A Template for Categorizing Business Processes in Empirical Research

Daniel Lübke[1,2], Ana Ivanchikj[3], Cesare Pautasso[3]

[1] innoQ Schweiz GmbH, Switzerland
[2] FG Software Engineering, Leibniz Universität Hannover, Germany
[3] Software Institute, Faculty of Informatics, USI Lugano, Switzerland

Abstract. Empirical Research is becoming increasingly important for understanding the practical uses and problems with business processes technology in the field. However, no standardization on how to report observations and findings exists. This sometimes leads to research outcomes which report partial or incomplete data and make published results of replicated studies on different data sets hard to compare. In order to help the research community improve reporting on business process models and collections and their characteristics, this paper defines a modular template with the aim of reports' standardization, which could also facilitate the creation of shared business process repositories to foster further empirical research in the future. The template has been positively evaluated by representatives from both BPM research and industry. The survey feedback has been incorporated in the template. We have applied the template to describe a real-world executable WS-BPEL process collection, measured from a static and dynamic perspective.

Keywords: Empirical Research · Meta-data Template · Business Process · Business Process Description · Business Process Metrics

1 Introduction

Empirical Research in the field of Business Process Management follows the increasingly wide adoption of Business Process Modeling practices and Business Process Execution technologies [9,17]. The validation of theoretical research, the transfer between academia and industry, and the quest for new research perspectives are all supported by empirical research, e.g., experiments, case studies, and surveys.

The goal of empirical research is to find repeatable results, i.e., observations that can be replicated thus providing results that can be combined and built upon. The more data points are available, the higher the significance of a study. One way to increase the number of data points is to perform meta-studies that combine results from multiple researchers (e.g., [14]). While this is common in other disciplines, such as ecology or medicine, business process-related data is usually not published in a comparable nor reusable way.

Additionally, the access to industry data is often restricted due to confidentiality requirements. Thus, publication of data sets must be done in an aggregated and/or anonymized manner.

To improve the reporting of empirical research concerning business processes, we propose a template that can be used to characterize processes in terms of their meta-data and (if applicable) their static and dynamic properties, without revealing confidential details. For example, business process models are used for different modeling purposes such as discussion, analysis, simulation, or execution. Processes are modeled using different languages (e.g., BPMN, BPEL, EPC). Process models also vary in terms of their size and structural complexity, which can be determined depending on the actual modeling language used to represent them.

The goal of the proposed template is to a) give readers the opportunity to "get a feeling" of a process (collection), and b) allow researchers to build on top of existing research by ensuring the presence of meta-data with well-defined semantics. Since, to the best of our knowledge, no such classification exists, in this paper we make an initial top-down proposal, intended as a starting point for extending and refining the template together with the research community.

In order to improve the reporting of research related to business process model collections (e.g., [6,20] as a starting point), we propose a set of meta-data described in tabular form. The meta-data template can be extended with other tables. For such extensions we initially propose static metrics for BPEL processes and some dynamic metrics, although further extensions for other modeling languages are welcomed.

We validate the meta-data template by a survey gathering the feedback of academic and industry professionals. Additionally, we apply the template in an industry case study to describe a large process collection.

The remainder of this paper is structured as follows. In Section 2 we motivate the need for such template, which we describe in Section 3. Section 4 depicts how we validated the template with a survey and a case study. Section 5 presents related work before concluding in Section 6.

2 Motivation

Models describing business processes contain sensitive information, making it difficult for companies to reveal how they use standard languages and tools, and rendering it challenging for empirical researchers to further improve the state of the art. As one of our survey respondents emphasized, much of the "research stops at the toy example level."

It is possible to anonymize process models, thereby limiting the understandability of what the process does and hiding their purposes and sources. Anonymized processes retain their entire control and data flow structure (which would be available for static analysis) while loosing important meta-data (which would limit the types of analyses that can be performed).

For example, Hertis & Juric published a large study with a set of over 1'000 BPEL processes [8]. However, they state that they "were unable to classify the processes into application domains since plain BPEL processes do not contain required information." This shows that researchers had to be aware when collecting the processes that they also need to collect associated meta-data.

Thus, whether or not a complete or anonymized process model is present, it is necessary to accompany it with a given set of meta-data. The meta-data has to be carefully selected and placed in a template to ensure that readers and other researchers can get an overall understanding of the discussed processes. Such a template needs to support the following goals:

1. Help researchers to collect data about processes that is relevant to others;
2. Help researchers to publish meaningful results by knowing which properties of the business processes can be anonymized and which should not;
3. Help researchers to report the important properties of business processes in their publications, so that their audience has sufficient details to evaluate the quality of the reported research;
4. Foster empirical research about business processes so that a body of knowledge can be accumulated based upon multiple, comparable works;
5. Enable meta-studies that combine, aggregate and detect trends over existing and future empirical research about practical use of business processes.

3 Template

Business Process Models can be created in many languages and can serve many purposes. Thus, it makes sense to report only values that have been actually measured in the specific usage context and are related to the conducted research. The templates are defined in a tabular format with a key/value presentation in order to allow quick digestion and comparison of reports. We understand that research publications need to present their results in a compact form. When space does not allow to use the tabular format, the tabular templates can be published together with the data, e.g., in technical reports and research data repositories.

The template we propose is built in a modular fashion. It consists of a required meta-data template that describes general, technology-independent properties of the process. The meta-data part can be extended by standardized templates for reporting different properties that have been analyzed. Researchers should re-use existing templates as much as possible in order to provide results that can be compared to previous works.

For instance, in this paper, two additional templates for executable BPEL processes are presented. The list of static and dynamic metrics proposed in the additional templates is not exhaustive and can be extended depending on the research needs. BPEL was chosen for convenience, as the case study in Section 4.2 uses BPEL processes. Support for other languages can be easily defined in additional templates.

3.1 Meta-Data Template

Table 1. The Meta-Data Template for describing Business Process Models

Process Name	Name or Anonymous Identifier of the Process
Version	Process Version (if available)
Domain	Business Domain of the Process
Geography	Location of the processes
Time	Period of data collection
Boundaries	Cross-Organizational / Intra-Organizational / Within-Department
Relationship	Calls another / Is being called / No call / Event triggered
Scope	Business Scope: Core / Auxiliary, or Technical Scope
Process Model Purpose	Descriptive / Simulation / Execution
People Involvement	None / Partly / No Automation
Process Language	e.g., WS-BPEL 2.0 / EPC / BPMN1 / BPMN2 / ...
Execution Engine	Engine used for running the Process Model if the model is executable
Model Maturity	Illustrative / Reference / Prototypical / Reviewed / Productive / Retired

The meta-data template, as shown in Table 1, is the only required part. It is designed to be applicable to any process model regardless of the modeling language used. This template contains the basic information necessary to obtain general understanding about a process model and the most important properties that can be of interest to filter and classify such process model. Its content has been updated with the feedback received during the survey described in Section 4.1. Following is a more detailed description of the categories and the classes included in the table:

Process Name: The process name as used in the organization. If the real name cannot be published, this field can be anonymized by providing an ID that can be used to reference the process from the text;

Version: If available, the name can be augmented with process versioning meta-data;

Domain: The business domain which this process is taken from. Existing ontologies like [7] can be used;

Geography: The geographical location where the process is used;

Time: The time period the process data refers to;

Boundaries: The organizational scope of the process: *cross-organizational* for processes that span across multiple legal entities, *intra-organizational* for processes that are conducted within one legal entity but across different departments/units in it and *within-department* for processes that are narrowed to a single organizational unit within one legal entity;

Relationship: The structural dependencies of the process with other processes: *calls another*, *is being called*, *no call*, *event triggered*;

Scope: The process model can have a horizontal, business scope, or a technical scope. In the business scope we can distinguish between: *End-to-end* processes for fully end-to-end descriptions like order-to-cash, and *auxiliary* processes for processes that do not contribute directly to the business purpose. Processes can have a pure technical scope instead, e.g., an event handling process that propagates permissions in the infrastructure;

Process Model Purpose: The purpose of a process model can be description, simulation or execution. A *descriptive* process is a model from a business point of view, which is more abstract in order to facilitate discussion and analysis among stakeholders, and also to prescribe how operations are carried out in an organization; a *simulative* process contains further details regarding resources, costs, duration, frequency, etc., while an *executable* process contains sufficient details to enable the automation of the process. Because a model can serve multiple purposes, this field is a list. The main purpose should be the first item in this list;

People Involvement: Classification of how much manual/human work is to be done. Ranges from *none* (fully automated) over *partly* to *no automation* (people involvement in each task);

Process Language: The process language used to create the process model. If a standard process language, such as BPEL, BPMN, etc., has been extended that should be specified in the meta-data;

Execution Engine: The execution engine(s) used to run the process model (if executable), including the exact version, if available;

Model Maturity: *Illustrative* for models which are not intended for industry use but to showcase certain modelling situations for educational purposes, *reference* for generic models which prescribe best practices and are used as starting point for creating other types of models, *prototypical* for models that are under discussion or are technical prototypes, *reviewed* for models that have been reviewed but are not yet in productive use, *productive* for models that are used productively in a real-world organization, with or without systems to enact them automatically, and *retired* for models which had been productive previously but have been replaced with other models.

The meta-data template is the main template that describes process characteristics regardless of the context and used technologies. In order to report details, additional templates should be used which often need to be language specific. Within this paper we define additional templates that describe different viewpoints of business processes, especially for those modeled in executable WS-BPEL.

3.2 BPEL Element and Activity Count Template

One of the interesting properties of processes are the various "size" metrics, with "size" being defined by Mendling [13] as "often related to the number of

nodes N of the process model." Since every process language provides different ways to express nodes and arcs for defining the control-flow, such template must be process language-specific. Thus, in this paper we define the template for measuring the size of BPEL processes by using activity and element counts, since BPEL is used in the case study that is presented in Section 4.2.

The template for reporting BPEL Element Counts is shown in the case study in Table 3. The values are merely the counts of different BPEL constructs as defined by the WS-BPEL 2.0 standard [10]. In addition, the total count of basic activities and structured activities is given because these are often used to judge the size of a process model. In the literature they are also called Number of Activities (NOA) and Number of Activities Complex (NOAC) [5]. In addition to activities, this table also contains the number of links, number of different sub-activity constructs (e.g., pick branches, if branches), and the number of partner links (service partners). To distinguish between the different BPEL constructs, basic activities are marked with a (B) and structured activities are marked with an (S) in Table 3.

3.3 BPEL Extensions Template

Although BPEL is a standardized language, it offers support for extensions. These extension points are used to extend the BPEL standard, e.g., the standardized extension BPEL4People to support human tasks, or to enable vendors to offer unique features that distinguish their products from their competitors'. BPEL defines a general facility to register extensions globally and the extension activity that can contain activities that are not defined in the core standard, or to use additional query and expression languages that are referenced by a non-standard URI. In contrast to [15] we think that the use of extensions is common. Also the case study has shown a high use of both vendor-specific and standardized extensions.

When reporting on BPEL processes, researchers can use the template as shown in the case study in table 4 that contains all declared extensions in the BPEL process and the extension activities used together with their activity counts.

3.4 Process Runtime Performance Template

For executable processes, it becomes possible to report their runtime performance. While a large number of metrics have been proposed (e.g., [18]), for space reasons, in this paper we propose to focus on reporting the number of process instances and their duration. These metrics can be described for each process instance or aggregated among multiple instances.

Counting the total number of process instances for a given process model gives an idea of its usage frequency relative to other process models.

Capturing the performance of individual process instances amounts to measuring their execution time ($T(finish) - T(start)$). Since the execution time of

every process instance is usually not of interest, we suggest to give statistical information about the distribution of the process instance duration for all process instances of a given process model as shown in Table 5.

4 Validation

To validate the usefulness of the proposed templates we combine an exploratory survey with researchers and industry experts (Section 4.1) and a case study of real-world BPEL business processes (Section 4.2).

4.1 Survey with Researchers and Industry Experts

To validate whether the proposed template fulfills the goals presented in Section 2 we have conducted an exploratory survey [19, Chap. 2][4]. The intention of this survey was not statistical inference of the results, but rather getting a deeper understanding of the surveyed field. We targeted audience from both academia and industry, i.e., both producers and consumers of empirical research. Thus, we used different social media channels and private connections to disseminate the survey.

Survey Design We organized the questions in five sections: Background, Meta-Data Template, Template Remarks, Template Extensions and Empirical Research in BPM. While the Background questions were mandatory to enable further classification in the analysis of the results, the remaining questions were optional to incentivize greater survey participation. In the Meta-Data Template section we showed the meta-data presented in Table 1 and asked the respondents to rate the importance of each of the proposed meta-data classes. In the Template Remarks section we focused on the perceived need of standardized reporting and asked suggestions for the appropriateness and completeness of the proposed process classification and meta-data. In the Template Extensions section we inquired about the relevance of reporting structure and performance metrics on process level, as well as on the usefulness of using the meta-data and metrics for describing entire collections of process models. Last but not least, in the Empirical Research in BPM section we asked for personal opinions on the state of the empirical BPM research.

Survey Sample Since we were not aiming at inferring statistical conclusions from the conducted survey, we closed the survey as soon as we considered the obtained feedback sufficient for improving the proposed templates. This has resulted with 24 respondents with diversified background. To obtain more insights into respondents' professional background, they could select multiple options between experience in academia (further divided into IT or Business Process Management), and in industry (further divided into IT or Business). While most

[4] The questionnaire is available at `http://benchflow.inf.usi.ch/bpm-forum-2017/`

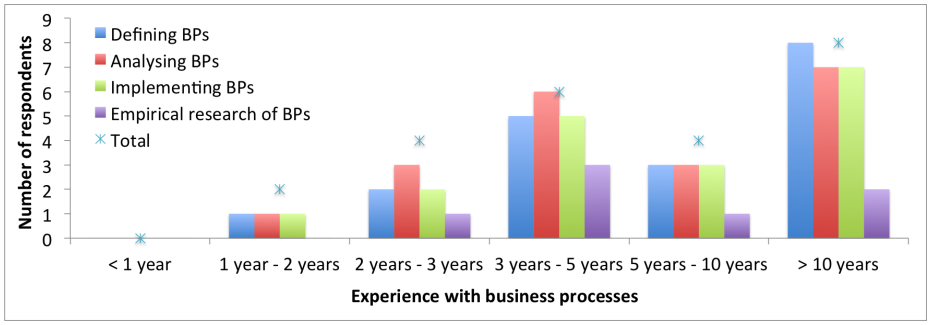

Fig. 1. Survey respondents: years of experience vs. business process areas expertise

of the respondents, i.e., 46% have experience only in academia, 21% have experience only in industry and 33% in both academia and industry. Most of them, i.e., 88% have IT background (16 respondents in academia and 12 in industry) and 63% have been dealing with the business perspective of process management (12 respondents in academia and 3 in industry).

Respondents participate in different phases of the business process life-cycle, and/or simply conduct empirical research on BPM. When asked what type of experience they have with business processes, the majority, i.e., 83% marked analyzing, while 79% marked defining, 75% implementing and 29% researching. These results could already indicate some lack of empirical research in this area.

All the respondents have more than one year of experience in working with business processes with 50% having up to 5 years and other 33% over 10 years of experience. Figure 1 shows the years of experience vs. the business process life-cycle experience of the survey participants. It is noticeable that people with longer experience have been more exposed to different phases of the business process life-cycle.

Survey Results We have presented the meta-data and process classifications as shown in Section 3.1 to the respondents, which in addition included the Modeling Tool category that we removed from the updated table as per respondents' feedback. We asked them to evaluate each proposed category on a scale from 1 (not important) to 5 (very important). As per the average score the Process Model Purpose is considered the most important with 4.38 points to be followed by People Involvement with 4.13 points. As mentioned previously, the Modeling Tool was considered as the least valuable with 3.17 points together with the Execution Engine with 3.38 points. Indeed in an ideal world, where the standards are correctly implemented, these two categories would not add to the understanding of the process model. In Figure 2 we stratify the importance rating of each proposed category per sector (industry, academia or both). It is interesting to notice that, even if those having experience only in industry allocate less importance to the meta-data on average, similar importance trends are evident between the different sectors. If stratified per years of experience, the highest

ratings are provided by respondents with 1 to 2 years of experience to be followed by those with over 10 years of experience.

Encouraging ratings were also obtained on the helpfulness of the standardized reporting approach for "getting a feeling" about the studied process (4.08 points on average) and for comparing different empirical reports (4.29 points on average). Based on the feedback on missing meta-data we have added the Version, Geography, Time, and Relationship categories to Table 1 as well as the Reference and Retired classes in the Model Maturity category.

In the next section of the survey we focused on the extended tables presented in Sections 3.2 and 3.4. Always on the same scale from 1 to 5, the respondents found the presentation of structure metrics and performance metrics sufficiently relevant, with average points of 3.40 and 3.57 respectively. We were curious to see whether priorities and interests change when using the meta-data and extended data presented in Section 3 on a collection of business processes. Thus, we asked respondents to rate them. While on process level, as mentioned earlier, Process Model Purpose and People Involvement were considered the most important, at collection level the Aggregated Structured Metrics (4.11 points) and the Domain (3.84 points) were considered the most important. As on process level, on collection level as well, the least important remained the Modeling Tool (3.11 points) and the Execution Engine (2.68 points).

As for the processes, also with the collections the responses followed similar trends among different sectoral experiences (academia, industry or both) evident from Figure 3, with industry always providing lower average scores than academia, while people with experience in both sectors tending to have opinions more aligned with academia. The greatest differences in opinions between industry and academia refer to the Model Maturity and Process Name where average academia's importance rating is around 4 while industry's importance rating is around 3 on process level and 2 on collection level. Significant differences in opinion are also noticed on collection level regarding the importance of the Structure Metrics which are rated at 2.5 by industry, 3.9 by academia and 4.9 by respondents with experience in both sectors. However, when aggregating among the importance rating of all proposed meta-data and extended data categories, the opinions are relatively positive with an average of 3.77 out of 5 points for data on process level and 3.53 out of 5 points for data on collection level.

We asked for additional properties that respondents would like to have in the template. Two recommendations, the connectedness of the model and a link to a process map, were made. However, connectedness is hard to define without requiring a special modeling language, while without standardized process maps, we think that the links are not helpful.

Last but not least, when asked whether they consider the existing empirical research in business process management (surveys, experiments, case studies) sufficient, out of the 16 respondents to this question only 4 answered positively.

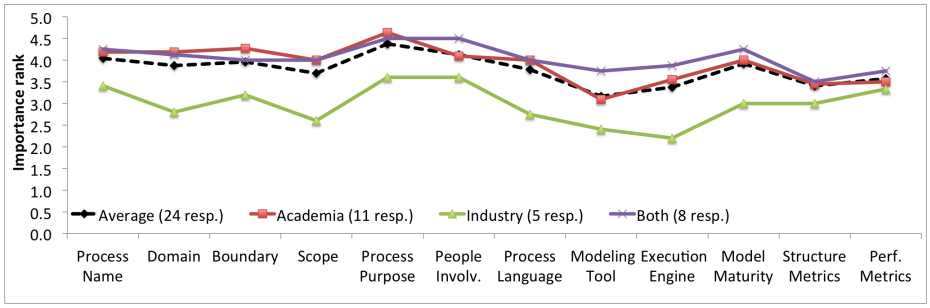

Fig. 2. Process Meta-Data Template validation (mean importance)

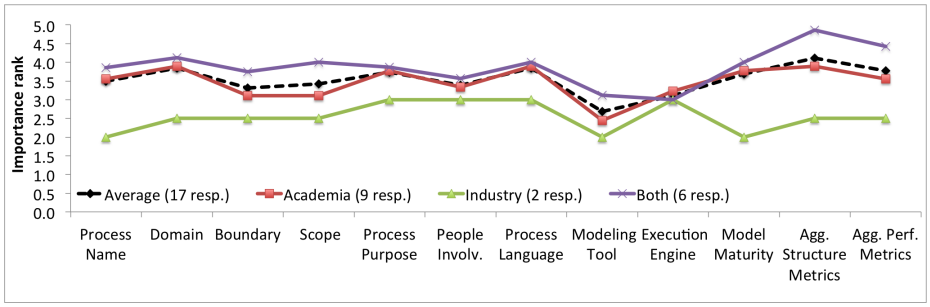

Fig. 3. Process Collection Meta-Data Template validation (mean importance)

4.2 Case Study with Industry Processes

We use the Terravis project as a case study for reporting process meta-data and metrics in a standardized fashion. Terravis [2] is a large-scale process integration project in Switzerland that coordinates between land registries, banks, notaries and other parties business processes concerning mortgages. In contrast to previous reportings of metrics [11], in this paper we apply our template and all additional templates as defined in this paper.

The Research Questions addressed by this case study are the following:

– Can the template be applied without problems? Especially are all category values clearly defined and applicable?

– Can all categories be measured? Which measurements can be automated?

– Is the categorization in the meta-data template beneficial when evaluating the process metrics?

The set contains 62 executable BPEL models that are executed on ActiveVOS 9.2. We could acquire a total of 918 versions of the process models and information for 435,093 process instances executed in Switzerland in the period between 2012 and 2016. To apply the templates we conducted the following steps:

1. For each process we assigned a value to each category of the general meta-data template, automating the assignment where possible;
2. Automatically measured the static metrics for the models;

Table 2. The Meta-Data Template for a Terravis Process

Process Name	Transfer Register Mortgage Certificate to Trustee
Version	26.0
Domain	Land Registry Transactions
Geography	Switzerland
Time	2016-08-30
Boundaries	Cross-Organizational
Relationship	Calls another / Is being called
Scope	Core
Process Model Purpose	Executable
People Involvement	None
Process Language	WS-BPEL 2.0 plus vendor-extensions
Execution Engine	Informatica ActiveVOS 9.2
Model Maturity	Productive

3. Validated the People Involvement assignment by cross-checking the value of the count of human activities in the static metrics;
4. Automatically collected the used BPEL extensions; and
5. Calculated the run-time metrics from the process logs.

In the first step we manually classified each process as per our meta-data template. In the People Involvement category we initially chose to offer more fine-grained values (partly, mostly). However, it was impossible to find a meaningful and objective threshold for these values. Thus, we opted to offer only one intermediate value, i.e., partly. To show-case the application of the meta-data template the meta-data of one process model is shown in Table 2.

Many static metrics, e.g., the static element counts [3,12] have been proposed and some tools have been developed for calculating them [1,8]. However, to our knowledge no working tool is freely available to calculate element counts and extract extension information from BPEL process models. Thus, we have built an open source implementation[5] to automatically calculate the data for the BPEL element and activity count template (Table 3).

To calculate the run-time metrics, the process logs were extracted and processed automatically. However, not all executable processes were configured with persistence and logging enabled. Thus, for some models we could not calculate any run-time metrics. Process instance run-time metrics are shown in Table 5.

After successfully applying the templates to all process models, an aggregation over the whole collection can be made. The results are shown in templated form in Table 6 with information on the percentage of models belonging to each class.

If the categorization in the meta-data template is meaningful, there should be no overlapping between classes in the same category and preferably each class should have some processes which pertain to it. We grouped the static metrics and process duration metrics of the latest version of every process model accord-

[5] Available at `https://github.com/dluebke/bpelstats`

Table 3. BPEL Element and Activity Counts for a Terravis Process

Transfer Register Mortgage Certificate to Trustee (Version 26.0)			
BPEL Element	**Count**	**BPEL Element**	**Count**
Assign (B)	79	OnAlarm (Pick)	0
Catch	4	OnAlarm (Handler)	0
CatchAll	2	OnMessage (Pick)	6
Compensate (B)	0	OnEvent (Handler)	0
Compensate Scope	0	Partner Link	15
Compensation Handler	0	Pick (S)	3
Else	13	Receive (B)	13
Else If	3	Repeat Until (S)	0
Empty (B)	42	Reply (B)	18
Exit (B)	9	Rethrow (B)	0
Extension Activity	1	Scope	74
Flow (S)	1	Sequence (S)	90
ForEach (S)	4	Throw (B)	0
If (S)	13	Validate (B)	0
Invoke (B)	37	Wait (B)	0
Link	2		
Derived Metrics:			
Basic Activities (B)	198	Structured Activities (S)	185

Table 4. BPEL Extensions for a Terravis Process

Extensions:	http://www.activebpel.org/2006/09/bpel/extension/activity http://www.activebpel.org/2009/06/bpel/extension/links http://www.activebpel.org/2006/09/bpel/extension/query_handling http://www.activebpel.org/2009/02/bpel/extension/ignorable http://www.omg.org/spec/BPMN/20100524/DI	
Activities:	**Type**	**Count**
	ActiveVOS Continue	1
	Total	1

ing to the different categories and their classes. The results are shown in Table 7. As can be seen, the distribution of the number of process models in the classes is different than the distribution of the number of activities. For example, only 37% of the process models describe cross-organizational processes but they contain 71% of the activities. This means that on average the cross-organizational models are larger than those in the different classes of the Boundaries category, and the within-system processes are the smallest on average. The distribution of the number of process instances and the distribution of the accumulated process duration among all executed process instances also differ. Only 14% of the process instances are cross-organizational but account for 68% of the overall process time spent. This means that cross-organizational and intra-organizational processes on average take longer to complete than within-system processes. Also technical process models have a very different distribution.

Table 5. Template for Capturing Run-time Performance Metrics of Process Instances

Transfer Register Mortgage Certificate to Trustee (Version 26.0)	
Number of Process Instances	13
Execution Time (min)	00h:00m:01s
Execution Time (med)	02h:33m:00s
Execution Time (mean)	12h:34m:39s
Execution Time (max)	64h:24m:14s
Execution Time (total)	163h:30m:32s

Table 6. Aggregated Meta-Data for the Terravis Process Collection

Collection Name	Terravis
Process Count	62 Models with 918 versions
Domain	Land Registry Transactions
Geography	Switzerland
Time	2012-03-09 – 2016-08-30
Boundaries	Cross-Organizational 37%, Intra-Organizational 13%, Within-System 50%
Relationship	Is being called 31%, Calls another 26% Is being called/Calls another 8%, Event triggered 24% No call 11%
Scope	Technical 52%, Core 39%, Auxiliary 10%
Process Model Purpose	Executable
People Involvement	None 79%, Partly 21%
Process Language	WS-BPEL 2.0 plus vendor-extensions
Execution Engine	Informatica ActiveVOS 9.2
Model Maturity	51 Productive, 11 Retired Models 51 Productive, 867 Retired Model Versions

The results support the classification categories because based on these values different characteristics of the processes in this collection are exhibited.

5 Related Work

The extensions to the meta-data template (Sections 3.2, 3.3 and 3.4) are language specific, and their aim is emphasizing the need of including structure and performance metrics, while not trying to be exhaustive in the list of metrics. Defining such metrics is out of the scope of this paper, and has already been addressed in existing work [13,4,5,18]. The main goal of this paper is standardizing the meta-data on process model and/or collection level. Thus, the related work we survey in this section refers to current availability and definition of such meta-data.

The need of extracting knowledge from business processes has been identified in literature and has led to the creation of business process repositories. Yan et al. [20] propose a Repository Management Model as a list of functionalities that

Table 7. Distribution of Terravis Process Models and Instances by Category

	#Model	#Activities	#Instances[a]	#Duration
Total	62	10'132	86'035	2'238'583 hours
Boundary				
Cross-Organizational	37%	71%	14%	68%
Intra-Organizational	13%	19%	8%	32%
Within-System	50%	10%	78%	0.1%
Relationship				
Is being called	31%	22%	19%	71%
Calls another	26%	55%	62%	9%
Is being called, Calls another	8%	12%	2%	20%
Event triggered	24%	3%	15%	0%
No call	11%	9%	2%	1%
Scope				
Technical	52%	10%	85%	0.2%
Core	39%	85%	13%	99%
Auxiliary	10%	5%	2%	1%
People Involvement				
None	79%	66%	86%	10%
Partly	21%	34%	14%	90%
Model Maturity				
Production	82%	84%	100%	96%
Retired	18%	16%	0.2%	4%

[a] Only for latest process model version

can be provided by such repositories and survey which of them are offered by existing repositories. Since what they propose is a framework, they emphasize the need of meta-data for indexing the processes, but do not define which meta-data should accompany each process. They have found that only 5 out of 16 repositories use a classification scheme based on part-whole and generalization-specialization relations. Vanhatalo et al. [16] built a repository for storing BPEL processes with the related meta-data, which in their usage scenario referred to the: number of activities, degree of concurrency, execution duration and correctness. Their flexible repository architecture could be used to store the templates proposed in our paper. The MIT Process Handbook project focuses on classifying the process activities and on knowledge sharing[6]. We focus on standardization of the reporting of such acquired knowledge.

The BPM Academic Initiative [6] is a popular process repository offering an open process analysis platform, aimed at fostering empirical research on multiple process collections. The meta-data required when importing processes refers to the process title, the collection it belongs to, the process file format and modelling language. Even though the data to be stored is not restricted only to these fields, no further standardization of the process classification is offered. In their survey on empirical research in BPM, Houy et al. [9] define a meta-perspective, a content-based and a methodological perspective for classifying the surveyed ar-

[6] `http://process.mit.edu/Info/Contents.asp`

ticles. Their content-based perspective refers to context (industry or public) and orientation (technological, organizational or inter-organizational). The standard meta-data we propose can offer a richer classification for meta-studies like [9,14] and more in-depth analysis performed using platforms like [6].

6 Conclusions and Future Work

Empirical research in BPM helps to close the feedback loop between theory and practice, enabling the shift from assumptions to facts and fostering real-world evaluation of so far untested theories. While the process mining research has benefited from the availability of large event log collections, the same cannot be claimed concerning process model collections [6]. As process models clearly represent trade secrets for the companies using them productively, in this paper we have proposed a language-independent template for describing them by focusing on key properties (classification meta-data, size & instance duration) which are useful for empirical analysis by the academic research community without revealing proprietary information. The template has been validated with an exploratory survey among 24 experts from industry and academia, who have positively commented on the choice of properties (no negative score was reported) and also made constructive suggestions that have already been incorporated in the template described in this paper. We have also demonstrated the applicability of the template in an industrial case study by using it to report on the Terravis collection of 62 BPEL processes and a subset of their 435,093 process instances executed across multiple Swiss financial and governmental institutions in the period between 2012 and 2016.

While the meta-data template presented in this paper is language independent, the extensions concerning static metrics are BPEL specific. Therefore, we plan to work on similar templates for other modeling languages in the future. Additionally, we plan to collaborate with modeling tool vendors to enable the automated collection of the meta-data described in this paper. The long-term plan is to grow the amount of available and well-classified process models to the empirical BPM community. One way to increase the number of classified processes is to auto-classify existing model collections. Future work will elaborate which properties can be inferred from existing data.

Most of the respondents of our survey said that there is not enough empirical research in the field of BPM. We hope that more empirical research will be conducted and that the meta-data presented in this paper will help researchers to improve the classifications of data collections and make them easier to compare and re-use across different publications.

Acknowledgments The authors would like to thank all of the participants in the survey for their time and valuable feedback.

References

1. Alemneh, E., et al.: A static analysis tool for BPEL source codes. International Journal of Computer Science and Mobile Computing 3(2), 659–665 (2014)
2. Berli, W., Lübke, D., Möckli, W.: Terravis – large scale business process integration between public and private partners. In: Plödereder, E., Grunske, L., Schneider, E., Ull, D. (eds.) Lecture Notes in Informatics (LNI). vol. P-232, pp. 1075–1090. Gesellschaft für Informatik e.V. (2014)
3. Cardoso, J.: Complexity Analysis of BPEL Web Processes. Software Process: Improvement and Practice Journal 12, 35–49 (2006)
4. Cardoso, J.: Business process control-flow complexity: Metric, evaluation, and validation. International Journal of Web Services Research (IJWSR) 5(2), 49–76 (2008)
5. Cardoso, J., Mendling, J., Neumann, G., Reijers, H.A.: A discourse on complexity of process models. In: International Conference on Business Process Management. pp. 117–128. Springer (2006)
6. Eid-Sabbagh, R.H., Kunze, M., Meyer, A., Weske, M.: A platform for research on process model collections. In: International Workshop on Business Process Modeling Notation. pp. 8–22. Springer (2012)
7. Executive Office of the President – Office of Management and Budget: North American Industry Classification System (2017), `http://census.gov/naics`
8. Hertis, M., Juric, M.B.: An Empirical Analysis of Business Process Execution Language Usage. IEEE Trans. Software Eng. 40(08), 738–757 (August 2014)
9. Houy, C., Fettke, P., Loos, P.: Empirical research in business process management-analysis of an emerging field of research. Business Process Management Journal 16(4), 619–661 (2010)
10. Jordan, D., et al.: Web Services Business Process Execution Language Version 2.0. OASIS (April 2007)
11. Lübke, D.: Using Metric Time Lines for Identifying Architecture Shortcomings in Process Execution Architectures. In: Software Architecture and Metrics (SAM), 2015 IEEE/ACM 2nd International Workshop on. pp. 55–58. IEEE (2015)
12. Mao, C.: Control and data complexity metrics for web service compositions. In: Proceedings of the 10th International Conference on Quality Software 2010 (2010)
13. Mendling, J.: Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness. Springer, 1 edn. (2008)
14. Mendling, J.: Empirical studies in process model verification. In: Transactions on petri nets and other models of concurrency II, pp. 208–224. Springer (2009)
15. Skouradaki, M., Roller, D., Pautasso, C., Leymann, F.: "bpelanon": Anonymizing bpel processes. In: ZEUS. pp. 1–7. Citeseer (2014)
16. Vanhatalo, J., Koehler, J., Leymann, F.: Repository for business processes and arbitrary associated metadata. In: Proc. of the Demo Session of the 4th International conference on Business Process Management (2006)
17. Weber, B., Mutschler, B., Reichert, M.: Investigating the effort of using business process management technology: Results from a controlled experiment. Science of Computer Programming 75(5), 292 – 310 (2010)
18. Wetzstein, B., Strauch, S., Leymann, F.: Measuring performance metrics of WS-BPEL service compositions. In: Proc. of ICNS. pp. 49–56 (2009)
19. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A.: Experimentation in Software Engineering. Springer (2012)
20. Yan, Z., Dijkman, R., Grefen, P.: Business process model repositories–framework and survey. Information and Software Technology 54(4), 380–395 (2012)