A CONCEPTUAL APPROACH TO DISCOVERING KNOWLEDGE FROM CONSTRUCTION WORKFLOWS

*A. J. Antony Chettupuzha[1], Dr Carl T. Haas[1], Joel Gray[2]

[1]*University of Waterloo*

200 University AvenueWest,

Waterloo, Ontario, Canada N2L3G1

(*Corresponding author: ajantony@uwaterloo.ca)

[2]*Coreworx Inc.,*

22 Frederick Street, Suite 800

Kitchener, Ontario, Canada N2H6M6

ABSTRACT

Electronic Product and Process Management [EPPM] systems provide the capability to establish and map information flow between different parties in a construction project as well as model project processes. Much of the information relating to processes and their structure, the actors that operate them, and the data associated with each instance of a process is encapsulated within workflows. The wealth of information contained in an EPPM system can be exploited to extract knowledge that can provide significant cost benefits to construction companies. Workflows provide an ideal medium for the capture of knowledge over the course of a project lifecycle. We demonstrate how this knowledge may then be used to analyze workflows and determine metrics that establish the suitability of workflows to construction processes. We present a conceptual approach to measure the performance of a workflow, its constituent components and evolution, which may be used to lay the foundation for establishing paragon workflows suited to classes of construction activities.

KEYWORDS

Construction workflows, knowledge discovery, suitability measures

INTRODUCTION

Electronic Product and Process Management [EPPM] systems permit the mapping of information flow, the modeling and implementation of project processes and record important participatory information between different project parties. Recognizing the utility of using these structured document management systems which incorporate process oriented workflows, several large corporations have begun implementing them for large scale global projects.

The wealth of information contained in an EPPM system can be exploited to extract knowledge that can provide significant benefits to construction companies. Much of the information relating to processes and their structure, the actors that operate them, and the data associated with each instance of a process is encapsulated within workflows. Workflows that are ideally matched to construction processes provide several advantages such as increased visibility of processes and resources, enforced compliance, and quicker execution over processes that are not automated.

However if a workflow is not matched suitably to a construction activity, there are several risks such as processes being executed out of sequence, inadequate detection of single points of failure and resource overload that may occur. It is therefore important to establish criteria that may evaluate the suitability of a workflow to a construction process. Workflows provide an ideal medium for the capture of knowledge over the course of a project lifecycle. This knowledge may then be used to analyze the performance of workflows and develop metrics that establish the suitability of a workflow template to a class of construction processes.

BACKGROUND

The Workflow Management Coalition [WfMC] Specification, which was founded in 1993 to establish workflow standards based on processes, defines a workflow as "the computerized facilitation or automation of a business process, in whole or part" and further note that workflows are "concerned with the automation of procedures where documents, information or tasks are passed between participants

according to a defined set of rules to achieve, or contribute to, an overall business goal." As such, workflows are essentially a product of process oriented modeling of business activities.

A workflow management system (WFMS), as defined by the WfMC, is "a system that completely defines, manages and executes 'workflows'through the execution of software whose order of execution is driven by a computer representation of the workflow logic." An EPPM system imbibes core functionalities of a workflow management system in addition to various other tools for analysis and document management.

The nature of a workflow management system makes it an excellent facilitator of knowledge management. A workflow system explicitly contains information about whom in the project should be working on what process operation. Hence they implicitly contain information about whom has the required knowledge over a particular domain and when during a project that person's know-how is required. This facility, as recognized by Stohr and Zhao (Stohr and Zhao, 2001), make workflow management systems "*a repository of valuable process knowledge*." Workflows enable the achievement of a core functionality of a knowledge management system; the delivery of pertinent information to the right person at the right time. The capture of workflow history can also be used as a training tool for new employees (Zhao, J.L., 1998) or participants working at a different location for a construction project on a similar module.

Aalst et al expound upon the potential of workflow mining for continuous improvement from the analysis of workflow transactions (Aalst, Dongen, Herbst et al, 2003) Data mining techniques may be applied to a database containing the transactions of workflow instances to assist with analysis of workflow functioning. Leymann and Roller (Leymann and Roller, 2000) also observe that data mining permits meticulous analysis of processes which could lead to the recognition factors inducing bottlenecks besides aiding continuous improvement. Thus contemporary literature is very clear on the plethora of knowledge available within workflows and there are several leading research initiatives that seek to capture this knowledge based both on manual feedback and computer aided analysis of data.

The design of workflows is a complex task that usually draws on an expert's experience (Aalst, Dongen, Herbst et al, 2003). Often construction workflows are designed based on project requirements and resources, but not checked for soundness or subjected to performance analysis. As several different workflows are designed and employed for essentially the same process by different construction companies, it would be interesting to note which of these workflows most improves the efficiency, reliability and effectiveness of that process and therefore is beneficial to any organization which is interested in employing that process. Also, during the course of the project lifecycle, workflows may evolve as employees deviate from strict guidelines, to either the benefit or detriment of a workflow. It therefore becomes important to determine performance metrics that link the suitability of a workflow to the task at hand.

Performance evaluation comprises measuring and modeling (Marsan, Bobbio and Donatelli 1998). Measurements require the formulation of metrics and benchmarks to be used for useful comparisons to be carried out on live systems or processes. The performance of a workflow depends on its constituent elements: (i) The control flow and behavior of the workflow, (ii) The actors or processors who execute the tasks and (iii) The resources and documents that make up the flow of information in a workflow (Aalst 1998). Therefore, if one is to analyze the workflow histories of specific workflow

instances and mine their associated logs, it is possible to formulate metrics and benchmarks which will enable measurements on the suitability of workflow templates to construction processes.

KNOWLEDGE DISCOVERY FROM CONSTRUCTION WORKFLOWS

We identify two methods for the discovery of metrics that demonstrate the effectiveness of workflows. The first method draws on feedback from the wealth of experience accumulated by expert users and designers of workflows, while the second method utilizes knowledge discovery techniques. As such the first method is expected to produce important qualitative metrics. However, it should be possible to substantiate the validity of the metrics determined from the expert feedback by correlation with the results of data-mining of the subsection of selected database fields based on the expert feedback.

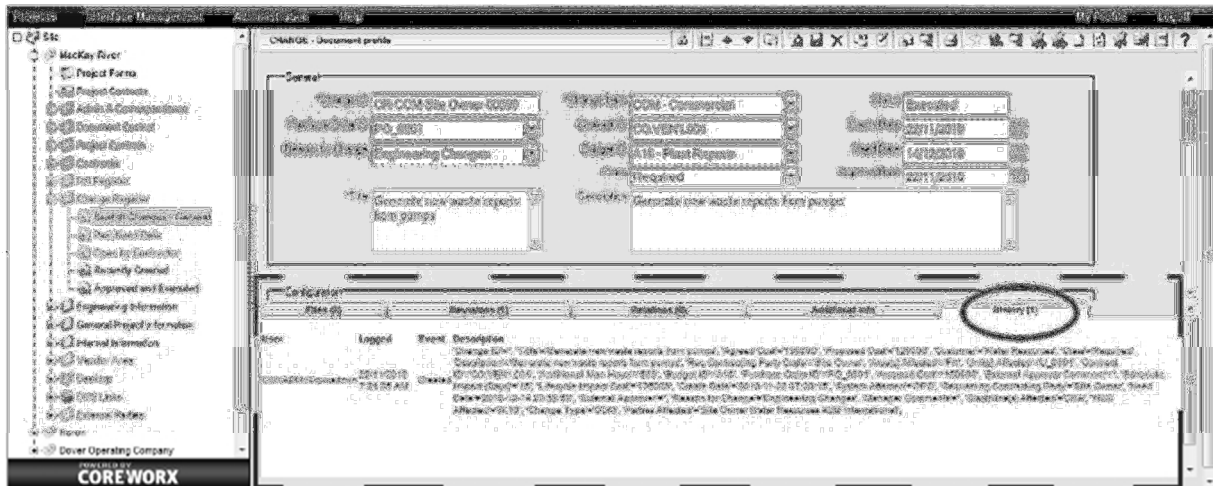### Determination of Metrics from Expert Feedback

To assist with the development of metrics from expert feedback, interviews and workshops were conducted with personnel closely involved in the creation of a workflow. Experienced EPPM system consultants who develop workflows for new projects were interviewed to assist with understanding the key components related to how a workflow is designed from the vendor's perspective. These interviews focused on important empirical design considerations based upon years of experience developing workflows for similar projects.

On the client side, most experts take part in the workflow design at the beginning of a project. They usually focus on modifying templates recommended to them by the consultants so as to imbibe the construction company's culture on the design of a workflow and the core principles behind these cultural considerations. As such these interviews focused on what factors affected existing workflows, why modifications were made, what could be done to improve the performance of a workflow and what factors were considered important in the effective functioning of a workflow.

Both sets of experts agreed that the performance of a workflow could not just be measured in terms of the time it took to initiate and complete a task. Additional factors that were adjudged to be crucial to functioning of an effective workflow included how (i) they ensured levels of compliance, (ii) the flexibility or adaptability of a workflow to accommodate minor changes during an activity or set of activities, (iii) enhanced visibility and auditability of work tasks and processes, and (iv) improved execution times due to the automation of tasks.

Determination of Metrics from Data Mining

It is important to identify tables and fields within an EPPM system's database that permit the extraction of data that corresponds to the factors identified by experts. This is a challenge as qualitative concepts such as compliance, visibility and flexibility of workflows are not factors that can be directly identified from existing EPPM system databases. In such cases where there is no specific field that corresponds to a metric, careful selection of a set of fields that must then be modeled to suitably reflect the metric is required. Traditional EPPM systems do contain explicit information about the execution times of workflow instances and changes are tracked in audit reports generated for every workflow instance as shown in Figure 1.

Figure 1 - An example of how changes are tracked via audit reports in traditional EPPM systems

The extraction, transformation and loading (ETL) of data that is required for analysis is usually the most time consuming task of data analysis. After pertinent data has been identified and extracted from the EPPM system, it must be pre-processed. This includes cleaning the data, i.e. the removal of noisy data and outliers, normalization and scaling of data to fit acceptable data ranges uniformly. Dimensionality reduction and feature selection algorithms may be employed to obtain subsets of data that reduce computation time.

The absence of class labels or ground truth data has resulted in employing pattern recognition or unsupervised learning algorithms, rather than machine learning or supervised learning techniques such as classification algorithms. Once the data has suitably been suitably transformed, it may be loaded as data sets for analysis. Hierarchical and centroid based clustering algorithms shall be employed to identify clusters of workflow instances. These clusters shall then be analyzed to determine the distributions of identified data that were modeled around the identified metrics.

Validating the Suitability of Performance Measures

We define a workflow template as the workflow structure that is modeled around a construction process. We further define a workflow implementation as the derivation of a modified model that is selected by a company or different companies to better reflect the policies that influence the way the particular activity is required to be executed at that particular company. For example a workflow template

may be designed for a construction process such as a Request for Information (RFI). However different companies may require different levels of verification depending on whether factors such as speed or thorough verification of particular processes are assessed to be of higher importance. Further, workflow templates are modified over the course of a project as the needs of an activity are better understood and change. Hence these derived new models or implementations may differ from the original template that was proposed. We define a workflow instance as a specific executed process based on the workflow implementation model.
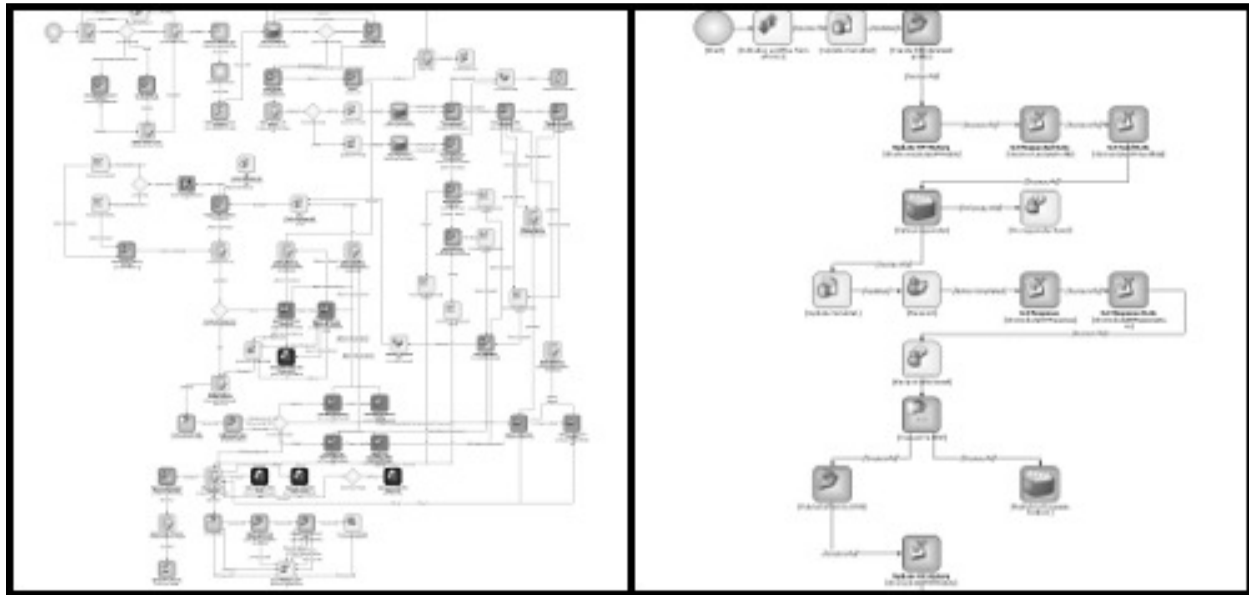

Figure 2 - Two different workflow implementations of the same construction process, Request for Information, by different construction firms

Subsets of workflow instances that are representative of a particular cluster will be analyzed to determine how closely they resemble specific workflow implementations. Where a cluster is representative of a particular workflow implementation, say when a threshold percentage of the instances is reached, it may be associated with that specific implementation. When sets of such implementation clusters have been derived, they may be presented to the company in order to ascertain which of these implementations were preferred. In addition, if certain implementations represent later revised versions of an implementation that were used for a high percentage of executions or instances, these may be assumed to be workflow implementations that were better suited to that particular construction process. The associated metrics can then be verified to have higher associative values for that class of workflow implementations, and may be used to measure the performance of successive workflow templates for specific construction activities.

SUMMARY

The wealth of information contained in an EPPM system can be exploited to extract knowledge that can provide significant cost benefits to construction companies. Workflows provide an ideal medium for the capture of knowledge over the course of a project lifecycle. It was demonstrated how this knowledge may then be used to analyze workflows and determine metrics that establish the suitability of workflows to construction processes. A conceptual approach to measure the performance of a workflow, its constituent components and evolution, was presented which drew perspectives of experts from both the vendor and client users of EPPM systems. By using cluster analysis, different implementations of workflows can be delineated and if there is a corresponding association to varying workflow implementations, it is possible to validate the metrics proposed by the experts to measure the suitability of workflows to their corresponding construction activity. Hence based on the discovery of knowledge related to constituents of the workflow, it is possible to establish benchmarks that can be used to determine workflows that are ideally suited to classes of construction activities.

REFERENCES

Leymann F, Roller D., (2000) "Production Workflow: Concepts and Techniques" Prentice Hall

Marsan M.A, Bobbio A., Donatelli S., "Petri Nets in Performance Analysis: An Introduction", Lecture Notes in Computer Science, 1998, Volume 1491/1998, 211-256

Stohr, E. A., Zhao, J. L., (2001) "Workflow Automation: Overview and Research Issues", Information Systems Frontiers 3:3, pp 281-296

Van der Aalst, WMP, "The Application of Petri Nets to Workflow Management", Journal of Circuit Systems and Computers, 1998, Vol 8, pp. 21-66

van der Aalst, W.M.P., van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G., Weijters, A.J.M.M, (2003), "Workflow Mining: A Survey of Issues and Approaches", Data and Knowledge Engineering 47(2), 237–267

WFMC [Hollingsworth, D.,] (1995) "The Workflow Management Coalition Specification", (WFMC – TC00 – 1003), Issue 1.1

Zhao, J.L., "Knowledge Management and Organizational Learning in Workflow Systems", Proceedings of AIS 1998