
# Newsletter Stream #1 (02/2020)

Over the past **two decades**, Process Mining evolved from a novel, exploratory field to an established discipline across process and data science. A plethora of methods, techniques and tools flourished in all aspects of **process discovery**, **checking**, **enhancement**, **analytics**. The maturity of the area opened up new challenges that are yet to be solved. In 2019, it turned out that the time was right to begin a new independent conference series: the **International Conference on Process Mining (ICPM)**, the first successful edition of which took place in Aachen, Germany. This triggered the creation of a new steering committee, and the definition of the newsletter you are reading now. The goal is to publish four issues per year to report on the relevant updates from the ever-growing process mining community and the related conference series – first and foremost, ICPM.

**In this issue**, **Hajo Reijers** illustrates his story and personal view on process mining. We talk about the process mining company Lana Labs with one of the founders, **Rami-Habib Eid-Sabbagh**. **Marlon Dumas** describes the emerging field of predictive process analytics and his ERC-awarded endeavour. **Dominic Giss** illustrates his viewpoint on process mining from the end-user perspective. **Manuel Resinas**, **Adela del Río** and **Cristina Cabanillas** describe the upcoming Business Process Management conference (BPM 2020) and the role that process mining researchers, practitioners and enthusiasts play in it. The call for participation in the Process Discovery Challenge (PDC 2020) is out, so we enter the details of it with the help of a former contestant and, now, organiser: **Eric Verbeek**. Last but not least, **Massimiliano de Leoni** and **Alessandro Sperduti** give a preview of ICPM 2020, to be held in Padua, Italy.

New issues are upcoming. Your valuable input is welcome!

# Headlines at a glance

# Articles

## Academic stories: Hajo Reijers

*Narrated by Hajo Reijers*

**My name is Hajo Reijers**. I work at Utrecht University (UU), which is the largest and one of the oldest universities in the Netherlands. In Utrecht, I am the chair of the **Business Process Management & Analytics group**, a group that consists of some 10 professors, postdocs, and PhD students. I also still work at Eindhoven University of Technology (TU/e) for a day per week. Before moving to UU, I worked for a couple of years at the Vrije Universiteit in Amsterdam. My wife suspects that I am trying to tick off all Dutch universities from my bucket list. She is often right, but not in this case. I believe I will be in Utrecht for quite some time. I really like the university: It is broadly oriented, with many opportunities for interdisciplinary collaboration, and the people that I met here are good at what they do. What the students are like is yet to be seen: the first edition of my BPM course is only starting in these days. A definite bonus of my work at UU is the botanical garden right in front of our building, which is a good place for taking a walk. I have not tried out the squash courts of the university sports centre yet, but suspect I will do so soon.

**My first encounter with process mining** was when I started my PhD studies with Wil van der Aalst, which was in the late 1990s. I remember that I went to a **summer school on Petri nets**, in Jaca, Spain, very close to the Pyrenees. In the daytime, I learned more about Petri nets than I ever thought was there to be known about them. In the spare time we had, Wil would take me and other PhD students out on hikes. This was a fascinating experience for me. Coming from the Netherlands, which is an incredibly flat country, I think this was the first time I actually set foot in the mountains. While making these hikes, Wil would be uttering traces and simultaneously inferring process structures from those traces. I was much impressed with that intellectual act. I was even more impressed that Wil would still climb faster than we did while working on such puzzles. Still, I did not really see the point of this whole "**workflow mining**" at that time. In the years before I started my PhD, I worked as a consultant and my mind was tuned to practical IT problems for a while. So, I only got interested in the topic of process mining when I started seeing its practical value.

That became clear during **one of the lengthiest projects I have been involved in**, which lasted from 2001 to 2010. During that project, I was trying to determine the **organizational value** of using a **workflow management system** – we would call that a BPM system today. The idea is that such a system will take care of coordinating a business process, which saves time and potentially speeds up the process. To establish such advantages empirically, we followed a dozen organizations that were implementing this technology. I vividly remember a Dutch municipality where a business process would take more time to complete once it was supported by such a system, compared to the original situation. That did not make any sense to me. Only when we started looking into the event logs that were generated by the system, we could see that people were using it to transfer work endlessly to each other. Chris would start a case, but Esther would send it back to him after checking his work.

And not just once, but sometimes a dozen times, making us suspect that Esther did not really like Chris much! In any case, rerouting work turned out to be so much easier with a workflow management system in place. Never mind the consequences for the cycle time of the overall process…

**Since that time, process mining as a field has much evolved**. There are so many powerful techniques available and so many valuable applications of the technology. I think it is a wonderful and dynamic field, where it is still the case that solving intellectual challenges go hand in hand with generating practical impact. One of the topics that I personally find interesting is how event logs can be put together in situations where they are not readily available. With my former PhD student Eduardo Gonzalez, I worked on the **creation of event logs** based on **historic database transactions**. Recently, I also participated in a project with Andrés Jiménez, from Seville, where we would try to put an event log together **based on screen captures and keystrokes**. These are good examples of situations where unconventional data can be used to put event logs together. This, in turn, helps to apply process mining in new settings. My feeling is that there are still many situations that can be investigated.

Another direction that interests me is **to use process mining in different ways**. While it is currently strongly connected to the notion of studying a single business process, I feel that the whole approach can help us to understand much better the nature and evolution of human work. This is a shift of focus. We have already started applying process mining techniques to **detect patterns of work** that can be **easily automated**, for instance. In addition, with my colleagues in Utrecht, I am trying to use process mining to detect at what point people start working around the rules and procedures within an organization, what has triggered these **workarounds**, and how these practices spread within an organization. If we are going to use even more fine-grained data on how people interact with IT systems, we may start understanding what makes some work complex in comparison to other work. While the study of work is now mostly a topic for labor economists and human interaction engineers, I think that computer scientists will come up with much more profound insights.

**My group in Utrecht** is quite young, but we are determined to make a name for ourselves and create some noise. One way of doing so is to be very **welcoming** to visitors: It helps us to learn from our visitors and visitors get in touch with the things we do. So, for example, we hosted Marcelo Fantinato and Sara Peres, both from USPE, for a year. We made sure that they felt very sorry when their sabbatical ended! Considering the climatological differences between Brazil and the Netherlands, that says something. Another thing that we are putting efforts in is the organization of a **summer school**. People who are starting to work on a topic that relates to **process mining** can come over to Utrecht in the summer of 2020 for a week to give their work a good push. Faculty members from my own group, researchers from the group at TU/e, and experts from industry will all provide perspectives on the field of process mining. They will also help in a practical sense, during workshops, to let participants make progress with their work. People who find that appealing should definitely check out the set-up: https://utrechtsummerschool.nl/courses/engineering-and-technology/process-mining-research.

For those who like to read a bit more about some of the works I referred to, here are **some pointers**:

- The project that looked into the advantages of workflow management systems: https://doi.org/10.1016/j.ijinfomgt.2015.08.003
- The work of Eduardo Gonzales and others on event log creation on the basis of redo logs of databases: https://doi.org/10.1007/978-3-319-23063-4_25 and https://doi.org/10.1007/s10115-019-01430-6
- The paper by Andres Jimenez and others on using process mining to create event logs in shielded settings and establishing the potential of automation: https://doi.org/10.1007/978-3-030-21290-2_28
- Work by Iris Beerepoot and others on detecting workarounds: http://www.irisbeerepoot.com/publications/

You can also see what I have to say on Twitter by looking up my handle, @profBPM.

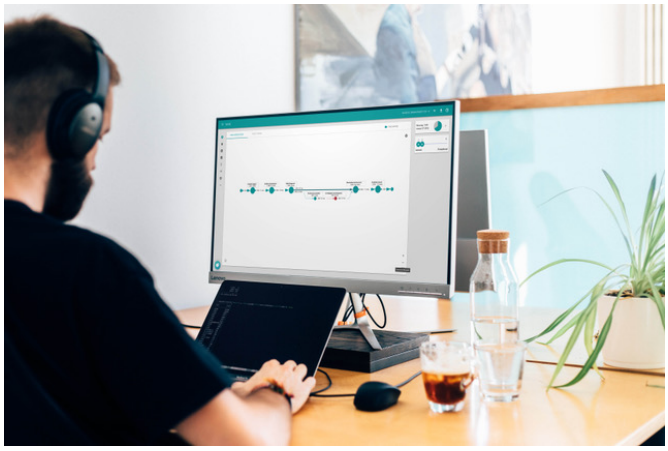
# Developers' point: Lana Labs

*Talking with Rami-Habib Eid-Sabbagh*

We have the pleasure to talk about the Process Mining Software company Lana Labs with one of the founders.

# Tell us a bit about you and Lana Labs!

Hi, my name is **Rami-Habib Eid-Sabbagh** and I'm one of the co-founders and managing directors of Lana Labs GmbH. Lana Labs is one of the leading providers for Process Mining Software and is headquartered in Berlin, Germany. My two co-founders **Karina**, **Thomas** and I founded Lana Labs in 2016 and aimed at providing an **entirely cloud-based process mining software from day one**. Today, we offer the most advanced process mining technology enabling global organizations across all sectors to continuously improve their value streams.



If you asked me to describe in a nutshell what LANA does, I would answer the following: We make companies better every day. LANA Process Mining allows organizations to **analyze and optimize business and production processes** of any complexity to sustainably improve business performance. With LANA Process Mining we developed unique capabilities to automatically analyze process data and **provide insights** to make companies smarter, more efficient and faster**.** The software helps organizations in any industry to get a clear picture of their process landscape, identify bottlenecks and highlight improvement potential.

We believe that process optimization has to be a continuous effort in any organization. Our team develops a process mining software that not only provides analysis for individual processes but generates a holistic view of the overall performance and optimization potential of the organization.

A big part of our Corporate Identity is our approach to **sustainability**. With our technology, we help organizations to increase efficiency, make better use of their resources and thus to get more sustainable and reduce their CO2 footprint. In addition, **we support a magellanic penguin colony in Chile** and the adherent research project by **adopting a penguin** for each license sold and each new employee onboarded. The company's penguin's name is "**LANA**".

### Where did the idea of building that process mining tool come up? Why did you start your endeavour in the first place?

Actually, if you dive deep into the history of Lana Labs, the foundation for the idea was already laid when my co-founder **Thomas** worked as Process Analyst at IBM and had to optimize a vast amount of business processes more than ten years ago. He soon found out that analyzing business processes with hundreds of different variants is not possible without the help of suitable software. He started to search for possibilities to automate process analysis and identified process mining as the technology to analyze complex business processes. Thomas later decided to write his **Master** at **TU Eindhoven**, Netherlands, on **Conformance Checking** which, by the way, later was one of the main process mining innovations that we brought into the process mining when we launched LANA Process Mining.

Now you are probably wondering **when Karina and I came into the game**. Let's get to that now :-)

Thomas and I first met at a BPM conference in **Clermont-Ferrand** in 2011. Back then I was doing my PhD studies at Hasso-Plattner-Institute (HPI) in Potsdam, Germany. After Prof. Jan Mendling, Thomas PhD supervisor changed to WU Vienna for his professorship, Thomas decided to join the business process technology group at the Hasso-Plattner-Institute. During many table soccer games we found out that we do not only share our **love for table soccer** but that we also shared the same **passion for process optimization** and for **founding our own company**.

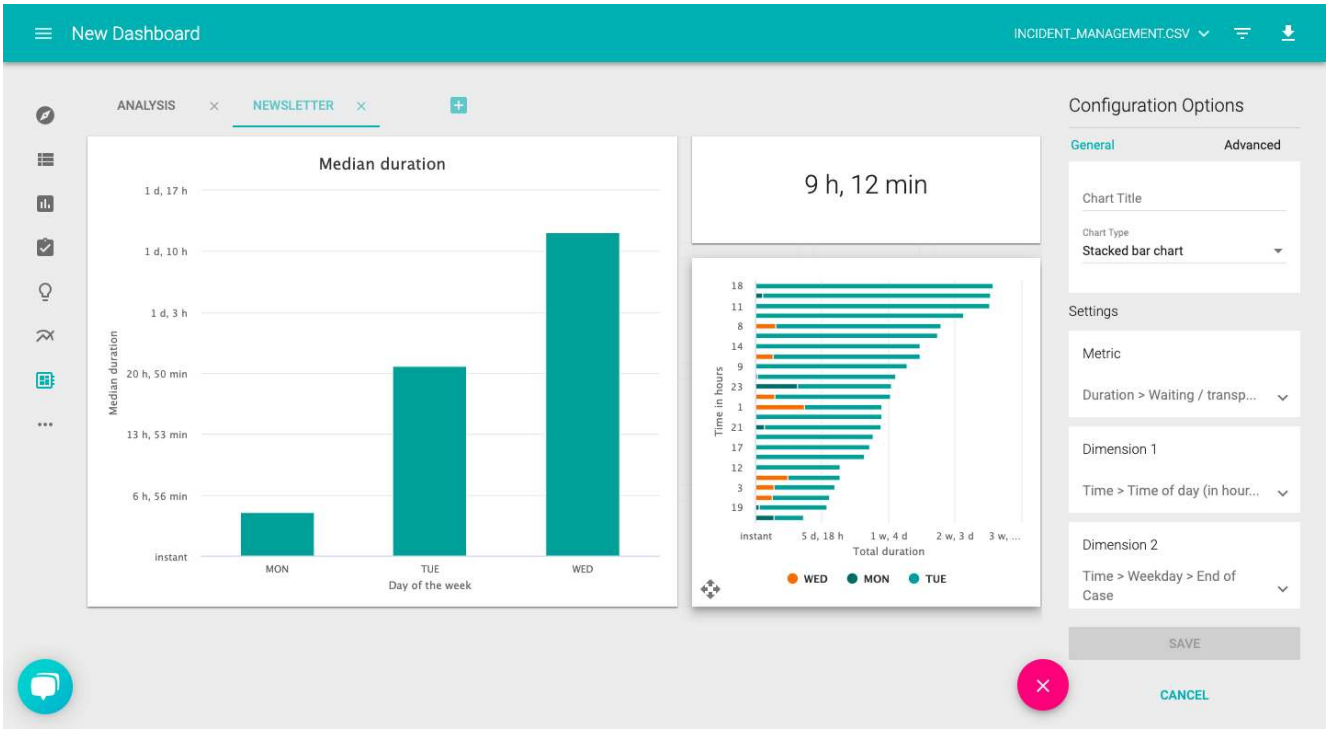
We both focused our PhD studies at the HPI on End-to-End process optimization and Process Mining. Of course, research is a big part when you are doing your PhD. Hence, both of us studied the existing process mining market and available process mining tools. We noticed that all existing process mining tools did not go beyond simple process discovery. Organizations using the available tools were only able to see the process flow of an existing process but not much more. Based on these findings **we developed the idea to create a new process mining software that goes beyond process mining** and enables organizations to automatically get insights on their business processes using **machine learning**.

During a workshop at the **Potsdam Entrepreneurship Experience Lab (PEEL)**, which was organized by HPI Design School and Potsdam Transfer, we met our third founder **Karina** who worked back then as Technology Scout at Inpro. Karina took part in the workshop to find new innovative technologies. She shared our **excitement** about our idea from the first minute and we soon found out that all three of us have a **mutual level of humour and fun** and that we were all **ready to kick off** something entirely new.

We then refined our idea together and decided to start Lana Labs to bring new innovations to the process mining market. We developed an innovative process mining software that enables organizations to automatically analyze and optimize complex processes and make the life of employees easier.

The first **innovations** in LANA Process Mining included for example **automated conformance checks** as well as **root-cause-analysis** which hadn't been available in any other tools before. We were the first ones to offer more than simple process discovery and LANA is today one of the leading process mining tools offering in-depth process analysis based on a high-performance technology stack. Coming in with a technical background it was important to us to ensure that LANA could be used by anyone with any other IT-system. For this reason, we developed the **open Lana REST API** which makes it easy for cloud developers and IoT developers to integrate and use the Lana API to connect LANA Process Mining to a worldwide ecosystem of other software and hardware providers and extend it with individual functionality.

Though we have grown the company to an international team of more than 40 process- and IT-experts, we are still working very closely with research institutions to stay on top of new innovations and offer our clients cutting edge capabilities even for complex process optimization projects. And of course, we also still get together regularly to continue our **tabletop soccer games** :-)




## LANA Process Mining: What is it about?

Today, there is massive amount of data available in most of the organizations. However, often the data is not used at all or only a small portion can be analyzed. Our software, LANA Process Mining, leverages all data available and provides the technical capabilities to automatically **analyze** this **data**, **identify patterns** and **recommend actions for improvement**.

We focused our development efforts from day one on building a process mining tool that enables a very granular analysis of business and production processes. The main concept behind LANA Process Mining is to automate and simplify the complex analysis and optimization of end-to-end processes so that anybody is able to use process mining.

Our software includes capabilities for **visualizing** business processes, automatic conformance checks and machine-learning-based root-cause analyses. In addition, we added integrated smart dashboards to provide a day-to-day management tool for complete KPI monitoring. The software can be used with any **common web-browser**, no installation required. However, we know that some clients prefer to have an **on-premises solution** instead of a SaaS subscription, hence we are offering both.

The tool is developed using technologies such as Scala, Elixir, and Angular being able to scale vertically as well as horizontally and easily cater to big data requirements. Through our open **REST-APIs** and our **LANA-R** and **LANA-Python** packages LANA Process Mining can be easily extended with additional functionality like prediction, simulation and cater to the individual needs of customers and consultancies as well as academia.

As our aim always was to create a software that does more than just process discovery we put a lot of effort into new, innovative capabilities. Hence, LANA Process Mining for example includes automatic target-actual analysis of process flows as well as waiting and execution times. Our integrated automatic root-cause analysis as well as further automated analyses insights give organizations all relevant information to make better informed decisions.

With our **upcoming** capabilities regarding **automated recommendation** and **predictions**, we will further extend the usage of LANA Process Mining. We are very proud that we were able to develop a complex machine learning algorithm that makes LANA Process Mining unique on the market and provides our clients valuable insights into previously unknown process flows. We are already excited to launch new innovations that we currently develop throughout the year. So watch out for LANA!

Our software is used by **healthcare** institutions to optimize the patient processes in hospitals, **manufacturers** use to increase process stability and the output of their production processes. In the **utilities sector**, our customers establish process controlling for continuous improvement and increasing automation of their processes, just to name a few examples of use cases.

### If you could characterise LANA Process Mining in one sentence without superlative adjectives (e.g., “best”, “fastest”, “most accurate”) and not mentioning competitors, this sentence would be…?

LANA Process Mining is easy to use and empowers organizations to have full control of their business processes and leverage data to continuously and sustainably optimize business performance, compliance and competitiveness.

# Exploring newland: From Process Mining to Automated Process Improvement

*An interview with Marlon Dumas*

The **European Research Council** (**ERC**) **Advanced Grants** are among the most prestigious research grants in Europe. An ERC Advanced Grant provides funding of up to 2.5 million euros to an individual researcher in order to implement a five-years fundamental research project that has the potential to lead to groundbreaking advances in its field.

In 2019, the ERC awarded its first-ever grant in the field of process mining to Prof. Marlon Dumas from the University of Tartu. This grant is a recognition of the growing importance of process mining as a field of research on its own. We asked Marlon to tell us about his trajectory in the field of process mining and where he expects this grant to take him.

### Tell us a bit about yourself and your research institute.

I am leader of the Business Process Management (BPM) research group at the University of Tartu. We are a team of five researchers and seven PhD students. We conduct research on process mining, predictive process monitoring, business process privacy, and verifiable process execution.

### When and why did you first come up with the idea to do research in process mining?

Between 2005 and 2013, my research agenda was focused on **business process modeling and analysis**. Together with colleagues at the Queensland University of Technology, we conducted research on process model verification, simulation, performance analysis, and process model versioning, comparison, and merging.

This research eventually led to the development of **Apromore**, an open-source toolset for managing collections of process models (apromore.org).

Back in 2013, we saw great potential in integrating process **modeling** and process **mining** in a **single platform**. So we pivoted Apromore in the direction of process mining. First, we integrated existing process mining techniques into Apromore. But when testing these techniques with real-life data and showing the results to end-users, we were not satisfied with the state-of-the-art in process mining. We noticed that existing automated process discovery techniques, such as heuristics miner and inductive miner, led to incorrect or imprecise process models when applied to real-life logs. Also, existing conformance checking techniques were too focused on measuring the level of conformance between a model and a log on a 0-to-1 scale, rather than presenting mismatches between an event log and a process model in a user-friendly manner.

As researchers, we wanted Apromore to go beyond the state of the art. So we designed new process mining algorithms, such as the **Split Miner** algorithm for automated process discovery, the Behavioral Alignment approach for conformance checking, as well as techniques for automated discovery hierarchical process models, for fast and accurate identification of business process changes, as well as techniques for comparison of business process variants based on event logs.

These techniques are nowadays packed in Apromore's Community Edition. Since early 2019, we have put together a professional software development team, which has re-implemented a lot of these techniques. We have recently created a spin-off to commercialize this technology. There will be more news about Apromore's commercial edition in the coming months.

## What is predictive monitoring and how can process mining inform it?

Process mining is a family of **tactical management tools**. The end-users of process mining tools are analysts and managers, who use process mining to identify performance and conformance issues in a process, to assess the impact of these issues, and to understand their root causes. The end result of process mining is insights that help analysts and managers to make tactical decisions on how to change the process in the medium-term in order to enhance its performance.

In contrast, **predictive process monitoring** is a family of operational management tools. The end-users of a predictive process monitoring tool are process workers and operational managers. A predictive process monitoring tool uses machine learning models to warn you, for example, that a particular case of your order-to-cash process will end up delayed, or that it will lead to a customer complaint, a returned product, or a refund claim from the customer. Using these predictions, workers and their direct managers can make near-real-time decisions as to when and how to intervene in order to prevent negative outcomes (for instance, preventing customer complaints).

Together with colleagues at the University of Tartu and FBK Trento, we have developed an open-source predictive process monitoring engine called **Nirdizati** (nirdizati.org). Nirdizati allows analysts to train **predictive models from event logs** with minimal knowledge of machine learning. The resulting machine learning models can then be used to produce predictive monitoring dashboards. The main functionality of Nirdizati has been ported into Apromore Community Edition.

## Tell us about your ERC Grant award. What are the expected outcomes of this project?

Current approaches to discover process improvement opportunities are expert-driven. In these approaches, data are used to assess opportunities derived from experience and intuition rather than to discover them in the first place. Moreover, since the assessment of opportunities is manual, an analyst can only realistically explore a fraction of the overall space of improvement opportunities.

The past two decades have seen an increased level of digitization of business processes, and with it, increased availability of fine-grained data about the execution of business processes. This **data availability** allows us to move from purely manual and expert-driven process improvement approaches, to more **automated** and **data-driven approaches** in which improvement opportunities are discovered and assessed systematically.

**In this ERC project**, we will build the foundations of a new generation of process improvement methods that do not exclusively rely on guidelines and heuristics, but rather on a systematic exploration of a space of possible changes derived from process execution data. Specifically, we will develop algorithms to analyze process execution data in order to **discover changes** to a process that are likely to improve its performance. This includes changes in the control-flow dependencies between activities, partial automation of activities (for example, using robotic process automation bots), changes in resource allocation rules, or changes in decision rules that may reduce wastes or negative outcomes.

The outputs of the project will be embodied in an open-source toolset called the **Process Improvement Explorer** (**PIX**). The PIX toolset will allow users to interactively explore spaces of process improvement opportunities with respect to a set of performance measures. By interactive, we mean that the user is able to start with a set of performance measures, provide the metadata required to construct the process improvement space, explore the Pareto-optimal groups of changes. A group of changes is Pareto-optimal if there is no other group of changes that is better along one performance measure without simultaneously being worst along another measure. For example, let's say that our objective is to reduce the cycle time of the process and to reduce its cost. A group of changes is Pareto-optimal if there is no other group of changes that can achieve both lower cycle time and lower cost simultaneously.

To illustrate the envisioned capabilities of PIX, let me take the following **scenario**. An analyst is tasked with analyzing an order-to-cash process due to rising customer dissatisfaction. The execution of this process is supported by an information system that records the creation of each case of this process as well as the execution of each activity. As in process mining, the data collected by the information system can be extracted as an event log, consisting of a collection of event records. Each record has a timestamp, a reference to an activity, a case identifier, and additional attributes. Let's assume also that the system records events that take place while an activity is performed, such as URLs accessed by a worker performing an activity, data entered into a field of a Web form by the worker, files and software applications opened (an event may indicate that a given Excel sheet is opened while the employee is verifying an order), and events occurring inside these applications (for example, when a cell in the spreadsheet is selected).

To explore the space of improvement opportunities using PIX, the analyst will specify the **performance measures** of interest (for instance, the cycle time or the defect rate) as well as the allowed changes, such as: Which activities may be re-ordered? Which activities may be selectively executed/skipped? Which resource allocation rules may be changed and in what ways? Which activities may be automated? She may also specify performance constraints, like: the cycle time of the to-be process should not exceed a given threshold.

Given this input, PIX will allow the analyst to **navigate** through the space of **possible process changes**, focusing on the most promising ones (the Pareto-optimal ones). The analyst will be able to alter the set of performance measures, constraints, and allowed changes, so as to enhance or prune this search space. Once she selects a group of changes, a specification of these changes is generated. If a change implies the automation of an activity, an executable model to automate the activity is generated. If it implies the redefinition of rules, the new ruleset is produced, and so on.

### Enabling data-driven process improvement is also one of the goals of process mining. How does the vision of PIX relate to process mining?

Process mining techniques focus on discovering process models, analyzing their performance, and comparing process models and event logs. While these techniques provide insight into the "as is" state of the process, they do not tell us **how to improve** the process in order to achieve a given set of performance objectives.

Process mining techniques are very suitable when we need to figure out what factors might be contributing to a given performance issue. But process mining techniques do not help us to identify, evaluate and explore possible improvement opportunities. Process mining is meant for "as is" analysis, not for "to-be" analysis. After using a process mining tool to identify bottlenecks, friction points, and root causes of performance issues, the analyst is left on her own. This is basically the **ATAMO approach** (**And Then A Miracle Occurs**). We hope that a light bulb will turn on in the analyst's head. In reality, though, performance issues are too complex for analysts to be able to manually explore the full space of remedial actions in its full extent.

The goal of the PIX project is to fill in the gap from "I have analyzed a performance issue and I know more or less where it comes from" to "**I have found a suitable set of changes** to address this performance issues".

For example, think about a handful of customers in a region who are experiencing delays in receiving the products they have ordered. Naturally, these customers will start visiting the company's website often to check the status of their delivery. PIX will detect that these customers are experiencing a delay and will trace down the reason for these delays. It might be that a supplier is waiting for missing information to fill out a customs clearance document. PIX will propose ways of preventing this problem from happening in future. For example, it might propose to add an item in the checklist used by employees in the warehouse, asking them to enter the required customs clearance information for every product that contains liquids.

## Do you have any ideas or results on how to develop this sort of magic "idea-generation box"?

So far, we have developed two components of PIX. The first one is a tool for **discovering simulation models**. It's called **Simod**. Simod is a box that takes as input an event log and produces a simulation model out of it. The simulation model is optimized for accuracy, which is essential because we will rely on the output of Simod to determine which process changes are better. In other words, Simod enables us to do a "what-if" analysis in a fine-grained and accurate manner.

The second ingredient is an **optimization engine** that takes as input an event log and **generates** all sorts of "**candidate changes**" that are likely to improve a process with respect to multiple performance measures (for example, processing time, cycle time, cost). This optimization engine invokes Simod in order to determine to what extent a given change improves (or degrades) the performance of the process along multiple dimensions. The optimization engine then builds a Pareto-optimal set of changes with respect to a set of performance measures.

It's early days. But one step at a time, we hope to make some progress towards the long-term vision of automated process improvement.

# End-user's corner: Dominic Giss

*An interview with Dominic Giss*

Dominic Giss is the Head of the Competence Center for Operational Excellence of the biggest Swiss Health Insurer. We ask him how he sees process mining in his experience.

## Dominic, where and when did you first hear about process mining?

I got introduced to process mining while working for a large consulting company. I then helped several clients to understand and improve their business processes, both in terms of process assurance and sustainable optimization.

## How do you use process mining in your organisation?

At my current employer, we use various data-driven methods for monitoring/controlling and to optimize processes. If process mining is applied, then primarily during early project phases (discovery) of larger process optimization projects.

This allows us to create a strong, precise data-layer upon which we build, on the one hand, dashboards containing the processes' "vital signs", on the other an interactive analytical instrument to search for undesired patterns.

## How do your counterparts perceive and adopt process mining?

They very much like the visual representation of the processes and the fact-based nature of the results. This allows for a much more objective discussion on the current state of the processes, as well as a single source of truth for impact quantification.

I am not saying that this makes the discussion more pleasant though :-)

## What is the major challenge while using process mining?

Challenge is twofold: first, getting (the right, complete) data and, secondly, translating the visual transparency into insights that can be incorporated into the overall project workflow. For the first leg of the challenge: I see process data coming from different systems at different levels of abstraction (e.g., workflow steps, web-service calls or clicks on a GUI), and processes triggered by an automated user (robot or batch) which may occur in random order. So, constructing a log that faithfully represents the complete process and without introducing unneeded complexity is tricky (see below my recommendation for research - Occam's razor!)

For the second leg of the challenge: data tells us how the processes are executed and recorded in the system, but it does not put it into the context in which they are executed. We need, then, to be careful when interpreting those "maps" and be very selective to what extend these "insights" are of value during root-cause and hypothesis validation step. This is why an effective process optimization project typically applies various tools and methods.

## What was your "ah-ha" moment while using process mining?

Laugh… a bit embarrassing. When I empirically discovered/understood the difference between an event- and a case attribute. So easy it might be or sound, the impact during the log-engineering phase is tremendous.

### How do you see process mining in the future?

As digitalization continues and thereby digital process footprints become more accessible, the foundation for process mining greatly improves. However, I believe that data-driven methods or data-driven decision making as such become more important and not a single method.

Finish with a freestyle, maybe provocative, but anyway to you relevant question. What would it be? "What should research in/around Process Mining stop doing/assuming?"

Stop thinking the process map or a visualization of a process is your result or of massive value for the business. Typically, there is hardly any value that goes beyond process discovery phase by just having a process map. The inputs created using process mining have to be incorporated into the overall process optimization methodology.

1. Stop thinking the process map or a visualization of a process is your result or of massive value for the business. Typically, there is hardly any value that goes beyond process discovery phase by just having a process map. The inputs created using process mining have to be incorporated into the overall process optimization methodology.
2. Stop thinking there are event logs ready to be used, at least until you know how they are built.
3. Stop assuming there is event data at all that could be transformed into event logs
4. Stop thinking event logs are all of the same structure (data model). Log engineering is key!
5. Stop thinking that there is just one process-object. Typically, there are several hierarchical levels and analytical questions that must be addressed can incorporate all of them simultaneously.
6. Stop thinking you can optimize processes solely on the grounds of process mining.

# Conference room: Business Process Management (BPM 2020)

*Presented by Manuel Resinas, Adela del Río, and Cristina Cabanillas*

This year the University of Seville hosts the 18th edition of the International Conference on Business Process Management (BPM 2020) from September 13th to September 18th. This conference is the **premium forum** for **researchers** and **practitioners** in **Business Process Management**. Over the past decade, it has built its reputation by showcasing leading-edge research of the highest quality together with talks, tutorials and discussions by the most renowned thought leaders and innovators in the field. The BPM conference series embraces the diversity and richness of the BPM field. It serves as a melting pot for experts from a mix of disciplines including Computer Science, Information Systems Engineering, and Management, thus covering a wide variety of topics. Among them, **process mining** is one of the most significant ones. As a matter of fact, every year at least **one third of the papers accepted** for publication at the conference focuses on aspects related to process mining. This has made BPM one of the principal venues for the process mining research community. Not only process mining scientists regularly attend the conference, also companies like Signavio

and Celonis contribute as regular sponsors and have a physical presence in it such that participants can get firsthand information about them. Therefore, BPM is an excellent venue to exchange ideas and to meet people working on this field.

In **this edition**, together with the three main research tracks, BPM 2020 encompasses a number of other components that are very attractive for process mining enthusiasts. They include a **Blockchain Forum** and a **Robotic Process Automation (RPA) Forum**, which, among others, welcome papers concerning the application of process mining to enhance these technologies; an **Industry Forum**, which serves to exchange experiences and to support networking among, among others, process mining experts around the world by inviting the submission of case studies and success stories of the application of process mining to the industry; a **Demonstration & Resources session**, which is intended to showcase innovative tools, services and applications related to process mining and other BPM-related topics, as well as resources like datasets, taxonomies, labelled event logs and annotated corpora alike, quantitative/qualitative data and benchmarks that may originate either from academic initiatives or industry endeavours; a **Doctoral Consortium**, which invites PhD students working in the area of BPM to submit their PhD proposal to get feedback and guidance about their planned research; a **Call for Tutorials**, which seeks to cover emerging topics and aspects of the BPM discipline, new technologies and opportunities, teaching issues and lessons learnt from industry experiences; and finally, 7 workshops, covering a broad set of topics related to BPM. Among the workshops, we want to highlight the workshop on **Business Process Intelligence (BPI)**, which currently counts with 16 editions, being one of the oldest venues mainly focused on process mining. Moreover, it hosted the BPI Challenge until 2018.

**The sessions involved in BPM 2020 will take place at the Royal Tobacco Factory**, an astonishing 18th-century stone building that was the first tobacco factory in Europe and the home of Carmen, the famous Bizet's opera. Currently, it is the seat of the rectorate of the University of Seville, which was established in 1505 and, with more than 70000 students, is the third-largest university in Spain. The **social events** also play an important role in BPM 2020. Besides enjoying the fantastic weather of September nights in Seville, the participants will be able to learn more about the history of the Royal Tobacco Factory, be amazed by the blend of Mudejar, Gothic and Renaissance styles of the Royal Alcazar and its marvelous gardens, and enjoy a wonderful sight of the Guadalquivir river while savouring the exquisite Spanish cuisine during the conference dinner.

**BPM 2020** is organised by the **Applied Software Engineering (ISA) Group** of the University of Seville. This research group is led by Antonio Ruiz-Cortés and composed of 21 members. Its research spans 6 areas of interest including BPM, Service Governance, Metaheuristics, Experiments Support, Search-based Software Engineering and Software Testing. The research on BPM aspects is led by Manuel Resinas and Antonio Ruiz-Cortés, and it occupies 9 group members. It addresses the development and application of software tools to improve performance and human resource management in business processes. Concerning the former, the group has a strong experience in the monitoring of business processes based on process performance indicators (PPIs). The current interests involve improving the modelling of PPIs by non-experts, the monitoring of decisions and unstructured processes based on event logs and the reliability and evolution of PPIs predictive monitoring models. Regarding the latter, the research covers several different angles. One stream of research focuses on the application of techniques like **mindfulness** or **time management** to improve personal **productivity**. Another research stream is focused on the configuration and use of **workstream collaboration** tools to improve the collaboration of people in the context of **digital transformation**. Finally, the third research stream tackles the **organisational perspective** of business processes pursuing the **optimisation** of the management of human resources along with process **modelling**, **execution** and **analysis**. Within the analysis, approaches for **mining process event logs** to discover resource assignment information have been developed.

## About Seville

For all its important monuments and fascinating history, Seville is universally famous for being a joyous town and one of the most beloved places by visitors to Spain. Seville is the capital and largest city of the autonomous community of **Andalusia**, in Southern Spain. It has a municipal population of about 700,000, and a metropolitan population of about 1.5 million, which makes it the fourth largest city in Spain. Its **Old Town** is the **third-largest in Europe** with an area of 4 square kilometres (2 sq mi) and contains **three UNESCO World Heritage Sites**: the Alcázar palace complex (you can see a part of its stunning beauty in the picture above), a royal Mudejar palace; the Cathedral; and the General Archive of the Indies, where the historical records of the American continent are kept. Other sights not to be missed include the Casa de Pilatos, a large sixteenth-century mansion where Mudejar, Gothic and Renaissance styles blend harmoniously amidst exuberant patios and gardens; La Maestranza bullring; the Tower of Gold; María Luisa park; and, crossing the Triana bridge over the large Guadalquivir river, the lively popular quarter Triana, with charming narrow streets and traditional ceramic factories. And, of course, one cannot leave Seville without tasting its "**tapas**" and enjoying a glass of **Sherry wine** in one of the probably thousands of bars in the city.

## Important dates

The closest due date is the submission deadline for the main conference: the 16th of March (strict!). Make sure to stay up to date with this and all the other key dates by checking the dedicated page on the BPM 2020 website: https://congreso.us.es/bpm2020/calls/keydates/

# Challenge and study: Process Discovery Contest 2020 (PDC 2020)

*Presented by Eric Verbeek*

The Process Discovery Contest (PDC) is the annual event that, since 2016, has been dedicated to the assessment of tools and techniques that discover business process models from event logs.To that end, synthetic data are used to have an objectified “proper” answer. Process discovery is turned into a classification task with a training set and a test set: the process model discovered from the training set is evaluated based on its ability to correctly decide which of the traces in the test set are fitting or not with the original (unknown) model. We talk about the PDC with a former contestant and now a newly acquired member of the organisation committee, **Eric Verbeek**.

## Eric, tell us a bit about yourself and your research institute.

My name is Eric Verbeek, and I work at **Eindhoven University of Technology (TU/e)**, which is one of four technical universities in The Netherlands. I've been working at TU/e for almost 30 years now as a scientific engineer, which is like an assistant professor but with “education” replaced by “support”. For the first ten years, I’ve worked at the Department of Mathematics and Computer Science, mostly on **ExSpect** (a modelling tool for coloured Petri nets, see www.exspect.com) and **Woflan** (a verification tool for sound workflow nets, see www.win.tue.nl/woflan). For the following six years, I’ve worked at the Department of Industrial Engineering, where I defended my PhD thesis on the verification of workflow nets. Meanwhile, the process mining field had started. I was there but only started working on our process mining tool **ProM** (www.promtools.org) later on. During the last fourteen years, I’ve worked again at the Department of Mathematics and Computer Science, with ProM as my main dish. My ingredients to this dish include the **framework** (the ‘plate’, if you like), the **transition system miner**, the **decomposed discovery and replay**, and, of late, the **log skeletons**.

## What is your history with the process discovery contests?

When the first process discovery contest was started in 2016, I decided to contribute using a collection of (decomposed) discovery algorithms. The idea was quite simple: Some discovery algorithms guarantee perfect fitness, hence the model they return can correctly replay all the traces of the original event log (no false negatives). False positives (traces that are accepted by the process model, though they shouldn’t) were then avoided as much as possible by using a knock-out system: If the result of one of the discovery algorithms was a no-go, then the final classification was negative. This was implemented in the **DrFurby** (which is how "Dr. Verbeek" is sometimes pronounced in English) classifier plugin in ProM. To my amazement, this was the **winning contribution**.

Of course, the DrFurby classifier did not really return very usable models, hence in the process discovery contest of 2017 a jury was introduced to rank the usefulness of the discovered models by different contributions. To contribute to that contest, I started the log skeletons. The main idea behind this was that many of us seem to have a strong bias towards Petri-net-like models. As a result, I assumed that the contest logs would typically contain constructs that are hard to model (or discover) in a Petri net. Thus, I started with a

**new modelling formalism** that would capture different properties than the properties as captured by Petri net. In the end, this resulted in the **log skeleton** formalism, which, as I learned later, was quite close to the declarative process modelling language of Declare. Although my contribution was the only one that classified all test traces correctly (after a lot of tuning, I must admit), I did not win the contest as the jury liked the interactively-discovered Petri nets from Alok Dixit over my automatically-but-tuned-discovered log skeleton models.

For the process discovery contest of **2019**, I used again the log skeletons, which I had improved by then, but I would use them only to gain sufficient insights into the model to be able to create a Petri net manually. Although the 2019 logs were much tougher to tackle with the log skeletons than the 2017 logs, I succeeded in creating 10 Petri nets that in the end **won** the contest.

## And now you are an organiser of the contest. What's new in PDC 2020?

The winners of the process discovery contests of 2017 and 2019 "kept the **expert in the loop**": Alok's discovery technique from 2017 requires an expert to make the necessary decisions along the way, and the conversion from the log skeletons to the Petri nets in 2019 requires an expert as well. But what about the **automated discovery algorithms**? What can we say about them? Obviously, they are not yet a match for an expert, but what should we do to close the gap between the algorithm and the expert? Which way to go to improve on these algorithms? To answer these questions, we should set up the contest a bit differently: Instead of asking the contributors to classify disclosed logs, we (the organizers!) should ask the contributors for their working discovery algorithms and do the classifications ourselves, on non-disclosed logs. This way, the contributors cannot tune their approach on the logs, as was done in earlier contests, and cannot take advantage of manual interventions. This, in the end, was the motivation for the current setup of the Process Discovery Contest 2020: The **automated contest** compares automated discovery algorithms, which shows the state-of-the-art among these algorithms, and the **manual contest** compares the algorithms to the experts, thus showing the gap we still need to bridge.

To implement this setup, this year, I joined the organization of the process discovery contest series. A configurable master model was created, from which event logs using different characteristics were generated: whether there are loops and of what complexity, whether there are optional tasks, duplicate tasks, noise, and so on. As the generation of the model and the logs requires precision and takes time, we created them **all over the summer** as that is a reasonably quiet period!

In the end, this has led to quite a number of event logs: 192 from which a model needs to be discovered, 192 that need to be classified by that discovered model, and 192 that contain the ground-truth classification. Every event log contains 1000 traces, of which at least 400 are positive and at least 400 are negative.

To score a discovery algorithm (we go a bit technical here), we use the F-score on the positive accuracy and the negative accuracy, where the positive/negative accuracy is the accuracy of the classification on the positive/negative (ground-truth) traces. We are thus looking for a balance between the correct classification of positive traces and the correct classification of negative traces. The discovery algorithm with the highest average score, wins the automated contest! Note that **the automated contest has already started**: You can submit your contribution today! *And* you can submit an updated version of your contribution tomorrow! Every submitted contribution counts.

After the automated contest has run, we will disclose the most complex log, that is, the one generated by setting all the characteristics I mentioned above to their maximum level. This will start the manual contest. The **manual contest** will run for about two weeks, during which the contestants can submit a classification for this

one log. How the classification was created is irrelevant to us. If needed, a contestant may just classify every trace in the log manually. However, only the last submitted classification counts! The contestant with the best score wins the manual contest.

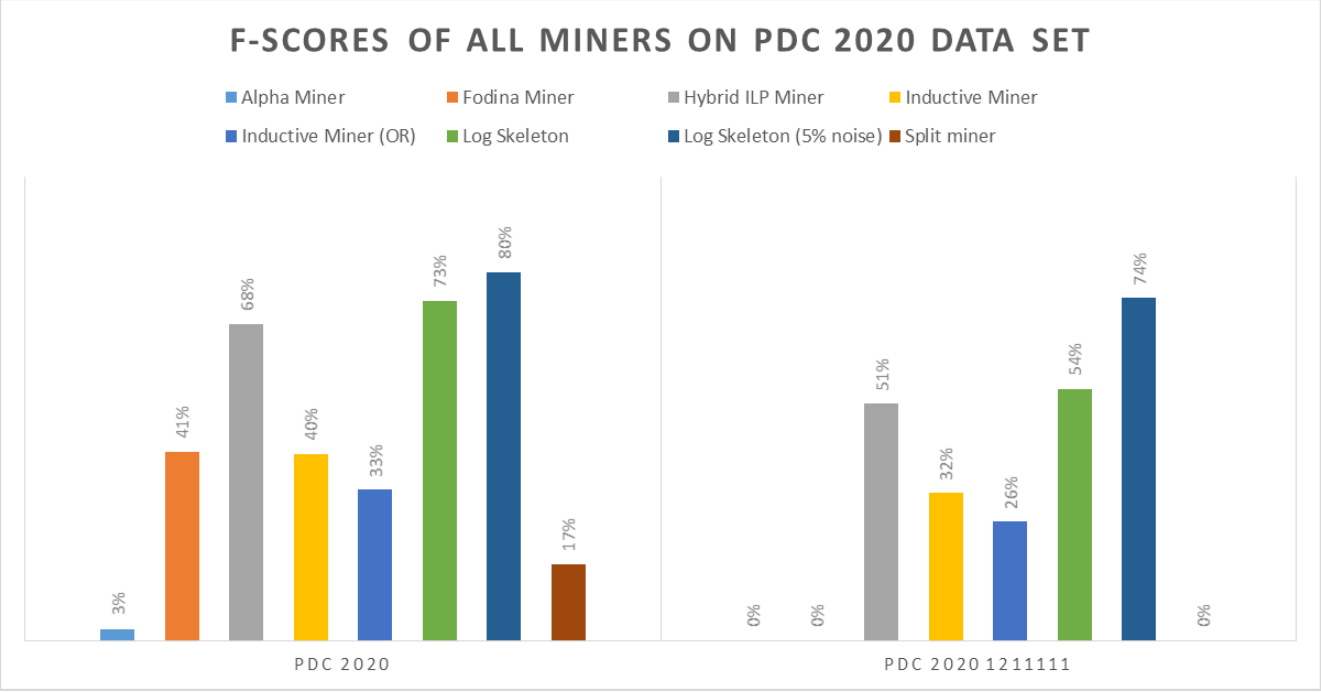
More details on the **PDC 2020** can be found on its **website**: [icpmconference.org/2020/process-discovery-contest](icpmconference.org/2020/process-discovery-contest)

### How do you see the future of the Process Discovery Contest?

With the current setup of the contest, we (the organizers) can build a body of working discovery algorithms. To start this, we have implemented eight sample working discovery algorithms, including the Alpha Miner, the Fodina Miner, and the Log Skeleton (details on these and the other implementations can be found on the website of the PDC 2020). Notice that those implementations are not contributions to the contest! If the authors of these miners want to contribute, they should submit them. As a result, we will be able to score these discovery algorithms in the future on new data sets as well, like for a next contest, and see whether new discovery algorithms have advanced over the old ones.

By adding the data set of every future process discovery contest, and by adding other interesting data sets as well, we can also build a body of data sets that we can use to test process discovery algorithms. Next to the PDC 2020 data set, we have already included a number of existing data sets in this body: PDC 2016, PDC 2017, PDC 2019, and aXfYnZ. The three PDC data sets are from the previous editions of the process discovery contest; the aXfNyZ data set was created by Laura Maruster, who was **the first PhD candidate** to defend a thesis **on process mining**, for as far as I know!

Together with the body of discovery algorithms, this body of data sets can provide a fair view on the current state of the field of process discovery algorithms. As an example, the figure below depicts the results of the eight sample discovery algorithms on the PDC 2020 data set. This shows that the Log Skeleton with 5% noise scores the best on the PDC 2020 data set with a score of 80%, and that experts need to score at least 75% to outperform this discovery algorithm. It also shows that on the PDC 2020 data set, the Hybrid ILP Miner is the Petri-net-based discovery algorithm that scores best. Other results can be found as well on the website of the PDC 2020.



# ICPM 2020: get ready!

*Presented by Massimiliano de Leoni and Alessandro Sperduti*

The **Second International Conference on Process Mining** (ICPM 2020) will take place in October 2020 in the city of **Padua**, Italy between **October 4th** and **8th**.

The first edition in Aachen in 2019 was a **large success** as 420 participants attended the event, 180 being consultants and end-users. If it was still necessary, this shows once more that Process Mining has a great appealing on the industry, and companies can clearly see the business value of the insights into the processes that process-mining techniques are able to generate. The first edition in 2019 was the start of a very successful conference series, a remarkable example of a venue where academia and industry meet, give feedback and provide a boost to each other. In fact, the memory of the first-edition success will likely provide leverage and a cascading effect. A large number of process-mining consultants, end-users, vendors and academics will participate and contribute to the second edition in Padua. You can see some pictures of the city here: from top to bottom, the **Padua's exhibition centre** (the venue of ICPM 2020), the beautiful Prato della Valle, and the breathtaking Scrovegni Chapel, at the earth of Padua.

The first premonition that ICPM 2020 will likely be successful, even perhaps more than last year, is evident when we look at the **sponsors** that have already indicated that they contribute. In fact, nine sponsors have so far indicated that they will contribute, of which some were not sponsoring last year; we are at the stage of finalizing the necessary formal agreements. Remarkably, we are still 6 months ahead, and we are confident that more than the current nine sponsors will follow in the months to come. As a matter of fact, if you are willing to sponsor, please **contact us** at icpm2020@math.unipd.it. There are four packages available that suit everybody's need.

Visit icpmconference.org/2020 to stay up-to-date about the conference. The web site is continuously updated as soon as new aspects of the conference organization are confirmed!

## About the program

The structure of the program of ICPM 2020 will not change drastically if compared with ICPM 2019 as the structure of last year has clearly demonstrated to be effective. The program of ICPM 2020 will be as follows:

- On **October 4th**, the **doctoral consortium** will take place where PhD students can present their plan of research and receive feedback from process-mining experts to better steer their research.

- **October 5th** is the day of the **co-located events**. In particular, the workshop chairs, Drs. Leopold and Leemans, perform a great job and selected six **workshops** on very relevant research topics in Process Mining (see icpmconference.org/2020/workshops). The program of the day will be complemented by the first **Italian Forum on Business Process Management** (see icpmconference.org/2020/itbpm), which will bring "under the same roof" the Italian community of BPM.
- **October 6th and 8th** will feature the **research track** with several sessions. Each session will focus on a different angle of the process-mining research: process discovery, conformance checking, predictive and recommender systems, and many more. During each session, prominent speakers will present their papers on cutting-edge research, which are carefully chosen by a program committee orchestrated by the conference chairs, Profs. van Dongen, Montali and Wynn. The call for research-track papers is available at icpmconference.org/2020/call-for-papers.

**Key Dates for Research Papers**
• Abstract Submissions: 24 June, 2020
• Full-Paper Submissions: 1 July, 2020
• Notification: 5 August 2020
• Camera-ready Submissions: 2 September 2020

**Key Dates for Workshops**
• Submission of Workshop-Paper Abstracts: 18 August 2020
• Submission of Workshop Papers: 25 August 2020
• Workshop-Paper Notification: 14 September 2020
• Camera-Ready Submissions of Workshop Papers: 22 September 2020

**Key Dates for Demonstration Papers**
• Submission of Demonstration Papers: 4 August 2020
• Notification of Acceptance: 31 August 2020
• Camera-Ready Submissions of Demonstration Papers: 15 September 2020

- **On October 7th**, the **Industry Day** will take place with several initiatives to bring together academia with industry. Most of the day will be devoted to several **process owners** and **analysts** who will present their **experience** of applying process mining in large **production** environments. In concert with the industry chairs, Dr. Marc Kerremans and Prof. Marcello La Rosa, we have just published the call for industry speakers (see icpmconference.org/2020/call-for-industry-day-talks). We aim to be as inclusive as possible: *You can **propose yourself**, if you are from the industry, and/or some of your industrial partners to talk at the industry day*! We need you!
  Also, the industry day will include a **discussion panel**, with companies and sponsors, to debate key questions that are currently gaining momentum. Last but not least, process-mining researchers and vendors will be given the chance in a dedicated **demonstration session** to showcase the last advances in process-mining tools: they will present the novel features that are available in commercial software suite and in academic tools, with dedicated stands to illustrate their achievements to interest conference attendees. This will be beneficial for all the actors involved: industry can take inspiration of the latest features development in academia, academia can reflect on which features were picked up by commercial process-mining software, and consultancy firms, process analysts and owners receive insights on repertoire of solutions available.

The program of ICPM 2020 does not finish here: different **contests** will be organized where researchers, vendors, consultants and, even, students will compete to win the challenges that will be proposed. Presentations of the winners will also take place on Monday, October 5th. Stay tuned on icpmconference.org/2020/contributions to see which contests will be present.

We are proud to confirm that the international value of the conference is widely recognized: the conference is technically co-sponsored by the **IEEE Computational Intelligence Society**, and supported by the IEEE Task Force on Process Mining. Also, **more details** will follow in the **next edition of this newsletter**.




- Membership
- Archives
- Contact


Follow the IEEE Task Force on Proess Mining on

- **WWW**: www.tf-pm.org
- **Twitter**: @ProcessMiningTF
- **LinkedIn**: linkedin.com/groups/13826719/