RePROSitory: a Repository platform for sharing business PROcess models and logS*

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

University of Camerino, School of Science and Technology, Computer Science Department, Via Madonna delle Carceri 7, 62032 Camerino, Italy
{flavio.corradini,fabrizio.fornari,andrea.polini,
barbara.re,francesco.tiezzi}@unicam.it

Abstract. The BPM community can certainly benefit from the adoption of open science principles. The availability of business process models and logs can make BPM research results more controllable, replicable, and comparable. Unfortunately, finding suitable collections of models and logs is pretty difficult to validate research proposals in the BPM field. To address this issue, we have developed a web-based repository, named RePROSitory, for sharing business process models and logs making them accessible to the community. We have started to systematically populate the repository with a collection of business process models, selected from the literature, and business process logs from an Italian company. The experience of models and logs retrieval from RePROSitory is enhanced by using metrics and metadata that allow researchers to select from RePROSitory a set of models or logs that they judge more suitable for the experiments they want to run.

Keywords: Business Process Repository · Process Model · Process Log

1 Introduction

RePROSitory was born [3] with the spirit of fostering open science principles [7] inside the BPM community. These principles aim at improving the capability of checking, and possibly re-validating, the results of a research effort. Referring to research on business processes, this demands for a common benchmark of models and logs to conduct research, to validate methodologies and techniques, and to compare tools performance. In this respect, several attempts have been made by the community to provide collections of accessible BPMN models and XES logs.

Referring to process models collections, the most known are BPM Academic Initiative Model Collection[1] and Camunda BPMN for Research.[2] These collections 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

[1] https://bpmai.org

[2] https://github.com/camunda/bpmn-for-research

the past, we used them for validating our research work (e.g., the framework in [2,4,5]). However, no platform is provided for easing the fruition of such models and no possibility to extend them with contributions from the community is provided. Recently, the authors of [6] proposed a technique to query github repositories searching for BPMN models that may be used for experiments. By scouting a part of all github repositories, they found over 8 thousands models over which they conducted some experiments. However, due to licensing issues, those models could not be freely re-distributed, which means that for replicating their experiments one has to undergo the entire procedure of mining github for gathering the same models they run the experiments on.

Referring to business process logs, few repositories are available and they mainly refer to the work carried out by the Process Mining Group at the Eindhoven University of Technology.[3] The main collection of logs they refer includes over 40 logs released by the IEEE Task Force on Process Mining on the *data.4tu.nl* platform.[4] Between the hosted logs we must mention those related to the International Business Process Intelligence Challenge,[5] a competition that provides participants with a real-life event log and, by applying any possible available technique, challenges them to analyze it and extract insights useful from a business perspective. With no doubt the *data.4tu.nl* platform is, and it has been, of great value for the research community; a simple query on Google Scholar, at the time of writing, results in a total of 562 scientific contributions mentioning and referring to data-sets available on such a platform. Not only data-sets related to business processes are available on *data.4tu.nl*, but also data-sets related to Chemistry, Earth Sciences, Biology and many other subjects, making it a general purpose repository.

What we present with RePROSitory is a dedicated platform for the sharing of business process related material, such as models and logs, with the possibility of taking advantage of specific functionalities that allow querying and filtering models and logs based on metadata and metrics. These functionalities allow therefore to define shareable collections of models and logs. In addition, we provide the possibility to correlate models and logs, in such a way to be able to inspect models related to specific logs, and logs related to specific models.

The rest of the paper is organized as follows. Section 2 describes the RePROSitory platform's main features. Section 3 reports about the maturity of the platform and Section 4 concludes the paper and provides indications for future work.

2 RePROSitory Main Features

In the following, we provide an overview of the main RePROSitory's functionalities that easily enable the sharing and the fruition of business process models and logs.

[3] http://www.processmining.org/logs/start
[4] https://data.4tu.nl/repository/collection:event_logs
[5] https://icpmconference.org/2020/bpi-challenge/

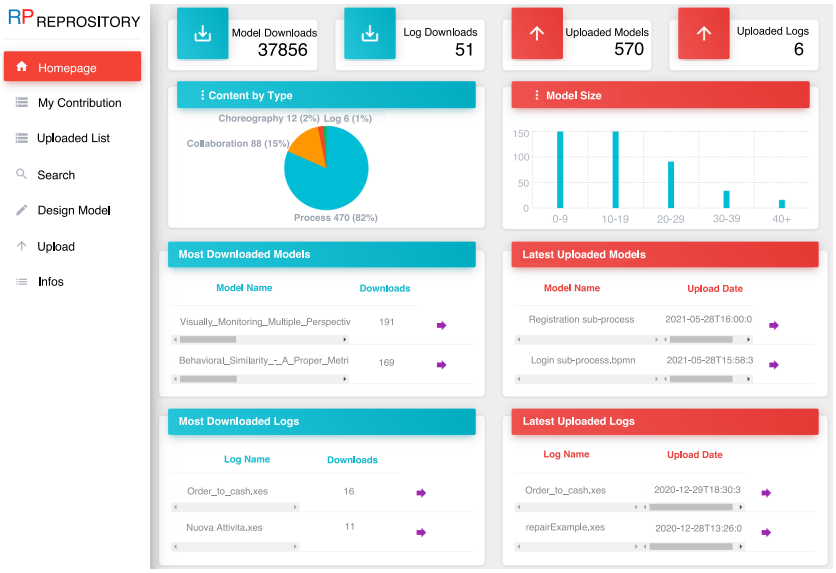

Fig. 1. RePROSitory homepage with content information.

RePROSitory's homepage (shown in Figure 1) provides a summary of the platform content (i.e., number of present models and logs, number of downloads, etc.) and allows user to access the platform functionalities by means of a sidebar. The platform provides two kinds of access: as guest and as registered user. A *guest* user can access functionalities such as: *Uploaded List* to see the lists of models and logs uploaded on the platform and eventually export them; *Search* to navigate models and logs; and *Info* to access the descriptions of metadata and supported metrics used to describe models and logs. A *registered user*, in addition to the guest's functionalities, can contribute to the platform by designing models and uploading models and logs, and by defining collections of models or logs shareable via URL addresses. The user can directly **Design a Model** by means of the integrated bpmn-js[6] library and decide to upload it, together with some metadata, directly on the RePROSitory platform. The user can choose to **Upload** models or logs together with related information (e.g., source, type, application domain). In the case of a *log*, when the log is uploaded a ***Log Metrics Extractor*** component is called. It computes the values for log metrics, which constitute the parameters a user can tune for filtering logs, and it shows the resulting values to the user. These results are also made available for the download in the form of a *.json* file. Up to now, the list of metrics for log includes: *Total Number of Traces*, *Average Week Duration*, *Median Week Duration*, *Start Date*, *End Date*, *Minimum Week Duration of a Trace* and *Maximum Week Duration of a Trace*. In the case of a *model*, two components are invoked when the model is uploaded: ***BPMN Metrics Extractor*** and ***BPMN Model Validator***. The former component computes values for business process model

[6] https://github.com/bpmn-io/bpmn-js

Metrics for Nuova RDA.xes

Metric Name	Value
Log Name	Nuova RDA.xes
Total Number of Traces	154
Number of Events	2522
Log Start	Mon Jun 05 16:57:00
Log End	Tue Feb 27 14:49:00
Min Week Duration	0
Max Week Duration	214.9
Avg Week Duration	2.9
Median Week Duration	0

Log's Information for Nuova RDA.xes

ID	3_1592583009243_1126183410912322
Uploaded By	fabrizio.fornari@unicam.it
Uploaded Date	2020-06-19T16:10:09.000Z
Name	Nuova RDA.xes
Description	This log refers to a purchasing order process of an Italian company.
Language	Italian
Scope	Research
Originality	New Log
Format	XES
Application Domain	Order Management
Type	Process
Origin	Real Case

Fig. 2. Metrics and metadata for the "Nuova RDA" process log.

metrics, whose results are made available for the download in the form of a *.json* file; the latter component checks if the BPMN syntax has been properly used, thus ensuring that no violation of the BPMN standard is present. The result of this syntactic check is stored into the database. Both results from the two components constitute, together with the information provided by the user, the parameters that can be tuned for filtering models.

It is worth noticing that with the **Search** functionality, RePROSitory provides three different ways of filtering models or logs: by metadata, by metric values, and by a combination of both metadata and metric values. ***Filtering by metadata*** allows the user to apply a filter based on information, such as *id*, *source*, *name*, *year*, *type*, *application domain*, etc. ***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 and a search is requested, the models or logs that satisfy the parameters are returned. The user can then inspect, download or remove models or logs from the resulting list. Upon pressing the download button, the user is able to download a *.zip* archive containing the selected material and the extracted metrics. A registered user can also define collections of models or logs from the result of a search operation and make them accessible via URL address.

The platform is accessible at `http://pros.unicam.it/reprosit ory` together with a detailed *User Guide* explaining how to use RePROSitory.A screencast is available on `http://pros.unicam.it/video/reprosit ory/rp2`; it shows a typical user experience on the platform.

3 Maturity of the Tool

Since the first release of RePROSitory (March 2019), the number of models shared on the platform has grown from 174 to 570. The number of registered users detected on July 2021 is 44; based on the users declared affiliation we estimated that over 80% of them are students and researchers of European Universities, the remaining 20% did not explicitly specified an affiliation.

In [3] we described the capability of RePROSitory of handling BPMN models, from the upload to the filtering and download of such models. Those functionalities are still supported, although we applied some enhancements. We modified the regulation for accessing the platform; we smoothed the possibility of uploading contribution by allowing to upload also syntactically invalid models, which are categorized by means of the boolean metadata *valid* (set to false) allowing to filter them. In addition, we introduced the possibility to manage process logs in the XES format, and to define a collection of models or logs in such a way to share them externally by means of a generated URL address.

For what concerns process logs, we started by distributing four real logs coming from an Italian consulting company. The log named *Help-Desk* reports the activities related to an help-desk process where a consultant provides support and indications to a client in a telematic way. The log named *Nuova RDA* reports about activities that an employee of the company performs for requesting the purchase of a new PC, a server, or a new hardware device to the company manager who is in charge of approving or denying the request based on the actual necessity and based on the budget. The log named *Nuova Attività* reports the process related to an employee of the company that needs to perform an internal activity (e.g., sending a fax, an e-mail, or performing an investment). The log named *Dismissione* reports the company internal activities that an employee performs for disposing of an old PC, a server or any other hardware device. By means of the new functionalities available on RePROSitory we have been able to upload such logs on the platform together with metadata describing the uploaded logs. The platform, by means of the *LogMetricsExtractor* component, automatically computes some metrics (e.g., *number of traces*, *time of start*, *time of end*, *min, max, average and median week duration of traces*) that can be used to filter logs stored on the platform. An example on how those metrics and metadata are displayed is reported in Figure 2. In addition, with the new possibility of creating collections of models or logs, we have been able to share the uploaded logs via a URL address.[7]

The possibility of uploading and sharing logs on RePROSitory enables new usage scenarios. In fact, the logs stored on RePROSitory can be for instance downloaded by a researcher who is conducting a study over BPMN process mining algorithms. The researcher can perform some tests and share on RePROSitory the resulting models documenting them appropriately filling the model upload form, linking them to the origin log, and making them available to the entire community. As an example, we applied the Split Mining algorithm (one of the many process mining algorithms [1]) and we uploaded and shared the resulting BPMN model on RePROSitory, by defining a link between the two of them. In this way, we are able to keep track of logs and related models, making it simple to navigate them. The generated model, together with a reference to the original log, are shown in Figure 3.

[7] https://pros.unicam.it:4200/guest/logcollection/fflogs062020

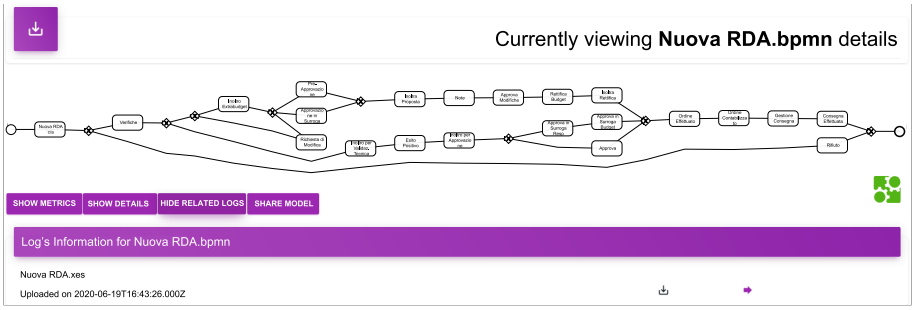

Fig. 3. Model obtained by applying the Split Miner algorithm over the *Nuova RDA* log, shared on RePROSitory.

4 Conclusion and Future Work

Business Process models and logs are not easily accessible. This hinders the possibility to validate and compare research approaches extensively. We developed RePROSitory, a platform for sharing and retrieving business process models and logs to overcome this issue. RePROSitory is on continuous development. We are working to improve the platform's usability and add new functionalities, especially related to process log visualization. We are also planning to extend the set of available models conducting harvesting procedures from the literature and conducting BPM projects with companies to derive and share real-life models and logs.

References

1. Augusto, A., Conforti, R., Dumas, M., Rosa, M.L., Maggi, F.M., Marrella, A., Mecella, M., Soo, A.: Automated discovery of process models from event logs: Review and benchmark. IEEE TKDE **31**(4), 686–705 (2019)
2. Corradini, F., Fornari, F., Polini, A., Re, B., Tiezzi, F.: A formal approach to modeling and verification of business process collaborations. Sci. Comput. Program. **166**, 35–70 (2018)
3. Corradini, F., Fornari, F., Polini, A., Re, B., Tiezzi, F.: RePROSitory: a Repository Platform for Sharing Business PROcess modelS. In: In BPM 2019 (Demos). CEUR Workshop Proceedings, vol. 2420, pp. 149–153 (2019)
4. Corradini, F., Fornari, F., Polini, A., Re, B., Tiezzi, F., Vandin, A.: BProVe: a formal verification framework for business process models. In: ASE. pp. 217–228 (2017)
5. Corradini, F., Fornari, F., Polini, A., Re, B., Tiezzi, F., Vandin, A.: A formal approach for the analysis of bpmn collaboration models. JSS **180**, 111007 (2021)
6. Heinze, T.S., Stefanko, V., Amme, W.: Mining bpmn processes on github for tool validation and development. In: Enterprise, Business-Process and Information Systems Modeling, pp. 193–208. Springer (2020)
7. Woelfle, M., Olliaro, P., Todd, M.H.: Open science is a research accelerator. Nature Chemistry **3**(10), 745 (2011)