

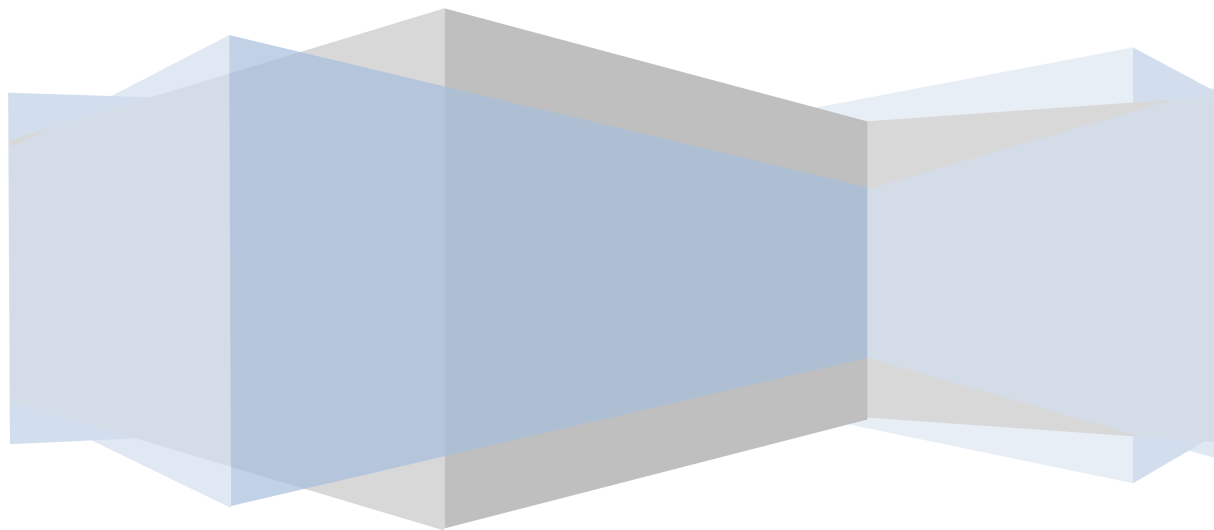
The secret of getting ahead is getting started. Mark Twain

However, don't let perfectionism become an excuse for never getting started. Marilu Henner

A journey of a thousand miles begins with a single step. Lao Tzu

How to get started with process mining?

Wil van der Aalst



Process mining provides new ways to utilize the abundance of event data in world surrounding us. These event data enable new forms of analysis facilitating process improvement. Process mining provides a novel set of tools to discover the real process, to detect deviations from some normative process, and to analyze bottlenecks and waste. Process mining should be in the toolbox of any data scientist! Process mining is as generic as a spreadsheet. Where spreadsheets work with numbers, process mining starts from event data with the aim to analyze processes. Process mining bridges the gap between traditional model-based process analysis (e.g., simulation and other business process management techniques) and data-centric analysis techniques such as machine learning and data mining. Process mining provides a new means to improve processes in a variety of application domains. The omnipresence of event data combined with process mining allows organizations to diagnose problems based on facts rather than fiction.

The interest in process mining is rapidly growing. This is reflected by the number of commercial vendors of process mining software, the number of process mining publications, and the number of process mining projects. As a result, many people approach me with the question:

How to get started with process mining?

This short note aims to answer this question. The bad news is that process mining cannot be understood in 5 minutes. Process mining is not a "one-trick pony" and includes a range of techniques and approaches. The good news is that there is a wealth of information available.

Process mining can be used and understood at different levels. Depending on your current level and your goals different pointers can be given.

Level 1: Understanding Process Mining

The best introduction to process mining is the book **Process Mining: Discovery, Conformance and Enhancement of Business Processes** by W.M.P. van der Aalst, Springer Verlag, 2011 (ISBN 978-3-642-19344-6, <http://www.springer.com/gp/book/9783642193446>).¹ The book will provide the basis for a broad understanding of the field. The book provides a comprehensive overview of process mining including topics such as process discovery, conformance checking, mining organizational aspects, performance analysis, and operational support using process mining (e.g., prediction). The book is self-contained and also provides a crash course into business process modeling and data mining.

The book is complemented by the **Coursera** course **Process Mining: Data science in Action** (<https://www.coursera.org/course/procmin>). This MOOC (Massive Open Online Course) on process mining is based on the book. This course starts with an overview of approaches and technologies that use event data to support decision making and business process (re)design. Then the course focuses on process mining as a bridge between data mining and business process modeling. The course is at an introductory level with various practical assignments. The course covers the three main types of process mining:

- The first type of process mining is **discovery**. A discovery technique takes an event log and produces a process model without using any a-priori information. An example is the Alpha-algorithm that takes an event log and produces a process model (a Petri net) explaining the behavior recorded in the log.

¹ A new version of the book called "Process Mining: Data science in Action" (<http://www.springer.com/gp/book/9783662498507>) will appear in June 2016.

- 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.** An example is the extension of a process model with performance information, e.g., showing bottlenecks. Process mining techniques can be used in an offline, but also online setting. The latter is known as operational support. An example is the detection of non-conformance at the moment the deviation actually takes place. Another example is time prediction for running cases, i.e., given a partially executed case the remaining processing time is estimated based on historic information of similar cases.

The course uses many examples using real-life event logs to illustrate the concepts and algorithms. After taking this course, one is able to run process mining projects and have a good understanding of the Business Process Intelligence field. Visit <https://www.coursera.org/course/procmin> to register for the course. Although not visible yet the Coursera course is also available as an **on-demand course**. Send your full name and e-mail to get registered.

Please do not send questions (on how to get started) by e-mail before studying the course and book!

Level 2: Using Process Mining to Improve Processes

The Coursera course and the Process Mining book, provide a solid basis for conducting process mining projects. However, next to learning basics and knowing about the theoretical foundations, it is important to experience the "magic of process mining" by conducting a range of process mining projects. Ideally, you use your own data sets. All process mining tools can read csv files, so the threshold is really low. Most tools also load XES logs making things even easier.

Having no data is no excuse! Lots of data sets are available via the websites such as www.processmining.org, <http://www.win.tue.nl/ieeetfpm>, and http://data.3tu.nl/repository/collection:event_logs. The last link refers to the repository of the 3TU data center. This site provides persistent data sets. Each event log has so called Digital Object identifier (DOI) reference. Clicking on a DOI reference in an article, report or website will seamlessly provide access to the event log. The collection of event logs includes several real-life logs, including the event logs used for the annual BPI challenge.

To perform process mining one needs **software**. The Coursera course uses the open-source tool ProM and a limited version of Disco. Since Feb. 2016, ProM comes in two flavors: the extensive version and a "lite" version (ProM Lite). The latter has been created to make it easier to get started. For regular ProM over 1500 plug-ins are available, ProM Lite only provides the most frequently used plug-ins. A subset of ProM is also available via the RapidMiner marketplace. ProM, ProM Lite, and RapidProM can be downloaded and used freely. About 15 commercial process mining tools are available on the market. Most of these are easy to use, lowering the threshold to get started.

Tool Name	Vendor	Website
ProM	Open Source	www.promtools.org
ProM Lite	Open Source	www.promtools.org
RapidProM	Open Source	www.rapidprom.org
Celonis Process Mining	Celonis GmbH	www.celonis.de
Disco	Fluxicon	www.fluxicon.com

Interstage Business Process Manager Analytics	Fujitsu Ltd	www.fujitsu.com
Minit	Gradient ECM	www.minitlabs.com
myInvenio	Cognitive Technology	www.my-invenio.com
Perceptive Process Mining	Lexmark	www.lexmark.com
QPR ProcessAnalyzer	QPR	www.qpr.com
Rialto Process	Exeura	www.exeura.eu
SNP Business Process Analysis	SNP AG	www.snp-bpa.com
webMethods Process Performance Manager	Software AG	www.softwareag.com

Everyone that wants to improve operational process is encouraged to experiment with multiple process mining tools using common data sets.

Ideally event data are in XES format (the standard format supported by most tools). However, it is also perfectly OK if the event data are in CSV (comma-separated values) format. Each event should correspond to one row in the CSV file. Each event should have a timestamp, refer to a case (e.g. a customer id), and refer to an activity (e.g., "send invoice"). Of course event may also refer to resources, values, amounts, types, etc.

An example:

```
case;activity;time;product;prod-price;quantity;address
187;pay;2015-02-16 17:11:56;APPLE iPhone 6s Plus 64 GB;969.0;2;NL-7887AC-13
226;prepare delivery;2015-02-16 17:12:18;SAMSUNG Galaxy S4;329.0;2;NL-7821AC-3
155;pay;2015-02-16 17:18:59;SAMSUNG Galaxy S4;329.0;2;NL-7751DG-21
129;make delivery;2015-02-16 17:23:34;APPLE iPhone 6 16 GB;639.0;4;NL-7887AC-13
224;send invoice;2015-02-16 17:30:05;APPLE iPhone 5s 16 GB;449.0;1;NL-9411ME-47
206;confirm payment;2015-02-16 17:31:59;SAMSUNG Galaxy 32 GB;543.99;5;NL-7828AM-11a
...
```

Try to get concrete data yourself and link it to performance and compliance problems and questions. Generic questions are:

- What is the process that people really follow?
- Where are the bottlenecks in my process?
- Where do people (or machines) deviate from the expected or idealized process?
- What are the "highways" in my process?
- What factors are influencing a bottleneck?
- Can we predict problems (delay, deviation, risk, etc.) for running cases?
- Can we recommend countermeasures?
- How to redesign the process / organization / machine?

You have reached Level 2 if you are able to extract such data from systems and use the process mining tools mentioned to answer a variety of questions.

Level 3: Process Mining Expert/Researcher

At Level 2 you are able to conduct process mining projects without any problems. However, analysis techniques like for example the inductive mining algorithms are applied as a black box. To get a deeper understanding of the techniques themselves it is important to start by reading the following papers (after reading the book and taking the course):

1. W.M.P. van der Aalst, A. Adriansyah, and B. van Dongen. [Replaying History on Process Models for Conformance Checking and Performance Analysis](#). *WIREs Data Mining and Knowledge Discovery*, 2(2):182-192, 2012.
2. W.M.P. van der Aalst. [Business Process Management: A Comprehensive Survey](#). *ISRN Software Engineering*, pages 1-37, 2013. doi:10.1155/2013/507984.
3. W.M.P. van der Aalst, K.M. van Hee, A.H.M. ter Hofstede, N. Sidorova, H.M.W. Verbeek, M. Voorhoeve, and M.T. Wynn. [Soundness of Workflow Nets: Classification, Decidability, and Analysis](#). *Formal Aspects of Computing*, 23(3):333-363, 2011.
4. W.M.P. van der Aalst. [Decomposing Petri Nets for Process Mining: A Generic Approach](#). *Distributed and Parallel Databases*, 31(4):471-507, 2013.
5. W.M.P. van der Aalst. [Business Process Simulation Survival Guide](#). In J. vom Brocke and M. Rosemann, editors, *Handbook on Business Process Management 1*, International Handbooks on Information Systems, pages 337-370. Springer-Verlag, Berlin, 2015.
6. W.M.P. van der Aalst. [Process Cubes: Slicing, Dicing, Rolling Up and Drilling Down Event Data for Process Mining](#). In M. Song, M. Wynn, and J. Liu, editors, *Asia Pacific Conference on Business Process Management (AP-BPM 2013)*, volume 159 of *Lecture Notes in Business Information Processing*, pages 1-22. Springer-Verlag, Berlin, 2013.
7. W.M.P. van der Aalst. [Extracting Event Data from Databases to Unleash Process Mining](#). In J. Vom Brocke and T. Schmiedel, editors, *Business Process Management Roundtable 2014*, Springer-Verlag, Berlin, 2015.
8. M. de Leoni, W.M.P. van der Aalst, and M. Dees. [A General Framework for Correlating Business Process Characteristics](#). In S. Sadiq, P. Soffer, and H. Voelzer, editors, *International Conference on Business Process Management (BPM 2014)*, volume 8659 of *Lecture Notes in Computer Science*, pages 250-266. Springer-Verlag, Berlin, 2014.
9. S.J.J. Leemans, D. Fahland, and W.M.P. van der Aalst. [Process and Deviation Exploration with Inductive Visual Miner](#). In L. Limonad and B. Weber, editors, *Business Process Management Demo Sessions (BPMD 2014)*, volume 1295 of *CEUR Workshop Proceedings*, pages 46-50. CEUR-WS.org, 2014.
10. W.M.P. van der Aalst. [Process Mining in the Large: A Tutorial](#). In E. Zimanyi, editor, *Business Intelligence (eBISS 2013)*, volume 172 of *Lecture Notes in Business Information Processing*, pages 33-76. Springer-Verlag, Berlin, 2014.
11. R.P. Jagadeesh Chandra Bose, W.M.P. van der Aalst, I. Zliobaite, and M. Pechenizkiy. [Dealing With Concept Drifts in Process Mining](#). *IEEE Transactions on Neural Networks and Learning Systems*, 25(1):154-171, 2014.
12. S.J.J. Leemans, D. Fahland, and W.M.P. van der Aalst. [Discovering Block-Structured Process Models from Event Logs Containing Infrequent Behaviour](#). In N. Lohmann, M. Song, and P. Wohed, editors, *Business Process Management Workshops, International Workshop on Business Process Intelligence (BPI 2013)*, volume 171 of *Lecture Notes in Business Information Processing*, pages 66-78. Springer-Verlag, Berlin, 2014.
13. R.P. Jagadeesh Chandra Bose, R. Mans, and W.M.P. van der Aalst. [Wanna Improve Process Mining Results? It's High Time We Consider Data Quality Issues Seriously](#). In B. Hammer, Z.H. Zhou, L. Wang, and N. Chawla, editors, *IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2013)*, pages 127-134, Singapore, 2013. IEEE.
14. R.P. Jagadeesh Chandra Bose and W.M.P. van der Aalst. [Discovering Signature Patterns from Event Logs](#). In B. Hammer, Z.H. Zhou, L. Wang, and N. Chawla, editors, *IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2013)*, pages 111-118, Singapore, 2013. IEEE.

15. W.M.P. van der Aalst. **[A General Divide and Conquer Approach for Process Mining](#)**. In M. Ganzha, L. Maciaszek, and M. Paprzycki, editors, *Federated Conference on Computer Science and Information Systems (FedCSIS 2013)*, pages 1-10. IEEE Computer Society, 2013.
16. M. De Leoni and W.M.P. van der Aalst. **[Data-Aware Process Mining: Discovering Decisions in Processes Using Alignments](#)**. In S.Y. Shin and J.C. Maldonado, editors, *ACM Symposium on Applied Computing (SAC 2013)*, pages 1454-1461. ACM Press, 2013.
17. S.J.J. Leemans, D. Fahland, and W.M.P. van der Aalst. **[Discovering Block-structured Process Models from Event Logs: A Constructive Approach](#)**. In J.M. Colom and J. Desel, editors, *Applications and Theory of Petri Nets 2013*, volume 7927 of *Lecture Notes in Computer Science*, pages 311-329. Springer-Verlag, Berlin, 2013.
18. D. Fahland and W.M.P. van der Aalst. **[Simplifying Discovered Process Models in a Controlled Manner](#)**. *Information Systems*, 38(4):585-605, 2013.
19. A. Rozinat, M. Wynn, W.M.P. van der Aalst, A.H.M. ter Hofstede, and C. Fidge. **[Workflow Simulation for Operational Decision Support](#)**. *Data and Knowledge Engineering*, 68(9):834-850, 2009.
20. A. Rozinat, R.S. Mans, M. Song, and W.M.P. van der Aalst. **[Discovering Simulation Models](#)**. *Information Systems*, 34(3):305-327, 2009.
21. A. Rozinat and W.M.P. van der Aalst. **[Conformance Checking of Processes Based on Monitoring Real Behavior](#)**. *Information Systems*, 33(1):64-95, 2008.
22. W.M.P. van der Aalst, H.A. Reijers, and M. Song. **[Discovering Social Networks from Event Logs](#)**. *Computer Supported Cooperative work*, 14(6):549-593, 2005.
23. C.W. Günther and W.M.P. van der Aalst. **[Fuzzy Mining: Adaptive Process Simplification Based on Multi-perspective Metrics](#)**. In G. Alonso, P. Dadam, and M. Rosemann, editors, *International Conference on Business Process Management (BPM 2007)*, volume 4714 of *Lecture Notes in Computer Science*, pages 328-343. Springer-Verlag, Berlin, 2007.
24. IEEE Task Force on Process Mining. **[Process Mining Manifesto](#)**. In F. Daniel, K. Barkaoui, and S. Dustdar, editors, *Business Process Management Workshops*, volume 99 of *Lecture Notes in Business Information Processing*, pages 169-194. Springer-Verlag, Berlin, 2012.
25. W.M.P. van der Aalst, A.J.M.M. Weijters, and L. Maruster. **[Workflow Mining: Discovering Process Models from Event Logs](#)**. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128-1142, 2004.

The above papers (co-)authored by me provide numerous pointers to the work of other researchers in the field and are all available online. Understanding these 25 papers implies that one is indeed a process mining expert! Next to reading the papers, it is important to use related process mining tools (typically exposed as ProM plug-ins).

Level 4: Process Mining Developer

The ultimate level is the level where one is able to develop novel process mining techniques and implement them. One can develop stand-alone tools or build on existing tools like ProM. You can contribute to ProM by creating your own ProM packages, see <https://svn.win.tue.nl/trac/prom/wiki/Contribute>. This requires experience in implementing in java. The best way to get started is to look at the source code of a related plug-in. Also see the forum <http://www.win.tue.nl/promforum/> and the mailing lists <https://svn.win.tue.nl/trac/prom/wiki/MailingLists>.

It is important to design plug-ins for real-life event logs having millions of events. Given the widespread availability of process mining tools it is important to implement new ideas and compare them with existing analysis techniques. Experimentation is compulsory. Process discovery algorithms need to be evaluated using performance indicators related to speed, fitness, simplicity, generalization, and precision.

One does not to reach Level 3 or Level 4 to enjoy the fruits of process mining. The first two levels allow you to apply process mining and improve processes. However, if you are interested in process mining R&D functions, you also

need to understand what is happening under the hood of tools like ProM, Disco, Celonis Process Mining, Minit, myInvenio, Perceptive Process Mining, QPR ProcessAnalyzer, etc.

I hope this helps you get started and reach your target level!

Cheers, Wil



Postal address: Prof.dr.ir. W.M.P. (Wil) van der Aalst, Eindhoven University of Technology, Department of Mathematics and Computer Science (MF 7.103), PO Box 513, NL-5600 MB Eindhoven, The Netherlands.

E-mail: www.vdaalst.com

Phone: +31 40 247.4295/2733

E-mail: w.m.p.v.d.aalst@tue.nl

Campus address: MF 7.103 (room 103 on level 7 in the "MetaForum" building), see [here for contact information and directions](#).