Towards a formal approach to a Knowledge Base supporting Business Process Analysis*

Antonio De Nicola[1], Anna Formica[2], Ida Mele[2], Michele Missikoff[2,**] and Francesco Taglino[2]

[1]*Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Casaccia Research Centre, Via Anguillarese 301, I-00123 Rome, Italy*

[2]*Istituto di Analisi dei Sistemi ed Informatica (IASI) "Antonio Ruberti", National Research Council (CNR), Via dei Taurini 19, I-00185 Rome, Italy*

Abstract

Business Process Analysis (BPA) is a strategic activity, necessary for enterprises to model their business operations. It is a central activity in information system development, but also for business process design and reengineering. Despite several decades of research, the effectiveness of available methods is still questionable. The majority of methodologies adopted by enterprises are rather qualitative and lack a formal basis, often yielding inadequate specifications. On the other hand, there are methodologies with a solid theoretical background, but they appear too cumbersome for the majority of enterprises. This paper proposes a knowledge framework, referred to as BPA Canvas, conceived to be easily mastered by business people and, at the same time, based on a sound formal theory. The methodology starts with the construction of natural language knowledge artifacts and, then, progressively guides the user toward more rigorous structures. The formal approach of the methodology allows us to prove the correctness of the resulting knowledge base while maintaining the centrality of business people in the whole knowledge construction process.

Keywords

Business Process Analysis, Business Model Canvas, Knowledge Representation, Formal Methods

1. Introduction

Business Process Analysis (BPA) [1] is a strategic activity for an enterprise, used for instance for organizational changes, Business Process (BP) reengineering, and information system development. BPA is positioned in the preliminary phase of a software project. Software projects are among the most difficult engineering undertakings. Despite the significant advances in Software Engineering and, specifically, Requirement Engineering, software projects still face a number of shortcomings. One of the major causes of software project failures is represented by the problem of business/IT misalignment [2], which happens when the services implemented

SEBD 2023: 31st Symposium on Advanced Database System, July 02–05, 2023, Galzignano Terme, Padua, Italy

* Under review.

** Corresponding author.

antonio.denicola@enea.it (A. De Nicola); anna.formica@iasi.cnr.it (A. Formica); ida.mele@iasi.cnr.it (I. Mele); michele.missikoff@iasi.cnr.it (M. Missikoff); francesco.taglino@iasi.cnr.it (F. Taglino)

0000-0002-1045-0510 (A. De Nicola); 0000-0002-7992-6898 (A. Formica); 0000-0002-3730-6383 (I. Mele); 0000-0002-7972-5201 (M. Missikoff); 0000-0001-8364-4407 (F. Taglino)

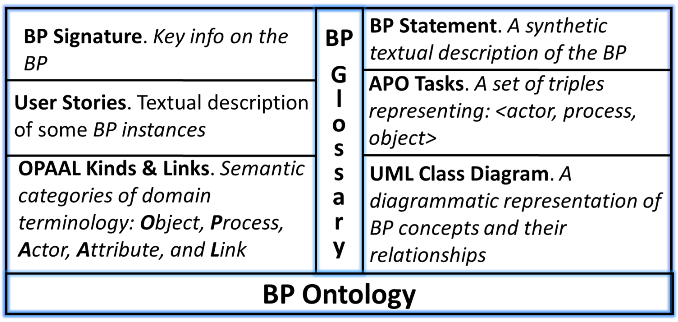

Figure 1: BPA Canvas layout

by the information system do not fully correspond to the business needs. Such a problem is mainly caused by difficulties in the communications between business people and IT specialists, yielding poor requirement specifications [3]. In this paper, we propose an evolution of the knowledge-driven BPA methodology, referred to as *BPA Canvas*, presented in its preliminary version in [4]. We present a formal foundation of the proposed methodology, keeping its user-friendly characteristics to be easily adopted by business people.

2. The Business Process Analysis Canvas

In this section, we introduce the main ideas of the BPA Canvas and the related methodology. It includes a set of knowledge artifacts and a procedure aimed at guiding business experts in collecting and organizing the knowledge of a business process.

With respect to the business process modeling methods available in the literature, the BPA Canvas has not the objective of drawing process diagrams, an activity that is postponed to the BP design phase. BPA Canvas is aimed at the careful collection of the knowledge necessary to build a first static model of a business process. The idea is that a rigorous and detailed knowledge base about a BP will substantially support the subsequent design task, improving the quality of the process flow diagrams and, therefore, of the produced information system.

The BPA Canvas is organized into eight knowledge sections that hold different kinds of knowledge artifacts, i.e., models of the given business process. The models can assume various forms, with different levels of details and formality. In particular, we have: (i) plain text, i.e., a narrative form of knowledge representation; (ii) structured text, e.g., itemized lists (bullet points) that collect and organize short statements; (iii) tables, typically providing a systematic visualization of knowledge items; (iv) diagrams, where the knowledge is graphically represented according to a given standard; (v) formal representation of the business domain by means of a BP Ontology. Figure 1 shows the layout of the eight sections of the BPA Canvas that are listed below.

- **BP Signature**. The first knowledge artifact, in the form of a list, aimed at providing a synthetic profile of the business process.

- **BP Statement**. This is a preliminary plain text description of the business process and its business scenario, described in general terms (i.e., at an intentional level).
- **User Stories**. One or more plain text descriptions of exemplar executions of the BP (i.e., at an extensional level). In essence, it represents one or more instances of the BP Statement.
- **APO Tasks**. This is a set of triples representing a first operational account of the business process, abstracting the actual sequencing of the tasks.
- **BP Glossary**. A collection of terms, with their descriptions, that characterize the BP domain.
- **OPAAL Kinds & Links**. This structure is composed of two parts. The first part, $Kinds$, provides a semantic tagging of the terms (concept) names used in the construction of the knowledge artifacts, according to the following categories: *Object*, *Process*, *Actor*, and *Attribute*. The second part, $Links$, represents semantic relations among concept names, i.e., *ISA* for subsumption relation, *PartOf* for composition relation, and *HasA* to relate an entity with an attribute.
- **UML Class Diagram**. This section reports a Class Diagram providing a static view of the BP. The Class Diagram is built by using tasks and links in *APO Tasks* and $OPAAL$ sections, respectively.
- **BP Ontology**. An encompassing representation of the knowledge collected in the previous sections, encoded in formal terms by using an ontology language (e.g., OWL).

Then, the methodology indicates how to proceed in building the above knowledge structures.

3. A running example

The example illustrates the construction of the Business Process Knowledge Base (BPKB) for a home delivery pizza shop, called $PizzaPazza$, achieved following the BPA Canvas methodology. We show how the knowledge artifacts are first built in a step-wise fashion, omitting, for sake of space, the successive refinement cycles.

BP Signature. It represents the first knowledge artifact of the pizza shop BP. This is a structure of seven pairs, where the first element is a label:

[**BP Name**, HomeDeliveryPizza], [**Trigger**, OrderArrived], [**KeyActors**, Customer, Cook, DeliveryBoy], [**KeyObjects**, Order, Dough, Pizza], [**Input**, PurchaseOrder], [**Objective**, "Cook and deliver pizzas to customers"], [**Output**, PizzaDelivered, CustomerHappy]

BP Statement. The BP Statement is the synthesis of an interview with a (fictitious) pizza shop owner, who describes how a customer order is handled by the shop.

> *My business, PizzaPazza, is a home-delivery pizza shop. The customer fills in the order, by using our Web site, and then submits it to the shop, together with the payment. Making good pizzas requires good quality dough, produced in-house, and careful baking of the pizza. To make clients happy, we need to quickly fulfill the order and the delivery boy needs to know the streets and how to speedily reach the customer's address.*

User Story. Here, the text describes a specific execution of the BP, i.e., it represents an instance of the BP. If necessary, more user stories are reported, to represent various use cases and the corresponding process instances.

> *Mary connects to the PizzaPazza Web site and places her order of two Napoli pizzas, also providing the payment. Upon the arrival of Mary's order at PizzaPazza, John, the cook, puts the order on the worklist. When Mary's turn arrives, John prepares the ordered pizzas, bakes them, and then alerts the delivery boy Ed to come and pick up the pizzas. Thus, Ed collects the pizzas and starts his delivery trip, eventually achieving the delivery to Mary's home.*

The first three knowledge artifacts, *Signature, Statement,* and *User Story*, represent an important, but informal, starting point easily managed by a business expert. The following BPA Canvas sections are built starting from the textual artifacts, moving toward the semantic analysis of the business scenario. The analysis starts from the above free-form texts to extract the following structured knowledge artifacts.

APO Tasks. This section contains a set of triples, each representing an element of an actionable knowledge item: (*actor*, *process*, *object*). E.g.:

[Customer, Submitting, Order], [PizzaShop, Receiving, Order], [Cook, Baking, Pizza], [DeliveryBoy, Delivering, Pizza], [Customer, Receiving, Pizza]

OPAAL Kinds & Links This section has two parts. The first part, $Kinds$, indicates concept categories (*Object*, *Process*, *Actor*, *Attribute*), e.g.:

[**Object**, (Order, Pizza, Dough, Margherita, ...)],
[**Process**, (Submitting, Baking, Delivering, Receiving, ...)],
[**Actor**, (PizzaShop, Customer, Cook, DeliveryBoy, ...)],
[**Attribute**, (Price, Quantity, Address, ...)]

and the second part, $Links$, indicates binary relations among them (*ISA*, *PartOf*, *HasA*), e.g.;

[Dough, *PartOf*, Pizza], [Customer, *HasA*, Address], [Margherita, *ISA*, Pizza], ...

BP Glossary. This section is dedicated to the free text description of the concept names. Each entry (in its simple form) is a pair *[ConceptName, Description]*, e.g.,

[Customer, *"One who buys goods or services from a store or business"*],
[Baking, *"To cook food with dry heat, especially in an oven"*],
[Order, *"A request made by a customer at a pizza shop for food"*], ...

The two final sections, the **UML Class Diagram** and the **BP Ontology** can be derived from the three central sections of the BPA Canvas. For the sake of space, they will not be described here.

4. A Formal Account of a Business Process Knowledge Base

The formal grounding of the BPA Canvas methodology aims at guaranteeing the quality of the released knowledge base, avoiding missing information, redundancy, and contradictions. In this section we first present the formal structure of the BPKB, with its components. Then we present the consistency rules.

4.1. The Business Process Knowledge Base

Given a terminology N (i.e., a set of terms), a BPKB is a complex structure organized according to the layout of the BPA Canvas, where the OPAAL section has been decomposed into two parts, $Kinds$ and $Links$, yielding the following 9-tuple:

$$BPKB = (P, S, U, K, L, T, G, D, O)$$

where:

- P is the **BP Profile**;
- S is the **BP Statement**;
- U is the set of **User stories**;
- K is the set of pairs representing the categorization of terms, referred to as **Kinds**;
- L is the set of triples representing the structural **Links**;
- T is the set of triples representing the **APO Tasks** belonging to the BP;
- G is the **Glossary** in the form of a set of pairs *(conceptName, description)*;
- D is a **UML Class Diagram**;
- O is the **Ontology** of the BP.

The following formalization focuses on the core of the *BPKB* represented by the four central components, i.e., *K*, *L*, *T*, and *G*, whereas the first three sections consist of unstructured knowledge artifacts expressed in natural language. The last two sections, the *UML Class Diagram* and the *BP Ontology*, are derived from the core and their formalization goes beyond the scope of the paper. Below, we report the formalization of the *Kinds*, *Links*, *APO Tasks*, and $Glossary$ sections.

Kinds. This component of the $BPKB$ is used to define the categories of the different terms. Given a terminology N, K is a set of pairs:

$$K \subseteq \{(n, k) \mid n \in N, k \in \{O, P, Ac, At\}\}$$

where O, P, Ac, At represent the categories a term can belong to, and:

- O stands for *Object*;
- P stands for *Process* (or *activity*);
- Ac stands for *Actor*;
- At stands for *Attribute*.

In our running example, for instance, the pairs: $(Cook, Ac)$, $(Pizza, O)$ state that the terms *Cook* and *Pizza* represent an *Actor* and an *Object*, respectively.

Structural Links. Given a terminology N, L is a set of triples:

$$L \subseteq \{(n_1, r, n_2) \mid n_1, n_2 \in N, r \in R, n_1 \neq n_2\}$$

where $R = \{$*ISA, PartOf, HasA*$\}$ defines the structural relations (links) in the *BPKB*.
A triple (n_1, r, n_2) is in L if n_1 and n_2 are related according to r.
For example: $(Cook, ISA, Person)$, $(Dough, PartOf, Pizza)$.

APO Tasks. This component of the *BPKB* represents the tasks of the BP as a set T of 3-tuple, defined as follows:

$$T = \{(ac, p, o) \mid \{(ac, Ac), (p, P), (o, O)\} \subseteq K, (ac, p) \in Inv, (p, o) \in Ach\}$$

where:
$Inv = \{(ac, p) \mid (ac, Ac), (p, P) \in K$ and ac is involved in $p\}$
$Ach = \{(p, o) \mid (p, P), (o, O) \in K$ and p achieves $o\}$.

Inv contains all the ordered pairs indicating an actor and a process it is involved in. Ach includes all the pairs whose first element is an activity, p, *achieving* or producing the second element that is an object, o.
For instance, in our business domain: $(Cook, Preparing, Pizza)$ is a possible task.

Glossary. The glossary G of the *BPKB* is a set of ordered pairs defined as follows:

$$G = \{(n, d) \mid n \in N, d \in D\}$$

where D is the set of all possible strings, standing for natural language descriptions.
In our running example, the pair: (*Pizza, "Italian open pie made of thin bread dough spread with a spiced mixture of, e.g., tomato sauce and cheese"*) is a possible element belonging to the glossary.

Although in this paper we do not elaborate on the *UML Class Diagram* and the *Ontology* details, we anticipate that the *UML Class Diagram* can be built starting from the *APO Tasks* and the structural *Links*. In particular, the built *UML Class Diagram* will consist of boxes (i.e., classes), named with *object* or *actor* names, connected by two types of arcs: functional and structural. The functional arcs (i.e., associations) will be labeled with *process* names connecting the actors with the objects, as reported in the *APO Tasks* triples. The structural arcs will be created from the triples in the structural *Links* where the label of the arc is the second element. For the ISA and $PartOf$ relations, the arc will connect two boxes labeled with the first and third elements. In the case of the $HasA$ relation, the first element will be a box name and the third element one of its attributes that will be listed within the box (according to the *UML Class Diagram* syntax).

At this point, the *Ontology* can be derived from the knowledge so far collected. Note that we illustrated the various knowledge artifacts in a sequence, however the actual construction of the knowledge base does not follow a 'waterfall' approach, but the Agile philosophy [5]. Therefore, its construction is achieved in a spiral fashion, and, at each cycle, it is possible to check and correct it, while enriching the overall content.

4.2. The Consistency rules

Now we introduce the consistency rules that will be used to accomplish the formal verification of the BPKB. Below, such rules are presented in an informal fashion, in this paper we omit their formal specification.

R1 – Definedness. All concept names in N need to have a description in G.
R2 – Uniqueness. Each concept name must be present only once in G.
R3 – Categorization. All concept names need to have a kind, i.e., to be categorized according to one of the $OPAAL$ categories.
R4 – Disjointness. Each concept name needs to be associated with only one kind.

R5 – Structural completeness. All the concept names need to participate in at least one triple in L.

R6 – Functional completeness. All the actor, object, and process names need to participate in at least one task, i.e., a triple in T. If a concept does not appear in a task, at least one of its subsumees or components or attributes (as declared in L) needs to participate.

R7 – Pragmatics. For all triples in T, the concept names need to belong to their respective categories, i.e., ac in the first place, p in the second place, and o in the third place.

Each time a BPKB is released, it can be checked for its correctness. To this end, the above rules are triggered and, in case of failure, a diagnostic message will indicate what is wrong, suggesting also where to intervene to repair the knowledge base.

5. Conclusion

In this short paper, we presented the BPA Canvas, a methodology for the acquisition, modeling, and management of business process knowledge. It has been conceived to be easily adopted by business people, offering at the same time, a solid formal grounding. The knowledge organization is guided by a canvas layout, structured according to eight sections representing a sort of knowledge dashboard and providing a synoptic view of the BPKB. With respect to previous proposals in the area of BPA, this methodology presents three key characteristics: (i) it starts with informal, intuitive models to grant business experts a central role; (ii) it adopts an Agile approach, with a cyclic progression of model building, with continuous releases and validity checks; (iii) it is characterized by a theoretical foundation for the *core* of the BPKB that represents its backbone.

Currently, we are working on a platform that, based on the formal part of the methodology, supports the knowledge acquisition task and checks the consistency as well as the completeness of the BPKB (under the Closed World Assumption (CWA) [6]). In the most popular BPA methodologies, all the checks need to be achieved manually.

Our work will continue along three main lines. The first consists of the development of a number of services aimed at supporting the BPKB construction. We will start with NLP services that analyze the first three canvas sections (BP Signature, Statement, and User Stories) to start populating the core of the BPKB. Then, we will provide semantic services aimed at enriching the BPKB by exploring existing terminological resources, such as DBpedia, Wikidata, WordNet, available on the Internet. We are aware that such resources do not guarantee the accuracy of their content, therefore they will be addressed with the necessary precautions. Finally, we will develop an automatic procedure that will use the consistency rules to check the soundness and completeness (within the CWA) of the BPKB.

The work presented in this paper is the continuation of the work carried out in the context of the European Project BIVEE (Business Innovation in Virtual Enterprise Environment) where a first proposal of knowledge-based enterprise analysis has been presented in [7].

References

[1] K. Vergidis, A. Tiwari, B. Maieed, Business process analysis and optimization: Beyond reengineering, IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews 38 (2008) 69 – 82.

[2] W. van Grembergen, S. De Haes, A research journey into enterprise governance of it, business/it alignment and value creation, International Journal of IT/Business Alignment and Governance (IJITBAG) (2010) 1–13.

[3] L. Aversano, C. Grasso, M. Tortorella, A literature review of business/it alignment strategies, Procedia Technology 5 (2012) 462–474. 4th Conference of ENTERprise Information Systems – aligning technology, organizations and people (CENTERIS 2012).

[4] M. Missikoff, A knowledge-driven business process analysis methodology, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 13427 LNCS (2022) 62 – 67.

[5] R. Prakash, N. Agarwal, Managing business analysis for agile development, International Journal of Modern Engineering Research (IJMER) 3 (2013).

[6] J. Álvez, I. Gonzalez-Dios, G. Rigau, Applying the Closed World Assumption to SUMO-Based FOL Ontologies for Effective Commonsense Reasoning, in: G. D. Giacomo, A. Catalá, B. Dilkina, M. Milano, S. Barro, A. Bugarín, J. Lang (Eds.), ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, volume 325 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2020, pp. 585–592.

[7] M. Missikoff, P. Assogna, The BIVEE Project: An Overview of Methodology and Tools, 2015.