---

summary

Special English Edition 2013-2014
Annual Selection of Articles

Anne Rozinat[1], Wil van der Aalst[2]

[1]*Co-founder of Fluxicon, The Netherlands;* [2]*Technical University of Eindhoven, The Netherlands*

<anne@fluxicon.com>, <w.m.p.v.d.aalst@tue.nl>

Process Mining: The Objectification of Gut Instinct - Making Business Processes More Transparent Through Data Analysis

Abstract: *Big Data existed in the 19th Century. At least that might be the conclusion you would draw by reading the story of Matthew Maury. We draw a parallel with the first systematic evaluations of seafaring logbooks and we show how you can quickly and objectively map processes based on the evaluation of log files in IT systems.*

Keywords: *Big Data, Case Study, Log Data, Process Mining, Process Models, Process Visualization, Systematic Analysis.*

Authors

Anne Rozinat has more than eight years of experience with process mining technology and obtained her PhD cum laude in the process mining group of Prof. Wil van der Aalst at the Eindhoven University of Technology in the Netherlands. Currently, she is a co-founder of Fluxicon and blogs at <http://www.fluxicon.com/blog/>.

Wil van der Aalst is a professor at the Technical University in Eindhoven and with an H-index of over 90 points the most cited computer scientist in Europe. Well known through his work on the Workflow Patterns, he is the widely recognized "godfather" of process mining. His personal website is <http://www.vdaalst.com>.

1. Introduction

The archive of the United States Naval Observatory stored all the naval logbooks of the US Navy in the 19th century. These logbooks contained daily entries relating to position, winds, currents and other details of thousands of sea voyages. These logbooks lay ignored and it had even been suggested that they be thrown away until Mathew Fontaine Maury came along.

Maury (see **Figure 1**) was a sailor in the US Navy and from 1842 was the director of the United States Naval Observatory. He evaluated the data systematically and created illustrated handbooks which visually mapped the winds and currents of the oceans. These were able to serve ships' captains as a decision-making aid when they were planning their route.

In 1848 Captain Jackson of the W. H. D. C. Wright was one of the first users of Maury's handbooks on a trip from Baltimore to Rio de Janeiro and returned more than a month earlier than planned. After only seven years from the production of the first edition Maury's Sailing Directions had saved the sailing industry worldwide about 10 million dollars per year [1].

The IT systems in businesses also conceal invaluable data, which often remains completely unused. Business processes create the modern day equivalent of "logbook entries", which detail exactly which activities were carried out when and by whom, (see **Figure 2**). If, for example, a purchasing process is started in an SAP system, every step in the process is indicated in the corresponding SAP tables. Similarly, CRM systems, ticketing systems and even legacy systems record historical data about the processes.

These digital traces are the byproduct of the increasing automation and IT support of business processes [2].

2. From Random Samples to Comprehensive Analysis

Before Maury's manual on currents and tides, sailors were restricted to planning a route based solely on their own experience. This is also the case for most business processes: Nobody really has a clear overview of how the processes are actually executed. Instead, there are anecdotes, good feeling and many subjective (potentially contradicting) opinions which have to be reconciled.

The systematic analysis of digital log traces through so-called Process Mining techniques [3] offers enormous potential for all

Figure 1. Matthew Fontaine Maury (Source: Wikipedia).

" The manual discovery through classical workshops
and interviews is costly and time-consuming,
remaining incomplete and subjective "
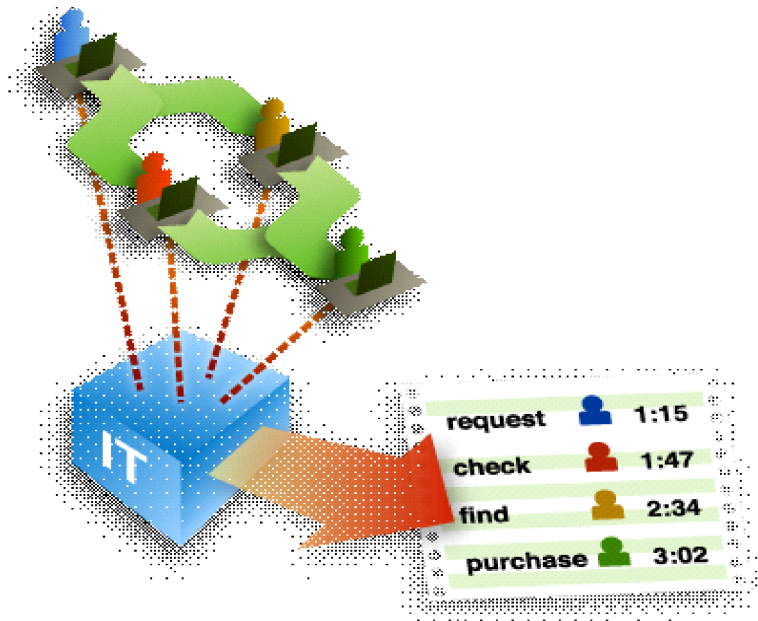
Figure 2. IT-supported Processes Record in Detail Which Activities Were Executed When and by Whom.

organizations currently struggling with complex processes. Through an analysis of the sequence of events and their time stamps, the actual processes can be fully and objectively reconstructed and weaknesses uncovered. The information in the IT logs can be used to automatically generate process models, which can then be further enriched by process metrics also extracted directly out of the log data (for example execution times and waiting times).

Typical questions that Process Mining can answer are:
■ What does my process actually look like?
■ Where are the bottlenecks?
■ Are there deviations from the prescribed or described process?

In order to optimize a process, one must first understand the current process reality - the 'As-is' process. This is usually far from simple, because business processes are

(a) Via **Callcenter**

(b) Via **Internet**

Figure 3. Process Visualization of the Refund Process for Cases Were Started Via the Call Center (a) and Via the Internet Portal (b). In the case of the internet cases missing information has to be requested too often. In the call center-initiated process, however, the problem does not exist.

> ❝Like Maury did with the naval log books, objective process maps can be derived that show how processes actually work in the real world❞

performed by multiple persons, often distributed across different organizational units or even companies.

Everybody only sees a part of the process. The manual discovery through classical workshops and interviews is costly and time-consuming, remaining incomplete and subjective. With Process Mining tools it is possible to leverage existing IT data from operational systems to quickly and objectively visualize the As-Is processes as they are really taking place.

In workshops with process stakeholders one can then focus on the root cause analysis and the value-adding process improvement activities.

3. A Case Study

In one of our projects we have analyzed a refund process of a big electronics manufacturer. The following process description has been slightly changed to protect the identity of the manufacturer. The starting point for the project was the feeling of the process manager that the process had severe problems. Customer complaints and the inspection of individual cases indicated that there were lengthy throughput times and other inefficiencies in the process.

The project was performed in the phases: First, the concrete questions and problems were collected, and the IT logs of all cases from the running business year were extracted from the corresponding service platform. Then, in an interactive workshop involving the process managers the log data were analyzed.

For example, in **Figure 3** you see a simplified fragment of the beginning of the refund process. On the left side (a) is the process for all cases that were initiated via the call center. On the right side (b) you see the same process fragment for all cases that were initiated through the internet portal of the manufacturer. Both process visualizations were automatically constructed using Fluxicon's process mining software Disco based on the IT log data that had been extracted.

The numbers, the thickness of the arcs, and the coloring all illustrate how frequently each activity or path has been performe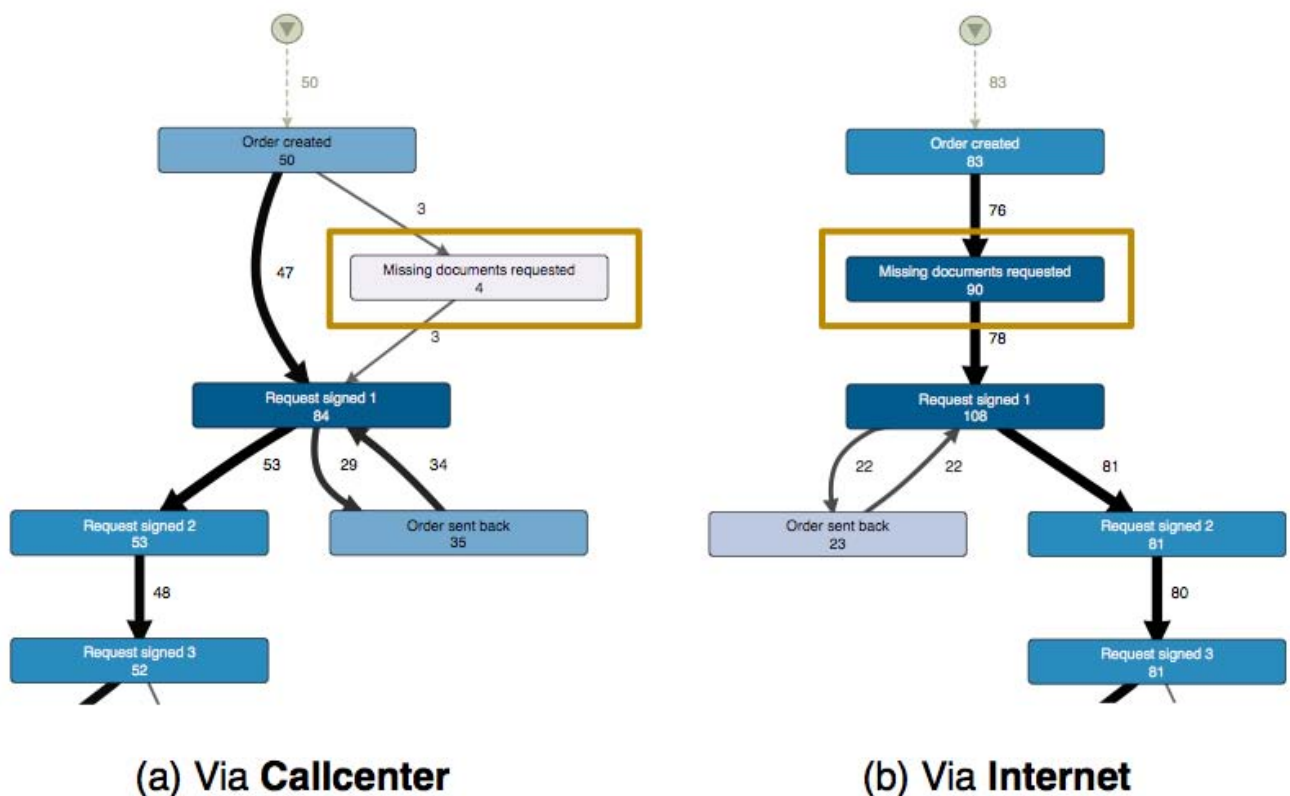d. For example, the visualization of the call center-initiated process is based on 50 cases (see left in **Figure 3**). All 50 cases start with activity Order created. Afterwards, the request is immediately approved in 47 cases. In 3 cases missing information has to be requested from the customer. For simplicity, only the main process flows are displayed here.

What becomes apparent in **Figure 3** is that, although missing information should only occasionally be requested from the customer,
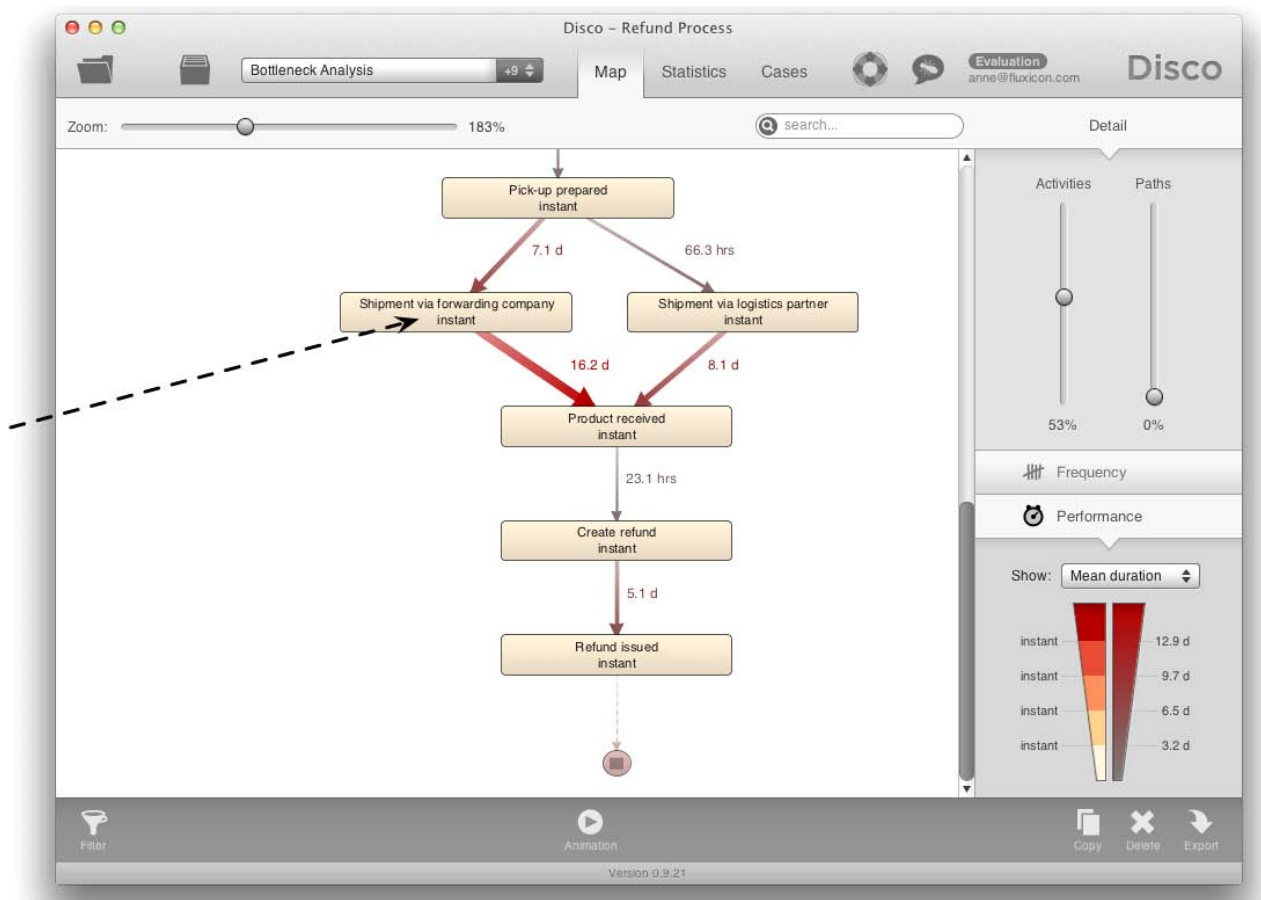
Figure 4. Screenshot of the Process Mining Software Disco in the Performance Analysis View. It becomes apparent that the shipment through the forwarding company causes a bottleneck.

this happens a lot for cases that are started via the internet portal: For 97% of all cases (77 out of 83 completed cases) this additional process step was performed. For 12 of the 83 analyzed cases (ca. 14%) this happened even multiple times (in total 90 times for 83 cases).

This process step costs a lot of time because it requires a call or an email from the service provider. In addition, through the external communication that is required, the process is delayed for the customer, who in a refund process has already had a bad experience. Therefore, the problem needs to be solved. An improvement of the internet portal (with respect to the mandatory information in the form that submits the refund request) could ensure that information is complete when the process is started.

Another analysis result was a detected bottleneck in connection with the pick-ups that were performed through the forwarding company. The process fragment in **Figure 4** shows the average waiting times between the process steps based on the timestamps in the historical data.

Also such waiting times analyses are automatically created by the process mining software. You can see that before and after the process step Shipment via forwarding company a lot of time passes. For example, it takes on average ca. 16 days between Shipment via forwarding company and Product received. The company discovered that the root cause for the long waiting times was that products were collected in a palette and the palette was shipped only when it was full, which led to delays particularly for those products that were placed in an almost empty palette. Also the actual refund process at the electronics manufacturer was taking too long (on average ca. 5 days). For the customer the process is only completed when she has her money back.

As a last result of the process mining analysis, deviations from the required process were detected. It is possible to compare the log data (and therewith the actual process) objectively and completely against required business rules, and to isolate those cases that show deviations. Specifically, we found that (1) in one case the customer received the refund twice, (2) in two cases the money was refunded without ensuring that the defect product had been received by the manufacturer, (3) in a few cases an important and mandatory approval step in the process had been skipped.

4. State of the Art

Process mining, which is still a young and relatively unknown discipline, is being made available by the first professional software tools on the market and supported by published case studies [4 |. The IEEE Task Force on Process Mining [5] was founded in 2009 to increase the visibility of process mining. In autumn 2011, it published a Process Mining Manifesto [6], which is available in 13 languages.

Companies already generate vast quantities of data as a byproduct of their IT-enabled business processes. This data can be directly analyzed by process mining tools. Like Maury did with the naval log books, objective process maps can be derived that show how processes actually work in the real world [7]. Developments in the field of Big Data are helping to store and access this data to analyze it effectively.

Matthew Fontaine Maury's wind and current books were so useful that by the mid-1850s, their use was even made compulsory by insurers [8] in order to prevent marine accidents and to guarantee plain sailing. Likewise, in Business process analysis and optimization, there will come a point when we can not imagine a time when we were ever without it and left to rely on our gut feeling.

▶ **References**

[1] Tim Zimmermann. *The Race: Extreme Sailing and Its Ultimate Event: Nonstop, Round-the-World, No Holds Barred*. Mariner Books, 2004. ISBN-10: 0618382704.

[2] W. Brian Arthur. *The Second Economy.* McKinsey Quarterly, 2011.

[3] Wil M.P. van der Aalst. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, 2011. ISBN-10: 3642193447.

[4] Alberto Manuel. *Process Mining - Ana Aeroportos de Portugal*, 2012. BPTrends, <www.bptrends.com>.

[5] IEEE Task Force on Process Mining. <http://www.win.tue.nl/ieeetfpm/>.

[6] IEEE Task Force on Process Mining. Process Mining Manifesto. Business Process Management Workshops 2011, *Lecture Notes in Business Information Processing, Vol. 99*, Springer-Verlag, 2011.

[7] Anne Rozinat. *How to Reduce Waste With Process Mining*, 2011. BPTrends, <www.bptrends.com>.

[8] Mark A. Thornton. *General Circulation and the Southern Hemisphere*, 2005. <http://www.lakeeriewx.com/Meteo241/ResearchTopicTwo/ProjectTwo.html>.