PROCESS DISCOVERY DRIVEN PROCESS OPTIMIZATION: A CASE STUDY TO THE RECRUITMENT OF STATE PERSONNEL IN CAMEROON

Gaëlle Laetitia Mouafo Mapikou
University of Yaoundé I
Cameroon
laetitia.mouafo@uy1.uninet.cm

Atsa Etoundi Roger
University of Yaoundé I
Cameroon
roger.atsa@uy1.uninet.cm

ABSTRACT
Process Optimization is a branch of business process management whose purpose is to reduce or eliminate time and resource wastage, unnecessary costs, bottlenecks, and mistakes while achieving the process objective. In the last decades, many practical contributions such as Six Sigma, Lean management, Lean Six Sigma and others have been proposed for optimization, and the associated tools have been widely used and appreciated within organizations. However, the main difficulty with these contributions is that it isn't clearly shown the scientific method used to understand or identify how exactly the actual processes are conducted before optimizing them. If the optimization process is carried out without really understanding how processes are actually going on, then the most likely outcome would be to continue executing outdated and incorrect processes. This paper try to address this difficulty by presenting a methodology for process optimization based on process discovery. Process discovery is a method concerned with extracting descriptive business processes from event logs within organization. The proposed methodology proceeds by comparing the process obtained after applying process discovery technique, with the process supposed to be executed. The result of this comparison combined with the way it had an impact on the business goals and customer's requirements are used to efficiently identify possible parts processes to be optimized. The utility of this methodology is demonstrated through its application to the Cameroonian state personal recruitment process. The descriptive process obtained showed that there were some unexpected tasks used in practice, and only 40.23% of files followed expected paths while executing the process. This analysis result has been used to efficiently define the control flow of the optimized absorption process.

KEYWORDS
Business Process, Process Discovery, Process Optimization, Event Log, Normative Process, Descriptive Process

1. INTRODUCTION
In the last decades, some advanced countries have enjoyed higher level of successes with the adoption and use of Information and Communication Technologies (ICT) while developing countries had reacted so slowly towards this adoption. Many developing organizations, even when provided with full access to information still prefer not to adopt the existing systems to provide modern solutions (Arreymbi et al., 2008). In the same order of ideas, it has even been stated that despite the great promise of technologies, the challenge of effectively deploying technologies for benefit learners, educators, and administrators is still great. Despite all the efforts being made to deal with this resistance to change, there is a high failure rate in Business management. For instance, if we consider the state personal recruitment process in Cameroon public services; beneficiaries of this process are really unsatisfied because of its slowness and the difficulty they have in following their files; Investigations on this situation (Etoundi et al., 2016) highlighted three reasons: (a)-corruption: agents in charge of the process are very corrupted and this has a real influence on the follow-up of the files insofar as it conditions the progress of the file. (b)-The lack of control over the process to follow: both

customers and agents in charge of the files processing are not aware of the real process to follow; and (c)-the ignorance of the file constitution. This generally leads to a bad quality of service, a considerable wastage of time, and a high cost; hence the importance to optimize processes in a way that results in cost reduction and awareness about what each process represents in the business.

Business Process Optimization is a branch of business process management whose purpose is to reduce or eliminate time and resource wastage, unnecessary costs, bottlenecks, and mistakes while achieving the process objective. Process Optimization focuses on the whole process, starting from product conceptual stage to final product design and provides the opportunity to rethink the process by reducing the number of resources involved in the process, or reducing the number of activities to be carried out with the help of advanced Information Technology (IT). Existing tools for process optimization globally proceed in three steps that include identifying a need for optimization within the organization, rethinking processes and automating new solutions. The main difficulty with these tools is that it isn't clearly shown the scientific method used to understand or identify how exactly the actual processes are conducted (Andersson et al., 2006; Salal et al., 2010; Chugani et al., 2016). If the optimization process is carried out without really understanding how processes are actually going on, then the most likely outcome would be to continue executing outdated and incorrect processes. It is therefore important to be aware of what are the real processes executed within the organization in order to better improve them. This weakness can be solved by using business process discovery (van der Aalst et al., 2004).

Business process discovery or simply process discovery is a discipline of Business Process Management which provides scientific methods to extract the implemented process, validate it with the process owners, and create a baseline that gives a nearly exact view of the really executed business processes. It also helps to identify the process hierarchy, process owners, process entities, business rules and process operations. Business Process Discovery can be done using two approaches: (a)-Manual - Process Analysts use interviewing techniques to derive the processes, (b)-Automated - Process Analysts use automated business process discovery tools that can extract the process from databases or logs available with IT.

The aim of this paper is to present a methodology for business process optimization that uses process discovery to better identify possible parts of processes to be improved. This is done by comparing the process obtained after applying process discovery technique, with the process supposed to be executed. The result of this comparison combined with the way it has an impact on the business goals and customer's requirements are used to efficiently identify possible parts of processes to be improved. The remainder of this paper is organized as follows: section 2 introduces process optimization, its objectives and its general methodology, section 3 presents some background concepts necessary to understand the proposed methodology, section 4 presents the process discovery technique and explains a main existing algorithm for this technique called the alpha-algorithm, section 5 presents the proposed methodology for process optimization, section 6 discusses the application of the proposed methodology on the state personal recruitment process, and finally section 7 concludes the paper and provides directions for future research.

2. Process Optimization

Organizations are faced every day with the **challenge to produce more, and spending less.** To grow, managers look for a way to improve its processes in a way that results in cost reduction and awareness about what each process represents in the business. What process optimization can bring to a company is a way to reduce money, time and resources spent in a process through a critical look at the operations of the organization, leading to better business results.

Optimizing a business process requires more effort than simply obtaining new software. This can be done from a variety of approaches. The most five commonly used methods are presented below.

Business Process Re-engineering (BPR): is known by many names, such as ‘core process redesign’, ‘new industrial engineering’ or ‘working smarter’. All of them imply the same concept which focuses on integrating both business process redesign and deploying IT to support the re-engineering work. Re-engineering business processes means tossing aside existing processes and starting over. In (Hammer & Champy, 2009), business process redesign (re-engineering) is defined as “the *fundamental* rethinking and *radical* redesign of *business processes* to achieve *dramatic* improvements in critical contemporary measures of performance such as costs, quality and speed”. This definition contains four key words: (1)-*Fundamental*: reevaluate the primary goals of the company, ignoring rules and assumptions formulated in the past; (2)-*Radical*: do not try to improve the existing situation, invent completely new ways of accomplishing work; (3)-*Dramatic*: do not use business process redesign to obtain marginal improvements, aim at 'order of magnitude' improvements; (4)-*Process*: focus on the business processes instead of organizational structures.

BPR has not been much appreciated in practice because of its high failure rate. Investigations on this failure allowed understanding and identifying some inconsistencies. First, the idea of cleaning the slate and throw everything away is considered as a fallacy because, according to (Davenport, 1993; Boudreau & Robey, 1996), the most radically redesigned business processes need to be implemented in real organizations that have histories and memories and existing constraints have to be taken into account during implementation of BPR plans. Second, the paradoxical role of IT, enabler and disabler of radical change, because revising the basic business processes using IT implies a new structure that may become even more difficult to change in the future. Third, the hypocrisy of empowerment which states that, behavior, values and attitudes of workers at all levels of the organization must be changed to hope success in a BPR project. Finally, the irony of employee commitment to re-engineering efforts; for (Boudreau & Robey, 1996), the commitment and positive attitude of most of the individuals in an organization towards BPR seems to be the most important condition for project success.

Total Quality Management (TQM): is a comprehensive and structured approach to organizational management that seeks to improve the quality of products and services through ongoing refinements in response to continuous feedback. TQM requirements may be defined separately for a particular organization and can be applied to any type of organization; it originated in the manufacturing sector and has since been for the quality management from the customer’s point of view. Even if TQM is a widely used improvement method, there is still not an unequivocal definition and understanding of its functioning (Andersson et al., 2006).

Lean Management: is a production practice focused on the idea of ‘removing all form of waste’ (Shah &Ward, 2003; 2007). The core principle behind lean is that customers do not pay for mistakes or waste but value. As such, companies need to increase the value of their products or services in order to maximize profit. Lean management offers an opportunity to drive up value and promote continuous improvement. An example is lean manufacturing with Kaizen (Brunet & New, 2003) that considers the expenditure of resources for any goal other than the creation of value for the end customer to be wasteful, and thus a target for elimination. Working from the perspective of the customer who consumes a product or service, "value" is defined as any action or process that a customer would be willing to pay for. This method has also been used in others fields like software improvement (Petersen & Wohlin, 2010); however its implementation remains difficult (Martínez-Jurado & Moyano-Fuentes, 2014; Jahdav et al., 2014).

Six Sigma: is a methodology for improvement whose purpose is to achieve savings by reducing the causes of defects and variability in manufacturing and business processes (Andersson et al., 2006). Six Sigma is a statistical method within the area of quality management using the approach Define- Measure-Analyze-Improve-Monitor. It is frequently used in manufacturing processes to increase the quality level. The improvement strategy focused on removing variability from a process. This method is used primarily by large manufacturing companies like manufacturing or automotive industry in order to become even more cost effective. Although originally developed for manufacturing processes, the Six Sigma methodology has been successfully applied to a wide range of processes. As a tool for process improvement and reduction of defects, Six Sigma compliments Lean (Zu et al., 2008; Salah et al., 2010) and is a component of many Lean programs.

Lean Six Sigma (LSS): is a combination of Lean management and Six Sigma (Salah et al., 2010; Snee, 2010; Pepper et al., 2010) that focuses on the elimination of waste and variation, following the Define- Measure-Analyze-Improve-Control structure, to achieve customer satisfaction with regards to quality, delivery and cost. The adoption of LSS in organizations rapidly evolved; Recent studies (Chugani et al., 2016) show that it tends to produce more success than using Lean or Six alone.

3. BACKGROUND

This section presents some background concepts necessary to understand the methodology proposed in this paper. These concepts include event logs and process modeling techniques such as BPMN and Petri-nets.

3.1. Event Logs

An event log is basically a table. It contains all recorded events that relate to executed business activities. Each event is mapped to a *case*. A process model is an abstraction of the real world execution of a business process. A single execution of a business process is called a *process instance*. They are represented in the event log as a set of events that are mapped to the same case. The sequence of recorded events in a case is called a *trace*. The model that describes the execution of a single process instance is called a *process instance model*. A process model abstracts from the single behavior of process instances and provides a model that represents the behavior of all instances that belong to the same process. Cases and events are characterized by classifiers and attributes. Classifiers ensure the distinction of cases and events by mapping unique names to each case and event. Attributes store additional information that can be used for analysis purposes. An example of an event log is given in Table 1.

Table 1- An Example of Event Log

Case ID	Event ID	Timestamp	Activity	Resource
1	100	10/10/2016	Order goods	Laetitia
	101	12/10/2016	Receive goods	Mickael
	102	23/10/2016	Receive Invoice	Frank
	103	24/10/2016	Pay Invoice	Armel
2	104	02/11/2016	Order Goods	Laetitia
	105	03/11/2016	Receive Invoice	Alain
	106	10/11/2016	Pay Invoice	Frank

3.2. Process Modeling

It is almost pointless to speak of optimization methods without clearly specifying what process modelling technique is being used. Instead of using the well-known Business Process

Modelling Notation technique (BPMN), process discovery uses petri-nets as the reference modelling technique because of its formal semantic which facilitates the verification of model properties. Nevertheless, it has been developed techniques to convert petri-nets diagrams into BPMN diagrams. Both models are used in this paper, and thus presented in this section.

BPMN 2.0: is a *de-facto standard notation* for modeling business processes understandable by a wide range of people and organizations. BPMN diagrams are composed of activities, events, and gateways, connected through connecting objects. Activities represent the actual work that is carried out in a process; events (start, end and intermediate events) are used to indicate that something occurs in a process; gateways (exclusive, inclusive, and parallel) are used to split and join the control flow within a process model through the sequence flow to which they are connected. It has a formal syntax for specifying the notation of process models, but the semantic is not clear.

Petri-Nets: is the oldest technique (Van Der Aalst et al., 1998) for specifying business processes. Originally, petri-nets consist of places, transitions and directed arcs connecting places and transitions. Connections between two nodes of the same type are not allowed. Places are represented by circles and transitions by rectangles. Once the network structure has been established, tokens move from one place to another adjacent, thereby forming a transition. Since transitions may change the network status, i.e. the number and the position of tokens, they are active components such as events or tasks; tokens represent physical objects or information objects.

Normative Process: the business process supposed to be executed as defined in the manual that describes the organization.

Descriptive Process: the real business process that is actually being executed within the organization.

4. PROCESS DISCOVERY

Process discovery was firstly evolved in the context of analyzing software engineering processes by Cook and Wolf in the late 1990s (Cook & Wolf, 1998). Agrawal, Gunopulos and Leymann (Agrawal et al., 1998) and Herbst and Karagiannis (Herbst & Karagiannis, 1998) introduced process discovery to the context of workflow management. Major contributions to the field have been added during the last decade by van der Aalst and other authors (Van Der Aalst, 2011, Van Der Aalst, 2016) by developing mature mining algorithms and addressing a variety of topic related challenges. This has led to a well-developed set of methods and tools that are available for scientists and practitioners. The purpose of process discovery is the extraction of knowledge from execution traces of business processes within organizations. Its starting point is an event log and, it is assumed that it is possible to sequentially record events such that each event refers to an activity and is related to a particular process instance. Based on this assumption, Van Der Aalst clearly defined the process discovery problem as "*to find an algorithm or a function that maps an event log onto a process model such that this model is representative for the behavior observed in the log*". According to this definition, several algorithms have been proposed in the literature, but we will just present the first and basic one (the alpha-algorithm) in the remainder of this section.

4.1. α-algorithm

This algorithm is an example of the function that maps an event log onto a process model (Workflow-net) such that this model is representative for the behavior observed in the log. It takes an event log as input and scans the event log, looking for particular patterns. For example, if an activity a is followed by an activity b and b is never followed by a, then it is assumed that there is a dependency between a and b. To reflect this, the corresponding petri-net has a place connecting a to b. There are four order relations used to extract data in the log:

(i)-*Direct precedence (>)*: a is in direct precedence with b iff there is a trace in the event log where b directly follows a; (ii)-*Causal dependency (->)*: a is in causal dependency with b iff a is in direct precedence with b and b is not in direct precedence with a; (iii)-*Freeness (#)*: a and b are free iff there is no relation between them ie a is not in direct precedence with b and b is not in direct precedence with a; (iv)-*Parallelism (||)*: a and b are parallel iff a is in direct precedence with b and b is in direct precedence with a.

The α-algorithm uses this four order relations to discover a workflow-net from an event log. The discover process embodies eight steps: (1)-identify the set of transitions T; (2)-identify initial transitions; (3)-identify ending transitions; (4)-build the set of places; (5)-deduct the set of maximal places from the set of places; (6)-build the final set of places P by adding the source and sink places to the set of maximal places; (7)- build the set of directed arcs F; (8)- return the workflow-net N = (P,T,F).

5. Proposed Methodology

The methodology for Process Optimization incorporates the well-known BPR methodology presented before, with six principal phases but the difference comes from the set of actions/steps embodied in the second and third phases. Table 2 summarizes this new methodology with its associated steps.

Table 2: Summary of Process Discovery Driven Process Optimization Methodology

1. ENVISION Identify a need for process improvement Gain management sponsorship Establish change leadership roles Introduce automated business processes management technology Build, Educate, and Train a Change Team Identify Business Cycles Create an Organizational Vision Analyze the Project Environment Implement a Change Management Program **2. INITIATE** Catalog Business Products Identify Business Processes Select a Process for Implementation and Improvement Construct a Work Breakdown Structure Define Task Components Specify Performance Measures Install and Test Required Infrastructure Implement and Monitor Automated business processes Record execution traces in event logs	**3. DIAGNOSE** Collect the event logs of automated business processes Discover business processes from these logs Compare automated processes with discovered processes Consider the Customers' Requirements Benchmark Against Industry Leaders Specify Performance Goals **4. MONITOR** Re-engineer the business Construct new business models Simulate each business process alternative Select the Most Efficient and Effective business processes **5. RECONSTRUCT** Identify Organizational Changes Specify the Required Infrastructure Gain Approval Institute Organizational and Infrastructure Changes **6. REDESIGN** Implement New business processes Manage the business process Configuration Perform continuous Improvement

The diagnosis, initiation and redesign phases are the only detailed below because they are the core phases of this methodology. More details on the other phases are given in (Dumas et al., 2013).

Initiate: the business products of the business cycle under review are identified. The component processes of the cycle are distinguished, and a single process is selected for improvement this process is modeled and implemented on the automated workflow management tool. This phase embodies nine steps:

1. *Catalog Business Products*: the change team identifies the products that result from the completion of the chosen business cycle;
2. *Identify Business Processes*: the business cycle is decomposed into its component processes by identifying how each business cycle product is created;
3. *Select a Process for Implementation and Improvement*: the change team selects a single business process for implementation based factors like changeability, performance, and business/customer impact (Harrington, 1991);
4. *Construct a Work Breakdown Structure:* the change team, with the help of the process owner, constructs a work breakdown structure i.e. a graphical depiction of the hierarchical structure of the business cycle decomposed into its component parts;
5. *Define Task Components*: the change team defines work objects, roles, rules, resources, time, and routing of each task;
6. *Specify Performance Measures*: the change team specifies performance indicators (effectiveness, efficiency and adaptability) to be recorded by the workflow tool;
7. *Install and Test Required Infrastructure*: with the help of the IT department, the change team examines the network infrastructure and determines if any additional hardware or software is required to connect all the workflow participants;
8. *Implement and Monitor Automated business processes*: the automated workflow system is implemented and allowed to operate for a certain period of time;
9. *Record execution traces in event logs*: once the automated business process has started operating, the change team needs to collect all execution traces and record it in some files called event logs. This is really important because, in order to the diagnosis phase to be successful, the change team needs to be aware of what is really going on within the organization not what they imagine or to be going on. These event logs will help to discover real processes being executed within the organization at the next phase.

Diagnose: this phase is essential for this work because it explains a more efficient way to identify possibilities for improvement. Once business processes are automated and their execution traces are being recorded in event logs, the change team must use process discovery techniques to extract real processes from these logs. This is very important because it would be unrealistic to intend to optimize business processes if the change team cannot first of all ensure that what is really going on is what was supposed to be. So, this primary work will help them in better detecting deviations, or bottlenecks for possible improvements. Clearly, this phase is divided into six steps summarized in Figure 1:

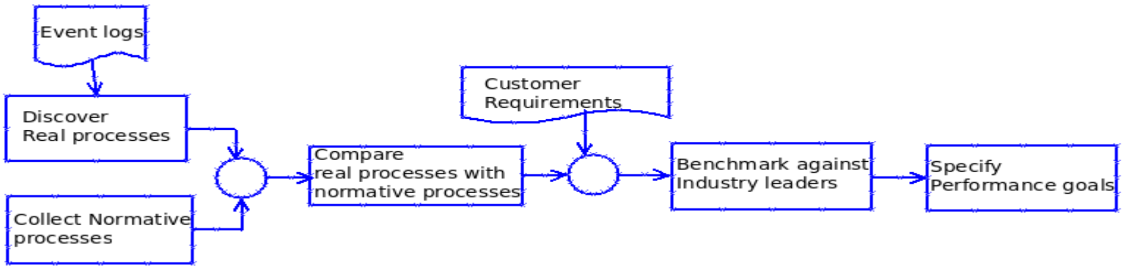

Figure 1: Identifying Process Improvements

1. *Consider the event logs of automated business processes*: the change team must collect all the event logs of the last step of the preceding phase. Each log is associated to one of automated business processes and must contain its execution traces during a certain period and respect the template of event logs used for process discovery. Recall that as each log is associated with one business process, the number of event logs is equal to the number of business processes executed within the organization.
2. *Discover business processes from these logs*: each log of the preceding step is used as input for a process discovery technique like the alpha-algorithm presented before, to extract the associated business process corresponding to the one which has been really executed within the organization. The discovered process will help the change team in understanding if what happened conforms to what was supposed to happen.
3. *Compare automated processes with discovered processes*: having discovered all the real processes and collected all normative process models (processes supposed to be executed), the change team must compare each discovered process with the associated normative process model. At the end of this task, the change team must come out with an analysis report indicating the performance of each executed business process in terms of where did the process deviate and where did it conform with what was supposed to be. This report will help in defining the goals and directives of new processes.
4. *Consider the Customers' Requirements*: the change team must consider the customer's requirements (what they require from the business processes and products)in defining the goals of new processes;
5. *Benchmark against Industry Leaders*: a business must be more innovative than its competitors to maintain its market share. For this reason, the change team must become aware of any new methods and technologies that are employed within its industry;
6. *Specify Performance Goals*: using the data gathered from analysis report, customers, and benchmarking process, and the performance objectives specified in the organization's vision statement, the change team specifies overall process performance goals.

Redesign: during this last phase of the methodology, the new business process is implemented and any changes to the process configuration are documented as it is maintained. After this, a continuous process improvement plan is establish as presented in the life cycle in Figure 2.

1. *Implement New business processes*: new business process models are implemented;
2. *Manage the business process Configuration*: manage the business process configuration to ensure that it remains standardized and unaltered;

3. *Perform continuous improvement*: the performance of the improved business process is continuously monitored and periodically analyzed, as required.

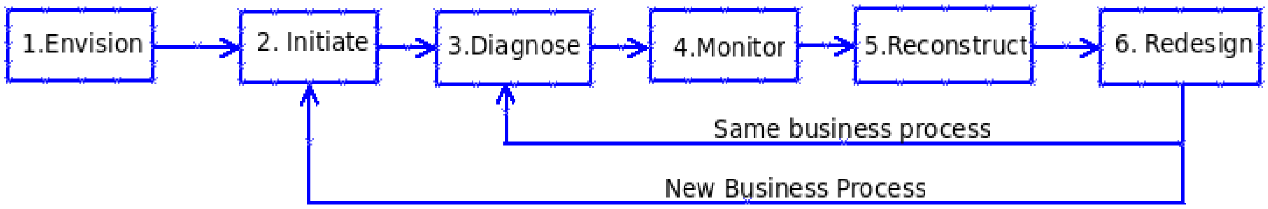

Figure 2: Process Discovery Driven Process Optimization Life Cycle

6. Case Study

A case study that included the application of a part of the proposed methodology was conducted using workflow data gathered from the ministry of public service and administrative reform in Cameroon. Civil servants of this sector are interested in understanding why some applications take longer to be processed than others. So, the application of the proposed methodology focuses on the second and third phases, namely the initiation and diagnosis phases. After collecting execution traces and storing it in event logs, it will be discovered the process that was followed in order to identify what are the causes of the problem and what can be done to improve the situation. The process followed to achieve this objective embodies four steps, (1)-model the normative process taken from the Manual of Administrative Procedures - Human Resources management (MPA-HRM); (2)-collect the trace data for all civil servant absorption files for the period of January 2011 to December 2016 and use it to discover the descriptive process of civil servants absorption;(3)-compare the descriptive and normative processes in order to detect errors, misbehaviors and deviations and; (4)- propose a possible optimized process.

6.1. Normative Process

This sub-section presents the BPMN model of the state personal recruitment procedure taken from the Manual of Administrative Procedures- Human Resources Management (MPA-HRM)

Given that the number of tasks (activities) performed in this process is very large, a management rule has been defined to facilitate and streamline the modelling. This rule states that *"at one step of the process, the processing performed by an organizational unit (resource) is atomic"*. In this way, all tasks performed by a unit at a given step are considered as a single task, and an activity now represents all the processing performed by a unit at this step. The normative absorption process contains 18 activities and five paths detailed in appendix. So, if T_N is the set of tasks in the normative process and C_N its set of paths, then:

$T_N = \{T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, T12, T13, T14, T15, T16, T17, T18\}$ and

$$C_N = \{C1, C2, C3, C4, C5\}$$

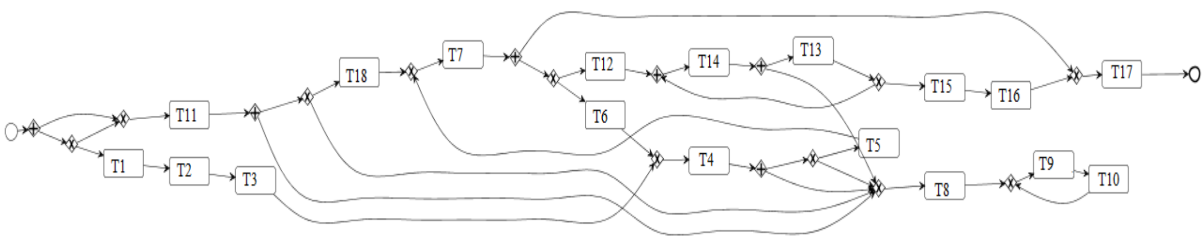

Figure 3: Normative Process

Figure 3 shows this process modeled with petri nets. The modeled process contains five paths (a path contains the set of tasks on a road going from the initial place T1 to the final place T17 or T20, and corresponds to an execution instance).

6.2. Descriptive Process

A statistical analysis was done on the data collected and it emerges that during the last five years there were 1496 applications registered. Among these, 527 were completed and 969 are still being processed. Since the descriptive process represents what really happened every time the process has been completely executed within the organization, only completed requests are considered. Thus, the associated discovered process will give the exactly process that was followed by completed files. The event log containing execution traces of the 527 completed files is used as input of the alpha algorithm presented section 3 to discover the descriptive process presented in figure 5. This process contains 33 paths and 19 activities detailed in appendix.

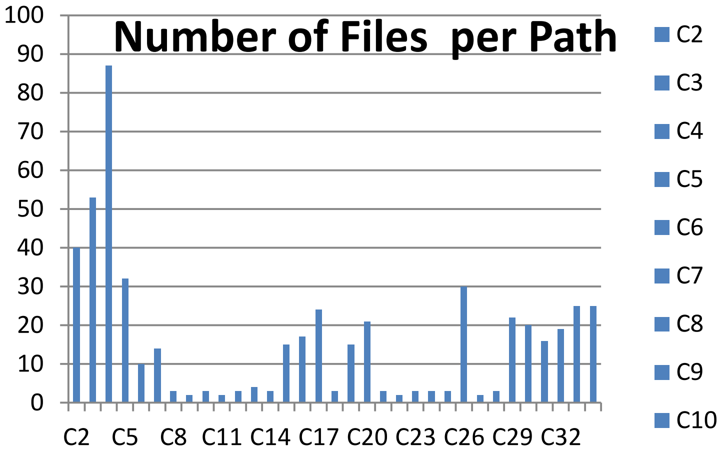

Figure 4: Statistics on Number of Files per Path

Let T_D and C_D respectively the set of tasks and paths in the descriptive process:

T_D = {T20, T11,T18, T7, T12,T14, T8,T17,T21,T9,T15,T16,T4,T6,T22,T13,T19,T15,T23}.

C_D={C6,C7,C8,C9,C10,C11,C12,C13,C3,C14,C15,C16,C4,C17,C18,C19,C20,C21,C22,C23 ,C24,C25, C2, C5, C26, C27,C28, C29,C30,C31,C32,C33}

Figure 4 presents statistics per path in terms of number of files and processing time. From this evaluation, it can be concluded that the most followed path in practice is C4 with 87 files. Only C2, C3, C4 and C5 are expected paths, which represent a total number of 212 files for a percentage of 40.23%.

6.3. Normative vs. Descriptive Processes

The previous statistics showed that, among the 527 successful applications for absorption filed, only 212 followed expected paths, i.e. a percentage of 40,33%, which foretells a problem in the business management. To identify the potential causes of this problem, the method used in this paper is a model-based comparison. Since the petri-net discovered by the alpha-algorithm only shows the control flow perspective (the flow between tasks) of the process, the model-based comparison can only focus on two criteria: tasks and paths. However, many other perspectives like resource, data and time perspectives can be discovered using advanced process discovery techniques (Van Der Aalst, 2016).

The comparison between both models allows to make the following observations: let $T_{N\backslash D}$ the set of tasks in the normative process that aren't in the descriptive process and $T_{D\backslash N}$ the set of tasks of the descriptive process that are not in the normative process:

- $T_N \cap T_D$ = { T4,T6,T7,T8,T9,T11,T12,T13,T14,T15,T16,T17,T18,T20}: set of tasks appearing in both processes; these tasks represent potential activities for the future re-engineered process ;
- $C_N \cap C_D$ = {C2,C3,C4,C5}: set of activities appearing in both processes i.e. expected paths which must necessarily appear in the re-engineered process;
- $T_{N\backslash D}$ = {T1,T2,T3,T5}: represents expected activities which are never used in practice. This means that these activities are useless for this process, and naturally must not exist in the re-engineered process;
- $T_{D\backslash N}$ = {T19,T21,T22,T23}: represents activities which were not expected but exist in practice. These activities must therefore be considered in the re-engineered process.
- $C_{N\backslash D}$ = {C1}: path expected by the normative process but never followed in practice. It must therefore not be considered in the re-engineered process because it is useless;
- $C_{D\backslash N}$={C6,C7,C8,C9,C10,C11,C12,C13,C14,C15,C16,C17,C18,C19,C20,C21,C22,C23,C24, C25,C26,C27,C28,C29,C30,C31,C32,C33,C34}: set of paths which were not expected but, sometimes used in practice. These paths represent 59.77% of the 527 files, which is really meaningful.

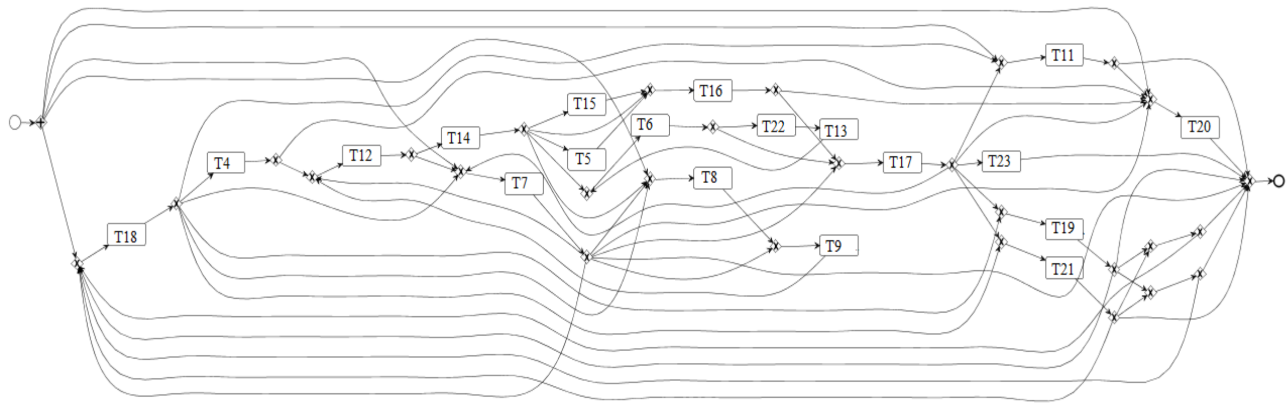

Figure 5: Descriptive Process

6.4. Optimized Process Proposal

Based on the analysis carried out in the previous section, the following observation is made: some tasks defined in the normative process model do not appear in the descriptive process model while some tasks that are not defined in the normative process model have been added by human actors involved in the processing of the files. The resulting instruments are signed by the officials and handled to their beneficiaries. It means that these instruments are valid, and therefore the corresponding new tasks can be added to the new process while non-used tasks presented in the normative process model will be removed in the definition of the new process model. We must note that even if some normative tasks are removed, there are some normative tasks kept in the new process. Thus, from the results obtained, the structure of the optimized process can be defined as follows: let T_R its set of tasks and C_R its set of paths.

- TR = (TN∩TD) ∪ TD\N = {T4,T6,T7,T8,T9,T11,T12,T13,T14,T15,T16,T17,T18,T19, T20,T21,T22,T23}: activities appearing in both processes plus activities which were not expected but which exist in practice; These activities are those that are actually used in practice during the execution of the process. They must therefore be retained.
- CR = (CN∩CD) ∪ CD\N: paths appearing in both processes plus paths which were not expected but which exist in practice. CR represent the sets of real paths which

must normally retained but given the average processing time of a file that is 2.5 months, and the proportion of files processed beyond these deadlines(8.5%), it has been removed all the paths for which the processing time is greater than 3 months. So CR = CR - {C8,C9,C10,C11,C12,C13,C14,C18,C21,C22,C23,C24,C25, C27,C28} = {C2,C3,C4,C5,C6,C7,C15,C16,C17,C19,C20,C26,C29,C30,C31,C32,C33,C34}.

The result of this research has been presented to the ministry of public service and adopted for civil servant file processing in various ministries. The use of the optimized process model allows a better follow-up of the processing of files by staff and beneficiaries. The computerized system for the integrated management of state personnel and the payroll (Etoundi et al., 2016) has been updated to support the result of this work. Based on this enhancement, the follow-up by stakeholders of the processing of file has been made available on line through the following link http://dossier.minfopra.gov.cm/index.php/fr/ which significantly reduced the corruption level in the processing. Indirect results are the reduction of the corruption level as new recruited state personal are no longer going to this ministry to follow-up their files which was conducted by a corruption act. Moreover, personal are more present in their office.

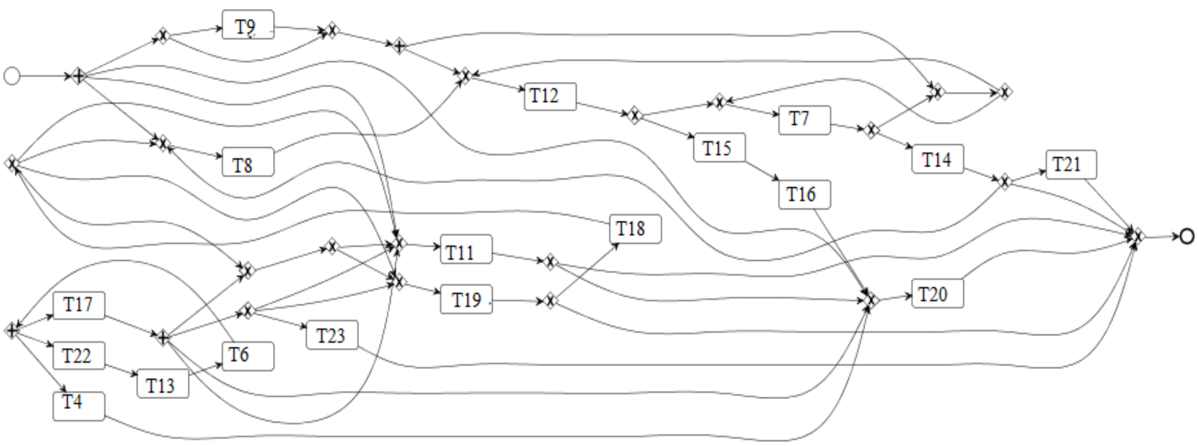

Figure 6: Optimized Process

7. Conclusion

The aim of this paper was to propose a methodology for process optimization within organizations. It was identified that existing methods do not provide scientific methods to identify real processes and to diagnose errors, deviations and bottlenecks in business processes in order to detect the process to be improved. It was mentioned that empirical research showed that this way of was an unreliable and incomplete approach because it didn't point out what exactly happened in the field and led to a high failure rate in many projects. This gap has been filled in the proposed methodology using the process discovery.

The proposed approach improved the existing methodology. Major improvements consisted in recording execution traces in event logs; discovering descriptive processes and establishing a comparison between normative and descriptive processes. The discovered process obtained at the end of the diagnosis stage allowed to understand what really happened while executing processes within the organization and by doing a model-based comparison with was supposed to be, it has been identified precisely where the process deviated in terms of tasks and paths followed. The result of this comparison was used to define what shall be done in terms of control flow in the optimized process.

A case study has been carried out on the absorption process in Cameroun public service ministry for the validation of the proposed approach. Data collected on this process contained 1496 files among which 527 were successful. Data concerning only successful applications have been used at the initiation stage to construct a log conforming to process discovery event logs. The discovered real process obtained at the diagnosis stage showed that there were some unexpected tasks used in practice, and only 40.23% of the 527 successful files followed expected paths while executing the process. This analysis result has been used to efficiently define the structure (set of tasks and paths) of the new process. Due to the lack of an installed workflow tool, the proposed methodology was not fully tested. For future work, an automated process discovery driven process optimization tool should be developed in order to entirely support the proposed approach. Also, the discovery step in the proposed methodology can be extended with advanced techniques to discover other perspectives like the resource and data perspectives. This will considerably improve the optimization task.

8. References

Agrawal, R., Gunopulos, D. & Leymann, F. (1998). Mining Process Models from Workflow Logs. In *Proceeding of 6th International Conference on Extending Database Technology: Advances in Database Technology* (469-483). London: Springer.

Andersson, R., Eriksson, H. & Torstensson, H. (2006). Similarities and Differences between TQM, Six Sigma and Lean. *The TQM Magazine*, 18, 3, 282-296.

Arreymbi, J., Agbor, E.A. & Adnan, A. (2008). Critical Analysis of Why ICT Uptake Is Slow in Emerging Economies - The Case of Cameroon. In *Proceedings of Advances in Computing and Technology*.

Boudreau, M.-C. & Robey, D. (1996). Coping with Contradictions in Business Process Re-Engineering. *Information Technology & People*, 9, 4, 40-57.

Brunet, A.P. & New, S. (2003). Kaizen in Japan: An Empirical Study. *International Journal of Operations & Production Management*, 23, 12, 1426-1446.

Chugani, N., Kumar, V., Garza-Reyes, J.A., Rocha-Lona, L. & Upadhyay, A. (2016). Investigating the Green Impact of Lean, Six Sigma, and Lean Six Sigma: A Systematic Literature Review. *International Journal of Lean Six Sigma*, 8, 1, 7-32.

Cook, J.E. & Wolf, A.L. (1998). Discovering Models of Software Processes from Event-Based Data. *ACM Transactions on Software Engineering and Methodology*, 7, 3, 215-249.

Davenport, T.H. (1993). *Process Innovation: Reengineering Work through Information Technology*. Boston: Harvard Business School Press.

Dumas, M., La Rosa, M., Mendling, J. & Reijers, H.A. (2013). *Fundamentals of Business Process Management*. Berlin: Springer.

Etoundi, R.A., Tobie, A.M. & Onana, B.S.M. (2016). Using ICT in Controlling Staff and Payroll in the Cameroonian Public Service: Lessons Learned and Perspectives. *The Electronic Journal of Information Systems in Developing Countries*, 76, 2, 1-18.

Hammer, M. & Champy, J. (2009). *Reengineering the Corporation: A Manifesto for Business Revolution*. New York: HarperCollins.

Martínez-Jurado, P.J. & Moyano-Fuentes, J. (2014). Lean Management, Supply Chain Management and Sustainability: A Literature Review. *Journal of Cleaner Production*, 85, 1, 134-150.

Pepper, M.P.J. & Spedding, T.A. (2010). The Evolution of Lean Six Sigma. *International Journal of Quality & Reliability Management*, 27, 2, 138-155.

Petersen, K. & Wohlin, C. (2010). Software Process Improvement through the Lean Measurement (SPI-LEAM) Method. *Journal of Systems and Software*, 83, 7, 1275-1287.

Salah, S., Rahim, A. & Carretero, J.A. (2010). The Integration of Six Sigma and Lean Management. *International Journal of Lean Six Sigma*, 1, 3, 249-274.

Shah, R. & Ward, P.T. (2003). Lean Manufacturing: Context, Practice Bundles, and Performance. *Journal of Operations Management*, 21, 2, 129-149.

Shah, R. & Ward, P.T. (2007). Defining and Developing Measures of Lean Production. *Journal of Operations Management*, 25, 4, 785-805.

Snee, R.D. (2010). Lean Six Sigma - Getting Better All the Time. *International Journal of Lean Six Sigma*, 1, 1, 9-29.

Van der Aalst, W.M.P., Weijters, A.J.M.M. & Maruster, L. (2004). Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 16, 9, 1128-1142.

Van der Aalst, W.M.P. (1998). The Application of Petri Nets to Workflow Management. *Journal of Circuits Systems and Computer*, 8, 1, 21-66.

Van der Aalst, W.M.P. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Berlin: Springer.

Van Der Aalst, W.M.P. (2016). *Process Mining: Data Science in Action* (2nd ed.). Berlin: Springer.

Zu, X., Fredendall, L.D. & Douglas, T.J. (2008). The Evolving Theory of Quality Management: The Role of Six Sigma. *Journal of Operations Management*, 26, 5, 630-650.

APPENDIX

LIST OF ABBREVATIONS

ABBREVIATION	DESCRIPTION
T1	tasks performed by the guidance service
72	tasks performed by the link service
T3	tasks performed by the arrival mail service
T4	tasks performed by the directorate of career management
T5	tasks performed by the compliance department
T6	tasks performed by the Directorate of Administrative Affairs
T7	tasks performed by the absorption office
T8	tasks performed by the recovery service
T9	tasks performed by the computer cell
T10	tasks performed by the Directorate of legal affairs
T11	tasks performed by the administrative and financial staff
T12	tasks performed by the general secretariat
T13	tasks performed by the secretariat of the prime minister
T14	tasks performed by the minister's office
T15	tasks performed by the recovery office
T16	tasks performed by the departure mail service
T17	classifying
T18	tasks performed by the sub-directorate for the family protection
T19	tasks by the absorption office of the prime minister
T20	tasks performed by the Absorption Office and Financial Affairs
T21	tasks performed by the social protection service
T22	tasks performed by the directorate of administrative affairs and assistant
T23	tasks performed by the technical administration personal service
MPA-HRM)	Manual of Administrative Procedures - Human Resources Management

SET OF PATHS

NORMATIVE PATHS

(C1):T1-T2-T3-T4-T5-T7-T6-T4-T8-T9-T10-T9-T11-T8-T12-T13-T14-T13-T15-T16-T17;
(C2):T11-T20-T11-T18-T7-T12-T14-T18-T7-T17-T20;
(C3): T11-T18-T7-T12-T14-T8-T7-T17;
(C4): T20-T11-T20-T11-T18-T12-T14-T8-T20-T17;
(C5): T11-T20-T11-T18-T7-T12-T14-T8-T7-T17;

DESCRIPTIVE PATHS

(C6):T20-T11-T18-T11-T20-T11-T18-T7-T12-T14-T8-T7-T17-T20;
(C2):T11-T20-T11-T18-T7-T12-T14-T8-T7-T17-T20;
(C7):T11-T18-T7-T12-T14-T8-T7-T17-T21;
(C8):T11-T18-T8-T7-T8-T12-T14-T15-T14-T16-T7-T17-T20;
(C9):T11-T18-T11-T20-T11-T18-T7-T12-T7-T18-T11-T20-T11-T18-T7-T12-T14-T8-T7-T17-T20; (C10):T11-T18-T11-T20-T11-T18-T7-T12-T14-T8-T7-T17-T20;
(C11):T11-T18-T11-T20-T11-T18-T7-T12-T14-T8-T7-T17;
(C12):T11-T18-T7-T8-T8-T12-T14-T15-T14-T16-T7-T17;
(C3):T11-T18-T7-T12-T14-T8-T7-T17;
(C13):T11-T18-T8-T7-T8-T12-T14-T15-T14-T16;
(C14):T20-T11-T20-T11-T18-T7-T12-T7-T18-T11-T20-T11-T18-T7-T12-T14-T8-T7-T17-T11;
(C15):T18-T7-T12-T14-T8-T7-T17;
(C4):T20-T11-T20-T11-T18-T7-T12-T14-T8-T7-T17;

(C16):T20-T11-T20-T11-T18-T4-T12-T14-T6-T22-T13-T6-T8-T4-T17;
(C17):T19-T18-T7-T12-T14-T8-T7-T18-T21-T18-T7-T12-T14-T6-T22-T13-T6-T8-T4-T17;
(C18):T20-T11-T18-T11-T20-T11-T18-T7-T12-T14-T8-T7-T17;
(C19):T19-T18-T7-T12-T14-T8-T7-T18-T21-T18-T7-T12-T14-T6-T22-T13-T8-T6-T17-T4;
(C20):T19-T18-T19-T18-T7-T12-T14-T8-T7-T18-T21-T18-T7-T12-T14-T6-T22-T13-T8-T6-T17-T4; (C21):T20-T11-T18-T8-T7-T8-T12-T14-T15-T14-T5-T16-T7-T17-T11;
(C22):T18-T8-T7-T8-T12-T14-T15-T16-T7-T17;
(C23):T23-T18-T7-T8-T8-T12-T15-T14-T16-T7-T17;
(C24):T19-T7-T8-T7-T8-T12-T14-T15-T14-T5-T16-T7-T17;
(C25):T20-T11-T20-T18-T11-T20-T11-T18-T7-T12-T14-T8-T7-T17;
(C26):T11-T20-T11-T18-T7-T12-T14-T8-T7-T17-T11;
(C5):T11-T20-T11-T18-T7-T12-T14-T8-T7-T17;
(C27):T20-T11-T20-T18-T8-T7-T8-T12-T14-T15-T14-T5-T16-T17-T7;
(C28):T18-T8-T7-T8-T12-T14-T15-T14-T5-T16-T7-T17;
(C29):T7-T8-T7-T8-T12-T14-T15-T14-T5-T16-T17-T7;
(C30):T11-T18-T8-T7-T8-T12-T14-T15-T14-T5-T16-T17-T7-T20;
(C31):T11-T20-T11-T20-T11-T7-T18-T8-T8-T12-T14-T15-T14-T5-T16;
(C32):T20-T11-T18-T7-T12-T14-T8-T7-T17;
(C33):T11-T20-T18-T7-T18-T11-T18-T7-T12-T14-T8-T7-T17;
(C34):T20-T11-T20-T11-T20-T11-T20-T11-T18-T7-T8-T12-T14-T15-T14-T5-T16-T7-T17;