# Enrichment of Business Process Management with External Data

Pierre-Aymeric MASSE

Orange Labs, 38-40 Rue du Général Leclerc, 92130 Issy-les-Moulineaux, France
pierreaymeric.masse@orange.com
Institute Mines-Télécom, Télécom Bretagne, UMR CNRS 6285 Lab-STICC, 29238 Brest, France
pa.masse@telecom-bretagne.eu
Université Européenne de Bretagne, 29238 Brest, France

**Abstract.** Organizations have an increasing need to adapt faster their Information Systems (IS) to technical, functional or legal changes. Orange, a French telecom operator, works on the adaptation and the improvement of their business processes (BP), especially those related to Customer Relationship Management (CRM). One of the challenges is to help a business expert to efficiently and quickly adapt a BP. Indeed, this challenge includes the need to understand the reasons why the execution of the BP does not satisfy the business needs and the business goals. In this research paper, we propose to study how to identify these reasons based on the analysis of relevant data which include process generated data (such as logs and database data), and contextual data. To address this research question we plan to explore two directions: semantic enrichment of BP in order to detect relevant data and BP optimization to align BP to business goals on the one hand and to the relevant data on the other hand.

**Keywords:** Business Process Management, Semantics, Process optimization

## 1 Introduction

The increasing adoption of Business Process Management (BPM) [14] in recent years has resulted in a large standardization of processes. Companies are confronted to frequent changes. Consequently, adapting and continuously improving BPs, in order to align them to the changes of the company, is a key challenge to stay competitive. In practice, companies aim to reduce the time needed to take into account these changes. Currently, this is done manually by a business expert who analyzes the various components (such as business goals [18], roles and actors related to BP activities, process execution, business data and external data) and correlate them each other in order to deduce relevant adaptations to apply on the BP.

One existing approach to speed up the adaptation and the improvement of BPs is to dynamically detect relevant changes to apply on the BP execution (e.g. select a specific path). Authors in [14] surveyed existing solutions in this field. One solution is the concept of flexible process. It consists in incorporating alternative execution paths within the BP model so that the selection of the most appropriate execution path can be done at the runtime for each instance (flexibility by design). Another solution is the concept of configurable process [5], which consists in designing a model that provides a complete and integrated set of all possible process configurations. Afterwards, before the runtime, such a model can be configured to a specific solution by restricting the behavior of the configurable process model. For example, activities may be skipped or blocked during the configuration time.

We identified two limitations of these approaches. Firstly, in both cases, all possible adaptations must be well defined at the BP design time. Consequently, it significantly limits the adaptation scope. Secondly, these solutions are well suited to handle known exceptional and temporary situations in which the adaptation of the BP execution is necessary, but not for long term changes, in which the adaptation of the BP model itself is required.

At Orange (a French telecommunication operator), where the PhD takes place, there are currently two research projects related to this research field. The first ongoing project, named PRODIA, is based on process mining techniques [13] to detect BPs as they are executed by involved parties. Its goal is to provide experts with a comprehensive view of the execution of BPs in order to continuously improve them. The second project is based on web semantic techniques to speed up the implementation of BPs by matching BPs activities with Web services through ontology concepts [17].

The aim of this PhD thesis is to support business experts in the adaption of the BPs. To achieve this goal, it is necessary to:

– firstly detect relevant data which could influence the BP execution;
– secondly correlate these data with the business goals and the BP model in order to understand how these data influence the BP execution;
– finally suggest to the business expert changes to operate on the BP model in order to align it to the business goals on the one hand and to the relevant data on the other hand;

Our work could be used by Orange to improve its CRM, and more specifically the customer subscription to a telecommunication service business process. Let's suppose that one business goal of this BP is to have the rate of customer who abandons the subscription process below 30%. Using traditional process mining techniques we can detect that the abandon rate is higher than our objective. Current techniques however do not provide additional support to the business expert to understand "why" the BP does not satisfy the business goal. The proposed thesis aims to fill this gap by not only providing solutions to support the expert in understanding the reasons that make a BP does not satisfy a business goal, but also to suggest BP adaptations accordingly.

The rest of the paper is structured as follows. Section two, we briefly narrates the closely related work. In section three we present the research problem. A research methodology is then detailed in the section four. Finally, the section five presents the summary of the paper.

## 2 Related work

BPM is characterized by the BP lifecycle definition [13], which contains seven stages: (re)design, analysis, implementation, (re)configuration, execution, adjustment, and diagnosis. Though the thesis is related to all phases of the BP lifecycle, a particular focus is given to the (re)design, the analysis, the execution and the diagnosis phases, as they directly impact the BP transformation. In the (re)design phase a new BP model is created or an existing BP model is adapted. In the analysis phase a candidate model and its alternatives are analyzed to validate the model (e.g. avoid deadlocking, detect dead paths, etc.). Then, after the implementation and the configuration phase, the execution phase orchestrates the different BP activities in accordance to the designed model. At the end of this lifecycle, in the diagnosis phase, the enacted BP is analyzed, which may trigger a new BP redesign phase. The diagnosis phase usually relies on the logs and data generated by the different instances of the BP.

Improving the BP adaptations process has been investigated extensively. We classify existing approaches into 3 different categories.

**BP variant solution**

This first category concerns existing solutions that adapt the model before its execution according to a particular situation.

The authors in [14] survey BPM research field. They review BP variant solutions where the process model is subject to continuous evolution. It broaches the difference between flexible (run-time decision) and configuration BP (configuration-time decision).

Configuration BP consists in incorporating alternative execution paths within the BP model so that the selection of the most appropriate execution path can be done before the runtime for each instance. It also enables to merge several BP model from the same family (all related to the same domain) into a configurable BP model in order to reduce the number of BP managed by the BPM system and team. Configuration BP consists in having a model providing a complete and integrated set of all possible BP configurations. Afterwards, such a model can be configured to a specific solution by restricting its behaviour. For example, activities may be skipped or blocked during the configuration time. On the opposite, variability by extension contains the most common behaviour. Afterwards, the model is extended (e.g. adding new activities) during the configuration time to serve a specific situation.

The paper [5] presents the configurable workflows approach that proposes to customize the BP model by applying, “hiding” or “blocking” operations to BP activities. “Hiding” operation makes an abstraction of the model and hides some activities, but these activities are still executed. “Blocking” operation removes a path from the model. [19] proposes a framework to capture the variability of a BP by processing a set of business rules. Business rules cover all aspects of the business logic in BPs. A non-deterministic goal-driven BP inference engine is used to create the BP model. Consequently, business expert will focus on the design of business goals instead of specifying the detailed control and data flows. Another approach to identify the variants of a BP is proposed in [9] which is based on applying a questionnaire by domain. Based on the answers of the business expert, the system generates the most suitable BP model variants.

These approaches can also be used as a context aware BPM solution. In this field, [12] provides a top layer approach making automatic context-based decisions. This context-aware approach takes the context as relevant data to dynamically configure the BP.

The first limitation of existing BP variant solutions is the effort involved in constructing and maintaining customizable BP models beyond trivial examples. Indeed, The amount of information required to construct and to maintain such a model grows exponentially with the complexity of the BP [14]. Consequently, it significantly limits the adaptation scope. These solutions enable business experts to find a tradeoff between the number of BP to design and their complexity. The second limitation resides in the diagnosis phase. Indeed, in practice, BPs variant do not enable an accurate analysis using traditional techniques based on event logs such as process mining. This is due to the generation of multiple instances which depends on contexts that is not always accessible or taken into account by these techniques.

**Semantic techniques**

The second category harnesses semantic techniques to improve the BPM lifecycle (semantics-based BPM (sBPM) [4]). It consists in adding semantic annotations to a BP model. Semantic techniques are based on ontologies. Authors in [8] investigate current approaches in sBPM, especially those related to the existing gap between the business community and the IT community (e.g. the finished European-funded project, FUSION[1], which worked on a semantic framework to easily allow collaborative work of several enterprises in a BP). Semantic based approaches also enable for instance:

- to propose an auto-completion mechanism to speed up the modeling process [2]. The recommendation system determines possible activities set based on models previously created and similarity computing;
- to accelerate the transformation of the model into a valid implementation using natural language processing techniques and semantic technologies. The authors in [1] study how to automatically match BP activities with Semantic Web Services (SWS) description in order to transform the BP model into an effective implementation. The proposed framework detects automatically the web services to use for each BP activity. This matching process is based on an ontology, built around the e-Tom Framework[2]. Based on semantic

[1] www.fusionweb.org
[2] www.tmforum.org

description of activities, in the BP model, and web services (SWS) in the service platform, the proposed framework detects automatically which web service to use to achieve a BP activity;
- to link BPs activities with BP data in order to perform better diagnosis [11]. This approach could be extended to handle additional data, which are not especially produced by the BP but still influence it. These data are interesting to correlate with the BP but difficult to identify by business experts because they are not directly related to the BP;

**Deep analysis and optimization**
Another research area related to our work is BP optimization and deep analysis. BP optimization aims to study how a BP can be improved. BP optimization based on quantitative measures of goals achievement is not yet well addressed in the literature [16]. Deep analysis refers to a set of techniques that apply sophisticated data processing techniques to extract information or knowledge from large data set.

For instance, the framework proposed by [7] analyzes BP data and operational data, in real-time, to detect a predicted metric deviation. It uses mining techniques to generate decision rules based on BP data and the accomplishment of the BP goals. A recommendation mechanism evaluates the most compliant rule to fix the BP instance deviation. However, this approach uses only data of the BP instances to dynamically fix the BP deviation though the deviation could be caused by other data (e.g. road traffic which causes additional delay in the delivery BP). Another approach that use data of the BP instances is detailed in [3]. This framework recommends to a user the next action based on:

- the identification of the data which provides information about intentions;
- the identification of the intentional cluster of events associated with an intention and its naming;

The authors in [11] show an approach to match BP data and operational data in order to make a deep business analysis. The proposed framework correlates two types of BP data: those stored in the BPM and those stored in the IS.

To improve process mining results, an analysis approach is proposed in [15] which takes into account the BP execution context. This technique uses event logs with a clustering algorithm to regroup closest BP execution.

Another solution that uses context is proposed by authors in [6]. AGENTWORK provides healthcare domain with a comprehensive support for automated BP adaptation. This framework is based on Event/Condition/Action rules to detect the execution of exceptional activities in the BP in order to suggest BP relevant adaptations. The actor can accept these adaptations to apply to the current BP instance and save the current context in the framework. In addition, it tries to apply predictive adaptations based on similarity between the current and the previously encountered contexts.

## 3 Research Problem

The aim of this thesis is to study how to accelerate the adaptation of BP to optimize them based on:

- business goals;
- operational data such as the number of achieved subscription;
- non-operational data (external data) such as the context in which the BP is instantiated and executed;
- BP data such as an activity duration;

We define Operational Data (OD) as any data processed within the BP but not stored directly in the BPM system; data stored in BPM system is then named BP Data (BPD) [11]. We define Non-Operational Data (NOD) as any data that are not directly generated or modified by the BP (e.g. Urban Traffic and Weather in Delivery BP), we also refer to these data as external data.

Unlike BPD and OD, NOD are not directly linked to the BP. Consequently, it is currently difficult for an expert, and even for a machine, to correlate BP execution with these data. Therefore, to investigate this issue, we define several research questions:

- How to detect relevant data that impact BP execution based on business goals and BP model?
- How to correlate these data with BP instances to explain business goals deviation (metric deviation)?
- How to identify BP adaptations that address the goals deviation?

## 4 Research methodology

In order to better respond to our research questions, we plan to explore two directions which are the result of the early work of the thesis.

### 4.1 Research directions

**Semantic Enrichment for Enhanced Diagnosis**
This first direction aims to set up the foundation for a solution for linking data (OD, NOD, and BPD) to a BP. We propose to add semantic annotations to a BP in order to detect relevant data which impact BP execution. This proposal impacts the following BP lifecycle phases:

*(Re)Design and analysis*
During the design and analysis phase, our proposal consists in enabling business experts to add semantic annotations to each activity of the BP on the one hand and to the associated business goals on the other hand. This implies the definition of a tool that supports the business experts in this task, as well as the corresponding methodology. Based on the initial semantic annotations (manually specified by the business experts), the BP model, and business goals, the

tool must be able to find out additional concepts that could impact the BP execution. This mechanism explores ontological and BP relationships to discover new relevant concepts. These new concepts are validated or not by business experts.

*Execution*
During the execution phase, our proposal consists in retrieving the data that could impact the execution of the BP. Based on semantic annotations discovered at the design and analysis phase, these data are clearly identified. Nevertheless, it still remains important, especially for volatile data, to store them and associate them to the BP instance for further diagnosis.

*Diagnosis*
Our challenge in the diagnosis phase is to detect how the data impact the BP execution. From the technical point of view, this consists in correlating the data related to the concepts inferred in the design phase, and retrieved during the execution phase with the different instances of the BP. Clustering techniques (unsupervised learning) could be applied to detect such correlations.

**BP Optimization**
Semantic enrichment aims to provide business experts with all necessary elements that could explain "why" a BP doesn't respect a business goal. The aim of BP optimization is to design and implement a solution that suggests BP adaptations to a business expert; adaptations that align the BP to the business goals, taking into account the data environment. To achieve this goal, we analyze all the data (BPD, NOD, and OD) to highlight possible adaptations and to align the BP as well as possible to business goals. A candidate approach is the model-transformation by applying the goal-model to the BP model [10]. Model-transformation is based on rules that transform a given source model to a target model, according to specified meta-models.

### 4.2 Evaluation

In order to evaluate our approach we are currently developing a proof-of-concept which implements our proposals. We look forward to apply this prototype to a customer relationship use case in order to evaluate it. As our proposals are intended for business experts, we plan to interview them based on qualitative evaluation questionnaire in order to validate the results.

### 4.3 Research method

We divide our research method to several steps. This method will lead the PhD thesis with a methodology to provide scientific results. First we study the problem which consists in:

- studying the state of art of the BPM and related research fields;
- defining the research questions and highlighting the state of the art limitations;
- proposing new concepts and continuously studying the state of the art accordingly;
- defining the evaluation criteria;

Then, we plan to design the proposed concepts that respond to the different limitations identified in the state of the art. At this stage of our research progress we identified the following items:

- designing a BPM solution based on semantic to detect relevant data (data that influence the BP);
- designing a goal-driven BPM to improve relevant data selection;
- designing BP analysis solution that highlights the reasons of a metric deviation regarding business goals;
- designing BP optimization method, optimization that aims to redress a metric deviation;

Finally, we plan to evaluate these solutions and raise their benefits and limitations. We intend to:

- implement the prototype and apply it to customer relationship BPs;
- evaluate the proposals according to identified criteria;

## 5 Summary

In this paper, we detailed and motivated our PhD thesis subject. The main research question we are trying to address is how to speed up the adaptation of BPs to optimize them based on data that influence them and their business goals? We subdivide the main research question into several elementary subquestions. Then, we reviewed the state of art to position our work. Finally, we elaborated the plan of the PhD and proposed to deeply study two themes: Semantic Enrichment of BP for Enhanced Diagnosis and BP optimization.

### 5.1 Acknowledgements

I want to thank my supervisors, Dominique Seminel (Orange), Jacques Simonin (Telecom Bretagne), Nassim Laga (Orange) and Patrick Meyer (Telecom Bretagne), for many hours of discussion and feedback regarding to the research topic and for their help to formulate the thesis subject.

## References

1. Assy, N., Yongsiriwit, K., Gaaloul, W.: A Framework for Semantic Telco Process Management - An Industrial Case Study. Intelligent Systems Design and Applications (ISDA) pp. 44–49 (2014)

2. Betz, S., Klink, S., Koschmider, A., Oberweis, A.: Automatic User Support for Business Process Modeling. In: Proceedings of the Workshop on Semantics for Business Process Management, pp. 1–12 (2006)
3. Epure, E.V., Hug, C., Deneckere, R., Brinkkemper, S.: What Shall I Do Next ? Intention Mining for Flexible Process Enactment. In: Advanced Information Systems Engineering, pp. 473–487. Springer International Publishing (2014)
4. Fleischmann, A.: What Is S-BPM ? In: Buchwald, Hagen and Fleischmann, Albert and Seese, Detlef and Stary, C. (ed.) S-BPM ONE Setting the Stage for Subject-Oriented Business Process Management, pp. 85–106 (2010)
5. Gottschalk, F., Van Der Aalst, W.M., Jansen-Vullers, M.H.: Configurable Process Models A Foundational Approach. In: Reference Modeling, pp. 59–77. Physica-Verlag HD (2007)
6. Greiner, U., Rahm, E., Robert, M.: AGENTWORK : a workflow system supporting rule-based workflow adaptation. Data & Knowledge Engineering pp. 223–256 (2004)
7. Gröger, C., Schwarz, H., Mitschang, B.: Prescriptive Analytics for Recommendation-Based Business Process Optimization. In: Business Information Systems, vol. 17th Inter, pp. 25–37 (2014)
8. Hoang, H.H., Tran, P.c.T., Le, T.M.: State of the Art of Semantic Business Process Management : An Investigation on Approaches for Business-to-Business Integration pp. 154–165 (2010)
9. La Rosa, M.: Managing Variability in Process-Aware Information Systems by. Ph.D. thesis (2009)
10. Popp, R., Kaindl, H.: Automated Adaptation of Business Process Models Through Model Transformations Specifying Business Rules. In: CAiSE-Forum-DC 2014. pp. 65–72 (2014)
11. Radeschütz, S., Mitschang, B., Leymann, F.: Matching of Process Data and Operational Data for a Deep Business Analysis. In: Springer London (ed.) Enterprise Interoperability III, pp. 171–182 (2008)
12. Tavares Nunes, V., Werner, C., Santoro, F.M.: Dynamic Process Adaptation : A Context-aware Approach. In: Computer Supported Cooperative Work in Design (CSCWD), 2011 15th International Conference, pp. 97–104 (2011)
13. Van Der Aalst, W.M., Al.: Process Mining Manifesto. Business Process Management Workshops 99, 169–194 (2012)
14. Van Der Aalst, W.M., All: Business Process Management: A Comprehensive Survey. ISRN Software Engineering 2013, 1–37 (2013)
15. Van Der Aalst, W.M., Bose, R.J.C.: Context Aware Trace Clustering. In: SDM, pp. 401–412 (2009)
16. Vergidis, K., Member, S., Tiwari, A., Majeed, B.: Business Process Analysis and Optimization : Beyond Reengineering. IEEE transactions on systems, man, and cybernetics 38(1), 69–82 (2008)
17. Wang, X.H., Zhang, D.Q., Gu, T., Pung, H.K.: Ontology Based Context Modeling and Reasoning using OWL 3. In: Pervasive Computing and Communications Workshop. pp. 18–22 (2004)
18. Yu, E.S.k.: Modelling strategic relationships for process reengineering. Social Modeling for Requirements Engineering 11 (1995)
19. Zeng, L., Flaxer, D., Chang, H., Jeng, J.j.: P LM f low Dynamic Business Process Composition and Execution by Rule Inference pp. 141–150 (2002)