescribed process.

In order to support the previous cases of uses, several workflow mining systems have been proposed. Unfortunately, a good/quality workflow mining system has never been proposed. Most of the proposed workflow mining systems do not meet the previous uses cases. Regarding the first use case of WM, all the existing contributions only focus on the process model discovering by providing some algorithms like the alpha algorithm [4], and the neural networks and markovian approaches [5]. The organizational model discovering is not considered in the existing WM systems (ProM [4], InWolvE [6] and Workflow Miner [7]) except the ProM system. This latter supports only the discovering of actors and their roles and five kind of social network (Handover of work, Subcontracting, Working together, Reassignments, Doing similar tasks) [8]. We show later that the discovering of organizational model is not limited to the previous elements but also it includes others kind of social networks (federation, coalition, market, hierarchy, etc.) and interaction protocols (contract net, vote auction, etc.) as well defended in multi-agent systems approach [9].

Regarding the latter uses cases, a few propositions which propose analysis techniques such as delta analysis and

performance analysis to validate the discovered process (4).

In this context, the definition of quality criteria for such WM system is needed in order to ensure the previous uses cases and especially all workflow models discovery and their analysis for Business Processes Reverse engineering.

The problem being addressed in this paper can be resumed according to the following questions: "*which quality criteria are considered to develop a good WM system*? *How do we compensate the main drawbacks of the existing systems?*

The contribution of the paper is to provide a framework for studying Workflow Mining (WM). First, it proposes a set of quality criteria for evaluating existing WM systems. Then, it identifies a set of requirements for describing workflow log model which is considered as the basic for any WM system. Finally, it explains how the MASs approach can help the discovery of organizational model thanks to its principal characteristics.

Organization of the paper. This paper is organized as follows. Section 2 recalls briefly the main uses cases of Workflow mining. Section 3 presents the quality criteria that we have defined. Section 4 describes three systems representative of the WM area. We compare them according to these quality criteria. Section 5 compares the Workflow log models of the studied systems according to the key requirements that we have identified. Finally, we will summarize this study and we will explain how multi-agent systems approach can compensate the main drawbacks of the studied systems.

II. MOTIVATIONS: WHY WORKFLOW MINING?

The workflow mining is mainly used for three reasons which are the following:

- The discovery of workflow models such as the informational, organizational and process models to design a workflow;

- The analysis of performances: it concerns mainly the analysis of performances of a workflow component such as activity, actors, etc. Then, to proceed to the modification of existing models. According to the literature [8], four performance metrics have been proposed for process perspective such as flow time, waiting time, processing time and synchronization time. Regarding organizational model, four metric have been also proposed such as frequencies, time, utilization and variability;

- The Delta analysis: it consists of comparing the prescribed processes and the deployed processes. This analysis of differences permits to proceed then to adjust and/or enhance the processes. It also allows the comparison of different implementations of process within various organizations.

III. QUALITY CRITERIA FOR WORKFLOW MINING SYSTEM

To define quality measuring criteria of a WM system we must know the functions which we await from this system

A. *WM system functions*

The most important functions are the following (see figure 2):

- **Pre-Handling of Workflow log;**
- **Support for Multi-Analysis of Processes;**
- **Support for Discovering of Workflow Models.**

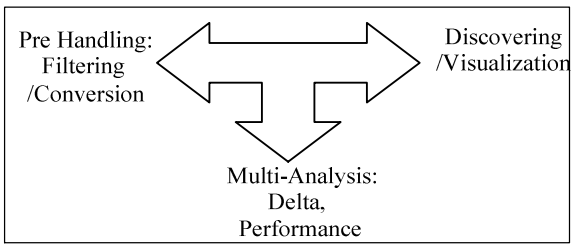

Figure 2: Principal functions of WM system

Let us detail each function.

***Pre-Handling of Workflow log*.** The WM system must allow, if necessary, the filtering of Workflow log at the first time in order to reduce the noise being in the log such as the separation between the principal activities and the optional activities known logistics activities and the taking into account only the completed activities. Often, the noise being in workflow log exists at ad-hoc workflow systems and groupware products which are based on unstructured process activities. At the second time, it must allow the conversion of the filtered Workflow log to known format like XML.

***Support for Multi-Analysis of Processes*.** As mentioned previously, a WM system must support various analyses of processes and namely the delta analysis and the performances analysis;

***Support for Discovering of Workflow Models*.**

Such workflow mining system must allow the discovering of the three workflow models which are in interaction. The *Organizational Model* (OM) has two objectives. First, it structures actors in classes playing each one a specific role (for instance subordinate and chief roles in strict hierarchical structure). A class is called an organizational structure defining the interaction space between actors. Second the OM describes the activities allocation policy between actors thanks to the interaction protocols. The *Informational Model* (IM) describes the structure of forms, documents and data which are consumed and produced by processes. The *Process Model* (PM) defines components activities, their coordination, information and actors involved in each activity.

B. *Quality criteria*

The functions presented above give a global view of what should be a WM system. Now let us present in detail possible quality criteria of a WM system.

Concerning the first function, we have defined criteria to measure ***the capacity to filter and convert a Workflow log***:

The possibility to **filter** a workflow log in order to reduce the noise elements mentioned above.

The possibility to **convert** workflow log from given format to other (for instance from text to XML) in order to ease the extraction process of information and then to analyze it.

Concerning the second function, we have defined criteria to measure ***the capacity to discover workflow models****:*

Informational Model: the ability to discover the consumed documents and produced documents by a workflow.

Process Model: The ability to discover clearly the activities and their coordination (sequential, parallel, iterative, etc.).

Organizational Model: The ability to discover clearly the actors, roles, their organizational units, the policy of activities allocation (i.e. the employed interaction protocols between actors and social networks describing the nature of collaborations between actors (federation, coalition, hierarchy, market,…).

Concerning the third function, we have defined the criteria to measure *the* ***capacity to support Multi-Analysis of processes:***

The possibility given to designers to make various analysis such as **Delta Analysis** and **Performance Analysis** in order to proceed to the adjustments of processes.

In addition to these criteria, other ones must be taken into account and which are related to the quality of software in general such as **Usability of interfaces**, **Portability** and **Extensibility**. Regarding the usability of interfaces, the ability of the system to allow designers (i) to model clearly the workflow models with graphic representation (Petri Nets for instance) and (ii) to make simulations and animations for detecting some errors and ambiguities.

IV. COMPARATIVE STUDY OF SOME EXISTING SYSTEMS

In this section we examine in detail some existing systems representative of this area. More precisely we describe the ProM [4], InWolve [6) and WorkflowMiner [7] systems, and we compare them according to criteria that we have defined. Note that the existing WM systems such as EmiT, Thumb, MinSocN and MiMo are merged within the ProM framework. For more information on comparative study of existing WM systems we refer to [10].

A. ProM system

ProM (Process Mining) is an extensible platform developed with java language within university of technology at Eindhoven. The purpose of this platform is to support the maximum of discovering techniques of workflow perspectives/models thanks to the offered plug-ins. It also ensures certain flexibility on the input and output formats thanks to the conversion tools. The figure 3 shows the ProM System architecture.

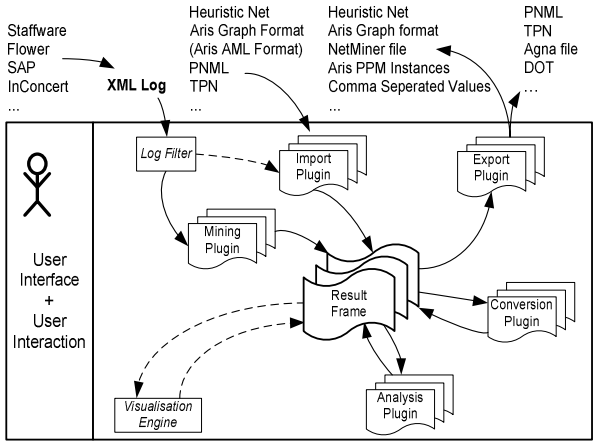

Figure 3: Architecture of the ProM system.

ProM takes in input a workflow log issued from several actual workflow management systems such as YAWL [11], Flower [12] and so on. The workflow log is often represented with XML format and filtered using the component LogFilter in order to reduce the noise as defined previously. It includes a set of plug-ins which is the following.

- *Import plug-in* to support a variety of graphic representation such as Petri Net, Social network…

- *Mining plug-ins* to allow the extraction of data from workflow log and then to store them in result entities or result frame. These latter can be used for visualization using import plugin.

- *Analysis plug-in* to analyse the discovered business processes like delta analysis and performance analysis.

- *Conversion plug-in* which converts a given model form format to another. For instance of EPC to the Petri Net

From a Pre-handling point of view, the ProM system supports the both functions filtering (LogFilter) and conversion.

From a capacity to discover workflow models point of view, the ProM system ensure the discovering of organizational and process models. Regarding the organizational model, it permits to discover some elements like actors, roles and a limited number of social networks. While the process model, it is discovered using the Alpha algorithm and visualized with Petri Net formalism.

From a capacity to support Multi-Analysis of business processes, ProM offers mainly two types of analysis such as Delta and Performances analysis.

From a system particularity point of view, it is portable since it has been developed in java. It is also extensible since it

is possible to add a new plug-in without changing the framework and offers sophisticated interfaces to allow simulation, checking and validation of business processes models.

B. *InWolve system*

The InWolve (**In**ductive **Wo**rkflow learning **v**ia **E**xamples) system has been developed by Joachim Herbst with C++ [6]. The figure 4 shows the architecture of the InWolve System.

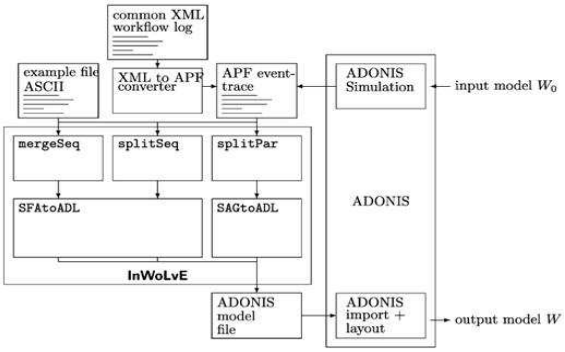

Figure 4: Architecture of InWolve system

The InWolve system supports three formats for Workflow log such as XML, ASCII and APF (Original format of Adonis WfMS). The proposed mining process by InWolve is composed of two steps.

a) *The Induction step* consists in analysing the workflow log in order to produce a stochastic activities graph.

b) *The transformation step* consists in refining the obtained graph according to the following sub steps: *Analysis of synchronization of processes instances structures*, (ii) *generating the synchronization of the process model structure* and (iii) *the generation of process model with ADL format*.

From a Pre-handling point of view, the InWolve system supports the filtering and conversion (for instance from XML to APF) of workflow log.

From a capacity to discover workflow models point of view, the InWolve supports only the discovering of process model. More precisely, the InWolve system creates in the first step a stochastic activity graph from workflow log and in the second step; it transforms the activities graph into a well defined process workflow model.

From a capacity to support Multi-Analysis of business processes, InWolve do not offer any technique of analysis of discovered processes.

From a system particularity point of view, InWolve is not portable since it has been developed with C++ and it does not offer useable interfaces and is not considered as an extensible system.

C. *WorkflowMiner System*

The WorkflowMiner system has been developed by Walid Gaaloul in ECOO team at University of Nancy-France. The aim of this system is to allow the discovering of workflow patterns from workflow log. This latter is generated by the Bonita WfMS in order to ensure a continuous design which answers to the business processes flexibility issue. The figure 5 shows the architecture of the WorkflowMiner System. More precisely, it includes the following components:

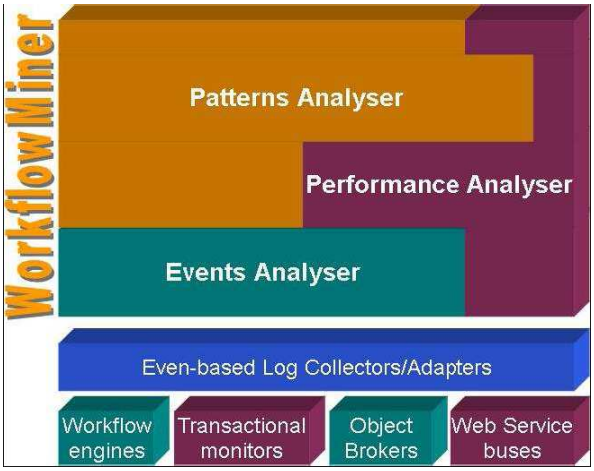

Figure 5. Architecture of WorkflowMiner System

Events based Log Collectors/Adapters consist in loading the execution traces from XML files and adapting them in Prolog predicates format;

Event Analyser defines the analysis engine of Prolog facts and discovers through a statistic technique the casuals' dependencies between events. It constructs the dependencies table which helps the construction of the workflow graph;

Patterns Analyser uses the predefined rules helping the discovery of workflow patterns. These rules are reproduced under form of predicates order first logic.

Performance Analyser uses the casuals' dependencies and discovered workflow patterns to measure the performance of workflow.

From a Pre-handling point of view, the WorkflowMiner supports only the conversion of workflow log form XML to Prolog events.

From a capacity to discover workflow models point of view, the WorkflowMiner supports only the discovering of process model by using a statistical technique. More precisely, it allows (i) the "local" workflow patterns mining that permits to cover partial results and (ii) the composing of those local discovered workflow patterns in iterative way until discovering the "global" workflow model.

From a capacity to support Multi-Analysis of business processes, WorkflowMiner offers only the analysis of performance of discovered processes.

From a system particularity point of view, it is portable since it has been developed in java but it does no propose useable interfaces as defended previously and is not considered as an extensible system.

D. *Comparative table*

In this section we have established a comparative table (see table 1) of the three described previously systems. The comparison is made according to the quality criteria that we have defined such as the capacity to filter and convert of Workflow log, the capacity to discover workflow models and the capacity to support Multi-Analysis of processes.

TABLE 1. Comparison of WM systems

Quality criteria \ Existing WM systems			ProM	InWolve	WorkflowMiner
Pre-Handling of Workflow log	Filtering		Log Filter	+	-
	Conversion		From EPC to Petri Nets	From XML to APF or ASCII	Adapters
Capacity to discover Workflow Models	Informational Model		-	-	-
	Organizational Model	Actors	Plug-in originator by task matrix)	-	-
		Roles	+	-	-
		Org Units	+	-	-
		Interaction Protocols	-	-	-
		Organizational structures	Plug-in "Social Network miner"	-	-
	Process Model		Plug-ins "Alpha algorithm"," Genetic mining", "Multi-phase mining"	Induction and tranformation methods	Patterns Analyser
Capacity to support Multi-Analysis of processes	Delta		Plug-in "Trace comparaison"	-	-
	Performance		Plug-in "a Basic statistical analysis	-	Performance Analyser
Particularities of system	Usability of interfaces		++	+/-	+/-
	Portability		Java	-	Java
	Extensibility		FrameWork	-	-

V. COMPARISON OF EXISTING WORKFLOW LOG MODELS

A. *Key requirements for a workflow log model*

Before detailing the key requirements that we propose for describing the workflow log model, let' us explain the reasons that justify their choice. First, according to the Workflow Management Coalition (www.wfmc.org), a Business process model is defined as a sequence of activities, their coordination, information and actors involved in each activity. In other terms, this model refers to the others models such as the informational model and the organizational model. For that we have chosen the requirement "refer to the workflow models" in order to have a comprehensive description of business processes. Second, it is important that the workflow log model must be simple with minimal concepts. Third, the workflow log model must answer to the two questions: What is the average time for performing an activity? How many activities are performed without interruption? Consequently, the workflow log must precise the TimeStamp and the EventStream for each activity. Finally, the workflow log model must be described with a standard representation in order to ensure the usability of the solution.

Let' us detail these requirements.

Usability: a Workflow log model must not only be easy to understand but also easy to be exploited by the user (i.e. based on standard format like XML).

Refer to the three workflow models: a workflow log model must refer clearly the three complementary workflow models as defined in section 3.

Simplicity: a Workflow log model must define the core concepts of the three workflow models. In other words, it must not be presented at a very detailed level (fine-grained) or very specific for a given process.

TimeStamp: a Workflow log model must precise start time and end time for each activity and as a consequence it is possible to measure services time and the utilization of the workforce (there are. performance metrics proposed by van der aalst).

EventStream: a Workflow log model must precise the event type for each activity such the aborted, failed, completed, compensated…

One of the most components specific to the WM systems is the Workflow log which is the basis for any WM system. We also compare their workflow log models by considering the key requirements below in the following section.

B. *A comparative study of existing Workflow log models*

Because of space constraint, we only give here a comparative study of these workflow log models according to the previous requirements (see table 2).

TABLE 2. A comparative study of existing Workflow log models

Requirements \ Workflow log models of studied WM systems			ProM	InWolve	Workflow Miner
Usability			+	+	+
Referencing the three workflow models	Informational Model		-	-	-
	Organizational Model	Organizational Structures	+/-	-	-
		Interaction Protocols	-	-	-
	Process Model		+	+	+
Simplicity			+	+	+/-
Event Stream			+	-	+
TimeStamp			+	-	+

Table 2 summarizes the qualities and the insufficiencies of

each studied Workflow log model.

The symbol +/- is interpreted as follows.

"+" If the model answers to the criteria.

"-" If the model does not answer to the criteria

This comparative study calls five remarks:

1) The usability criterion is assured by all the workflow log models. Indeed, all the models are described around XML standard.

2) The process model is supported by all the existing workflow log models while the other models are not supported by these models except ProM which considers only some elements like actors (or originator) and roles of the organizational model.

3) Any workflow log model supports the organizational structures and interaction protocols discovering as defined in introduction..

4) All the workflow log models are simple.

5) The two last requirements EventStream and TimeStamp are included in the majority of workflow log except the Workflow log of InWolve system.

VI. CONCLUSION

This paper has presented a Framework for studying Workflow Mining. First, it detailed the main uses cases of Workflow Mining. Second, it compared three existing systems representative of this area according to quality criteria that we have defined. Third, it studied the workflow log models of these systems according to the key requirement that we have also proposed. Even if these systems are powerful, they are not, in our opinion, completed since they neglect the important point that workflow is much more that process model. We believe that the main issue is due to the use of Workflow log having the following faults:

i) limited to activities execution,

ii) limited to actors whose execute these activities,

iii) no traces on the interactions among actors,

and (iv) semantic ambiguity in Workflow log (i.e. many interpretations). The figure 6 shows a fragment of a classical workflow log and explains in detail the last drawback.

Fragment of a classical Workflow log

Case	Activity	Actor
Case 1	Activity A	A1
Case 1	Activity B	A4
Case 1	Activity C	A1
Case 1	Activity D	A5

Many interpretations

1. A1 executes the Activity A, then delegates B to A4, who delegates Activity C. Finally, A1 delegates D to A5
2. the system automatically assigns Activities A, B, C and D respectively to actors A1, A4, A1 and A5 ;
3. the Activity B was assigned to the actor A4 by a call for tender;
4. Each actor should win an auction to execute an Activity.

Figure 6. Semantic ambiguities in Workflow log (many interpretations)

We think that Multi-Agent Systems (MASs) approach can help for mining interactions among actors and notably the organizational structures and interactions protocols thanks to the following high level properties:

Natural abstractions to deal with cooperation. A lot of sophisticated protocols like Contract-Net Protocols, Auctions and Negotiation mechanisms are available and could be used with benefits to coordinate processes. Agent technology also provides organizational concepts to abstract and structure a system as a computational society made of groups, roles and interaction.

Pro-active and reactive attitudes of agents ease the control and the enactment of business processes, reactions to events and consequently the synchronization of its activities. Being able to exhibit goal-directed behaviour, they can for example take the initiative to select and engage cooperation with others actors.

Social abilities of agents also ease the cooperation needed for enacting complex workflows and providing abstraction to high-level concepts like commitments, reputations and so on.

In our work, we concentrate on the social abilities and notably on the three following multi-agent concepts:

- FIPA-ACL (Foundations of Intelligent Physical Agents-Agent Communication Language, http://www.fipa.org) performatives that define clearly the semantic of messages and namely the agent's intentions (delegate, subcontract, negotiate…);
- Interaction protocols that enable to identify some dynamic coordination structure where each actor plays a role obeying to precise rules (call-for-proposals, auctions, and so on);
- Organizational structures (hierarchy, federation, team etc.) that model the behaviour of the actors group. I.e. they describe the macro-level dimension of the coordination among actors in terms of externally observable behaviour, independent of the internal features of each participating component.

To the best of our knowledge, the MASs approach has been widely used to study and implement business processes but has never been used in workflow mining. The interested reader can find more information about the contribution of MASs in Workflow Mining in [13].

Our future objective consists in enrichment of the workflow log by integrating key concepts like FIPA-ACL performatives, roles and organizational units in order to facilitate besides the process and informational models the organizational model (i.e. the organizational structures and the interaction protocols). Then, we plan to develop our Workflow mining tool which meets the previous criteria that we have defined.

REFERENCES

[1] Workflow Management Coalition.: The Workflow Reference Model. Technical Report WFMC-TC-1003, November, 1994.

[2] W.M.P. van der Aalst, H.A. Reijers, and M. Song. Discovering Social Networks from Event Logs. Computer Supported Cooperative work, 14(6):549-593, 2005.

[3] W.M.P. van der Aalst.: Process mining in cscw systems. In Proceedings of the 9th IEEE International Conference on Computer Supported cooperative work in design (CSCWD 2005), pages 1_8, London, UK, Coventry University/IEEE Computer Society Press, 2005.

[4] B. van Dongen, A.K. Alves de Medeiros, H.M.W. Verbeek, A.J.M.M. Weijters, and W.M.P. van der Aalst.: "The ProM framework: A New Era in Process Mining Tool Support". In G. Ciardo and P. Darondeau, editors, Application and Theory of Petri Nets 2005, volume 3536 of Lecture Notes in Computer Science, pages 444-454. Springer-Verlag, Berlin, 2005.

[5]. J.E. Cook and A.L Wolf.: Discovering models of software processes models from Event-based data. ACM transactions on software Engineering and Methodology , 7(3):215-249, 1998.

[6] J. Herbsta, D. Karagiannis, Workflow mining with InWoLvE, Computers in Industry. Special Issue: Process/Workflow Mining 53(3) (2003) 245264.

[7] K. Baïna, W. Gaaloul, R. El Khattabi, A. Mouhou.: "WorkflowMiner: a New Workflow Patterns and Performance Analysis tool". CAiSE Forum, 2006.

[8] W.M.P. van der Aalst, BF. Van Dongen, BF, J. Herbst, L. Maruster, G. Schimm, AJMM, Weijters.: Workflow mining: A survey of issues and approaches. In Data &Knowledge Engineering, volume 47(2), pages 237 - 267, 2003.

[9] C. Hanachi, C. Sibertin-blanc.: "protocol Moderators as Active Middles-Agent in Multi-agent System". In International Journal of Autonomous Agents and Multiagent Systems (JAAMAS'04), volume 8, n 2, March, 2004.

[10] L. Bouzguenda, D. Bousabeh, M. Abdelkafi.: "A comparative study of existing workflow mining systems. Technical report01-11/Miracl Laboratory, 2010.

[11] W.M.P. van der Aalst, L. Aldred, M. Dumas, A.H.M. ter Hofstede.: "*Design and Implementation of the YAWL system*". In the 16th International Conference on Advanced Information Systems Engineering, (CAiSE 04), LNCS 3084, Riga-Latvia, 7-11 June, 2004.

[12] H.M.W. Verbeek, A. Hirnschall, W.M.P. van der Aalst.: "XRL/Flower: Supporting Inter-Organizational Workflows Using XML/Petri-net Technology. WES in conjunction with CAiSE 2002, May 27-28, 2002, Toronto, Ontario, Canada.

[13] C. Hanachi, I. Khaloul. : "Découvertes de protocoles et de structures organisationnelles dans le workflow". In international conference NOuvelles TEchnologie de la Repartition, NOTERE, Lyon, 23-27 June, 2008.