Intra- and Inter-Organizational Process Mining: Discovering Processes within and between Organizations

Wil M.P. van der Aalst

Eindhoven University of Technology, P.O. Box 513,
NL-5600 MB, Eindhoven, The Netherlands
w.m.p.v.d.aalst@tue.nl,
vdaalst.com

Abstract. Due to the availability of more and more event data and mature process mining techniques, it has become possible to discover the actual processes within an organization. Process mining techniques use event logs to automatically construct process models that explain the behavior observed. Existing process models can be validated using conformance checking techniques. Moreover, the link between real-life events and model elements allows for the projection of additional information onto process models (e.g., showing bottlenecks and the flow of work within an organization). Although process mining has been mainly used within individual organizations, this new technology can also be applied in cross-organizational settings. In this paper, we identify such settings and highlight some of the challenges and opportunities. In particular, we show that cross-organizational processes can be partitioned along two orthogonal dimensions. This helps us to identify relevant process mining challenges involving multiple organizations.

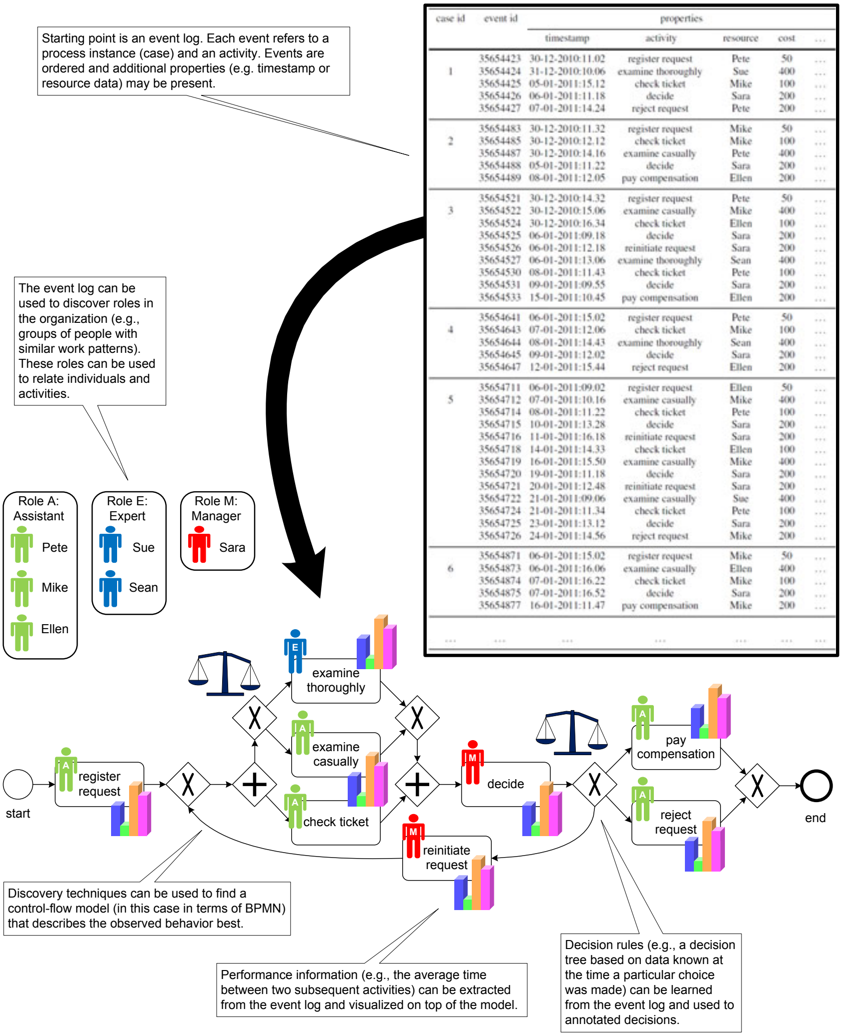
Keywords: process mining, cross-organizational mining, business process management.

1 Process Mining

We have applied process mining in over 100 organizations [4]. Our experiences show that process mining is a new and exiting technology that can be applied in a variety of domains (healthcare, governments, banking, insurance, education, retail, production, transportation, high-tech systems, etc.). However, lion's share of today's process mining projects are conducted within a single organization, whereas many processes are distributed over multiple organizations and different organizations are executing similar processes. Therefore, this paper aims to describe the various cross-organizational settings where process mining can be used. Before doing so, we provide a brief overview of the state-of-the-art in process mining.

Process mining provides a new means to improve processes in a variety of application domains. There are two main drivers for this new technology. On

P. Johannesson, J. Krogstie, and A.L. Opdahl (Eds.): PoEM 2011, LNBIP 92, pp. 1–11, 2011.



case id	event id	timestamp	activity	resource	cost	...
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
	35654427	07-01-2011:14.24	reject request	Pete	200	...
2	35654483	30-12-2010:11.32	register request	Mike	50	...
	35654485	30-12-2010:12.12	check ticket	Mike	100	...
	35654487	30-12-2010:14.16	examine casually	Pete	400	...
	35654488	05-01-2011:11.22	decide	Sara	200	...
	35654489	08-01-2011:12.05	pay compensation	Ellen	200	...
3	35654521	30-12-2010:14.32	register request	Pete	50	...
	35654522	30-12-2010:15.06	examine casually	Mike	400	...
	35654524	30-12-2010:16.34	check ticket	Ellen	100	...
	35654525	06-01-2011:09.18	decide	Sara	200	...
	35654526	06-01-2011:12.18	reinitiate request	Sara	200	...
	35654527	06-01-2011:13.06	examine thoroughly	Sean	400	...
	35654530	08-01-2011:11.43	check ticket	Pete	100	...
	35654531	09-01-2011:09.55	decide	Sara	200	...
	35654533	15-01-2011:10.45	pay compensation	Ellen	200	...
4	35654641	06-01-2011:15.02	register request	Pete	50	...
	35654643	07-01-2011:12.06	check ticket	Mike	100	...
	35654644	08-01-2011:14.43	examine thoroughly	Sean	400	...
	35654645	09-01-2011:12.02	decide	Sara	200	...
	35654647	12-01-2011:15.44	reject request	Ellen	200	...
5	35654711	06-01-2011:09.02	register request	Ellen	50	...
	35654712	07-01-2011:10.16	examine casually	Mike	400	...
	35654714	08-01-2011:11.22	check ticket	Pete	100	...
	35654715	10-01-2011:13.28	decide	Sara	200	...
	35654716	11-01-2011:16.18	reinitiate request	Sara	200	...
	35654718	14-01-2011:14.33	check ticket	Ellen	100	...
	35654719	16-01-2011:15.50	examine casually	Mike	400	...
	35654720	19-01-2011:11.18	decide	Sara	200	...
	35654721	20-01-2011:12.48	reinitiate request	Sara	200	...
	35654722	21-01-2011:09.06	examine casually	Sue	400	...
	35654724	21-01-2011:11.34	check ticket	Pete	100	...
	35654725	23-01-2011:13.12	decide	Sara	200	...
	35654726	24-01-2011:14.56	reject request	Mike	200	...
6	35654871	06-01-2011:15.02	register request	Mike	50	...
	35654873	06-01-2011:16.06	examine casually	Ellen	400	...
	35654874	07-01-2011:16.22	check ticket	Mike	100	...
	35654875	07-01-2011:16.52	decide	Sara	200	...
	35654877	16-01-2011:11.47	pay compensation	Mike	200	...
...	...	...	...	...	...	...

Fig. 1. Process mining techniques extract knowledge from event logs in order to discover, monitor and improve processes [4]

the one hand, more and more events are being recorded thus providing detailed information about the history of processes. On the other hand, there is a need to improve and support business processes in competitive and rapidly changing environments.

Process mining is a relative young research discipline that sits between computational intelligence and data mining on the one hand, and process modeling and analysis on the other hand. The idea of process mining is to *discover, monitor and improve real processes* (i.e., not assumed processes) *by extracting knowledge from event logs* readily available in today's systems (see Fig. 1). Note that process mining includes (automated) process discovery (extracting process models from an event log), conformance checking (monitoring deviations by comparing model and log), social network/organizational mining, automated construction of simulation models, case prediction, and history-based recommendations. Process mining provides an important bridge between data mining and business process modeling and analysis. Over the last decade, event data have become readily available and process mining techniques have matured. Moreover, process mining algorithms have been implemented in various academic and commercial systems. Today, there is an active group of researchers working on process mining and it has become one of the "hot topics" in Business Process Management (BPM) research. Moreover, there is a huge interest from industry in process mining. More and more software vendors started adding process mining functionality to their tools. Examples of software products with process mining capabilities are: ARIS Process Performance Manager, Enterprise Visualization Suite, Interstage BPME, OKT Process Mining suite, Process Discovery Focus, ProcessAnalyzer, ProM, Rbminer/Dbminer, Reflect|one, and Reflect.

Starting point for process mining is an *event log*. All process mining techniques assume that it is possible to *sequentially* record *events* such that each event refers to an *activity* (i.e., a well-defined step in the process) and is related to a particular *case* (i.e., a process instance). Event logs may store additional information about events. In fact, whenever possible, process mining techniques use extra information such as the *resource* (i.e., person or device) executing or initiating the activity, the *timestamp* of the event, or *data elements* recorded with the event (e.g., the size of an order).

Basically, there are three types of process mining. The first type of process mining is *discovery*. A discovery technique takes an event log and produces a model without using any a-priori information. Process discovery is the best-known process mining technique. For many organizations it is surprising that existing techniques are able to discover the real process based on the example executions in the event log. The second type of process mining is *conformance*. Here, an existing process model is compared with an event log of the same process. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa. The third type of process mining is *enhancement*. Here, the idea is to extend or improve an existing process model using information about the actual process recorded in some event log. Whereas conformance checking measures the alignment between model and reality, this third type of process mining aims at changing or extending the a-priori model. For instance, by using timestamps in the event log one can extend the model to show bottlenecks, service levels, throughput times, and frequencies.

Process mining may cover different perspectives. The *control-flow perspective* focuses on the control-flow, i.e., the ordering of activities. The goal of mining this perspective is to find a good characterization of all possible paths, e.g., expressed in terms of a Petri net or some other notation (e.g., EPCs, BPMN, and UML ADs). The *organizational perspective* focuses on information about resources hidden in the log, i.e., which actors (e.g., people, systems, roles, and departments) are involved and how are they related. The goal is to either structure the organization by classifying people in terms of roles and organizational units or to show the social network. The *case perspective* focuses on properties of cases. Obviously, a case can be characterized by its path in the process or by the originators working on it. However, cases can also be characterized by the values of the corresponding data elements. For example, if a case represents a replenishment order, it may be interesting to know the supplier or the number of products ordered. The *time perspective* is concerned with the timing and frequency of events. When events bear timestamps it is possible to discover bottlenecks, measure service levels, monitor the utilization of resources, and predict the remaining processing time of running cases.

Moreover, process mining can be used in *online* and *offline* settings. The results of process mining may be used to reason *about* processes (redesign) and to make decisions *inside* processes (operational support).

For a more comprehensive introduction to process mining, we refer to [4].

2 Intra- and Inter-Organizational Processes

Although most applications of process mining have been conducted inside a particular organization, there is no foundational reason why the technology cannot be applied across different organizations. Of course there may be issues related to confidentially, privacy, and data heterogeneity. In this paper we abstract from such problems and simply explore the possibilities of intra- and inter-organizational process mining. For this purpose, we consider two basic settings: (a) *collaboration* and (b) *exploiting commonality.*

In a collaborative setting, different organizations work together to handle process instances. A *process instance*, often referred to as *case*, corresponds to the "thing" that needs to be handled (e.g., a customer placing an order, a patient having a decease that needs to be treated, or a citizen applying for a building permit). The work associated to a case may be distributed over different organizations in a collaborative setting.

In the other basic setting (i.e., exploiting commonality) there are different organizations essentially doing the same thing. For example, there are 430 Dutch municipalities handing out building permits. Here the goal is not to distribute the work associated to a case as different organizations can do (more-or-less) the same thing. Organizations that have processes in common may be in competition, however, they can also learn from one another and share experiences and infrastructures. For example, Dutch municipalities are not competing with respect to handing out building permits. Although they may be competing for new citizens, they can still share a common IT infrastructure and share experiences to better (e.g. faster or more efficient) handle requests for building permits.

2.1 Collaboration: Distributing the Work among Different Organizations

First of all, we consider the collaborative setting where different organizations work together to handle process instances. This requires that the different parties are able to inter-operate, i.e., coordinate their activities. In [1], we identified five forms of *interoperability*. These are depicted in Fig. 2 and described next.

- The first form of interoperability is *capacity sharing*. This form of interoperability assumes centralized control, i.e., the routing of cases is under the control of a single organization. The execution of tasks is distributed, i.e., resources of different organizations may execute tasks.
- The second form of interoperability is *chained execution*: the process is split into a number of disjoint subprocesses which are executed by organizations in a sequential order. This form of interoperability requires that a partner transfers or initiates the flow for a case after completing all the work. In contrast to capacity sharing, the control of the workflow is distributed over the different organizations.
- The third form of routing is *subcontracting*. In this setting, one organization subcontracts subprocesses to other organizations. Consider for example Fig. 2(c) where two subprocesses are subcontracted. For the top-level organization the two subcontracted subprocesses appear to be atomic. For the two organizations executing subcontracted work, the subprocesses can be very complex. Note that the control is hierarchical, i.e., although there is a top-level actor, control is distributed in a tree-like fashion.
- The fourth form of interoperability is *case transfer*. Each organization has a copy of the same process description, i.e., the process specification is replicated. However, at any time, each case resides at exactly one location. Cases (i.e., process instances) can be transferred from one organization to another. A case can be transferred to balance the workload or because tasks are not implemented at all organizations. Note that in Fig. 2(d) it is essentially assumed that each of the organizations uses the same process definition (although some may implement only a part of it).
- The last form of interoperability is shown in Fig. 2(e): *loosely coupled*. For this form of interoperability the process is cut in pieces which may be active concurrently. Moreover, the definition of each of the subprocesses is local, i.e., the environment does not need to know the process. Only the protocol which is used to communicate is public for the other parties involved.

Note that chained execution and subcontracting can be seen as loosely coupled processes. One can think of such processes as “jigsaw puzzles”, i.e., the overall process is cut into parts that fit well together. Case transfer uses a different kind of partitioning: cases rather than process fragments are partitioned. Capacity sharing is the only form of interoperability which does not require some partitioning of the process and its instances. We will not consider this for of interoperability as conventional process mining techniques can be used.

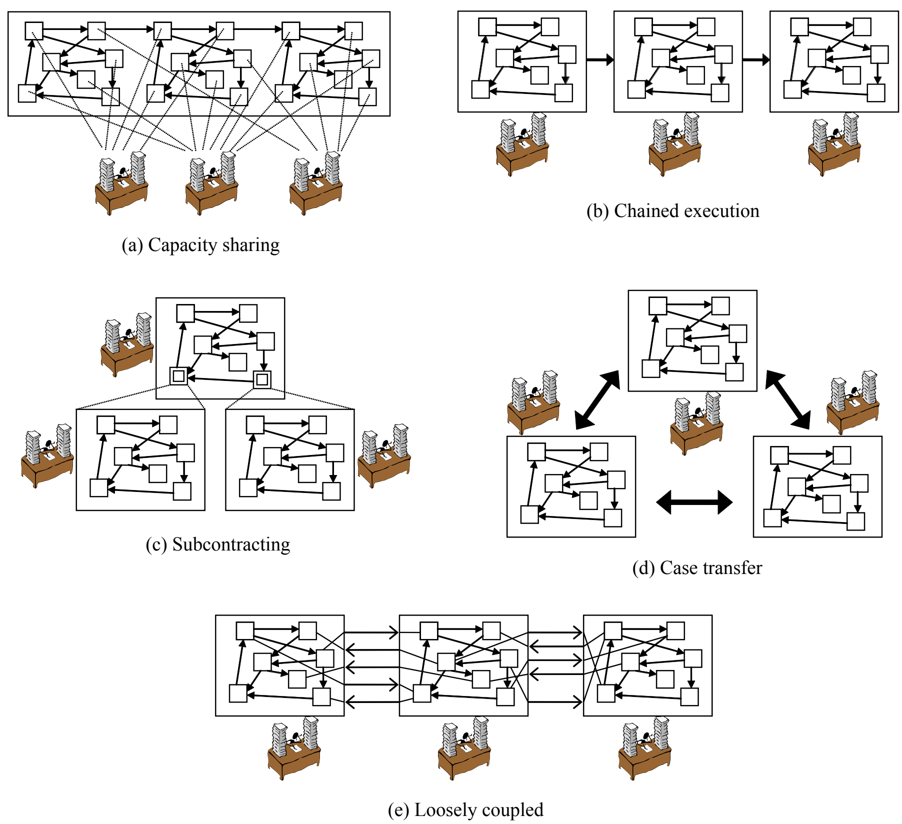


Fig. 2. Different ways of distributing work in a collaborative setting [1]

2.2 Exploiting Commonality: Sharing Knowledge and Infrastructures

As indicated earlier we consider two basic settings: (a) collaboration (cf. Fig. 2) and (b) exploiting commonality. Now we focus on the latter one. This type of cross-organizational processes does not involve interoperability, i.e., there is no explicit distribution of work. Instead, *organizations are executing essentially the same process while sharing experiences, knowledge or a common infrastructure.* To better understand such cross-organizational processes, we consider some examples taken from [2,3,7].

- There are about 430 *municipalities* in The Netherlands. In principle, they all execute variants of the same set of processes. For example, they all support processes related to building permits, such as the process handling applications for permits and the process for handling objections against such permits.
- *Suncorp* is the largest Australian insurance group. The Suncorp Group offers various types of insurance using brands such as Suncorp, AAMI, APIA, GIO,

Just Car, Bingle, Vero, etc. There are insurance processes related to different types of risks (home, motor, commercial, liability, etc.) and these processes exist for the different Suncorp brands. Hence, there are up to 30 different variants of the process of handling an insurance claim at Suncorp.
- *Hertz* is the largest car rental company in the world with more than 8,000 locations in 146 countries. All offices of Hertz need to support the same set of processes, e.g., how to process a reservation. However, there are subtle differences among the processes at different locations due to regional or national variations. For example, the law in one country or the culture in a particular region forces Hertz to customize the standard process for different locations.
- The sales processes of many organizations are managed and supported by *Salesforce*. On the one hand, these organizations share an infrastructure (processes, databases, etc.). On the other hand, they are not forced to follow a strict process model as the system can be configured to support variants of the same process.
- *Easychair* supports the review processes of many conferences. On the one hand, conferences share common functionality and processes. On the other hand, many variations are possible.

Organizations such as Suncorp and Hertz need to support many variants of the same process (*intra-organizational variation*). Different municipalities in a country need to offer the same set of services to their citizens, and, hence, need to manage similar collections of processes. However, due to demographics and political choices, municipalities are handling things differently. Sometimes these differences are unintentional; however, often these differences can be easily justified by the desired “Couleur Locale” (*inter-organizational variation*).

The cross-organizational processes mentioned above refer to a different type of cooperation than the different ways of distributing work depicted in Fig. 2. Organizations can learn from one another. For example, one municipality may improve its processes by learning from experiences of a better performing municipality. Moreover, if there a sufficient commonalities, organizations may want to share *configurable* processes and infrastructures [2,3,7].

2.3 Horizontal and Vertical Partitioning

After discussing the two basic forms of cross-organizational processes (collaboration and exploiting commonality), we conclude that there are two partitioning dimensions: the *case* dimension and the *process* dimension. *Vertical partitioning* uses the case dimension to partition work, i.e., the cases are distributed over several organizations but the process is not cut into pieces. *Horizontal partitioning* is based on the process dimension, i.e., the process is cut into pieces and organizations are responsible for specific parts of the jigsaw puzzle. The partitioning dimensions are in principle orthogonal but combinations are possible.

Chained execution, subcontracting, and loosely coupled, as described using Fig. 2, correspond to horizontal partitioning. Case transfer (Fig. 2(d)) and exploiting commonality (Section 2.2) correspond to vertical partitioning. Figure 3

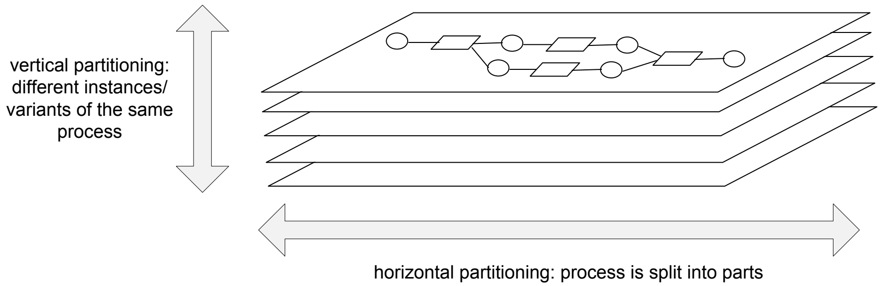


Fig. 3. Two partitioning dimensions: (a) horizontal partitioning and (b) vertical partitioning

illustrates these two partitioning dimensions. Traditionally, process mining has been focusing on processes that are not partitioned, i.e., all process instances belong to the same monolithic process. As will be shown in the reminder, the two dimensions shown in Fig. 3 can be used to structure the different process mining challenges.

3 Challenges for Process Mining

In Section 1 we introduced process mining as a new technology to analyze operational processes based on the footprints they leave in event logs. Subsequently, we provided a classification of intra- and inter-organizational processes in Section 2. Based on this we identified two main partitioning dimensions as shown in Fig. 3. These two partitioning dimensions serve as the basis for discussing various process mining challenges.

3.1 Horizontal Partitioning

A process that is partitioned horizontally can be seen as a jigsaw puzzle. Each "puzzle piece" corresponds to a fragment of the overall process and is under the control of one organization. In the classical setting (i.e., single process/organization) it is possible to capture all events relevant for a particular process, e.g., one can extract data from an SAP system clearly showing all steps taken in a particular process. However, when a process is partitioned horizontally one cannot assume this. An organization can only see some of the "puzzle pieces". It can see all events related to the puzzle pieces it is responsible for. Moreover, it can see the interactions with other puzzle pieces.

In a horizontally partitioned process there needs to be interaction between the different pieces. Typically, messages are exchanged between the puzzle pieces controlled by different organizations. Consider for example SOAP or EDI messages. SOAP (Simple Object Access Protocol) is a protocol specification for exchanging messages between web services. It uses a rather generic XML format which specifies a SOAP envelope consisting of a header and body.

Electronic Data Interchange (EDI) standards such as UN/EDIFACT (United Nations/Electronic Data Interchange For Administration, Commerce and Transport) impose more constraints on the messages being exchanged. The different fields in an EDI message have a predefined meaning. Note that messages can be exchanged in a synchronous or asynchronous manner. In some cases there may also be a party that is able to observe all message exchanges without being able to look "inside the puzzle pieces".

Hence, the main challenge is to conduct process mining *while only seeing a part of the overall process* [5]. Typical questions are:

- How to discover a process model when only seeing message exchanges and/or local events?
- How to check conformance when only seeing message exchanges and/or local events?
- How to identify bottlenecks when only seeing message exchanges and/or local events?
- How to correlate messages to process instances? When sending a message from one organization to another it needs to be routed to the appropriate instance of the process. This is a problem that is often underestimated and most researchers simply abstract from it [6].
- How to deal with many-to-many relationships across different organizations? One customer order may correspond to many order lines that may or may not be combined in different deliveries. Besides the problem of correlating messages there is the problem that one instance in one organization may refer to multiple instances in another organization and vice versa [4].

Since more and more processes are distributed over multiple organizations, it is important to address the above questions.

3.2 Vertical Partitioning

When a process is partitioned vertically, cases are distributed over several organizations each using their own variant of the process. These organizations may collaborate (see case transfer style of interoperability illustrated by Fig. 2(d)) or simply share knowledge and infrastructures. The metaphor of the jigsaw puzzle is not applicable anymore. A better metaphor is the "spot the difference game" children like to play (i.e., looking at two figures to find the differences between both). The bottom line is that there are different events logs all referring to some variant of the same process. The challenge is to *analyze commonalities and differences* between these processes based on the different event logs [2,3,7].

Besides a pair-wise comparison of logs and models, we can also use supervised learning to explain differences. For example, we can use classification techniques such as decision tree learning. For this purpose we need to label the data at the level of cases or at the level of event logs. Classification is based on a selected *response variable* and a set of *predictor variables*. For example, the response variable could be the (average) flow time or costs of a case or log. The fitness of an event log or case with respect to some reference model can also be taken as a response variable. Predictor variables are other properties of cases, event logs, or process models. For example, the complexity of the process model and the number or resources involved. Based on such information one can construct a decision tree that aims to *explain the response variable in terms of predictor variables.* This assists in understanding the essential differences between different organizations. For example, classification based on logs of different municipalities may reveal that (a) larger municipalities tend to have fewer deviations, (b) allowing for more concurrency results in shorter flow times but more deviations, and (c) a pre-check of building permits results in shorter flow times and a higher acceptance rate.

In [7], we provide some initial results obtained in the CoSeLoG project. In this project, 10 of the 430 Dutch municipalities are participating to investigate how process mining, configurable process models, and cloud technology can be used to reduce costs and improve service. All Dutch municipalities need to offer the same services to their citizens, and need to manage similar collections of processes. However, due to demographics and political choices, municipalities are handling things differently. The ten municipalities involved in CoSeLoG are eager to learn "proven best practices" from one another. This can be operationalized using cross-organizational process mining.

4 Conclusion

Although process mining is often applied within the boundaries of individual organizations, there are many process management questions that transcend the level of a single organization. Different organizations need to cooperate to realize a process, share an infrastructure, or may want to learn from one another. However, very few applications of process mining have been documented in literature. Therefore, this paper aims to structure the different cross-organizational settings in which process mining can be applied. Based on this, we highlighted some of the key questions. Currently, we are involved in several research projects that aim to address these questions:

– The EDImine project (`http://edimine.ec.tuwien.ac.at`) seeks to extend current process mining approaches in order to apply them to inter-organizational business processes while building on the additional information provided by traditional Electronic Data Interchange (EDI) standards. Advantages of using EDI technology are its widespread use and standardization of message content.
– The CoSeLoG project (`http://www.win.tue.nl/coselog/wiki/start`) focuses on a particular application domain: Dutch municipalities. Since all of these municipalities need to execute the same collection of processes, it is interesting to analyze differences and commonalities. The goal is to let these municipalities learn from one another and share a common (configurable) infrastructure.
– The ACSI project (`http://www.acsi-project.eu/`) uses artifact-centric modeling approaches to support service collaborations in open business networks. Process mining is used to understand such collaborations and to improve performance.

References

1. van der Aalst, W.M.P.: Process-oriented Architectures for Electronic Commerce and Interorganizational Workflow. Information Systems 24(8), 639–671 (2000)
2. van der Aalst, W.M.P.: Configurable Services in the Cloud: Supporting Variability While Enabling Cross-Organizational Process Mining. In: Meersman, R., Dillon, T., Herrero, P. (eds.) OTM 2010. LNCS, vol. 6426, pp. 8–25. Springer, Heidelberg (2010)
3. van der Aalst, W.M.P.: Business Process Configuration in The Cloud: How to Support and Analyze Multi-Tenant Processes? In: Zavattaro, G., Schreier, U. (eds.) Proceedings of the 9th IEEE European Conference on Web Services (ECOWS 2011). IEEE Computer Society Press, Los Alamitos (2011)
4. van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Berlin (2011)
5. van der Aalst, W.M.P., Dumas, M., Ouyang, C., Rozinat, A., Verbeek, H.M.W.: Conformance Checking of Service Behavior. ACM Transactions on Internet Technology 8(3), 29–59 (2008)
6. van der Aalst, W.M.P., Mooij, A.J., Stahl, C., Wolf, K.: Service Interaction: Patterns, Formalization, and Analysis. In: Bernardo, M., Padovani, L., Zavattaro, G. (eds.) SFM 2009. LNCS, vol. 5569, pp. 42–88. Springer, Heidelberg (2009)
7. Buijs, J.C.A.M., van Dongen, B.F., van der Aalst, W.M.P.: Towards Cross-Organizational Process Mining in Collections of Process Models and their Executions. In: International Workshop on Process Model Collections (PMC 2011). Springer, Berlin (2011)