Semantic based auto-completion of business process modelling in eGovernment

Eliane Maalouf, Maria Sokhn

Institute of information systems
University of applied sciences of western Switzerland, Valais-Wallis
Technopôle 3
3960 Sierre
eliane.maalouf@hevs.ch
maria.sokhn@hevs.ch

Abstract: So far, the semantic business process modelling is gaining in importance mainly due to the advances in the domain of semantic web technologies, which help enhancing, the modelling issues of the business process management life cycle. The linked data approach becomes interesting to apply on Business Processes to enrich them with information from diverse sources and link them to those sources. The aim of our project is to enhance the phase of process modelling of Swiss eGovernment public processes by suggesting to the modeller process fragments for auto-completion from a semantic repository. Created processes are linked and semantically enriched with ontology concepts related to the domain of public sector and eGovernment that will enhance query results in the repository as well as improve process documentation. We focus on linking the processes to the public sector standards.

1 Introduction

Business Process Management is the discipline concerned with business processes optimization. The BPM life cycle is generally divided into four phases: design, deployment/configuration, execution and analysis. Process modelling in the design phase plays a central role in this life cycle. Process models assist in managing complexity by providing insight and by documenting procedures. Van der Aalst survey on Business Process Management (BPM) [WAM03] states that the notion of a process model is foundational for BPM. He mentions that a process can be modelled from control-flow perspective (modelling the ordering of activities), resource perspective (roles, organizational units, authorizations, etc.), data perspective (decisions, data creation, forms, etc.), time perspective (durations, deadlines, etc.) and the function perspective (activities and related applications). In a recent work, the authors of [Ni13] stated that existing process knowledge in order to simplify the modelling process has not yet been applied. Van der Aalst categorized the main activities related to the management of large process model collections as search (given a query a set of models is returned), merge (combine a set of models into a single one), cluster (identify a set of related models), unify/refactor (improve the models by aligning them, removing redundancies, etc.), and convert (map from one notation to another). Van der Aalst mentions that a general problem affecting all activities is the use of informal text. The same activity may be

labelled differently in different process models thus hindering the detection of redundancies and inconsistencies. To avoid this problem, he suggests using a common ontology. In his opinion, semantic technologies aim to address obvious limitations related to string edit distances and linguistic similarity. However, in practice, few process model collections use a common ontology. Therefore, in most cases, semantic annotations still need to be added to process models before being able to use semantic technologies.

Work on Semantic Business Process Management have proliferated in the last few years, following the European Framework 6 funded project SUPER[1]. This project defined a Semantic Business Process life cycle adding automation to the bi-directional translation between the business and the systems spheres. The fundamental approach is to represent both the business perspective and the systems perspective of enterprises using a set of ontologies, and to use machine reasoning for carrying out or supporting the translation tasks between the two worlds. The project resulted in a semantic services architecture connecting semantic modelling and configuration tools, semantic execution and analysis tools and semantic repositories over a semantic service bus. It also resulted in a set of ontologies to support the design, resumed by [Di07] as follows:

- Upper Process Ontology (UPO), defining top-level concepts such as task, goal and condition
- Business Process Modelling Ontology (BPMO), extending the UPO into a full process ontology, providing abstractions over different business process modelling notations such as BPMN and EPC
- sBPMN , sEPC and sBPEL – ontologised versions of subsets of the BPMN, EPC and WS-BPEL respectively. sBPEL is additionally enriched with extensions from the Web Services Modelling Ontology (WSMO) for goal-oriented discovery, mediation and execution of services

The SUPER project allowed automatic discovery of services using semantic web technologies and linked them to process activities to ease configuration before execution. Our approach to the problem is different in the sense that we focus on the business process design phase only as will be described later on.

2 Swiss E-Government Public Business Processes Context

Switzerland is investing more and more in E-Government initiatives to modernize the electronic platforms to support citizens' access to services. The "eCH association" is responsible for elaborating the standards for cyber administration adoption in Switzerland. Those standards are emitted as recommendations and are freely available [eCHa]. The standard documents eCH-0140 and eCH-0158 defined the Business Process Model and Notation (BPMN 2.0) as the recommended notation to model Swiss administration public processes and defined some best practices for modelling inspired from the notation [eCHb] [eCHc].

[1] http://www.ip-super.org/content/view/45/69/

To complete those standardization efforts, the Swiss confederation along its State Secretariat for Economic affairs (SECO) started multiple priority projects to lead the research in cyber administration. One of those projects is the "B1.13: eCH-platform for process exchange between communities and cantons". The aim of this project is to ease knowledge transfer on business processes from one entity to another, share documentation about processes, surpass the languages barriers and offer standard procedures to external partners among other goals. This platform will play the role of a global process documentation platform and process repository. The different municipalities and cantons that form the administration body of the country provide processes. The platform is currently in its Beta-Version and accessible through http://www.ech-bpm.ch. Since it is a process documentation platform, execution data about processes are not collected and configuration of processes does not fall in its requirements. The modelling environment, subject of this paper, can be seen as the modelling tool on top of this public repository and provides process editing interface to the modellers. Hence, for modelling we are only concerned by the process model elements and its documentation without configuration details.

In their study about "Semantic Business Process Management for e-Government", Liu et al. [Zh13] concluded that in order to improve the degree of automation processing in business process management, techniques from semantic web such as ontology languages and reasoners have to be transferred to the business process world. An improved semantic management is thought to fix some of the inherent problems related to semantic deficiencies in government documents, inconsistencies of semantic information among different agencies, complex change management activities and poor process management practices. The authors mention that those problems seem to be more complex in Switzerland given its structure of 26 cantons and multiple municipalities with different degrees of independence in decision-making and the use of four different official languages.

3 Background

3.1 Business Process Modelling Notation

The BPMN language provides a graphical representation for the process designers to organize and model their business processes. Its recent 2.0 version defines the formal semantics of its elements. BPMN graphical objects are divided into three main categories: Flow Objects (Activities, Gateways and Events), Connecting Objects (Sequence Flows) and Swimlanes (Pools and Lanes) [Ch12]. Additionally, BPMN2.0 introduces extension capabilities to Flow Objects and Artifacts to allow process designers to express additional features of process models.

The authors of [Ch12] provide a formal definition of the BPMN meta-model syntax and specifications written in natural language. The ontology is called BPMN 2.0 Ontology, written in the Web Ontology Language (OWL) and has 85 classes, 178 object properties

and 59 data properties. This ontology will be further used to assert the structural elements of the process model.

3.2 Linked Data and Linked BPM

Linked Data is a set of best practices for publishing and interconnecting structured data on the web. The authors of [Ch09] provide an overview of its principles. Linked Data provides explicit links between data from diverse domains with the goal of increasing machine-readability of published data. Linked Data is published using the Resource Description Framework where URIs refer between various entities on the web.

The work of F.Gao in [Fe11] provided a case to implement the concepts of Linked Data with BPM. It presented BPMN2.0 extension examples to link a process to different views of the Architecture of Information System (ARIS) framework: organization view, functional view, data view, control view and control/service view. The process model in this framework is the controlling dispatcher of resources and actions. Extensions to BPMN2.0 are typically created and annotated in the design phase. The views are modelled with linked data, they are interlinked with each other and can be further linked to the external open data. BPMN 2.0 defines Extension element that can extend Flow Objects (e.g. Sub-processes, Activities, Tasks) and Relation element that can extend Artefacts (e.g. Group, Data Object, Text Annotation). Extensions in the different views are suggested as follows:

- Organization view: defines the roles, the participants, the organizational entities and the relationships between them. BPMN2.0 Activities should be annotated with organizational information like: who is responsible with executing the activity (individual or group), which organization/group does he belong to and what role does he have. BPMN 2.0 PartnerRole, Participant and PartnerEntity should be connected to concepts from dedicated organizational ontologies for roles, participants and organizational entities respectively.
- Data view: describes the information objects, the concepts in the messages exchanged by process tasks. Data encapsulated in BPMN ItemAwareElement should be annotated with concepts from domain specific ontologies.
- Control view: describes the control flow of business processes. BPMN Activity, Event and EventDefinition should carry information about occurrence time and other execution related data.
- Function view: defines the functions required to satisfy the objective of the enterprise. BPMN Activity should be defined as a domain specific business function instead of just a label.

3.3 Semantic Business Process Modeling with BPMN

Process modelling in eGovernment has been mainly been performed with standard modelling languages such as Activity Diagrams (AD) [Uml], Business Process Modelling Notation (BPMN) [Bpmn], or Event Driven Process Chains (EPC) [Sc00]. These approaches do not take into account the context of the eGovernment process

where activities may depends on legal regulations. Some works have been conducted to brdge this gap such as the PICTURE approach proposed by [Jo07]. This approach represents not only isolated processes but a complete landscape of the organisation taking into account technical and organisational measures. In their work, the authors of [Vo13] proposed a modelling approach that consists of building blocks within BPMN.

Requirements for a semantic modelling environment, were described by [Di07], we retain the following:

- Adding semantic annotation of existing process models, i.e. adding references to ontology elements.
- Storing the semantic process models into a Semantic Business Process Library
- Querying of the library for discovery of existing semantic process models or fragments for reuse through auto-completion, which will decrease the effort and time required for modelling of new processes

In their work the authors of [Ag09] define interesting scenarios for the annotated business models: decision making support, reuse BP patterns, reuse BP models/fragments, auto-completion, querying policies, goal-based BP analysis. In order to enable those scenarios, the paper suggested the use of organizational ontologies: Organizational Structure Ontologies (OSO), Organizational Units Ontology (OUO), Business Roles Ontology (BROnt), Business Functions Ontology (BFO), Business Resources Ontology (BRO), Business Goals Ontology (BGO). At the moment of writing of this paper, we were not aware of organizational ontologies representing the Swiss public administration.

In [Ch09], the authors present three types of assertions that can be carried out on annotated business processes: BP model type assertions to store information on the type of a graphical object; BP model structural assertions to store information on how the graphical objects are connected and BP model semantic assertions to represent annotations of graphical objects.

The authors of [MP07] provide a model for semantically annotating business process models stored in a repository to enhance process lookup, navigation in the repository and understanding of process models. It suggests annotating processes with the following perspectives: functional (process type, process area), behaviour (timing of activities), organizational (actor, organization level), informational (exchange process phase) and business process context (process relationship, business context, and goal).

Given the time allocated for this project, missing organization and domain ontologies cannot be engineered. We favour the use of an annotation scheme similar to the last model presented above combined with the ARIS views annotation model.

3.4 Business Process Auto-completion

The author in [Th07] suggests to complete processes at modelling time with fragments following a defined set of rules: Constraint rules concerning the integrity of semantic business process models; Event-Condition-Action rules having a syntax described by IF,

THEN, AND, OR and Dynamic Rules applied during process modelling by the modeller.

On the other hand, [Ag10] presents a modelling environment that takes the user's modelling context and the modelling history of a community of users into account. Rankings of the retrieved processes from the repository are based on similarity between the query and the model retrieved, patterns observed in other users' preferences and implicit user feedback.

Born et al. [Ma09] proposes a weighted average of three auto completion criteria. The criteria are process-context based analysis, pre-and post-condition analysis and non-functional property analysis.

In our project, we will base auto-completion algorithm on similarity with the process part that was modelled so far based on structural similarity constructed by querying with SPARQL the business process RDF definition. Context information are also included in the modelling interface to narrow the lookup space. Those annotations will be described later on.

4. Aim

From the context information provided above, we notice that managing a large process repository is a complex task especially in the case of the Swiss administration. First, managing a large amount of processes requires automations at some leves of the life cycle to avoid time consuming. Second, this administration is multi-leveled (cantons, communes) with relative degrees of freedom in decision making added to that the use of four national languages. The main issue of this context can be described as followed:

- Each commune can model its processes in its own language (i.e. German, French, Italian or Romansh). The level of language comprehension by the analyst hinders sharing knowledge about processes in the repository.
- Modelling procedures using BPMN is not an easy task especially for small communes who do not necessarily have a specialized analyst in process modelling. Humans modelling new processes do not have access to existing models in the repository through the modelling tool to lookup similar processes or fragments that can be reused.
- Business Processes are labelled and annotated using natural language that introduces inconsistencies in the usage of the language (e.g. the same activity can be named differently by different modellers). This makes the handling of the processes by machines more complex to extract context information.
- Documentation of processes is limited to the explicit information that can be added by a human and mapping a process to other processes or domain is yet to be done manually. In large repositories this element is crucial.

We aim to prove the concept that semantic web technology and linked Data can help address the mentioned issues in an efficient way.

5. Method

In order to attain the objective described earlier, we target the development of a semantic process modeling environment on top of a process repository containing a set of 50 processes from a local commune. In order to address the issues above the tool makes available multiple functionalities briefly described below. The development is taking place in iterations, at this stage of the work, some functionalities may be more detailed than others.

5.1 Modelling Interface

We favoured the reuse of existing Open source modelling tool over redeveloping a new one. “Signavio Core Components” code is being used for the initial development phase since it supports BPM 2.0 elements, offers a user-friendly interface and is based on JavaScript. The code is light and does not ship with any necessary prior database configuration, which leaves the choice open for us. New features can be developed in the form of plugins. Since this code is integrated in multiple other open source tools like jBPM[2] and Activiti[3] plugins developed in our platform can be easily reintegrated in others.

5.2 RDF Representation of Processes

When a new process is added to the repository, we store a representation of it in Resource Description Framework format in a triplestore. Based on the BPMN 2.0 Ontology, an XSLT transformation file is used to transform the BPMN process. The initial process file is saved without modification in a separate traditional transactional database.

For example, we might get the following triples to describe the process file we are working on:

bpmnFile:sid-ae602356-3fea-4054-abde-6a72aeba08fa rdf:type bpmn20base:Process .
bpmnFile:sid-2ad23ad3-68de-4eed-b3cb-907162de7da7 bpmn20base:rootElements bpmnFile:sid-ae602356-3fea-4054-abde-6a72aeba08fa .
bpmnFile:sid-ae602356-3fea-4054-abde-6a72aeba08fa bpmn20base:id "sid-ae602356-3fea-4054-abde-6a72aeba08fa"^^xsd:string .
bpmnFile:sid-ae602356-3fea-4054-abde-6a72aeba08fa bpmn20base:isClosed "false"^^xsd:boolean .
bpmnFile:sid-ae602356-3fea-4054-abde-6a72aeba08fa bpmn20base:isExecutable "false"^^xsd:boolean .
bpmnFile:sid-ae602356-3fea-4054-abde-6a72aeba08fa bpmn20base:processType bpmn20base:None .

[2] https://www.jboss.org/jbpm
[3] http://activiti.org/

5.3 Process Auto-completion

The auto-completion functionality in the modeller environment is being developed as an extension plugin. When the modeller adds an element in the interface an event is caught and a query is formed to the SPARQL endpoint that will query the structure of the processes for a similar construct. To narrow the query space filtering can be done based on annotation elements.

5.4 Process Translation and Vocabulary Management

The BPMN 2.0 standard enables the serialization of processes in XML format. Tools that support BPMN 2.0 support also serialization. In the XML representation of processes, name attributes of elements are the labels provided by the modeller for the given element. In our work, we only consider translating these attributes. For this proof of concept implementation we care to translate form German to French. Given the time we have on this project, implementing an advanced machine translation algorithms is not possible. We aim to offer an initial help to the modeller in the translation. The responsibility falls back on the human to confirm that the tool expressed correctly the concept in the destination language. In order to enable translation, we form a dictionary in RDF from a subset of the multilingual terminology database TERMDAT[4], a database for recording Swiss legal and administrative terminology along other terminology from the public sector. Then we populate GATE[5] Gazetteers with the dictionary concepts. The translation procedure would go as follows:

- The user activates the translation function for a process
- The process is annotated with GATE based on the Gazetteers. GATE enables the lookup of words in a given text using the Gazetteers and those matches are linked to dictionary concepts. This method only works to find exact matches.
- In no match is found, an external web translation API is called and requested a translation for the missing match. Meanwhile, partial matches are also looked up in the dictionary.
- The results from the local dictionary, the API are presented to the user is a form along the partial matches found. The user needs to confirm the translated text coming from the API or modify it.
- Once the user confirms the presented translation, this translation is added to the dictionary and will become the preferred translation to be used in subsequent translations when needed. A new process file is created where labels are in the destination language. If the user does not modify an API translation, the original text will save in the new process file to avoid having uncontrolled labels.

We favoured the construction of the dictionary with TERMDAT since the e-CH 0158 standard [eCHc] recommends the use of a standardized vocabulary and advised the use of this terminology database. To implement this recommendation furthermore, we

[4] http://www.bk.admin.ch/themen/sprachen/00083/00854/index.html?lang=en
[5] https://gate.ac.uk/

modify the modelling tool to allow the suggestion of words from the dictionary to label process elements and annotations.

5.5 Repository Querying

We defined a set of lookup queries to implement in order to question the repository for relevant processes. The queries are being constructed with SPARQL. The result is an identification of processes that answer this query. The queries are then displayed to the interface from the relational database. The queries being implemented are: lookup a match of a label in the repository (results include equivalent match in other languages), lookup based on a graphical input query, lookup for processes that start with an event E1 and followed by Activity A1, lookup all processes owned/executed/published by a person/role/entity, lookup for all processes that include a given fragment or sub-process, lookup all processes that include interaction between two named participants, lookup all process related to a given service/mission, lookup all process written in a given language.

The query interface will enable the user to enter a graphical query by constructing a part of the process he is looking for as well as a form to enter more semantic annotations to be looked up.

5.6 Linking Processes

In order to define interesting meta-data to add to processes in the repository of the B1.13 project, business consultants defined a set relevant to the domain. A basic ontology will be constructed using those concepts. These meta-data will be used to annotate the process with context information. A subset of those annotations included and their relation to the ARIS framework: process type (functional), total time (functional), owner (organizational), required documents (data), etc. Only annotations that are related to specific elements and not the whole process can be attached following the extension model presented by of F.Gao in [Fe11]

Two required annotation types can be linked to an external ontology: mission id and service id. More precisely, the eCH association emitted eCH-070, a standard listing the administration services [eCHd]. This standard was transformed into an ontology and is published under http://logd.ch/voc/service.html . For missions, there is a similar standard that is not published yet for which an internal ontology was developed and will be linked to the processes in our proof of concept.

6. Conclusions and future work

The current stage of the development is working on the interfaces, to integrate the modules of auto-completion and the translation presentation. In parallel, the translation module is being developed and dictionary management is being set. The results that will come out of the first tests will be compared to graph based techniques currently used in

modelling tools. Links to more data sources will be prepared, especially to link participants in processes to ontology instances in Friend of a Friend (FOAF) ontology for example.

In order to showcase the potential of Linked Data with business processes, a set of electronic public sector processes (e.g., eService suivi patrouilles) were annotated with information about the commune that offers it (e.g., Fribourg), the standard to which it is related (e.g., 400-10-012-006), the development language with which it was developed (e.g., Java). Then a visualization of this database was done using Linkurious. Such visualization is interesting in the case of a process repository where we try to make sense of the processes' relations between each other and their relations to the standards that regulate them.

References

[Uml] Object Management Group. UML 2.0 Supers tructure Specification. http://www.omg.org/cgi-bin/doc?formal/05-07-04

[Bpmn] Object Management Group. BPMN Final Adopted Specification 1.0. http://www.bpmn.org/Documents/OMG%20Final%20Adopted%20BPMN%2010%20Spec%2006-02-01.pdf

[Sc00] Scheer, A.-W.: ARIS - Business Process Modeling. 3 edn. Springer Publishing, Heidelberg et al. (2000)

[Jo07] Jörg Becker, Daniel Pfeiffer, Michael Räckers: PICTURE - A new Approach for Domain-Specific Process Modelling. CAiSE Forum 2007

[Vo13] Volker Schmitz, externer Berater des Projekts FIM, Föderales Informationsmanagement: Das FIM Vorhaben und seine Berührungspunkte zur XÖV Standardisierung 6. XÖV Anwenderkonferenz Bremen am 06. November 2013

[WAM03] Wil M. P. Van Der Aalst, Arthur H. M. Ter Hofstede, and Mathias Weske. 2003. Business process management: a survey. In *Proceedings of the 2003 international conference on Business process management* (BPM'03), Wil M. P. Van Der Aalst, Arthur Ter Hofstede, and Mathias Weske (Eds.). Springer-Verlag, Berlin, Heidelberg, 1-12.

[Di07] Dimitrov, M.; Simov, A.; Stein, S. & Konstantinov, M. (2007), A BPMO Based Semantic Business Process Modelling Environment., *in* Martin Hepp; Knut Hinkelmann; Dimitris Karagiannis; Rüdiger Klein & Nenad Stojanovic, ed., 'SBPM' , CEUR-WS.org, .

[eCHa] «eCH E-Government Standards,» Available: http://www.ech.ch/vechweb/page?p=categoryList&site=/documents/Standards/A%20bis%20Z [Last accessed on 06.05.2014].

[eCHb] F. Geschäftsprozesse, «eCH-0140 Vorgaben zur Beschreibung und Darstel-lung von Prozessen der öffentlichen Verwaltung der Schweiz,». Available:

http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0140&documentVersion=1.00. [Last accessed on 06.05.2014].

[eCHc] F. Geschäftsprozesse, «eCH-0158: BPMN-Modellierungskonventionen für die öffentliche Verwaltung,» 26 06 2013. Available: http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0158&documentVersion=1.0. [Last accessed on 06.05.2014].

[Zh13] Zhan Liu et al., Semantic Business Process Management for e-Governement, Proceedings of the IADIS International Conference WWW/INTERNET 2013, ISBN 978-989-8533-16-6.

[Ch12] Christine Natschläger-Carpella, Extending BPMN with Deontic Logic ISBN 978-3-8325-3225-3 313 Seiten, Erscheinungsjahr: 2012

[Ch09] Christian Bizer, Tom Heath, Tim Berners-Lee: Linked Data - The Story So Far. Int. J. Semantic Web Inf. Syst. 5(3): 1-22 (2009)

[Fe11] Feng Gao et al., Extending BPMN2.0 to Enable Links between Process Models and ARIS Views Modeled with Linked Data, Lecture Notes in Business Information Processing, 2011, p41-52

[Ag09] Agata Filipowska, Monika Kaczmarek, Marek Kowalkiewicz, Ivan Markovic, and Xuan Zhou. 2009. Organizational ontologies to support semantic business process management. In Proceedings of the 4th International Workshop on Semantic Business Process Management (SBPM '09), Martin Hepp, Knut Hinkelmann, and Nenad Stojanovic (Eds.). ACM, New York, NY, USA, 35-42. DOI=10.1145/1944968.1944975 http://doi.acm.org/10.1145/1944968.1944975

[Ch09] Chiara Di Francescomarino, Chiara Ghidini, Marco Rospocher, Luciano Serafini, Paolo Tonella, Semantically aided business process modelling, The Semantic Web - ISWC 2009. Lecture Notes in Computer Science Volume 5823, 2009, pp 114-129

[MP07] Mturi Elias, and Paul Johannesson. A context-based process semantic annotation model for a process model repository. Business Proc. Manag. Journal 19(3):404-430 (2013)

[Th07] Thomas Hornung , Agnes Koschmider , Andreas Oberweis, A.: Rule-based Autocompletion Of Business Process Models. In: CAiSE Forum 2007

[Ag10] Agnes Koschmider, Andreas Oberweis, Designing Business Processes with a Recommendation-Based Editor. Handbook on Business Process Management International Handbooks on Information Systems 2010, pp 299-312

[Ma09] Matthias Born et al. Auto-completion for Executable Business Process Models. Business Process Management Workshops. Lecture Notes in Business Information Processing Volume 17, 2009, pp 510-515

[eCHd] http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0070

[Ni13] Nico Clever et al. Towards auto-suggested process modeling – prototypical development of an auto-suggest component for process modeling tools. 5th International Workshop on Enterprise Modelling and Information Systems Architectures, 2013.