

PROCESS MINING: NOT ONLY BRIDGING DATA AND PROCESSES, BUT ALSO INDUSTRY AND ACADEMIA

by prof.dr.ir. Wil van der Aalst (24-9-2019)

At the end of the 1990-ties, I worked one year at the University of Colorado in Boulder (USA) in the context of a longer sabbatical. This was the time that I started to work on process mining. It is exciting to see that many of the things I envisioned at the time ended up in today's process mining tools. Process mining is one of the rare examples where a new category of software tools can be directly linked to university research. Therefore, it is interesting to reflect on the relationship between industry and academia. Experiences in process mining show that both can benefit from each other. Initiatives like the Celonis Academic Alliance are instrumental to this.

STARTING POINT: BRIDGING THE GAP BETWEEN DATA AND PROCESSES

At the beginning of 1999, I wrote a research proposal with the title "Process Design by Discovery: Harvesting Workflow Knowledge from Ad-hoc Executions". In the proposal, I defined process mining as "distilling a structured process description from a set of real executions". At the time, there were over 200 workflow management systems that all had problems supporting the dynamism of real-life processes involving people. Only highly structured and stable processes could be supported properly. Therefore, adding flexibility to workflow systems was one of the main research topics. Instead of adding more flexibility mechanisms, I proposed to use information about the actual process executions to infer a workflow model (expressed in terms of a Petri net) and use the model to automatically improve processes while they are running. Initially, the focus of process mining was on workflow automation. Therefore, we often used the term "workflow mining" in the first couple of years. Only later, we realized that the principles can be applied to any operational process in production, logistics, healthcare, learning, government, finance, etc. However, at the time event data were only available in administrative or financial settings, i.e., the natural habitat of workflow management systems. Re-reading the 20-year-old research proposal "Process Design by Discovery: Harvesting Workflow Knowledge from Ad-hoc Executions" makes me realize that many of today's ideas in process mining and Robotic Process Automation (RPA) have been around for quite some time.

My main motivation to start working on process mining at the end of the 1990-ties was my disappointment in the practical use of simulation software and workflow management systems. Simulation and workflow management have in common that they rely on humans to make process models. However, such models typically describe only the "happy flows" and fail to capture the less frequent executions that generate most of the problems. At the time, many workflow management projects failed. Together with Hajo Reijers and several PhD and MSc students, I did a longitudinal study on the effects of workflow management systems. Only half of all implementation projects in the study succeeded in taking the workflow system into operation. When using simulation software, the process of making the model was often more insightful than the actual results. Simulations rely on too many

simplifying assumptions. All these experiences show that process orientation is important, but one should connect process management to the actual evidence recorded in databases, audit trails, etc.

The above experiences naturally led to scientific challenges such as discovering process models (e.g., Petri nets) from event data. However, experts in process modeling and analysis (e.g., the Petri net, concurrency, and model checking communities) were not interested in actual data, and experts in data analysis (e.g., the statistics and data mining communities) were not interested in processes. For a very long time, there were very few researchers working on the interplay between "data science" and "process science". The situation changed only recently. The success of the First International Conference on Process Mining (ICPM) in June 2019 in Aachen illustrates this change (see <https://icpmconference.org/2019/>).

FROM WFM TO BPM TO PM

The main focus of Workflow Management (WFM) in the 1990-ties was on automation. The ultimate goal was Straight-Through-Processing (STP) by removing process logic from applications and using WFM systems to fully orchestrate processes. WFM evolved into Business Process Management (BPM) which had a much broader focus. In 2003, we organized the first international BPM conference in Eindhoven. Initially, the focus was still on automation and process modeling. Flexible workflow systems, modeling notations, workflow verification, process model repositories, reference models, and service orientation were typical research topics. However, over time the focus shifted to exploiting data to improve processes. As a result, many BPM papers written over the last decade are actually process mining papers. This is unsurprising given the availability of event data in today's information systems.

COLLABORATION BETWEEN INDUSTRY AND ACADEMIA

Although I started working on process mining in the late 1990-ties, it took about 10 years for the first tools to become available. Futura Reflect was the first commercial process mining tool (2007) followed by Disco (2009), Celonis (2011), and many more. Today there are over 30 commercial process-mining tools next to open-source tools like ProM, PM4Py, Apromore, and RapidProM. Celonis has been the most successful commercial process-mining tool with many larger users (e.g., Siemens, BMW, Edeka, Uber, Vodafone, etc.). Therefore, it is interesting to mention some of the features in Celonis Process Mining directly linked to research results presented earlier: process discovery inspired by the heuristic miner (2002) and the fuzzy miner (2006), token-based conformance checking (2005), token animation and sliders (2006), process-based root cause analysis (2006), and process discovery based on the inductive miner (2013). The years indicate when results were first published. These features are all included in the latest versions of Celonis. Some innovations were adopted in less than five years (e.g., inductive mining) while other innovations were adopted more than ten years later (e.g., token-based conformance checking) or are still waiting to be adopted.

Clearly, process-mining vendors can benefit directly from research. It is relatively easy to build a process-mining tool to discover a simple Directly-Follows Graph (DFG) where the arcs show frequencies and times. However, more advanced features like conformance checking, better discovery techniques, prediction, process improvement, etc. all require expert knowledge and cannot be copied easily. Therefore, leading process-mining vendors will need to invest in R&D and collaborate with process mining researchers.

Process-mining research and education also benefit from process-mining vendors. First of all, vendors provide easy-to-use tools lowering the threshold to get started. Several vendors provide an academic program allowing students and researchers to use their software freely. Examples are the long-running academic initiative of Fluxicon and the recently launched Celonis Snap. Snap can be freely used by researchers and students. Next to software, process mining vendors provide interesting use cases that drive new research, e.g., the Celonis Academic Alliance also provides data sets and lecture material. Applications in an expanding number of domains, trigger novel research questions. Enterprise-wide usage of process mining sets new requirements for scalability and usability. Moreover, practical use cases are very inspiring and motivating for students. This is important because both industry and academia need many well-educated process miners ready to transform processes all over the globe.