Discovery and Analysis of E-Government Business Processes with Process Mining: a case study

Andrea Delgado, Daniel Calegari
Instituto de Computación, Facultad de Ingeniería,
Universidad de la República, Uruguay
{adelgado, dcalegar}@fing.edu.uy

Abstract

One of the biggest challenges in performing a Process Mining (PM) initiative is the data availability since business processes (BPs) are usually implicit within the organization's information systems supporting them or by BPMS platforms with the process and organizational data distributed across heterogeneous databases within the organization. Moreover, in e-Government, inter-organizational collaborative business processes have traces of execution spread across several organizations. The main objective of this paper is to describe a case study on applying PM to e-Government business processes traced by an existing centralized traceability system, using our methodology for organizational data science. We provided a step-by-step analysis answering fundamental questions about their execution and evaluating improvement opportunities of the traceability system to strengthen PM initiatives.

1. Introduction

Business Process Management (BPM) [1] has gained importance within organizations for the explicit management and improvement of their business processes (BPs) according to their organizational needs. However, most BPs continue being implicit embedded in the information systems that support them. The application of mining techniques can provide organizations with the evidence-based information they need to improve their operations.

Process Mining (PM) [2] provide means for analyzing runtime events from information systems to discover corresponding BPs models (process discovery). Also, it allows verifying the compliance of the enacted BPs concerning the expected one by the organization (conformance checking) and analyzing key execution measures, e.g., bottlenecks, used resources, time duration, etc. There exist several tools, e.g., ProM [3] and Disco [4], which provide automated support to perform PM-based analysis of systems behavior. These well-known tools within the PM community provide the means to import event logs in XML format (XES) [5] containing the execution traces of the BPs and apply a vast collection of PM algorithms.

Even when there are BPM platforms in place to enact BPs, organizational data is not registered within the process engine database, and it is distributed across heterogeneous databases within the organization. When dealing with inter-organizational collaborative business processes, e.g., in an e-Government initiative, BPs execution traces can be spread across several organizations, making it more challenging to collect trace information to apply mining techniques.

Beyond getting the data, there must be an appropriate methodology for guiding the PM initiative. In [6, 7] we proposed the PRICED framework (Process and Data sCience for oRganIzational improvEment) for organizational data science. It is intended to help reduce the effort to identify and apply techniques, methodologies, and tools for organizational data science, i.e., from a traditional PM project to a more complex project requiring integrating process and organizational data. The classical data-centric analysis is most commonly guided by methodologies such as CRISP-DM [8], which does not include detailed guidelines on identifying and incorporating data that is useful to analyze organizations' processes and improve them. Other methodologies defined for process mining projects, e.g., PM^2 [9], does not provide specific guidance on the process and organizational data integration. A complete description of the methodology and related work can be seen in [7].

The main objective of this paper is to validate our previously defined methodology. For this, we performed a case study concerning e-Government BPs from our country, traced by an existing traceability system [10]. Since this is the first PM initiative within such an e-Government context, a second objective is to evaluate its feasibility, providing initial but valuable results from a non-traditional perspective. For this, we

URI: https://hdl.handle.net/10125/79629
978-0-9981331-5-7

observe the "as-is" situation, focusing on discovering process models that are not explicit today and answering fundamental questions about their execution. We also defined a third objective: evaluating the traceability data registered in the e-Government platform to assess improvement opportunities to strengthen PM initiatives.

The rest of the paper is structured as follows. In Section 2 we introduce the research methodology we followed. In Section 3 we describe the e-Government context in which we carried out the initiative. In Section 4 we present the case study. Finally, in Section 5 we provide conclusions and an outline of future work.

2. Research methodology

To guide our research work, we followed the Design Science methodology [11, 12]. Two main processes can be identified: building and assessment regarding the application of an artifact built to solve a given problem, and its usefulness is evaluated. An artifact can be a methodology, algorithm, a tool, among others. Design Science research tackles issues not yet solved or solved but using more effective or efficient approaches, contributing to the knowledge on fundamentals (theories, frameworks, constructions, models, methods, instantiations) and methodologies (data analysis techniques, formalisms, measures, validation criteria) [11].

Validation of artifacts can be carried out following different suitable approaches regarding the artifact under evaluation. Empirical methods, such as Case Studies or experiments [13, 14], can be used to assess artifacts' usefulness. For example, a methodology or tool within an organization, with human support on the assessment, i.e., carrying out surveys, interviews, or experiments, or technology-oriented, i.e., using benchmarks or theoretical validations over algorithms or tools. In this work, we carried out a case study integrating our previously defined methodology over e-Government BPs execution data from our country.

2.1. Case study design

To present the case study, we followed the guidelines and protocol definition from [13, 14]. We carried out a single-case study in a single organization and within a single project. The object of the study is the application of our methodology for integrated process and data mining in the context of an e-Government organization from our country and our BPM laboratory, and the unit of analysis the centralized BPs traceability system.

The main research question for the case study was defined as: *Do the activities, models, methods, process mining approach, techniques, and tools described within our methodology provide appropriate and valuable support for analyzing BPs execution as registered within the e-Government traceability system?*

We aimed to answer this question from two points of view: i) an evaluation of our methodology within a real context of e-Government BPs execution, and ii) to answer fundamental questions about BPs execution and the traceability system to business stakeholders.

2.1.1. Case selection, roles, and procedures As the case, we selected the BPs centralized traceability system of the e-Government organization and two BPs based on several characteristics they present. These BPs are representative of the BPs registered in the traceability system (c.f. Section 4). The main roles that participated in the case study were the organizational unit from the e-Government organization and a working team on BPM, including the two authors of this paper. The procedures for the case study were defined in the process and data mining methodology, as indicated within each phase of the lifecycle by following the defined activities (c.f. Section 2.2 and Section 4).

2.1.2. Data collection and analysis Data regarding the case study execution was recorded from its inception with the organizational unit of the e-Government organization, during the application of the methodology over the BPs dataset from the traceability system, and within the presentation of results to the organizational business unit where two domain experts participated. After the presentation, we asked them questions regarding the results obtained both of the approach and BPs analysis. Several internal documents were generated for the case study, including the clean dataset of the selected BPs, the corresponding process models discovered, presentations, and reports, mainly in Spanish internal to the research project. The dataset from the traceability system is not available due to confidentiality issues as defined in the project agreement by the University and the e-Government organization. The context of the study and other details regarding the e-Government traceability system are described in 3.

2.1.3. Threats to validity We identified threats to validity that could affect the case study:

Construct validity: we selected BPs from the centralized traceability system for the case study based on several characteristics that make them representative of the BPs carried out within e-Government settings.

Internal validity: the BPs data analyzed covered an extended period, and the research project itself was

started in April 2019. Authors' and business people's points of view can affect credibility and conclusions validity, which was mitigated by no previous knowledge on the selected BPs by the authors and no prior knowledge on the methodology by business people. Feedback from business people was asked to evaluate the methodology application and the BPs analysis.

External validity: the e-Government organization in which we carried out the case study presents common characteristics of e-Government settings: i) several organizations interacting within each other and with citizens, ii) deals with a significant number of BPs that are spread within different organizations, iii) supported by heterogeneous technological infrastructure in which process and organizational data are registered both in centralized and distributed databases.

Reliability: we clearly described the step-by-step execution of the case study for the application of the methodology and data collection for replication.

2.2. Process and data mining methodology

The PRICED framework [6, 7] provides guidance and support for organizational data science projects. Within its static view, the methodology for process and data mining defines five disciplines: **Process & Data Extraction and Integration (PDE)** dealing with the definition of goals and the extraction and integration of process and organizational data from associated sources; **Process & Data Quality (PDQ)** dealing with the selection, evaluation, and improvement (cleaning) of quality characteristics of process and organizational data; **Process & Data Preparation (PDP)** dealing with the preparation of the integrated data to be used as input for the mining/analysis effort; **Process & Data Mining and Analysis (PDMA)** dealing with selecting, executing, and evaluating approaches and tools for the mining/analysis effort; and **Process & Data Compliance (PDC)** dealing with selecting, specifying, and evaluating compliance requirements.

Within its dynamic view, the methodology defines four iterative phases: Enactment, Data, Mining/Analysis, and Improvement. The Enactment phase consists of the actual execution of processes and registration of process and organizational data, and the Improvement phase corresponds to the organization's improvement efforts after the analysis. In this context, our case study (described in Section 4) was focused on the Data and the Mining/Analysis phases. We customized the methodology application since we did not have organizational data to integrate to process traces. The traceability system does not provide it, and we did not consider compliance requirements.

3. e-Government scenario

This section introduces the e-Government context in which we carried out our case study, describing the e-Government platform and the traceability system for BPs. We obtained the actual data of procedures carried out within the e-Government organizations.

3.1. e-Government Context

The context of the e-Government approach in which our case study was carried out includes a centralized e-Government Platform for government organizations to interact with each other through it. Each organization can expose its services within the platform middleware, an Enterprise Service Bus (ESB), for other organizations to access. In this way, instead of directly invoking the services from one organization to another, all invocations go through the interoperability platform, providing access to the required service. Regarding the security component, tokens are provided for authentication and authorization of users to invoke services through the platform. The interoperability platform registers all interactions passing through it.

All BPs that interact within the interoperability platform are collaborative BPs. A subset of process tasks is carried out within each organization (i.e., orchestration), and organizations (participants) interact using messages. The platform only registers the interaction within organizations, i.e., the message flows defining the choreography of the collaborative BP, but the internal tasks carried out within each organization.

The traceability system's main objective is to provide information to the citizen regarding the execution of BPs where they are involved, i.e., which tasks were already executed and in which organization and the organizational unit is the BP currently running. It supports registering the BP tasks that each organization carries out to get such a view. The system is already integrated into some centralized components organizations use, and a traceability connector is also provided. Organizations need to invoke such connectors from their BPs execution to register each step of the collaborative BP.

The traceability system establishes key generic steps that are mandatory to trace within the system, such as when: the procedure is started, canceled, or finished, the citizen schedules a meeting within an organization or uses electronic signature (including success or failure), payment is issued through payment gateways, relevant approvals within the organization, notifications to the citizen or other participants, interactions between different organizational units within the same

organization, interactions within various organizations (which are also registered within the interoperability platform), when sub-processes are generated, or when the procedure is waiting for an action by the citizen or by third parties. In addition to the data currently loaded in the traces (transaction ID, timestamp, etc.), it is mandatory to register the name of the office executing the process activity, the description of such activity, and the process state at that point (Running, Canceled, or Finished).

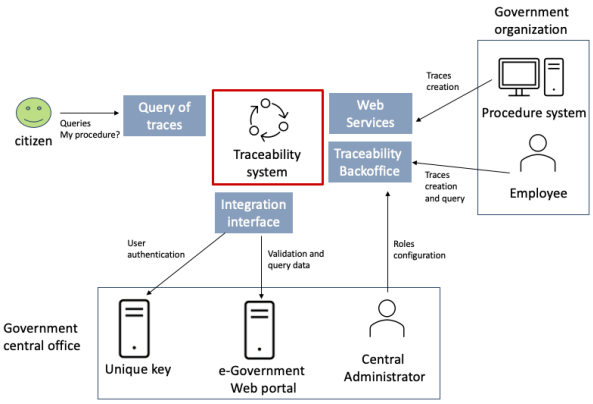


Figure 1: Traceability system interfaces

4. Case study

This section presents the case study we carried out following the methodology described in Section 2. BPs were performed within the e-Government context during the Enactment phase, and the traceability system registered their events. In what follows, we focus on the Data and Mining/Analysis phases.

4.1. Data phase

PDE1 – Select business processes. From the complete catalog of e-Government BPs, we selected the following ones considering a combination of interests: (i) a different number of cases and events within each case, (ii) diversity of event types defined for traceability such as payment, schedule a meeting, cancellation, evaluation, etc., (iii) the existence of previous reference models, (iv) the interest of the e-Government agency that sponsors the project.

- Small and Medium Enterprises (SMEs) Certificate: aims to accredit the SMEs status in front of any public or private institution.
- Passport Application: aims to prove the holder's identity and enables him to travel outside the country. We have previously analyzed its choreography model in [15].
- Permission for minors: aims to grant travel permission to minors with domicile or habitual residence in the country.

As we plan to analyze all BPs, this selection only provides an initial view of the traceability database. It allows us to discuss the preliminary findings with the e-Government agency and further apply the methodology considering the other processes.

PDE2 – Define mining/analysis goals. As mentioned in Section 2, we work together with business experts to define the mining/analysis goals. Since the project is intended to analyze both the traceability data and some concrete processes to observe the "as-is" situation, our mining/analysis goals were defined in two dimensions. First, we considered several e-Government agency guidelines for tracing processes and process patterns concerning the kind of process, e.g., benefit request or authorization request as the SMEs Certificate, among other types. Second, we considered general and specific business questions of interest for the business experts about the selected business processes, such as:

- (General) How are the cases actually being executed?
- (General) What is the most frequent path for every process model?
- (General) How is the distribution of all cases over the different paths through the process?
- (General) What is the average/minimum/maximum throughput time of cases?
- (SMEs Certificate) What percentage of requests are canceled?
- (Passport Application) In which period is there the highest number of requests?
- (Permission for minors) Are there paths for different situations on the request evaluation?

PDQ1 – Specify Data Quality Model. We selected quality characteristics from the Business Process and Organizational Data Quality Model (BPODQM) [7]. It defines specific dimensions, factors, and metrics such as the following ones of interest:

- Dimension: *Accuracy*, Factor: *Syntactic accuracy*, Metric: *Format*

- Dimension: *Completeness*, Factor: *Density*, Metric: *Not null*
- Dimension: *Uniqueness*, Factor: *Duplication-free*, Metrics: *Duplicate attribute/event*

We also considered common data issues [2] such as occurred events that were not captured by the traceability system or records that did not correspond to business activities.

PDE3 – Identify process and data sources. We accessed the traceability database with registers from April 2016 to December 2019. It consists of:

- 57 tables with size 20Gb
- 1.527 business processes
- 12.324.074 events
- 65 participants and 156 roles

The database contains business information and operational information, which is not of interest to the PM effort. Thus, we identified the tables that register the valuable information. The database contains several tables recording the static data of processes, i.e., process id and name, participants, and roles. It also includes several tables registering the traceability information, i.e., cases and their respective events.

PDE4 – ETL process data. We defined a SQL query that extracts the data we need for the selected processes, i.e., case id, activities, timestamps, and resources. Table 1 shows the main characteristics of the raw data we have extracted. The figure shows the number of traces and events for each process and the maximum and the average number of events within a trace.

PDP1 – Build event log. We generated a CSV file with the data extracted in the last step. We do not need to perform any additional transformation since the tools we use accept this input format.

PDQ2 – Evaluate quality characteristics. We checked some of the primary factors selected, such as date format, not null for timestamps, not null, and no duplicates for event names. We did not found registers with null timestamps. In contrast, we found many events with wrong dates (years 0006, 0007, etc.) when the data started in 2016. These cases correspond to initial tests of the traceability system. Within the *Permission for minors* log, we found activities with empty names. We also found in some processes activity names that have changed over time. We also discovered some character encoding problems on all activity names, probably due to different settings (i.e., UTF-8 and ISO-8859-1) when registering and reading data.

PDQ3 – Improve quality characteristics. We eliminated all null activity names and standardized the names that presented character encoding problems.

4.2. Mining/Analysis phase

PDMA1- Select mining/analysis approach. We use a traditional PM approach focused on process discovery to explicitly represent the e-Government BPs and reach the mining/analysis goals. We also perform some fundamental analysis based on previous experience with the *Passport Application* [15] processes.

PDMA2 – Select mining/analysis tools. We selected Disco[1] and ProM[2] to analyze the logs.

PDP3 – Filter event log and data We inspected the event log using the process mining tool Disco with an academic license. We analyzed the cases in the log, the variants identified by the tool, i.e., different paths over the control flow of the process execution. The activities within the log identify the most common ones for the start and end of the process. Since the event log included registers from not processes ended, we defined a general strategy for filtering those cases using the provided filter for selecting start and end activities. In the following, we present the analysis for each selected process:

SMEs Certificate In this process, we had previously corrected the encoding problems it presented regarding some activity names. It has 810 variants and 16 different activities after the correction of data.

We detected several variants where cases were not finished, i.e., only a few initial activities were present, e.g., in variant 1, only the two first activities, “Start procedure” and “Start procedure request” with 2.048 cases corresponding to 27,36% of the cases, in variant 3 also 545 cases present only this two activities and a third one “Load data”. We applied the filtering facility the tool provides to select the start and end activities to take into account, which was: “Start procedure”, “End procedure” and “Cancel procedure” which defines the two main paths for the execution of the process. Applying this filter, we obtained 2713 cases (36%) and 54.396 events, i.e., activities (69%) which correspond to

[1]Disco: https://fluxicon.com/disco/
[2]ProM: https://www.promtools.org/

Table 1: Raw data information

ID	Name	Owner	#traces	#events	max	avg
2559	SMEs Certificate	Ministry of Industry, Energy and Mining	7.485	78.577	94	11
2505	Passport Application	Ministry of Interior	94.239	521.558	100	6
2376	Permission for minors	Ministry of Interior	301.098	1.076.603	76	4

the same period, containing 16 different activities and 600 variants.

Passport Application In this case, before the filtering, we had eliminated the activities which had a null name, obtaining 94.239 cases (100%) and 428.314 events (82%) and 29 different activities (the original had 30 activities being one the null name). We carried out the selection of start and end events to filter the log as defined, but in this case, selecting as end the activities "End of: Passport request" and "End of: End Passport", which seem to be the two possible endings, leads to only 8 cases, so we decided to expand the selection.

We detected 11.192 cases (11%) with 144 variants that ended in the activity "Check payment gateway status", which could indicate problems with the registers. We decided to export these cases to be analyzed on its own. For the rest of the event log, we selected as end events all the ones including the word End in the name of the activity. These events were registered as: "End of: Request adjustment", "End of: Complete request", "End of: coordinate and pay audience", "End of: claim waiting", "End of: passport ending", "End of: request notification", "End of: manual reschedule", "End of: audience reschedule", "End of: request review", "End of: Passport request". Applying this filter leads to 16.562 cases (17%) and 100.507 events (19%) and 29 different activities, corresponding to cases ended in the period and 232 variants.

Permission for minors This case also presented null activity names and encoding problems, which we corrected. The process execution data were finally 301.098 cases (99%) with 1.041.064 events (96%) and 30 different activities. We found a variant with 143.923 cases which correspond to a 48% of the cases, with only the two initial activities registered: "Start procedure" and "Start of: Citizen task", which correspond to cases not finished.

We filtered the log using these two start events as before "Start procedure" and "Start of: Citizen task", and identified several possible end events for this process, which we selected to apply the corresponding filter: "Cancellation", "Ending", "End of: Permission for minors", "End of: Permission for minors v1", "End of: Permission for minors new", "End of: Citizen task", "End of: Employee task", "No attendance", "Procedure payment", "Reservation" and "Check payment gateway status". Although some of them are not proper endings, it seems that they are not running, so it would be of interest to the business to know what happened to them. In this process, we decided not to analyze them in a separate log. We also identified changes of activity names over time, such as "End of: Permission for minors" and "End of: Permission for minors v1" which correspond to different process versions. The event log from applying this filter has 132.888 cases (44%) with 701.864 events (65%), 29 different activities, and 1160 variants.

In table 2 we show the data analyzed for the three processes, including raw traces, events and steps, and the filtered traces, events, and steps which we used for the mining execution. In Figure 2 we present an example of the process of filtering the event log in Disco tool for the SMEs Certificate process. In a), the initial map model with all cases and events is shown, and in b) the resulting map model for the filtered log containing only complete cases. In a), we marked with a colored line the paths corresponding to incomplete cases, which were eliminated in b).

PDMA3- Execute mining/analysis approach. After filtering the event logs, as mentioned before, we analyze them from multiple perspectives, also using the tools Disco and ProM. In addition to analyzing the basic information provided by the tools about the cases, their variants, and the events they contain, we perform process discovery and replay the log to observe the different characteristics of the cases.

PDMA4- Evaluate mining/analysis results. We look to answer the general and specific questions for each process identified in the previous activity, "PDE2 – Define mining/analysis goals".

SMEs Certificate The filtered SMEs Certificate event log we analyzed consists of 2.713 cases with 600 variants, from which 2.527 (93,1%) were ended correctly with the "End procedure" event, and 189 (6,9%) were canceled, ending with the "Cancel procedure" event. To answer the specific question regarding the percentage of canceled requests, they

Table 2: Filtered logs

ID	Name	#traces raw	#traces filtered	#events raw	#events filtered	#steps raw	#steps filtered
2559	SMEs Certificate	7.485	2.713	78.577	54.396	23	16
2505	Passport Application	94.239	16.562	521.558	100.507	31	29
2376	Permission for Minors	301.098	132.888	1.076.603	701.864	36	29

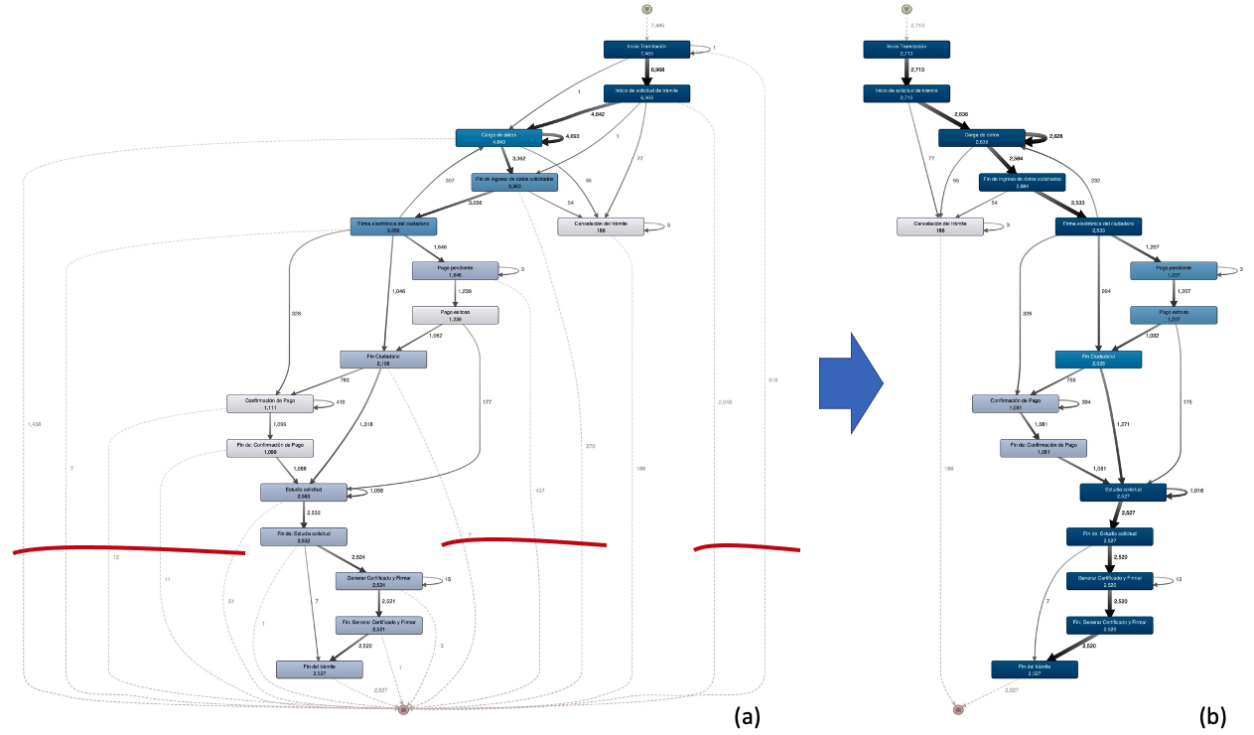


Figure 2: Filtering the SMEs Certificate process: a) all cases complete and incomplete b) only complete cases

represent only 2,5% of the total cases, which a priori does not indicate any problem. In Figure 3 the map model process corresponding to the filtered log of the SMEs Certificate process is presented.

It can be easily noticed that the right path of the model corresponds to the successful path that ends with the "End procedure" after executing all activities to carry out the process. On the left side, the paths ending with the activity "Cancel procedure" are shown, where it can be seen that there are variants with two activities and the cancellation, three, and four, and the cancellation, but no more paths. So it can be concluded that after executing the fourth activity ("End of data entry requested") with no cancellation and the fifth activity ("Citizen electronic signature") with no returning to activity "Load data", the process will successfully end. It is also important to notice that the successful path reaches the end of the process with 2.527 cases and the cancel path with 186, making a total of 2.713 cases which were the ones that initiated the process.

The execution of this process in the event log corresponds to the period 11/6/2018 to 16/12/2019, with a mean duration of 6.2 days. The minimum throughput time corresponds to a canceled case with 16 secs, and several cases corresponding to variant 6, which only includes the three activities "Start procedure", "Start procedure request", and "Cancel procedure" present short times of seconds or minutes. On the other hand, the maximum throughput time for the process is variant 285 with one year and 46 days of duration, starting on 30/08/2018 and ending on 16/10/2019 and including 94 events. It also presents several executions (i.e., a loop) of the activity "Load data", which is the activity that is executed most times in the event log with a relative frequency of 38,87%.

It should be analyzed within the organization in two

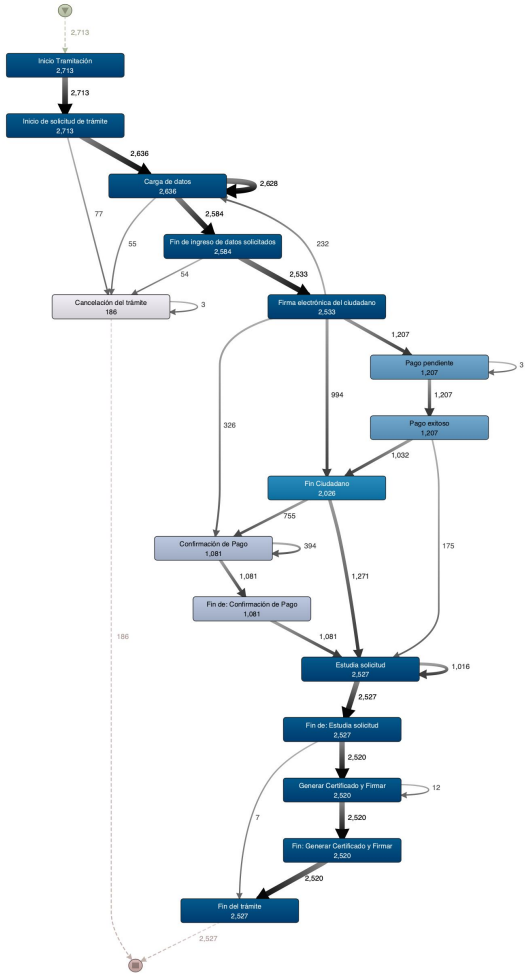


Figure 3: Model process for the filtered event log of SMEs Certificate process

dimensions: i) to detect why it is executed so many times over each case; is it the common path? Or maybe the input form that the citizen and employees have to fill is not completely clear? and ii) the data being loaded since it is hidden within the activity. Therefore the corresponding data is not registered in the traceability system but only in the organization's system.

Regarding the most frequent path for the process model, variants 1 to 6 account for near 50% of the cases, 1.035 cases, and the first 130 variants account for 80% of the behavior of the process. The first five variants (35.35%) are similar paths over the successful path on the right, including, among others, start procedure, load data, citizen electronic signature, payment confirmation, generation of the SME certificate and signature, and ending the procedure. Variant 6 (2,8%) corresponds to the cancel path with only the three first activities. Other variants ending with cancellation include more activities and loops overloading data.

We also analyzed the distribution of cases over time, looking for periods where SMEs Certificate requests could be concentrated due to specific context for the process. In the data we had, a peak of requests can be observed at the end of the year near December and around May, coinciding with the year's tax closing and the annual tax settlement, respectively. In Figure 4 we show a screenshot of the Active cases over time for the SMEs Certificate process in the Disco tool.

Passport Application The filtered Passport Application event log we analyzed consists of 16.562 cases with 232 variants. The execution of this process in the event log corresponds from 19/02/2018 to 16/12/2019, with a mean duration of 37,3 days.

Regarding the most frequent path for the process model, variants 1 and 2 accounts for near 90% of the cases. The most frequent variant (85% of the cases) is shown in Figure 5(a) and represents the online request, the coordination of an audience date, and the corresponding payment. The second most frequent variant (5% of the cases) only omits the payment since it could be done offline. Once the audience is confirmed, the citizen must attend the audience in person to process the request. These other steps are not registered within the traceability system but only in the organization's system. It is explained by the fact that the traceability initiative required implementing only the online start of the processes for 2020, leaving the complete tracing of processes for a later stage. The filtered event log has 29 different steps, but the most frequent variants show only 6 of them. The other steps are part of unusual behavior, such as reschedules and claims.

When analyzing the period in which there is a peak of requests, it can be seen that there are two clear periods: from mid-may 2018 to the end of October 2018, and from mid-January 2019 to mid-June 2019. The peaks occur a month and a half, two months before the summer and winter holidays, respectively. Figure 5(b) shows the tokens that are moving between events on June 7, 2019. It clearly shows the massive number of requests that start practically at the same time.

Concerning the 11.192 cases that ended in the "Check payment gateway status" activity, we confirmed that it represents that after scheduling the audience, the system requires an electronic payment with a deadline. If the citizen does not make the payment, the system automatically checks the status multiple times (a mean of 7-8 times) every 30 minutes. It is shown in Figure 5(c), in which can be seen the loop in the "Check payment gateway status" event.

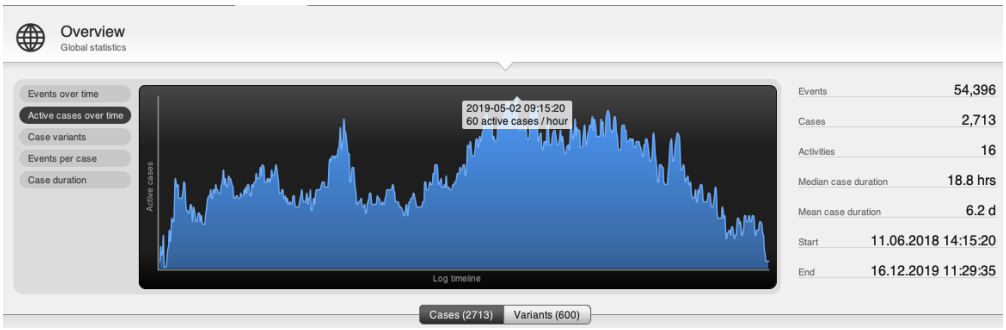
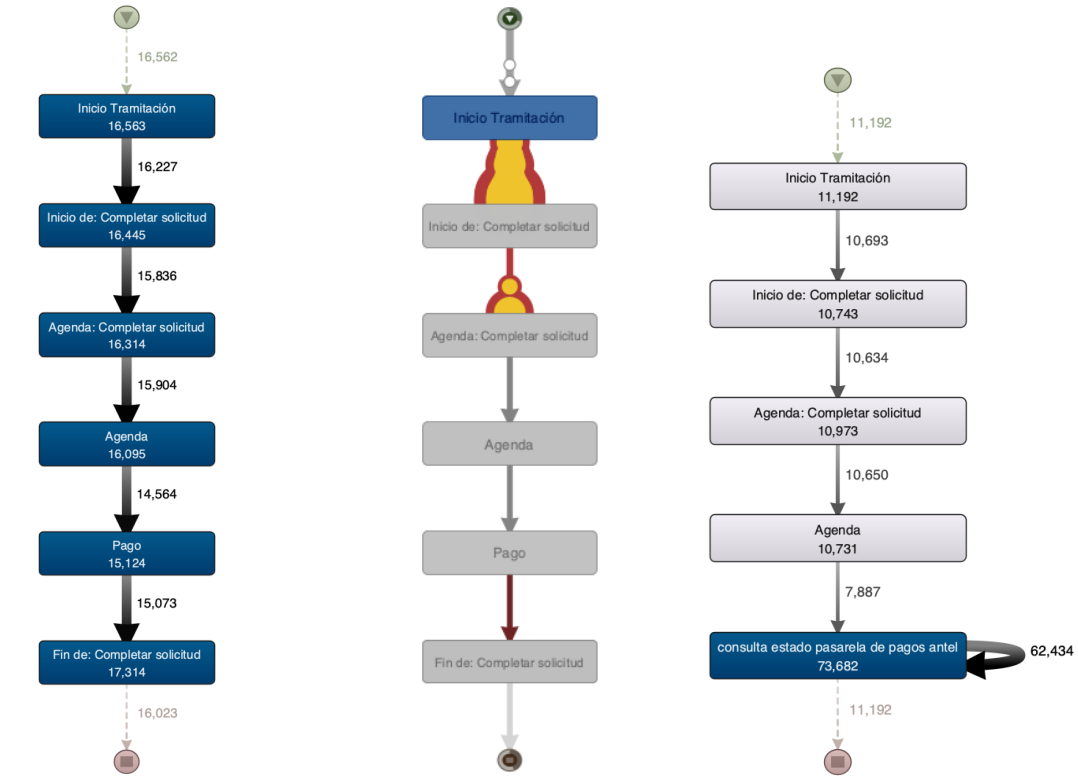


Figure 4: Active cases over time for SMEs Certificate process



(a) Most common variant (b) Token saturation (c) Check payment gateway version

Figure 5: Passport Application process

Permission for minors The Permission for minors process consists of 132.888 cases with 701.864 events, 29 activities, and 1161 variants. The process execution period goes from 23/02/2017 to 16/12/2019, so since the period is larger than the previous process expanding almost three years, several changes may have affected the registers. The mean duration of the process is of 44.4 days. It suffers the same problem as the Passport process since the first part was entirely online and registering activities before the rest of the process. Thus, the online agenda "Reservation" activity followed with the corresponding "Cancellation" or "Citizen attention" or "No assistance" and then "Ending" activity of the process are the most executed ones.

In this process we detected an exceptional path that occurs when it includes a Judicial audience, possibly due to problems between divorced parents or one parent not wanting the minor to travel with the other parent or relative. This exceptional path includes scheduling and having a judicial audience, issue minor certificates, controlling the permission delivery, and a new ending for the path named "End of:Permission for minors new".

5. Conclusions

This paper described a case study on applying PM to reference e-Government BPs to validate a previously defined methodology guiding organizational data science projects. We evidenced that the methodology can be used even in the context of partial requirements, e.g., when there is a classical PM project without integrating organizational data.

We also evaluated the feasibility of applying a PM initiative within the e-Government context. We selected three from more than 1.500 BPs. We worked with data from April 2016 to December 2019 traced by a centralized traceability system in operation. Our main findings allowed answering fundamental questions about such processes, evidencing the usefulness of the information registered by the traceability system.

This project also allowed us to identify improvement opportunities to strengthen PM initiatives since we also evaluated the traceability data registered in the e-Government platform. We have evidenced various problems with the recorded data during the project, both in form and content. In the first place, the quality of the recorded data is an aspect that should be studied in greater depth, mainly analyzed concerning the information that is currently registered, given that the traced processes have been increasing in number and quantity of traced events in the last years. Moreover, as part of the data quality, it seems convenient to force certain records that today are not strict, which generates, for example, null records.

Although there is a guide defining how to trace BP events, some records are not as expected. In events that mark the beginning or end of a task, in some cases appear both and in others only one, causing the logs to be inconsistent. An improvement could be to have a more specific criterion for the recording of events and precise terminology. For example, it could be possible to define a standard way of expressing the end or cancellation of a process, similar to the single task that initiates every BP within the traceability system.

The e-Government platform considers processes as orchestrations, being the citizen one role within. However, many of these processes are collaborative, where parts are carried out in different organizations. For example, the passport application BP involves getting a judicial background certificate, which is considered a standalone process. Given that the interoperability platform records all the interactions between organizations, future work links the traceability system with the interoperability information to analyze collaborative processes, extending the analysis.

Acknowledgements

Supported by project "Minería de procesos y datos para la mejora de procesos en las organizaciones" funded by Comisión Sectorial de Investigación Científica, Universidad de la República, Uruguay.

References

[1] M. Weske, *Business Process Management - Concepts, Languages, Architectures, 3rd Ed.* Springer, 2019.

[2] W. M. P. van der Aalst, *Process Mining - Data Science in Action, 2nd Ed.* Springer, 2016.

[3] B. F. van Dongen, A. K. A. de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters, and W. M. P. van der Aalst, "The ProM framework: A new era in process mining tool support," in *Applications and Theory of Petri Nets 2005*, pp. 444–454, Springer, 2005.

[4] C. W. Günther and A. Rozinat, "Disco: Discover your processes," in *Proc. of the Demonstration Track of the 10th Intl. Conf. on BPM (BPM 2012)*, vol. 940 of *CEUR*, pp. 40–44, CEUR-WS.org, 2012.

[5] IEEE, "IEEE standard for extensible event stream (XES) for achieving interoperability in event logs and event streams," *IEEE Std 1849-2016*, pp. 1–50, 2016.

[6] A. Delgado, A. Marotta, L. González, L. Tansini, and D. Calegari, "Towards a data science framework integrating process and data mining for organizational improvement," in *15th Intl. Conf. on Software Technologies, ICSOFT 2020*, pp. 492–500, ScitePress, 2020.

[7] A. Delgado, D. Calegari, A. Marotta, L. González, and L. Tansini, "A methodology for integrated process and data mining and analysis towards evidence-based process improvement," in *16th Intl. Conf. on Software Technologies, ICSOFT 2021*, pp. 426–437, SCITEPRESS, 2021.

[8] C. Shearer, "The CRISP-DM model: The new blueprint for data mining," *Journal of Data Warehousing*, vol. 5, no. 4, 2000.

[9] M. Eck, van, X. Lu, S. Leemans, and W. Aalst, van der, "PM2 : a process mining project methodology," in *Advanced Inf. Systems Engineering: 27th Intl. Conf., CAiSE 2015*, LNCS, pp. 297–313, Springer, 2015.

[10] AGESIC Uruguay, "Processes Traceability in the State." `https://bit.ly/3EwZVfH`, 2016.

[11] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, p. 75–105, 2004.

[12] R. J. Wieringa, *Design Science Methodology for Inf. Systems and Software Engineering*. Springer, 2014.

[13] R. K. Yin, *Case Study Research: Design and Methods, 5th edition*. SAGE Publications, Inc., 2014.

[14] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, and B. Regnell, *Experimentation in SW.Eng.* Springer, 2012.

[15] L. González and A. Delgado, "Towards compliance requirements modeling and evaluation of e-government inter-organizational collaborative business processes," in *54th Hawaii Intl. Conf. on System Sciences, HICSS 2021*, pp. 2079–2088, ScholarSpace, 2021.