RePROSitory: a Repository platform for sharing business PROcess modelS

Flavio Corradini, Fabrizio Fornari, Andrea Polini,
Barbara Re, and Francesco Tiezzi

School of Science and Technology, University of Camerino, Italy

Abstract. The BPM community can certainly benefit from the adoption of open science principles. The availability of business process models can make BPM research results more controllable, replicable, and comparable. Unfortunately, in our experience, it is quite difficult to find open collections of models suitable to effectively validate research proposals in the BPM field. To address this issue, we have developed a web-based repository of process models, named RePROSitory, for sharing BPMN models, making them accessible to the community. We have started to systematically populate the repository with a collection of BPMN models, manually selected from the literature. The experience of models retrieval from RePROSitory is enhanced by the implementation of more than two hundreds quality metrics. These allow researchers to select from RePROSitory a set of models that they judge more suitable for the experiments they want to run.

1 Introduction

Open science principles [9] ask for reproducibility of experiments reported in published research works. They can certainly contribute to enhance the quality and relevance of the research carried out by the BPM community. These principles intend to improve the capability of checking, and possibly re-validating, the results of a reported research effort. Referring to research on business processes, this demands for a common set of models to conduct research, validate methodologies and techniques, and compare tools performance.

We focus our contribution on models designed using the BPMN 2.0 notation, which has acquired a clear predominance among the various proposals. Up to now, the community can benefit from few available BPMN model repositories for conducting experiments. The most important ones are "BPM Academic Initiative Model Collection"(`https://bpmai.org/`) and "Camunda BPMN for Research" (`https://github.com/camunda/bpmn-for-research`). These repositories are of great value for the entire BPM community, as they make available a huge amount of models that anyone can access to support their studies. In the past, we used those repositories for validating our research work (e.g., the framework in [1,2]). Despite this, these model repositories present several issues. First of all, from a study we conducted on the two repositories, we discovered that around 14% of the models present issues related to the usage of the BPMN

syntax, leading then to models that do not conform with the standard. Moreover, most of the models coming from the two repositories get low results in relation to common quality attributes. Typical issues concern, the usage of sequence and message flows. These elements are often drawn without a specified source or target, as well as some sequence flows cross pool boundaries, while message flows are used to connect elements inside the same pool. In addition, these repositories do not provide any advanced mechanism for filtering the stored models.

In this paper, in line with the European vision fostering open science,[1] we aim at promoting it in the BPM research field. With a web-based and open repository, named RePROSitory, we want to offer to the members of the BPM community the possibility of sharing and retrieving BPMN models of interest for their research works. The experience of models retrieval from RePROSitory is enhanced by a component that, for each uploaded model, computes more than two hundred well-established metrics coming from the literature (e.g., [3–8]), providing users with an identikit of the stored models. The metrics range from what we call "basic metrics", concerning the number of BPMN elements present in a model (e.g., occurrences of tasks, of a given kind of gateway, or of pools), to "advanced metrics" (e.g., control-flow complexity, cross-connectivity, and layout complexity). We have also contributed in enriching the collection of publicly available BPMN models by releasing on RePROSitory, 174 models coming from previous editions of the BPM conference, since the release of the BPMN 2.0 standard. The access to these models will constitute the basis for speeding up experimentation activities and will allow researchers to compare results of their approaches over a shared common benchmark of models.

2 RePROSitory Main Features

The RePROSitory homepage is the user's main access point. It provides the possibility to register a new account or to log in using an already existing one, so to fully access the platform functionalities.

The RePROSitory interface for logged users is shown in Fig. 1. It reports a list of information about the latest models that have been uploaded, and the most downloaded ones. Furthermore, it gives the user the possibility to customize available graphs according to different kinds of parameters, thus allowing to have a more impacting view on the models contained in the repository (i.e., the user can modify what is displayed on the x- and y-axes to compare different metrics of the stored models). The interface presents an interactive sidebar that allows the user to navigate RePROSitory functionalities; a user can access: the *Uploaded Model List* to see the list of all the models uploaded on the platform and eventually export them, the *Upload Model* feature to upload models to the platform, the *Search Model* feature to navigate the models, and the *Metrics List* for getting an insight on the supported metrics and their source. Notably, *Upload Model* and *Search Model* constitute the main functionalities provided by RePROSitory and described below.

[1] Open Science: `https://ec.europa.eu/research/openscience/index.cfm`

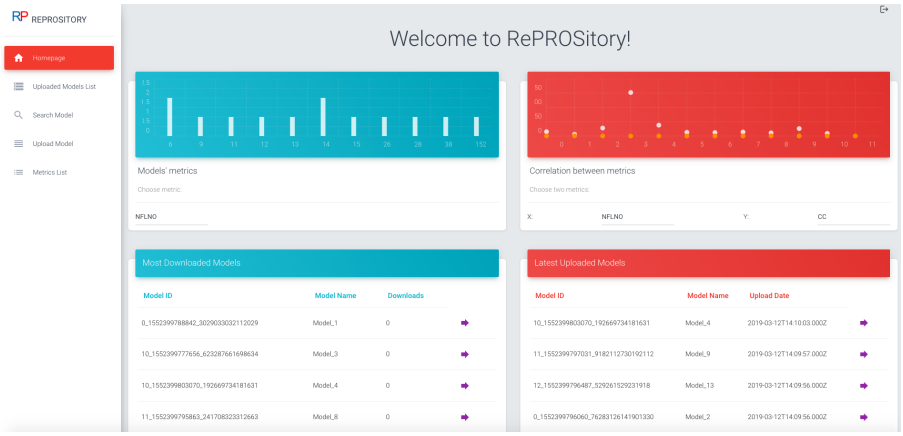

Fig. 1. Registered user homepage.

Upload Model. When a user uploads a model on RePROSitory, the model is sent to the ***BPMN Model Validator*** component, which checks if the BPMN syntax has been properly used, thus ensuring that no violation of the BPMN standard is present. In case a violation is detected, the model is automatically rejected. Instead, if the model passes the validation, it goes through the ***BPMN Metrics Extractor***. This component computes the values for the 245 supported metrics, which constitute the parameters a user can tune for filtering models, and it shows the resulting values to the user. This result is also made available for download in the form of a *.json* file. The model, together with the extracted metrics and the model information provided by the user, is then stored in a staging area of the RePROSitory platform, waiting for the quality and conformity check by an administrator and for the subsequent approval/rejection decision for sharing it on the platform. We consider valuable the involvement of a human administrator in the initial process of populating the RePROSitory platform, in order to avoid the incorporation of somehow problematic models. For the future, we intend to leave the repository completely open so that researchers will be able to share and retrieve also low-quality models.

Search Model. It is one of the richest and complex features of RePROSitory. It provides two different ways of filtering the models in the repository: *by models details* and *by metric values* as in Fig. 2. ***Filtering by model details*** allows the user to apply a filter based on models information, such as *model id*, *source*, *name*, *year*, *paper*, *doi*, *type*, *application domain*, *business process life-cycle phase*, and *modeling tool*. ***Filtering by metrics*** allows the user to specify customized parameters based on model metrics. A combination of comparison operators and values is used for each considered metric. Once all the desired filters have been applied, by pressing the Filter button the models that satisfy the parameters are returned. When selecting one model, the user is provided with a preview of the model. Upon pressing the download button, the user is able to download a *.zip* archive containing the selected model and the extracted metrics.

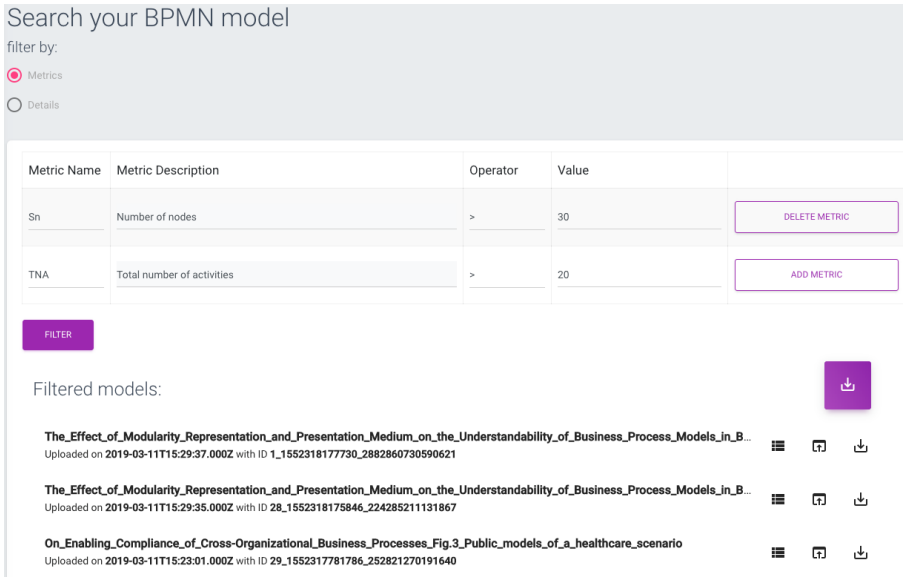

Fig. 2. Filtering customization based on metrics and list of resulting models.

3 Maturity of the Tool

The maturity of RePROSitory can be analyzed from a technological perspective and from that of the quality and quantity of its content.

The technologies at the basis of the RePROSitory architecture are well-established frameworks for the development of web-based applications: Angular, a Javascript framework that we have used for the development of the front-end, and Node.js, a cross-platform JavaScript runtime environment that we used for the back-end. The use of these state-of-the-art frameworks ensures high stability and performance of the tool.

In addition, to check that the models shared on RePROSitory are compliant with the BPMN standard, we have implemented a **BPMN validator**. As basis for this component we have exploited the Camunda BPMN Model API[2], which are largely used in BPMN tooling. Finally, to enable models filtering on RePROSitory, we have integrated a **BPMN Metrics Extractor**[3]. This is a Java web-service that we have developed to calculate metrics values for BPMN models. It is based on a collection of 245 metrics derived from the literature [3], and again takes advantage of the functionalities provided by the Camunda BPMN Model API. Calculation of some business process metrics is already provided by well-known tools, such as: ProM (`http://www.promtools.org`), APROMORE (`https://apromore.org`), PromniCAT (`https://bpmai.org/BPMAcademicInitiative/BpmTools`). However, the number of metrics they are able to calculate

[2] `https://github.com/camunda/camunda-bpmn-model`
[3] `https://github.com/PROSLab/BPMNMetricsExtractor`

is significantly lower then ours and, most of all, they are not leveraged as a means for classifying models as we are promoting them.

Let us focus now on the maturity of the content of RePROSitory considering quality and quantity of the models currently available on our platform. The models have been selected together with relevant information (title, publication year, model type, etc.), after a careful screening of the main track proceedings of the last seven editions of the *BPM Conference*. Our manual selection of models should ensure a certain quality of the shared models. Since each model has been reproduced by us, to avoid introducing syntactical errors in this phase, each model has been drawn by using Camunda BPMN Modeler, and validated using our BPMN Validator. From the quantity point of view, up to now we have collected and uploaded on RePROSitory 174 models, retrieved from 56 papers, which are available for being accessed and downloaded with the respective calculated metrics. The amount of models we have gathered does not pretend to be exhaustive; however, we are committed to extend the model retrieval also to other conferences and journal papers. In addition, we hope to receive in the future the contribution from other members of the community who perceive, as we do, RePROSitory as a valuable resource for their research.

4 Screencast and Website

RePROSitory is accessible at `http://pros.unicam.it/reprository` together with a detailed *User Guide* explaining how to use RePROSitory. A screencast is available on the PROS Lab channel at `https://youtu.be/MCYmV9sCREc`; it shows a typical user experience on the platform.

References

1. Corradini, F. et al.: BProVe: a formal verification framework for business process models. In: ASE. pp. 217–228 (2017)
2. Corradini, F. et al.: A formal approach to modeling and verification of business process collaborations. Sci. Comput. Program. **166**, 35–70 (2018)
3. Corradini, F. et al.: A guidelines framework for understandable BPMN models. Data & Knowledge Engineering **113**, 129–154 (2018)
4. Khlif, W., Makni, L., Zaaboub, N., Ben-Abdallah, H.: Quality metrics for business process modeling. In: ACS. pp. 195–200. WSEAS (2009)
5. Kluza, K., Nalepa, G., Lisiecki, J.: Square Complexity Metrics for Business Process Models. In: Advances in Business ICT, AISC, vol. 257, pp. 89–107. Springer (2014)
6. Lassen, K.B., van der Aalst, W.M.: Complexity metrics for Workflow nets. Information and Software Technology **51**(3), 610–626 (2009)
7. Mendling, J., Reijers, H.A., Cardoso, J.: What makes process models understandable? In: BPM, LNCS, vol. 4714, pp. 48–63. Springer (2007)
8. Reijers, H.A., Vanderfeesten, I.T.: Cohesion and coupling metrics for workflow process design. In: BPM, pp. 290–305. Springer (2004)
9. Woelfle, M., Olliaro, P., Todd, M.H.: Open science is a research accelerator. Nature Chemistry **3**(10), 745 (2011)