UNIVERSIDAD COMPLUTENSE MADRID

Facultad de Informática

Kybele Research Group

Process Mining

Closing the gap between Data Science and BPM

Conferencia de Posgrado

Universidad Rey Juan Carlos

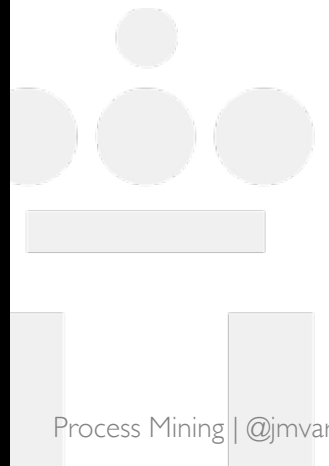
Juan Manuel Vara [Juancho]
juanmanuel.vara@urjc.es

@KybeleResearch | @jmvara

Intro

This aims to be a very applied talk

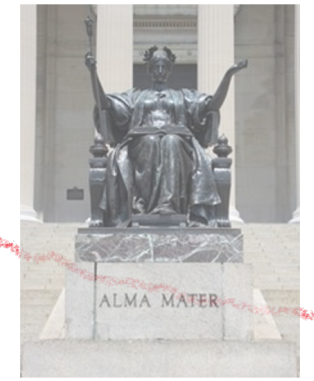
- The focus is not (only) to introduce the scientific basis of Process Mining
- But to show the potential of the topic
- Foster interest
  - To be aware that practical application is not only feasible but almost direct
  - To go deep into its application to your field of study
  - To have a first glance at the theory behind



The speaker

The beginning

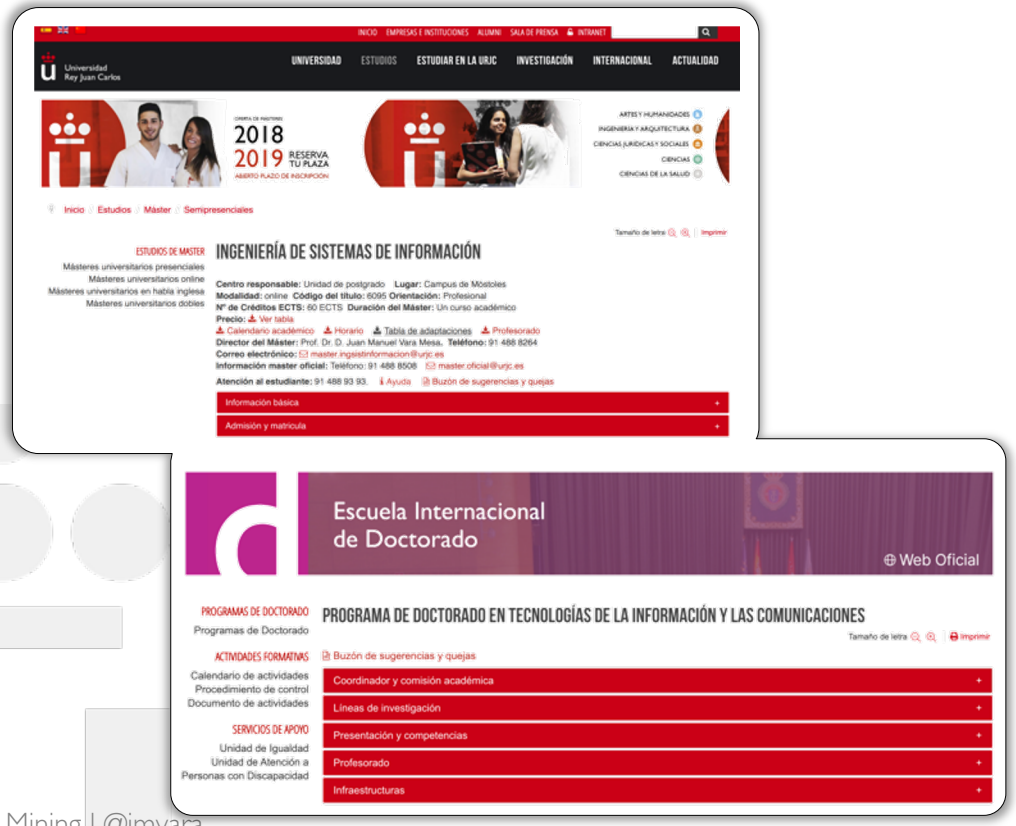
- Ingeniería Aeronáutica - UCM
- Ingeniería Técnica en Telecomunicaciones - UPM
- Ciencias Ambientales - URJC
- Ingeniería Informática – URJC
- IBERIA – Viva Tours



The speaker

Life at @URJC

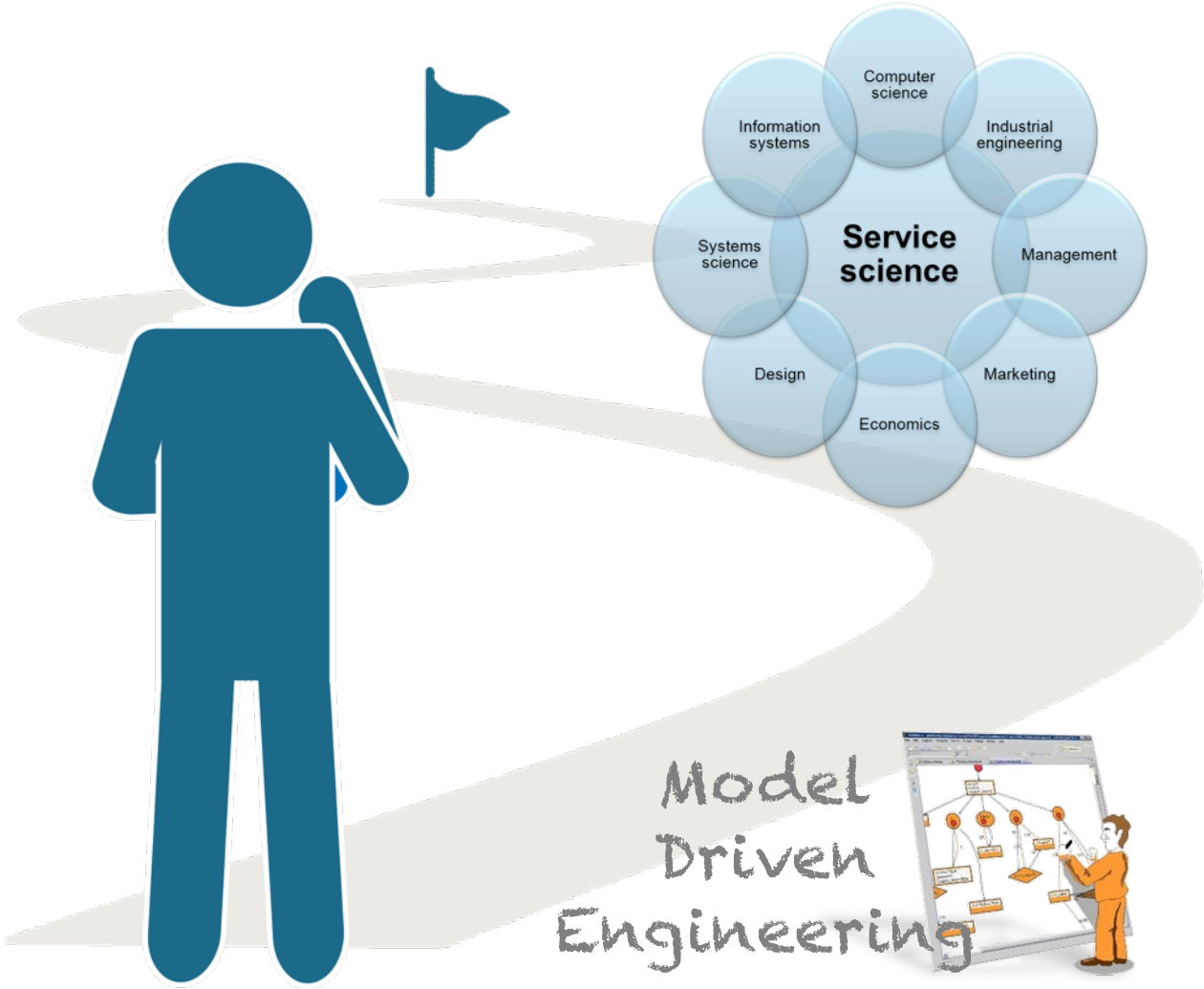
- Head of the MsC in Information Systems Engineering
- Head of the SE, Informations Systems and Service Engineering at the PhD Program on TIC
- Head of the BsC in Services Engineering



The speaker

My life as a researcher

- Pre-doc at U. Nantes (Jean Bézivin)
- Post-doc at Tilburg University (Mike Papazoglou)
- 2 Sexenios / 1 Transferencia
- 4 Docentia
- 7 Tesis Doctorales en los últimos 10 años
- 18 artículos JCR – 16 congresos CORE
- IP proyectos MINECO
- Investigador H2020
- …



Agenda

Motivation

What is in Process Mining?

Quick glance at Data Handling

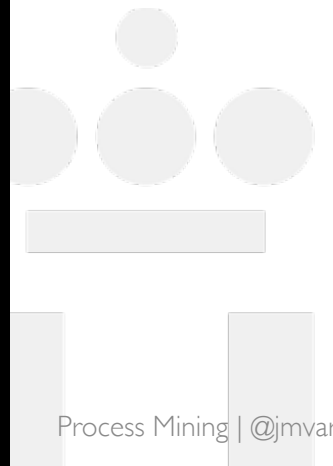
A glance at Process Discovery

Quick tour with Disco

- Process Mining tools overview

Challenges & Open Issues

Motivation



Business Process Management

One straightforward way of looking at BPM is to think of 10 W's associated with a business

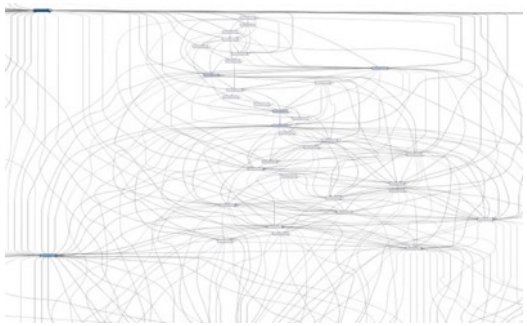
By understanding WHAT is done, WHY, WHERE and WHEN it is done, WHO does it, the WAY it is done, and WITH what resources, all while keeping a WATCH on performance, identifying what to WORRY about and how to mitigate those worries, a range of methods and tools can be applied to achieve the best results with the least cost and achieving the most WEALTH.

Business Process Management (BPM) is the art and science of overseeing how work is performed in an organization to ensure consistent outcomes and to take advantage of improvement opportunities.

A collection of discrete activities or events we perform in order to achieve a certain **goal**

https://www.igrafx.com/articles/why-business-process-management/

BPM Issues

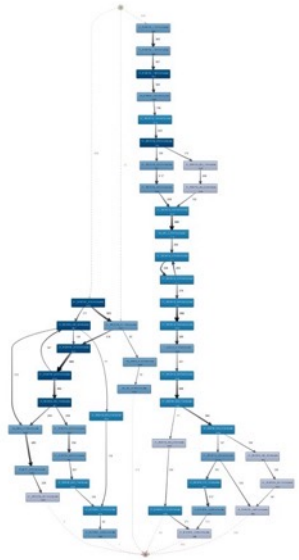


Creation of "current state" processes

- Business Process reengineering focus on an improved to-be process
- Little interest in exploring the as-is one, which is the key to:
  - Know whether it's worh to invest in improvements
  - Detect performance problems
  - How much variation there is in the process across the organization
- Skip or subcontract process analysis
- Spend too much time in interviews and sticky notes based process analysis

80/20

- 80% of the process instances are described by 20% of the variants
- The remaing 20% of process instances make up for 80% of the problems (deviations, delays, complaints, etc.)



Lack of connections between BPs and ISs

- Gathering information about how your process is performing entailed manual and complex steps to collect and synthetize data
  - E.g.: Lean & Six Sigma have not emphasized technologies as enables of Process Management

(Davenport & Spanyi, 2019)

- Data-based as-is model
- BPM initatives connected to ISs

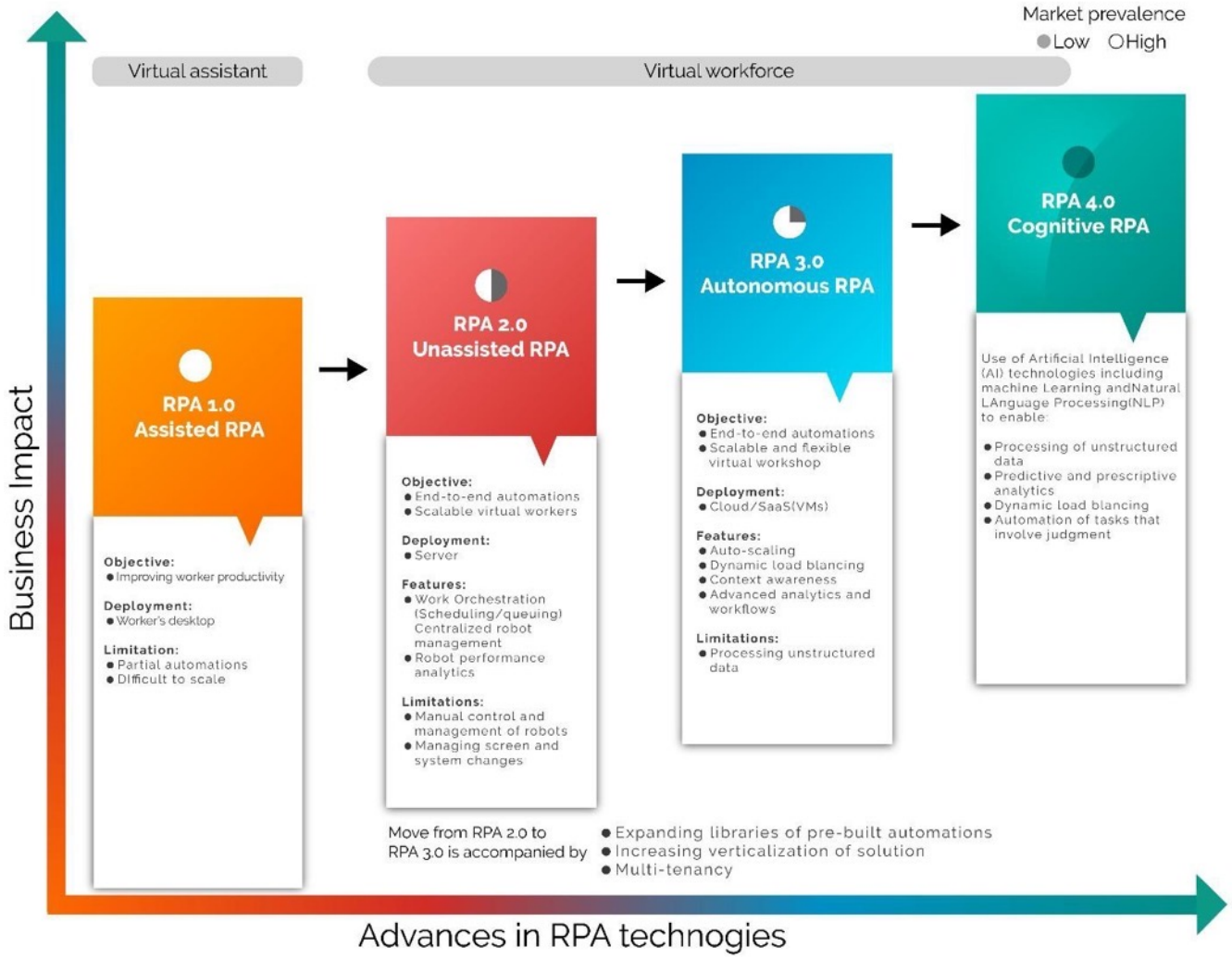
The promise of RPA

Robotic Process Automation

- APIfication
- Incorporación de ML: cognitive RPA
- Low-code tools

> Companies will flesh out their stacks by combining **RPA**, traditional **BPM**, workflow, and cloud services, layering in **process documentation**, **process discovery and analytics**, and looking to enable a 'virtuous cycle' of process improvement.
>
> Ryan Duguid, Nentix. From: RPA (Robotic Process Automation): What's In Store For 2020?. Forbes, Dec.2019

- Identify where to implement bots
- Calculate the impact of the implementation

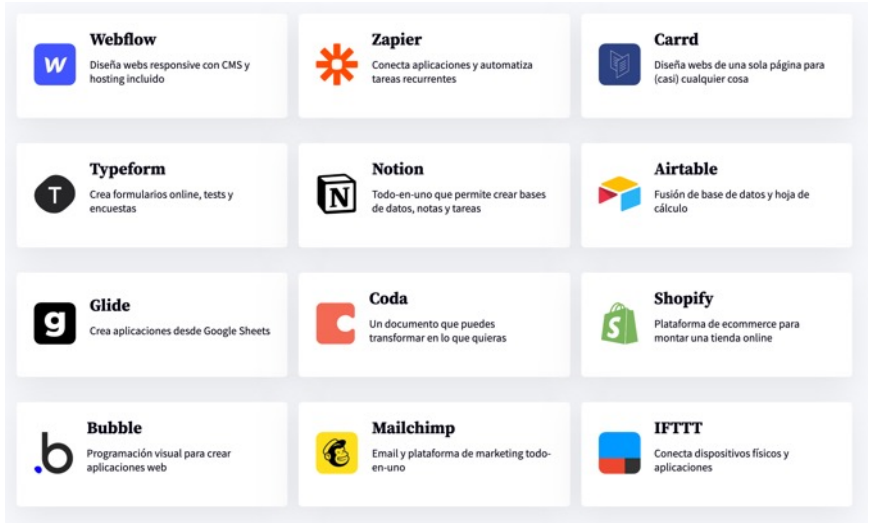


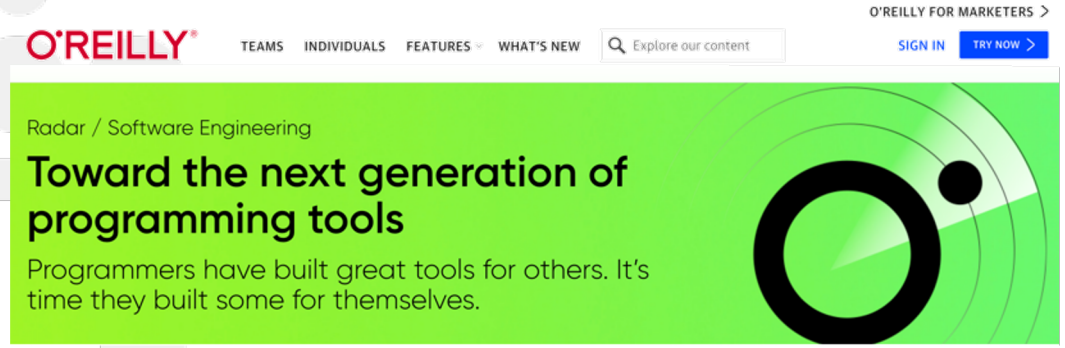
The Hype of low-code

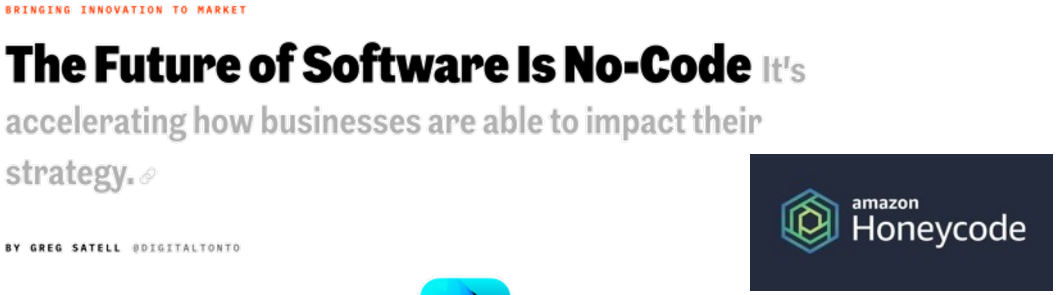
Sortable list of low-code platforms
https://www.managedfunctions.com/pirate

The rise of low-code & no-code tools

- Comunity or Citizen Developers
- Low-code for *pro-developers*







amazon Honeycode

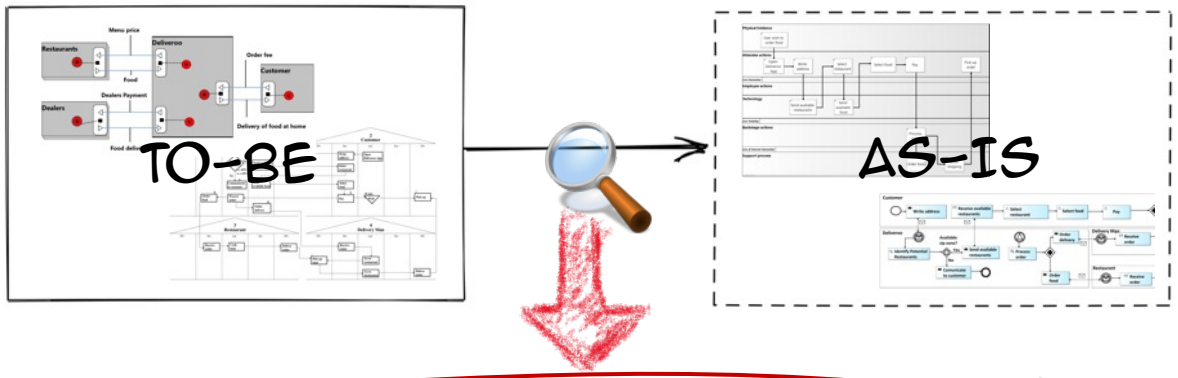
SwiftUI
Better apps. Less code.

Google App Maker

Microsoft | Power Apps

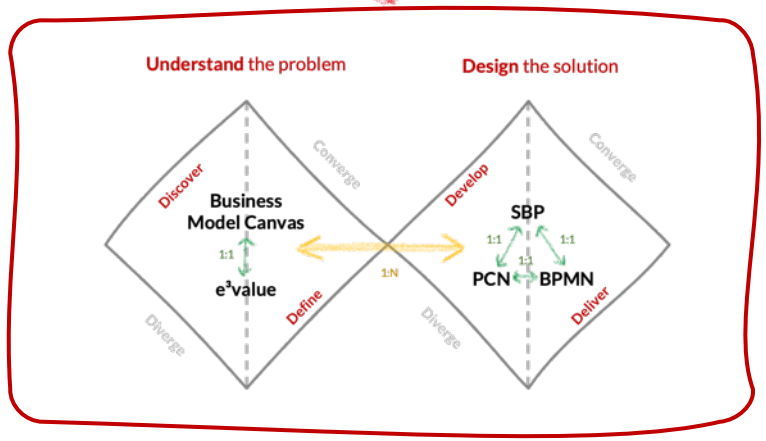
(Bridgewater, 2021)

Our interest on Service Design



How do we know what was really going on there?

How do we know what is really going on after our intervention?



INNoVaServ: model-based tool support for Service Design

Pérez-Blanco, F. J., Vara, J. M., Gómez, C., De Castro, V., & Marcos, E. (2020, April). Model-based tool support for Service Design. In *International Conference on Fundamental Approaches to Software Engineering* (pp. 266-272). Springer, Cham.

ETAPS EUROPEAN JOINT CONFERENCES ON THEORY & PRACTICE OF SOFTWARE

Gómez, C., Pérez Blanco, F. J., Vara, J. M., De Castro, V., & Marcos, E. (2021, January). Design and development of Smart Contracts for E-government through Value and Business Process Modeling. In Proceedings of the 54th Hawaii International Conference on System Sciences (p. 2069).

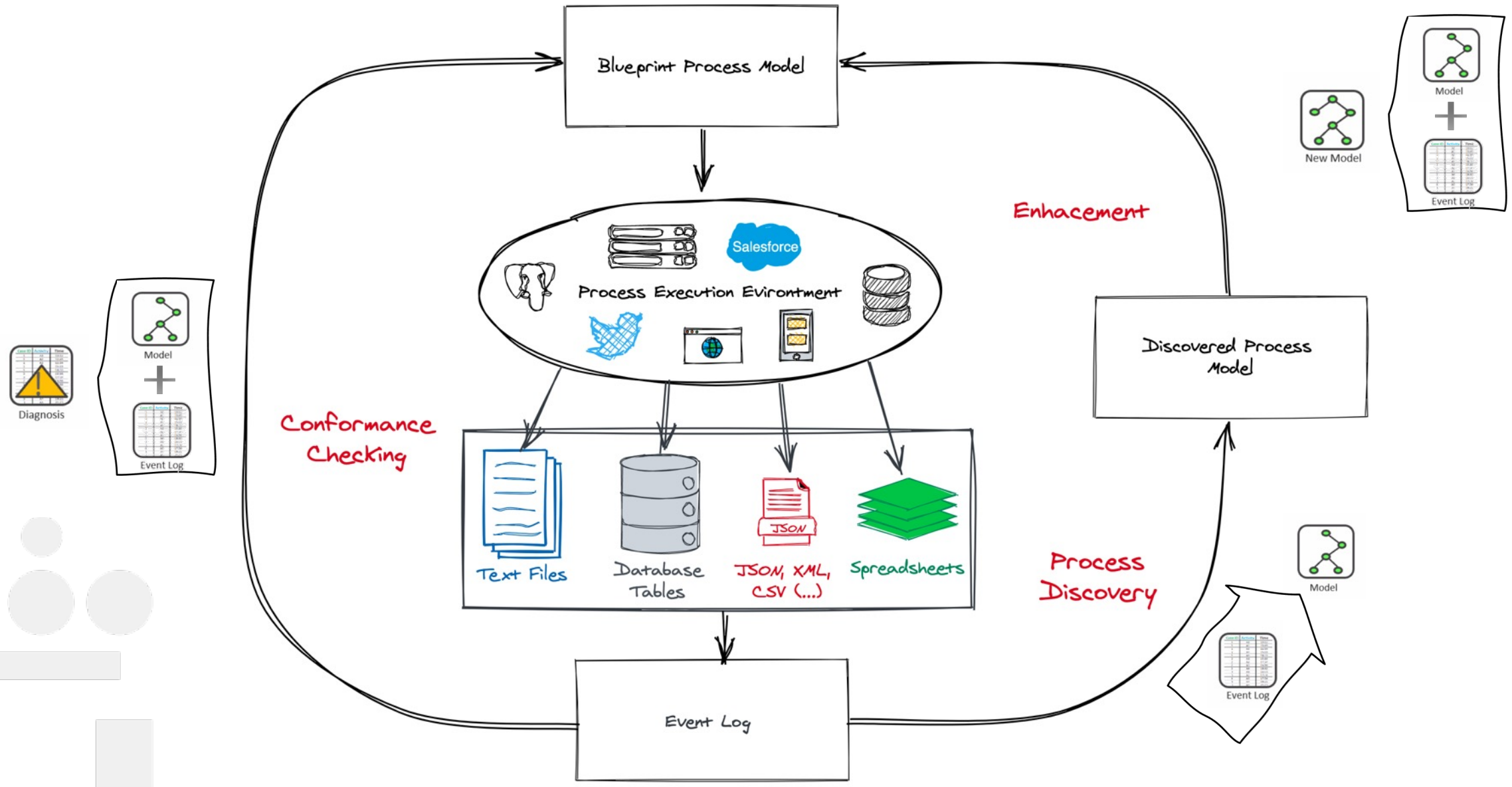
HICSS

https://modeling-languages.com/model-based-tool-for-business-service-design/

No one wants to ends like this

Process Mining
Data-driven approach to discovering gaps in your business process

Process Mining Overview

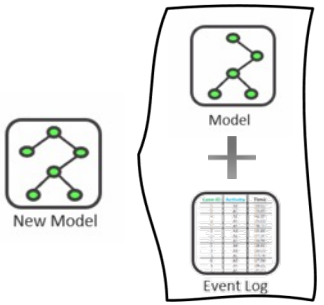


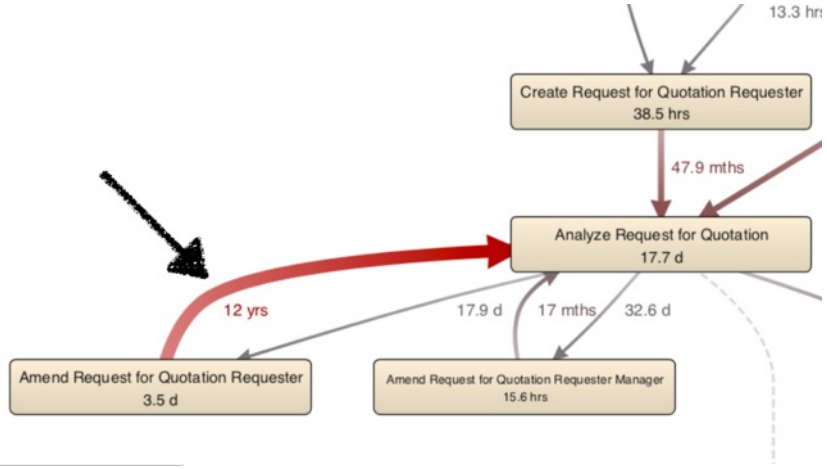
Adapted from @mathias_weske:
https://twitter.com/mathias_weske/status/1388036059871784962?s=20

A more pragmatical contribution

I'm personally not sure about the *enhacement* issue

More prone to think about



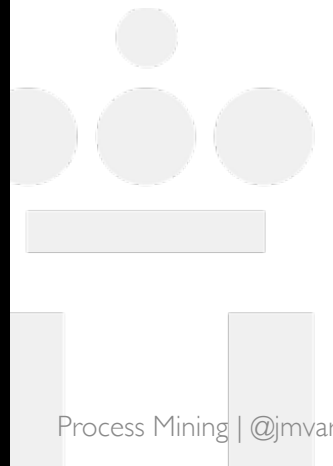


Throughput analysis/bottleneck detection

Accounting for the intensity of events' execution (measured by time spent to complete a particular event) in order to determine potential bottlenecks.

This kind of analysis can be used to improve time-related KPIs by minimizing throughput/overhead time

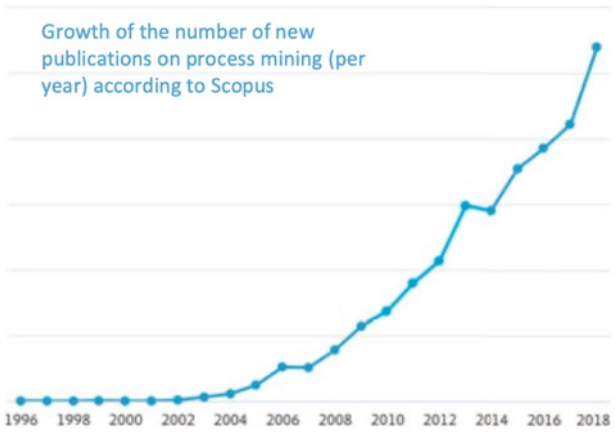
What is in Process Mining?



A bit of history

One man's Dream

- Wil van der Aalst dissapointment on the gap between **modelled processes and reality**
  - Late 90's sabatical at U. of Colorado in Boulder
  - Early 2000s the **α**-algorithm
  - 2004: first **ProM** release
  - 2007: foundation of *Futura PI* (van de Brand)
    - **Fluxicon**, **Celonis**, ProcessGold
  - 2014: Coursera Process Mining course (120k participants)
  - 2015: practical adoption of Process Mining in the industry



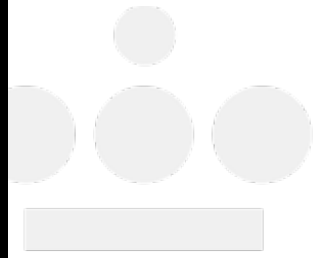
- 1999 Start of process mining research at TU/e
- 2000-2002 Development of the first process mining algorithms (Alpha and Heuristic miner)
- 2004 Release of the first version of the ProM framework
- 2004-2006 Pioneering work on token-based conformance checking, organization mining, decision mining, etc.
- 2007 First process mining company (Futura PI)
- 2010 Pioneering work on alignment-based conformance checking and operational support
- 2009-2011 Founding of Fluxicon, Celonis, and ProcessGold
- 2011 First process mining book
- 2014 Coursera process mining MOOC
- 2016 "Process mining data science in action" book
- 2018 First Market Guide for Process Mining by Gartner
- 2018 Over 30 software vendors offer process mining tools
- 2018 Celonis becomes a Unicorn
- 2019 First international process mining conference takes place in Aachen (ICPM 2019)

(Van der Aalst, 2020)

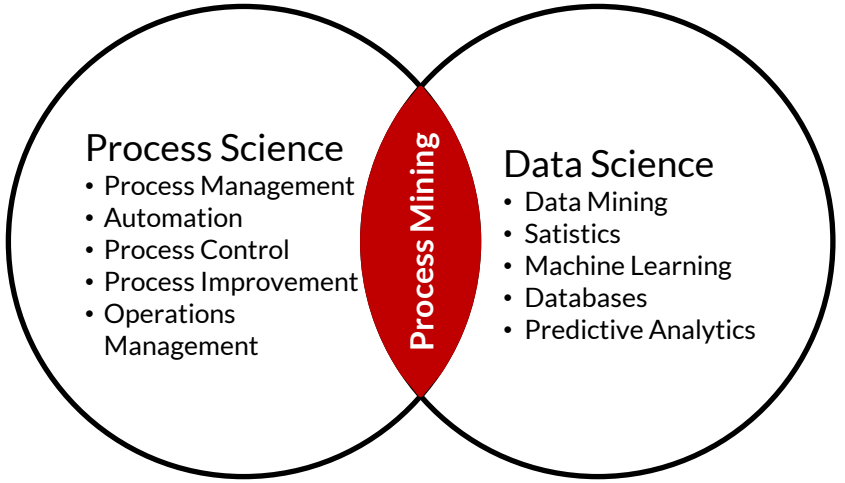
What is Process Mining?

Process Mining

- Lies at the intersection of process-science and data-science, bridging the gap between model-based process analytics and data-centered analytics.
- It allow users to reconstruct, analyze and improve business processes based on event logs from transactional IT systems like SAP, Oracle, Salesforce, etc.
- Undesired process patterns, bottlenecks and compliance issues can be detected and reduced in that way. This allows us to understand processes as they really happened based on their digital footprint.
- This reflects a major advantage to previous methods of process mining and modeling as findings are no longer based o assumptions but on data-based evidence reflecting real-world events.



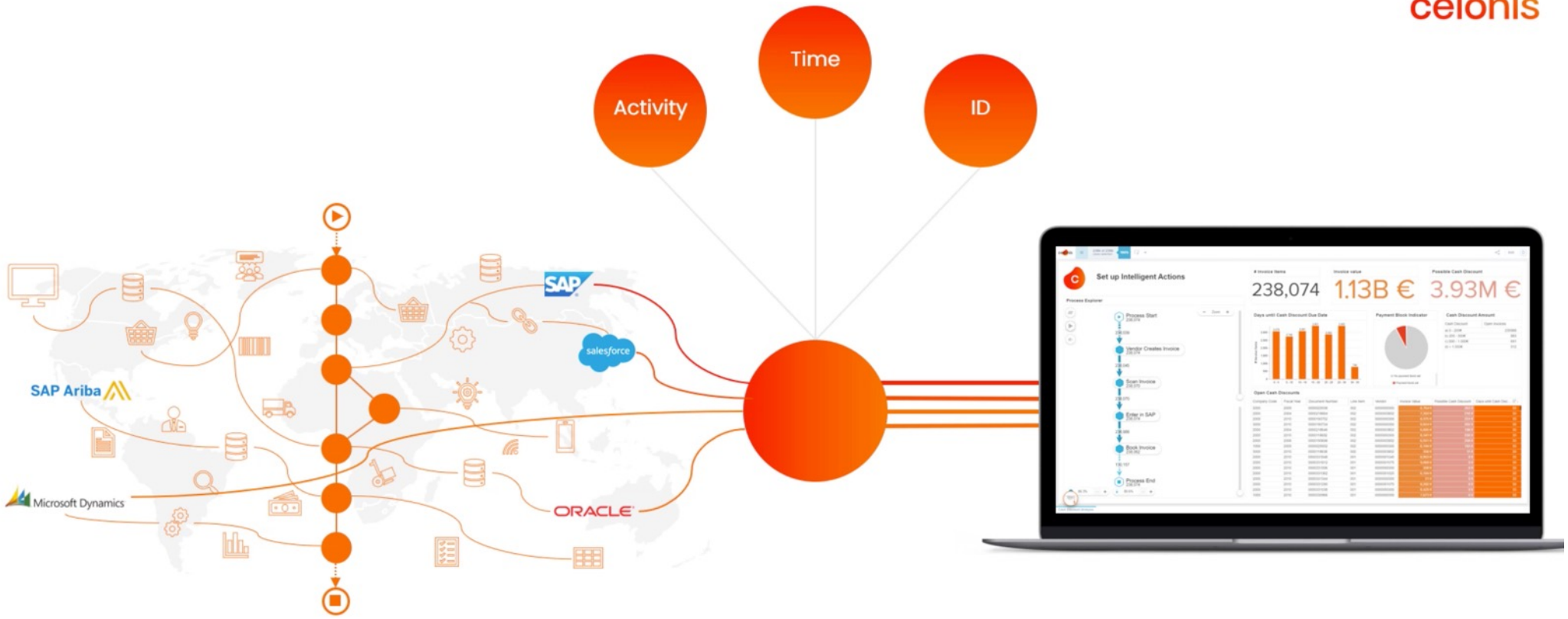
> The end goal of process mining is to discover, model, monitor, and optimize the underlying processes by analyzing the event data, generated during the execution of those processes.



Adapted from (van Der Aalst, 2016)

What is in Process Mining

It's basically about analysing **digital footprints**, don't matter what are they



IT-based work
Every workflow is supported by IT systems, like SAP, Oracle or Salesforce

Digital Footprints
Process Minig algorithms detect & reconstruct workflow traces

Full Transparency
Actual process executions are visualized in the tool

What do we need for Process Mining

Data generation

- Business process data is stored in a variety of IT systems, suh as workflow management systems, ERP systems, CRM systems, supply chain management systems, etc.

Celonis (2020)

Order No.	Activity	Time	User	Quantity
10001	Create purchase order	01-01-2009, 8:35 am	Sara Jones	1
10001	Print and send purchase order	03-01-2009, 12:13 am	Sara Jones	1
10001	Goods receipt	07-01-2009, 07:01 am	Pete Scott	1
10001	Scan invoice	09-01-2009, 2:00 pm	Sara Jones	1
10001	Book invoice	10-01-2009, 10:30 am	Carol Hope	1
10002	Create purchase requisition	02-02-2009, 1:17 pm	John Farmer	15
10002	Create purchase order	04-02-2009, 9:15 am	Sara Jones	15
10002	Print and send purchase order	07-02-2009, 4:41 pm	Sara Jones	15
10002	Goods receipt	27-02-2009, 6:53 am	Frank Miller	15
10002	Scan invoice	28-02-2009, 1:00 pm	Sara Jones	15
10002	Book invoice	13-03-2009, 11:59 am	Carol Hope	15
10003	Scan invoice	13-04-2009, 10:00 am	Sara Jones	23
10003	Create purchase order	17-04-2009, 3:47 pm	Sara Jones	23
10003	Print and send purchase order	17-04-2009, 5:30 pm	Carol Jope	23
10003	Goods receipt	27-04-2009, 4:23 pm	Pete Scott	23
10003	Book invoice	30-04-2009, 8:50 am	Sara Jones	23
...	...	...	...	...

Event

Trace

Case ID | Activity name | Timestamp | Resource | Other data

Event logs

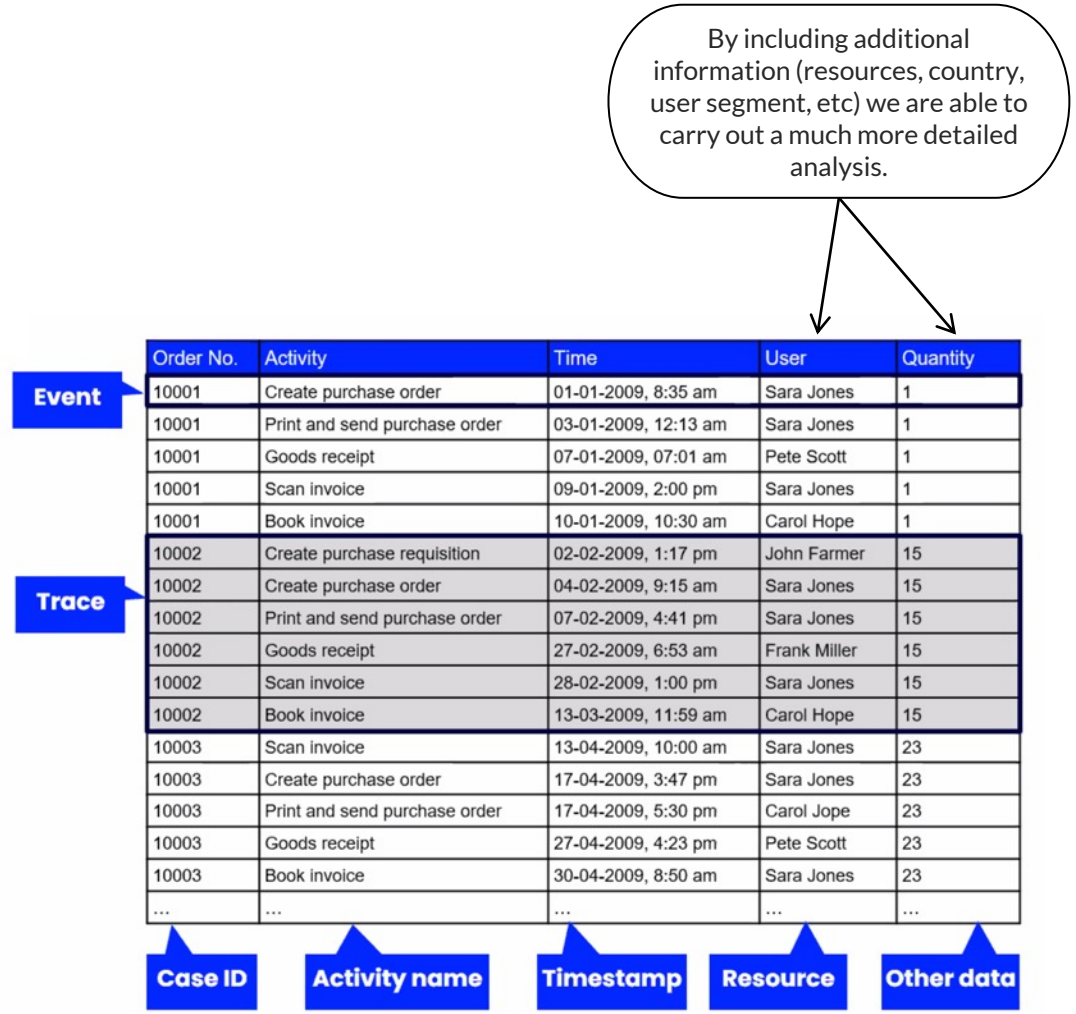
- Store the data that is required for Process Mining.
- At the minimum, the event log has to include three different fields:
  - Case ID
  - Activity Name
  - Timestamp

  Digital Footprint
- There may be other optional fields

Adapted from: van Der Aalst (2016)

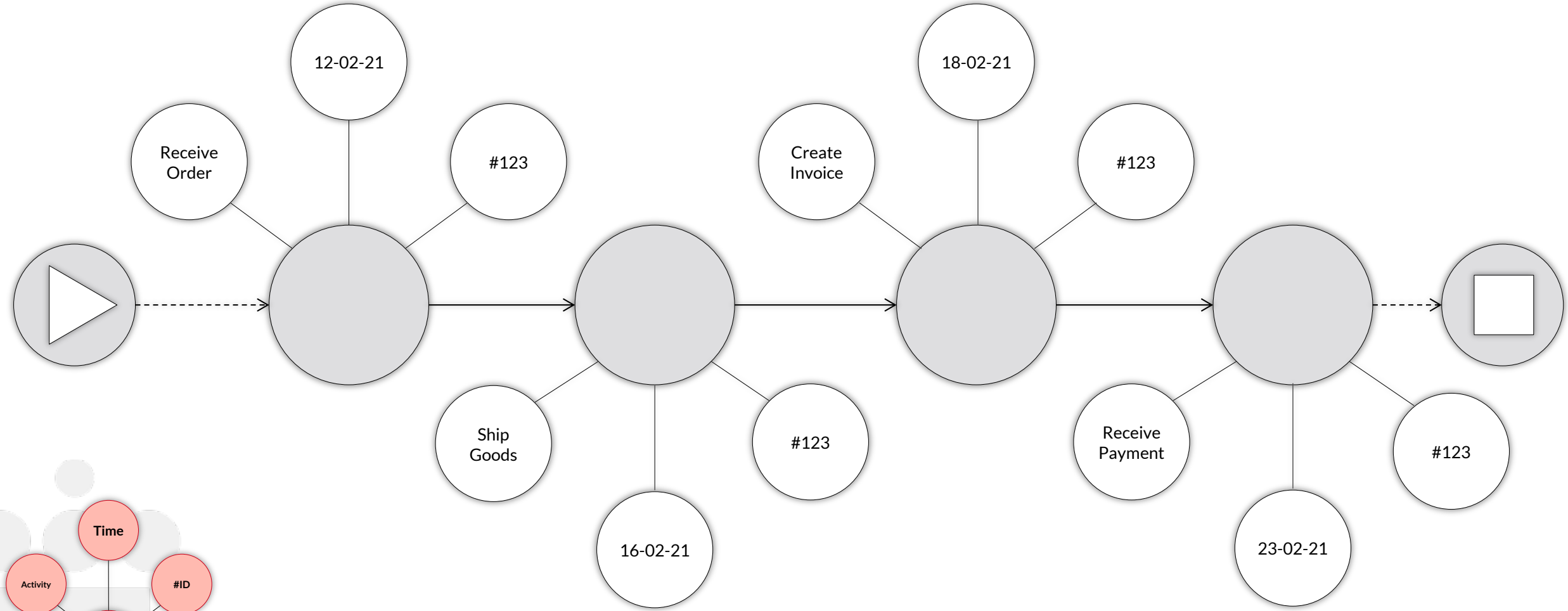
Digital Footprints

- Digital Footprint (key parts of an event)
  - Case ID: indicates which process instance the event belongs to.
    - A unique identifier of an entity going through the process
  - Activity / Event: describes the action that is captured by the event
    - A step of the process, any activity that is a part of the process we are analyzing
  - Timestamp: indicates the time when the event took place
    - used for performance evaluation and determining the order of events, can be the time when the user entered/exited the given event (or both actually)



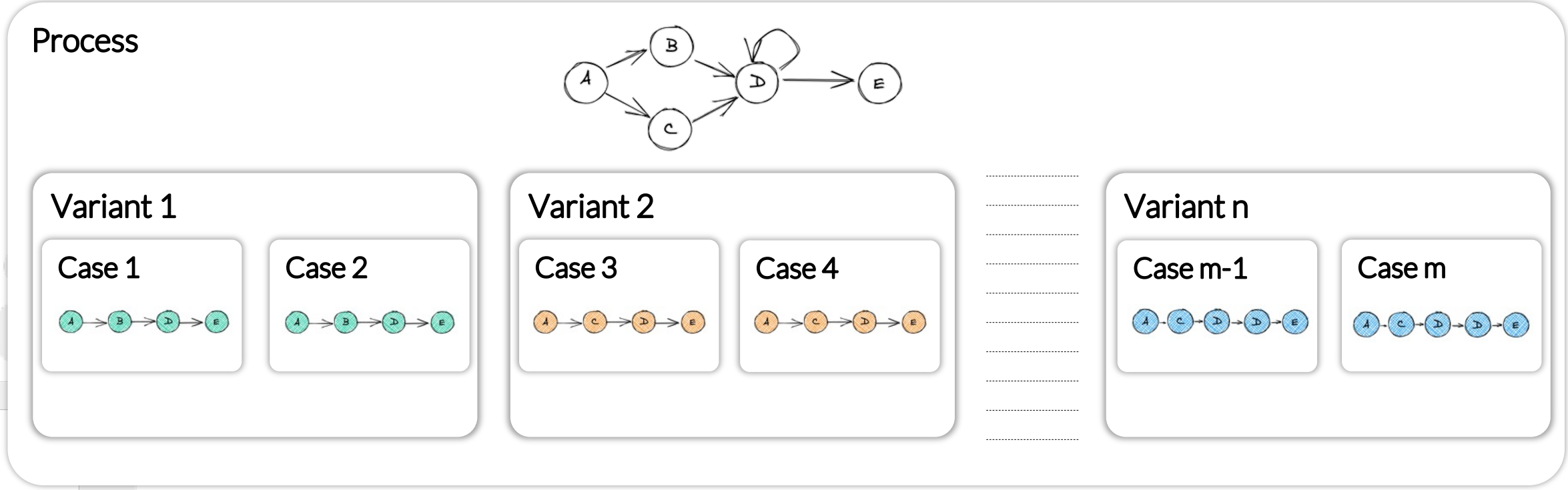
Order No.	Activity	Time	User	Quantity
10001	Create purchase order	01-01-2009, 8:35 am	Sara Jones	1
10001	Print and send purchase order	03-01-2009, 12:13 am	Sara Jones	1
10001	Goods receipt	07-01-2009, 07:01 am	Pete Scott	1
10001	Scan invoice	09-01-2009, 2:00 pm	Sara Jones	1
10001	Book invoice	10-01-2009, 10:30 am	Carol Hope	1
10002	Create purchase requisition	02-02-2009, 1:17 pm	John Farmer	15
10002	Create purchase order	04-02-2009, 9:15 am	Sara Jones	15
10002	Print and send purchase order	07-02-2009, 4:41 pm	Sara Jones	15
10002	Goods receipt	27-02-2009, 6:53 am	Frank Miller	15
10002	Scan invoice	28-02-2009, 1:00 pm	Sara Jones	15
10002	Book invoice	13-03-2009, 11:59 am	Carol Hope	15
10003	Scan invoice	13-04-2009, 10:00 am	Sara Jones	23
10003	Create purchase order	17-04-2009, 3:47 pm	Sara Jones	23
10003	Print and send purchase order	17-04-2009, 5:30 pm	Carol Jope	23
10003	Goods receipt	27-04-2009, 4:23 pm	Pete Scott	23
10003	Book invoice	30-04-2009, 8:50 am	Sara Jones	23
...	...	...	...	...

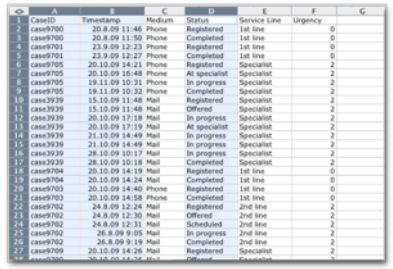
Digital Footprints



Cases & Variants

- Case / Trace: a particular sequence of activities
  - A case usually consists of multiple events
- Variant: the set of cases/traces that perform the very same sequence of activities ≅ Process Pattern

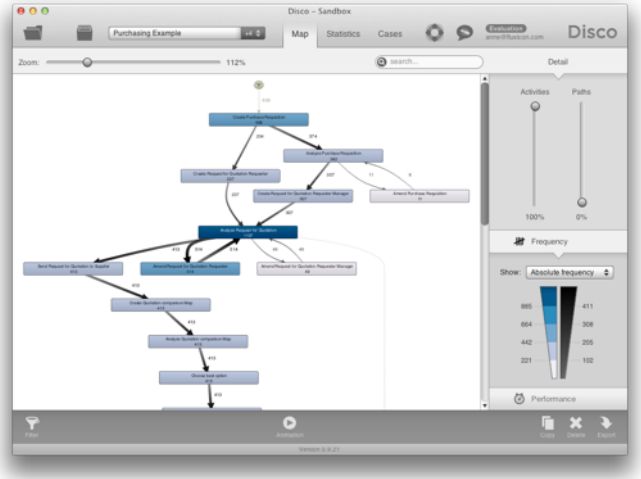




	A	B	C	D	E	F	G
1	CaseID	Timestamp	Medium	Status	Service Line	Urgency	
2	case9700	20.8.09 11:46	Phone	Registered	1st line	0	
3	case9700	20.8.09 11:50	Phone	Completed	1st line	0	
4	case9701	23.9.09 12:23	Phone	Registered	1st line	0	
5	case9701	23.9.09 12:27	Phone	Completed	1st line	0	
6	case9705	20.10.09 14:21	Phone	Registered	Specialist	2	
7	case9705	20.10.09 16:48	Phone	At specialist	Specialist	2	
8	case9705	19.11.09 10:31	Phone	In progress	Specialist	2	
9	case9705	19.11.09 10:32	Phone	Completed	Specialist	2	
10	case3939	15.10.09 11:48	Mail	Registered	Specialist	2	
11	case3939	15.10.09 11:48	Mail	Offered	Specialist	2	
12	case3939	20.10.09 17:18	Mail	In progress	Specialist	2	
13	case3939	20.10.09 17:19	Mail	At specialist	Specialist	2	
14	case3939	21.10.09 14:49	Mail	In progress	Specialist	2	
15	case3939	21.10.09 14:49	Mail	In progress	Specialist	2	
16	case3939	28.10.09 10:17	Mail	In progress	Specialist	2	
17	case3939	28.10.09 10:18	Mail	Completed	Specialist	2	
18	case9704	20.10.09 14:19	Mail	Registered	1st line	0	
19	case9704	20.10.09 14:24	Mail	Completed	1st line	0	
20	case9703	20.10.09 14:40	Phone	Registered	1st line	0	
21	case9703	20.10.09 14:58	Phone	Completed	1st line	0	
22	case9702	24.8.09 12:24	Mail	Registered	2nd line	2	
23	case9702	24.8.09 12:30	Mail	Offered	2nd line	2	
24	case9702	24.8.09 12:31	Mail	Scheduled	2nd line	2	
25	case9702	26.8.09 9:05	Mail	In progress	2nd line	2	
26	case9702	26.8.09 9:19	Mail	Completed	2nd line	2	
27	case9709	20.10.09 14:26	Mail	Registered	Specialist	2	

Quick glance

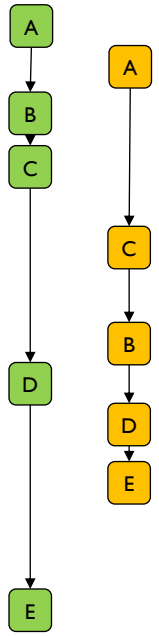
(Data import)



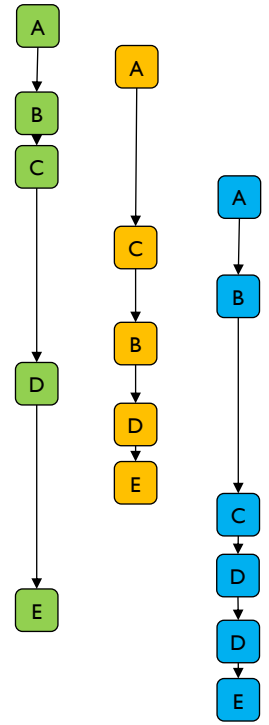
How does it work?

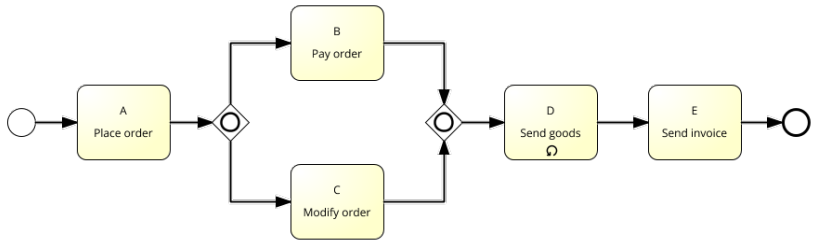
Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E

Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E



Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E

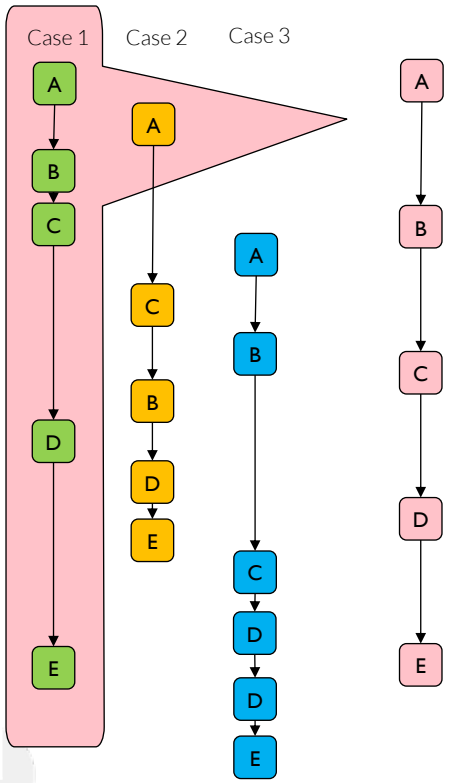


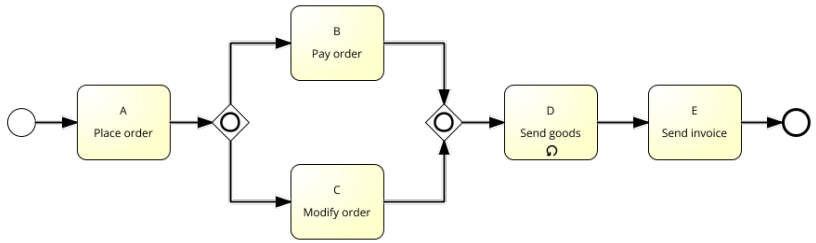


https://fluxicon.com/book/read/intro/#how-does-it-work

How does it work?

Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E

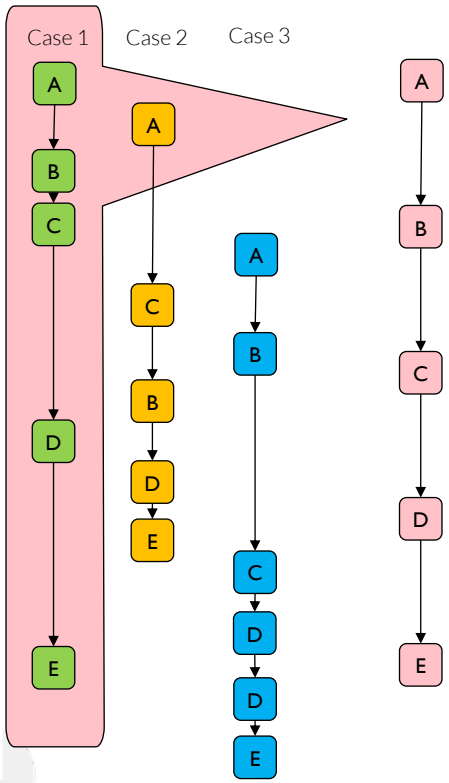


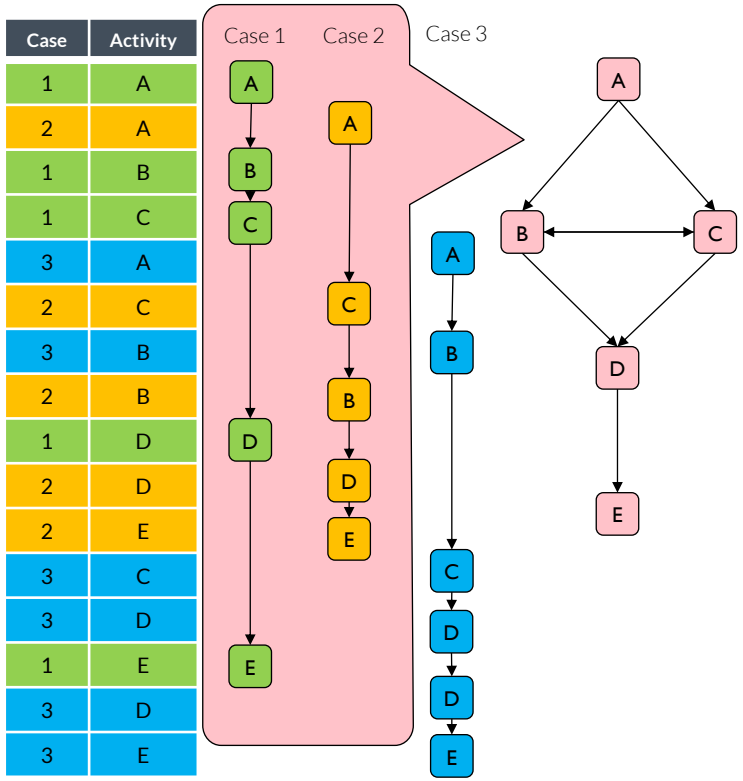


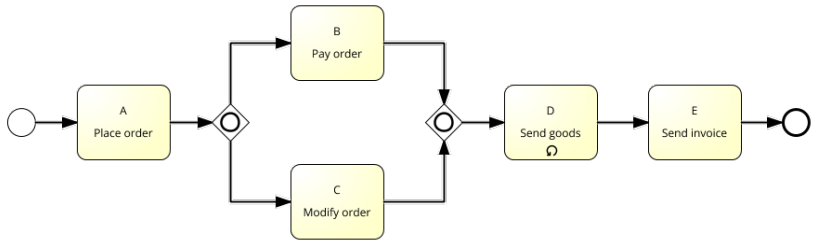
https://fluxicon.com/book/read/intro/#how-does-it-work

How does it work?

Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E



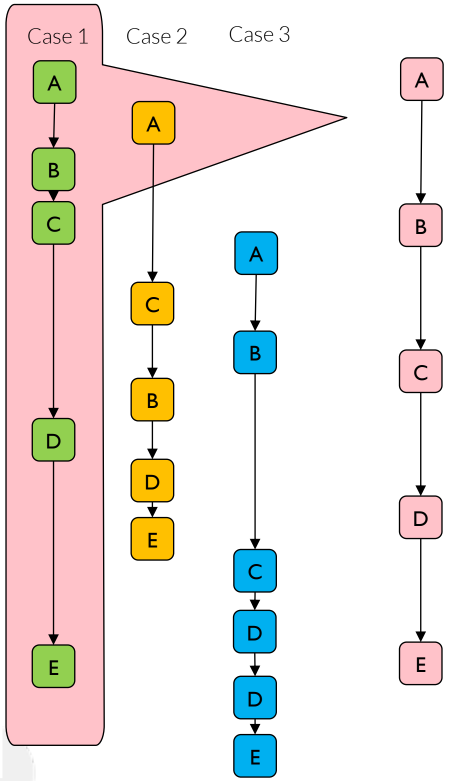




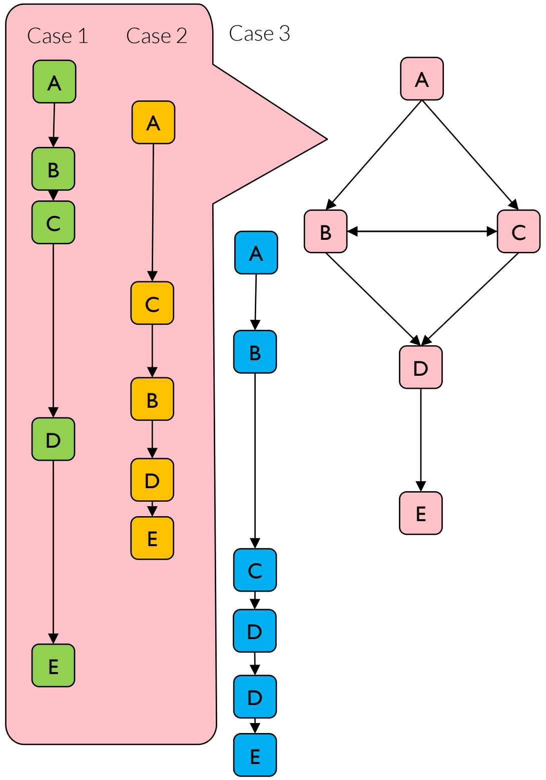
https://fluxicon.com/book/read/intro/#how-does-it-work

How does it work?

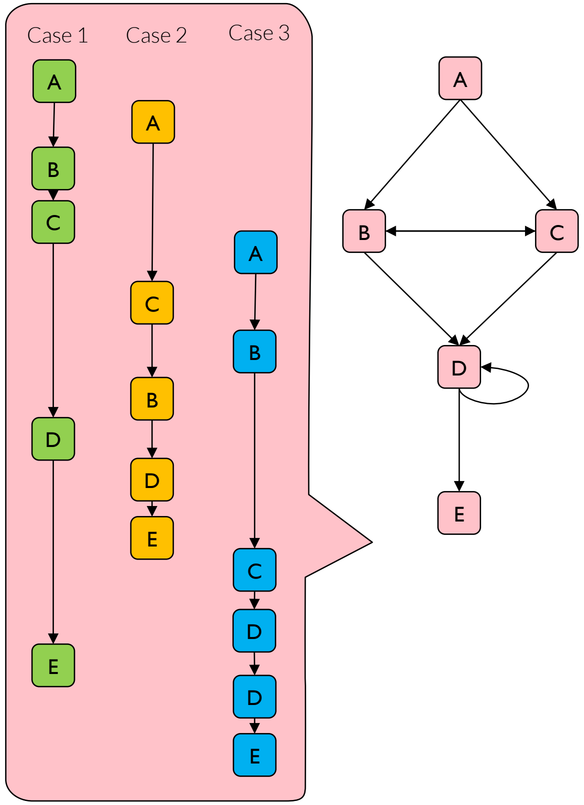
Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E

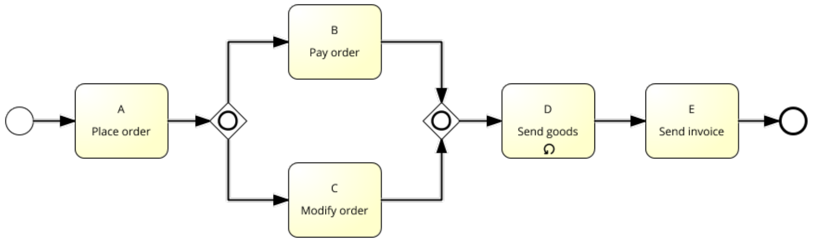


Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E



Case	Activity
1	A
2	A
1	B
1	C
3	A
2	C
3	B
2	B
1	D
2	D
2	E
3	C
3	D
1	E
3	D
3	E





https://fluxicon.com/book/read/intro/#how-does-it-work

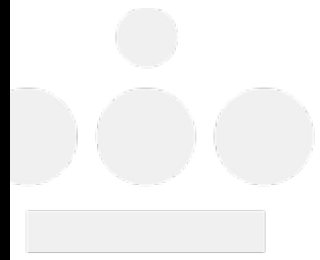
A glance at process discovery

(Van der Aalst, 2016)

The Alpha Miner

Algorithm

- Scan traces for ordering relations
  - Build footprint matrix
- Convert footprint matrix to Petri Net



The Alpha Miner

Build footprint matrix

- Four ordering relations
  - $>$ Directly follows — a > b — a is directly followed by b
  - $\rightarrow$ Sequence — a $\rightarrow$ b — if a > b and not b > a
  - $\|$ Parallel — a $\|$ b — if both a > b and b > a
  - # Choice — a # b — if neither a > b and b > a

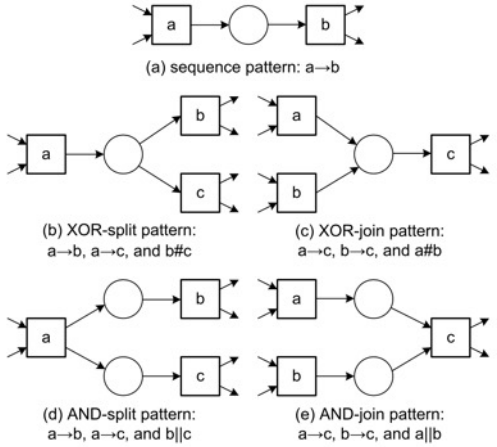
$L_1 = [\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle^2, \langle a, e, d\rangle]$

$\langle a, b, c, d\rangle$
$\langle a, b, c, d\rangle$
$\langle a, b, c, d\rangle$
$\langle a, c, b, d\rangle$
$\langle a, c, b, d\rangle$
$\langle a, e, d\rangle$

	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\|_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\|_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

The Alpha Miner

Convert footprint matrix to Petri Net



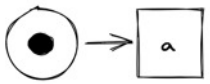
$L_1 = [\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle^2, \langle a, e, d\rangle]$

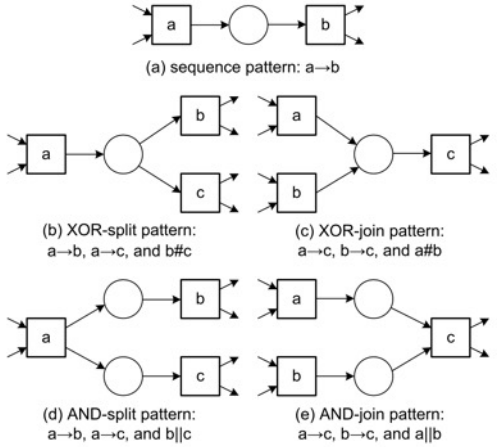
⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, c, b, d⟩
⟨a, c, b, d⟩
⟨a, e, d⟩

	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\|_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\|_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

The Alpha Miner

Convert footprint matrix to Petri Net





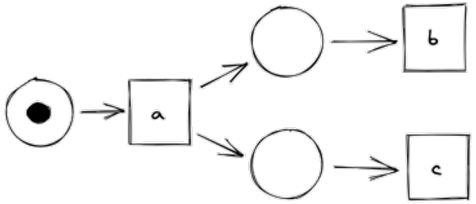
$L_1 = [\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle^2, \langle a, e, d\rangle]$

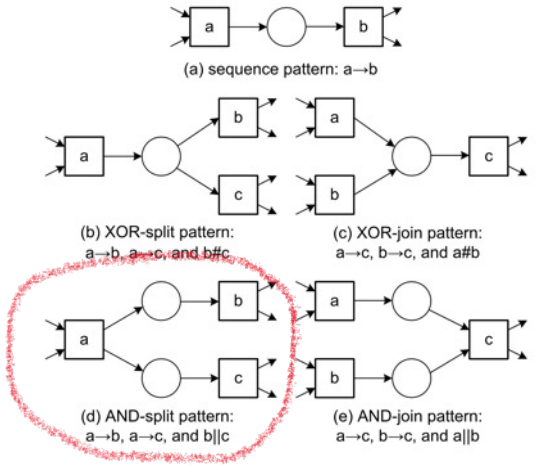
⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, c, b, d⟩
⟨a, c, b, d⟩
⟨a, e, d⟩

	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\|_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\|_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

The Alpha Miner

Convert footprint matrix to Petri Net





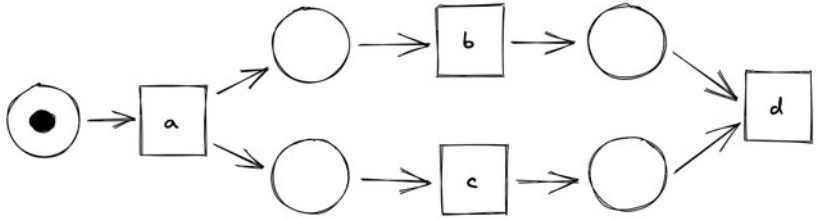
$L_1 = [\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle^2, \langle a, e, d\rangle]$

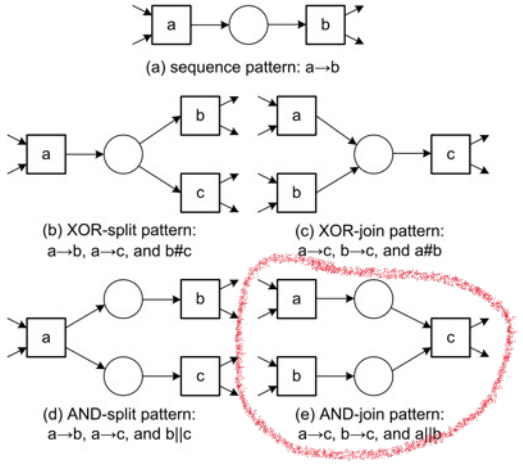
⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, c, b, d⟩
⟨a, c, b, d⟩
⟨a, e, d⟩

	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\|\|_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\|\|_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

The Alpha Miner

Convert footprint matrix to Petri Net





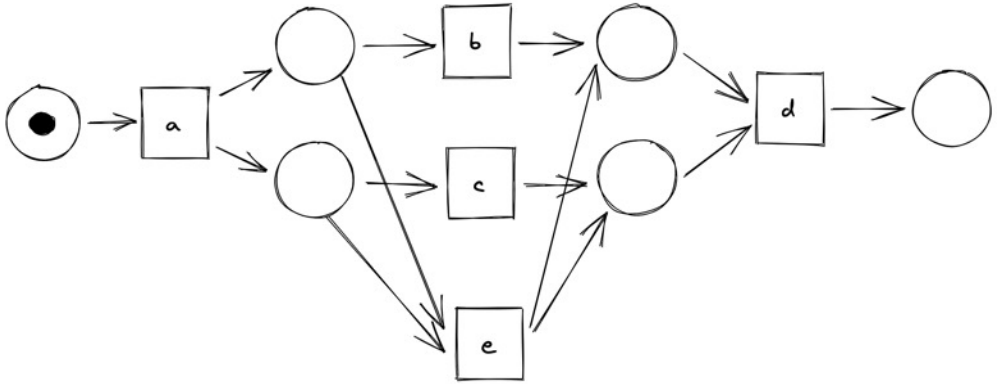
$L_1 = [\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle^2, \langle a, e, d\rangle]$

⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, c, b, d⟩
⟨a, c, b, d⟩
⟨a, e, d⟩

	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\|\|_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\|\|_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

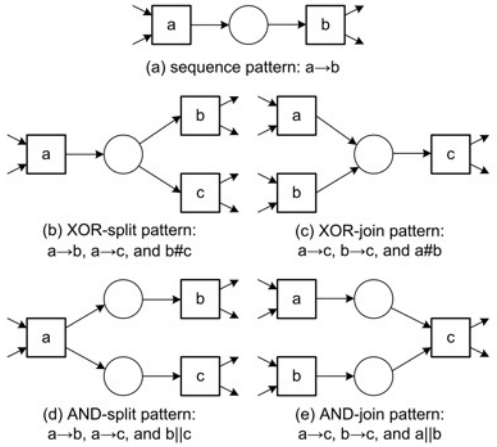
The Alpha Miner

Convert footprint matrix to Petri Net



$L_1 = [\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle^2, \langle a, e, d\rangle]$

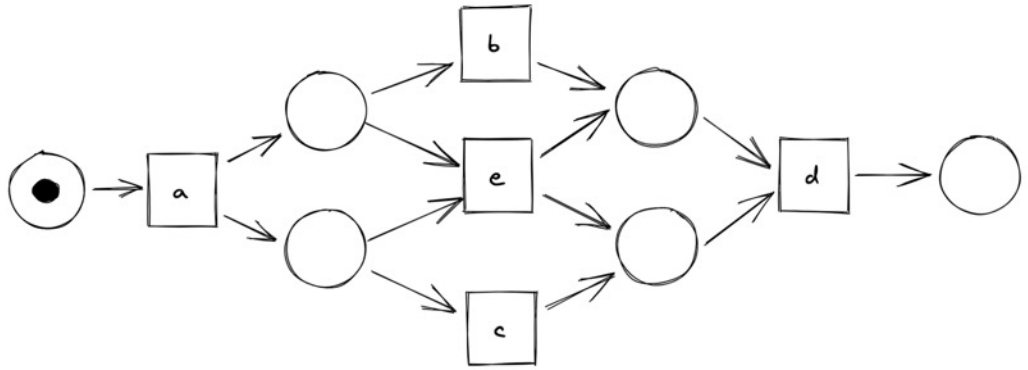
⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, c, b, d⟩
⟨a, c, b, d⟩
⟨a, e, d⟩

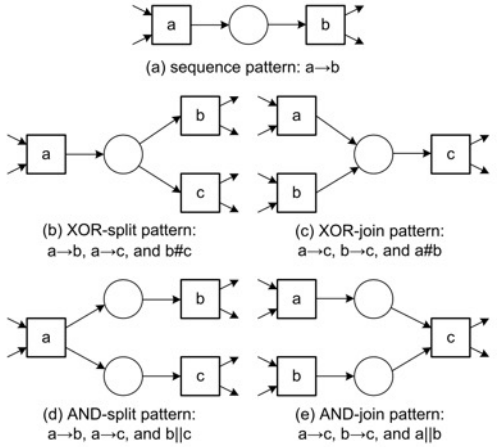


	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\|\|_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\|\|_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

The Alpha Miner

Convert footprint matrix to Petri Net





$L_1 = [\langle a, b, c, d\rangle^3, \langle a, c, b, d\rangle^2, \langle a, e, d\rangle]$

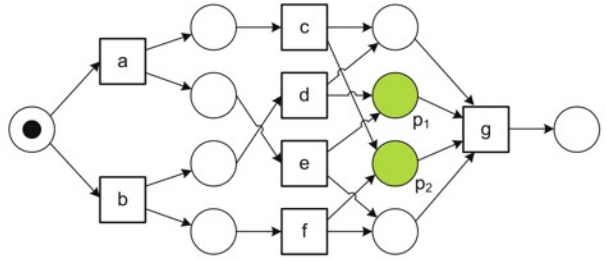
⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, b, c, d⟩
⟨a, c, b, d⟩
⟨a, c, b, d⟩
⟨a, e, d⟩

	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\|\|_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\|\|_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

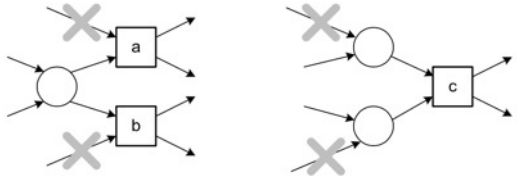
The Alpha Miner

Limitations

- Implicit places: places that can be removed without changing the behavior
  - Harmless, but the model becomes unnecessarily complicated.

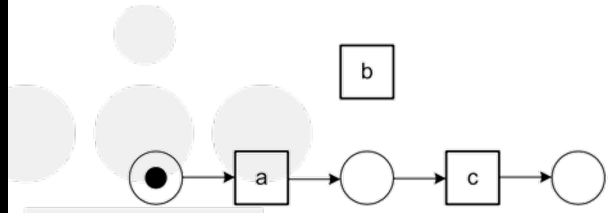


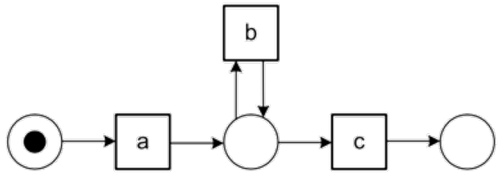
- Representational bias
  - Mixture of choice and synchronization



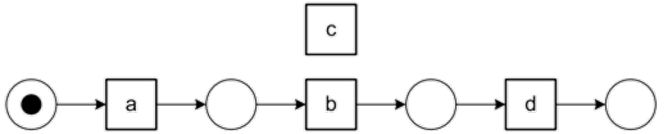
- Does not deal well with loops of length 1 and 2

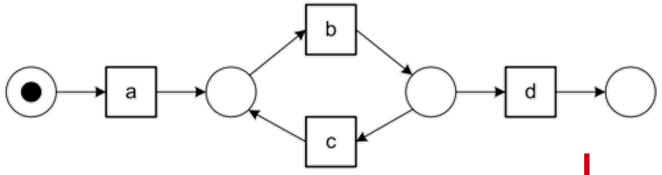
$$L = [\langle a, c\rangle^2, \langle a, b, c\rangle^3, \langle a, b, b, c\rangle^2, \langle a, b, b, b, b, c\rangle^1]$$



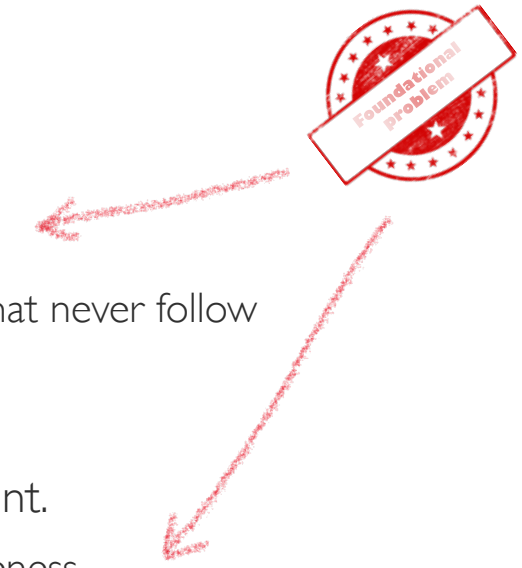


$$L = [\langle a, b, d\rangle^3, \langle a, b, c, b, d\rangle^2, \langle a, b, c, b, c, b, d\rangle]$$





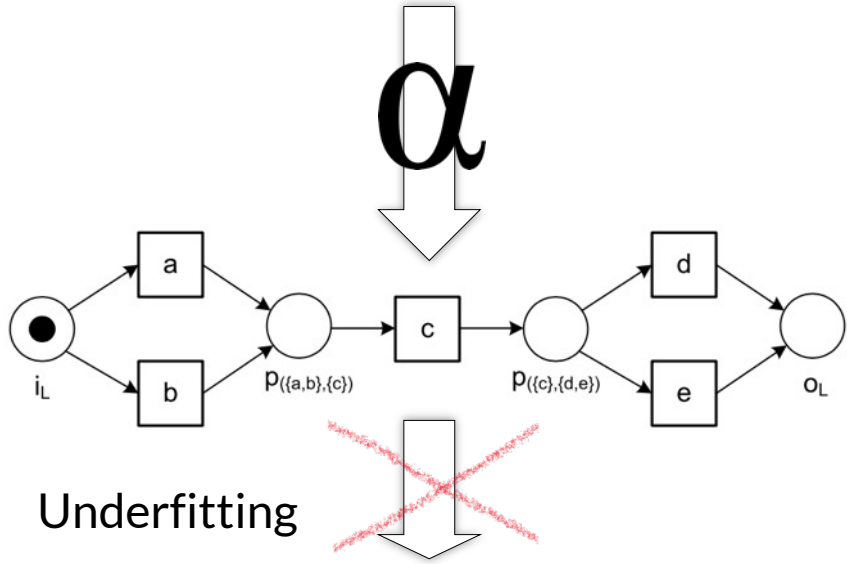
The Alpha Miner



Limitations

- Non-local dependencies
  - Dependency between two activities that never follow one another directly
- Frequencies are not taken into account.
  - Very sensitive to noise and incompleteness
    - Foundational problems for process mining
  - Don't have negative examples
- Soundness
  - Safeness: places cannot hold multiple tokens at the same time
  - Proper completion: if the sink place is marked, all other places are empty
  - Option to complete: it is always posible to reach the final marking
  - Absence of dead parts: for any transitions, there is firing sequnce enabling it

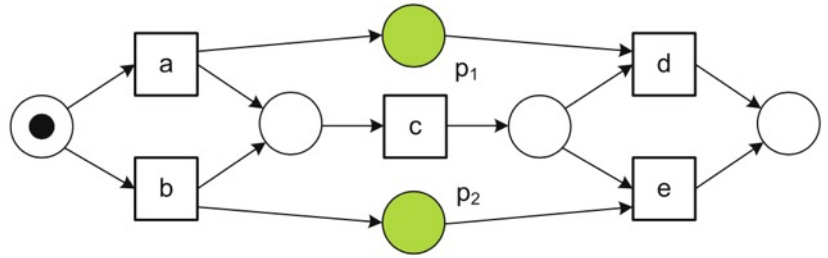
$$L = [\langle a,c,d\rangle^{45}, \langle b,c,e\rangle^{42}]$$



Underfitting

$$L' = [\langle a, c, d\rangle^{45}, \langle b, c, d\rangle^{42}, \langle a, c, e\rangle^{38}, \langle b, c, e\rangle^{22}]$$

Desired Model



Other discovery approaches

Heuristic mining

- Similar representation to Causal Nets
- Focus on dealing with **noise** & **incompleteness**
- Take frequencies and sequences into account
  - Avoid incorporating infrequent paths in the model
- Non-fitting model

Fuzzy mining

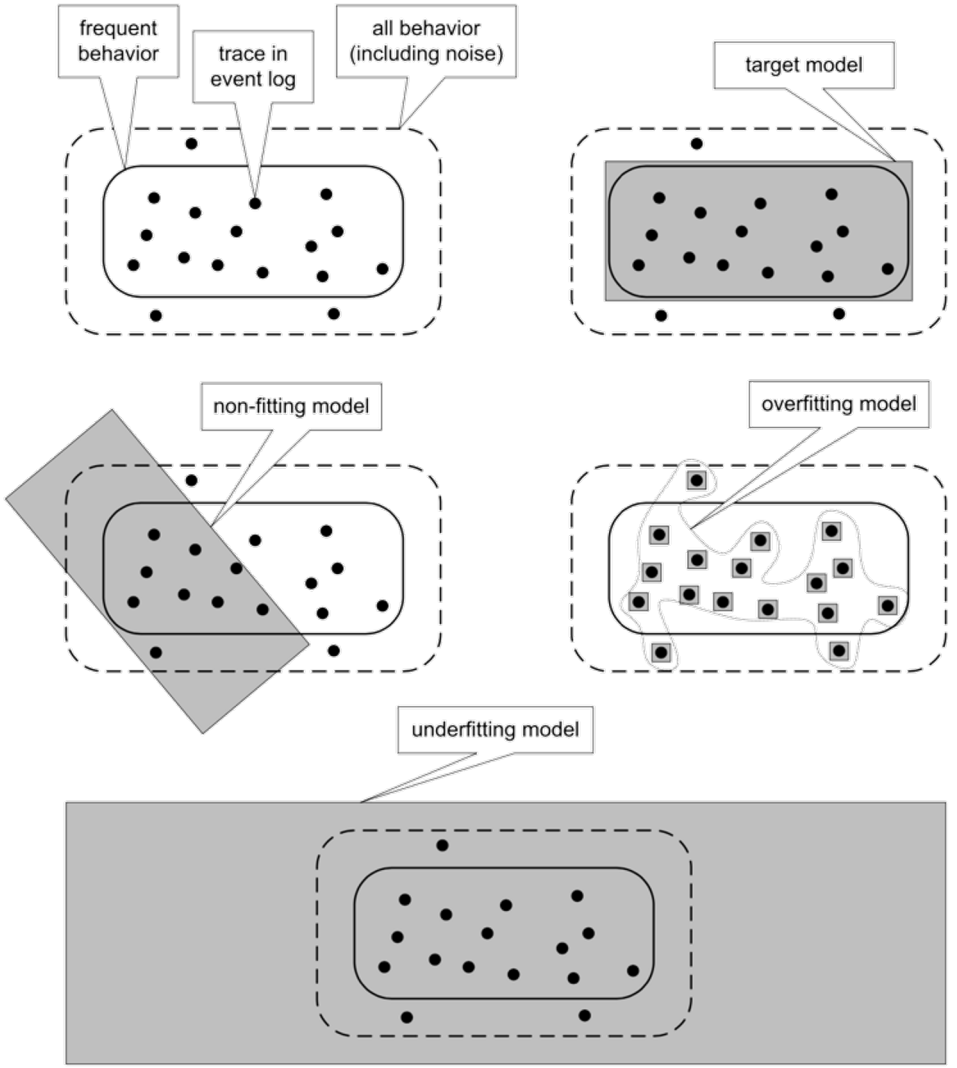
Genetic process mining

Region-based mining

- Able to express more complex control-flow structures without underfitting

Inductive mining

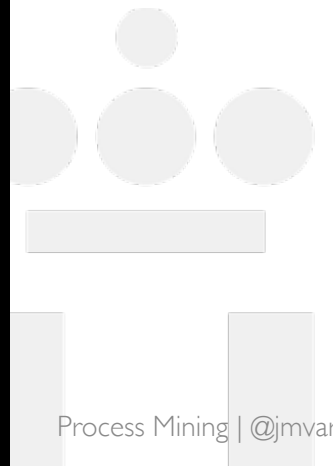
- Sound models & highly scalable



(van der Aalst, 2019)

Quick tour with Disco

a simple yet powerful tool for Process Mining



Adapted from (Fluxicon, 2020)

Example Scenario

Purchasing process

Objectives

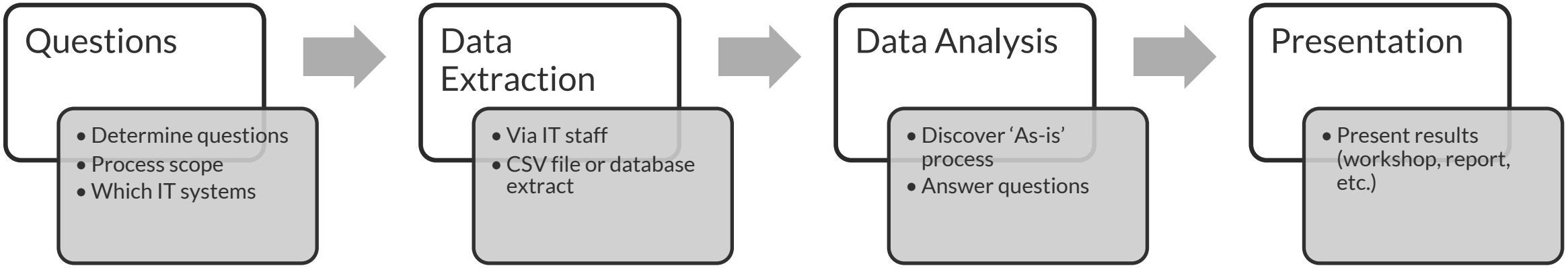
Problems

1. Inefficient operations
2. Need to demonstrate compliance
3. Complaints about process duration

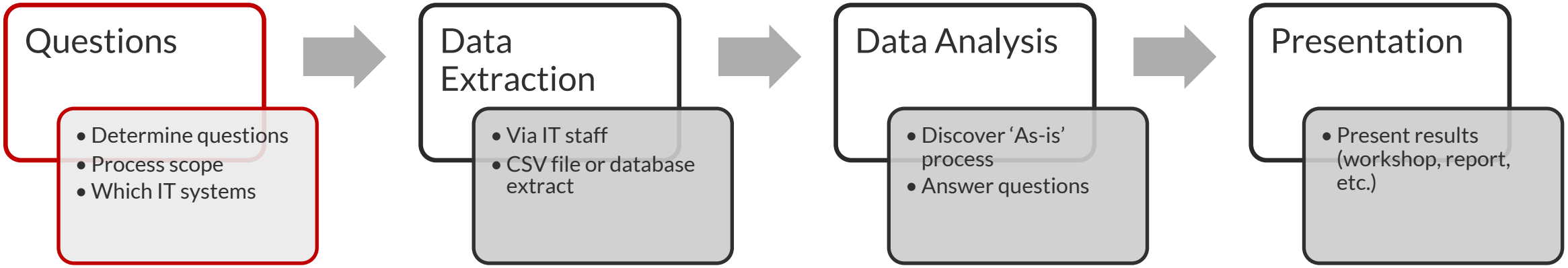
Goals

1. Understand the process in detail
2. Control performance targets (21 days)
3. Check whether there are deviations from the payment guidelines

Process Mining Projects Lifecycle - Roadmap



Roadmap - Questions



1. Understand the process in detail
2. Control performance targets (21 days)
3. Check whether there are deviations from the payment guidelines

1. How does the process actually look like?
2. Do we meet the performance targets?
3. Are there deviations from the prescribed process?

Roadmap – Data Extraction

Questions
- Determine questions
- Process scope
- Which IT systems

Data Extraction
- Via IT staff
- CSV file or database extract

Data Analysis
- Discover 'As-is' process
- Answer questions

Presentation
- Present results (workshop, report, etc.)

IT staff extracts history logs from the ERP system

Case ID	Start Timestamp	Complete Timestamp	Activity	Resource	Role
339	2011/02/16 14:31:00.000	2011/02/16 15:23:00.000	Create Purchase Requisition	Nico Ojenbeer	Requester
339	2011/02/17 09:34:00.000	2011/02/17 09:40:00.000	Analyze Purchase Requisition	Maris Freeman	Requester Manager
339	2011/02/17 21:29:00.000	2011/02/17 21:52:00.000	Amend Purchase Requisition	Elvira Lores	Requester
339	2011/02/18 17:24:00.000	2011/02/18 17:30:00.000	Analyze Purchase Requisition	Heinz Gutschmidt	Requester Manager
339	2011/02/18 17:36:00.000	2011/02/18 17:38:00.000	Create Request for Quotation Requester Manager	Francis Odell	Requester Manager
339	2011/02/22 09:34:00.000	2011/02/22 09:58:00.000	Analyze Request for Quotation	Magdalena Predutta	Purchasing Agent
339	2011/02/22 10:50:00.000	2011/02/22 11:03:00.000	Amend Request for Quotation Requester	Penn Osterwalder	Requester Manager

Things to consider for your data extraction

Which process

- Availability of data
- Champion support
  - A sponsor who wants a "surprise me" analysis is a red flag
- Potential for Improvement

RQ

- Focus your analysis
- Pose aditional requirements for data extraction

IT Systems

- CRM, ERP, SAP
- Workflow, ticketing systems, ad-hoc systems

Case ID

- Combination from multiple fields
- Different Case IDs in different parts of the process
- N:M Relationships

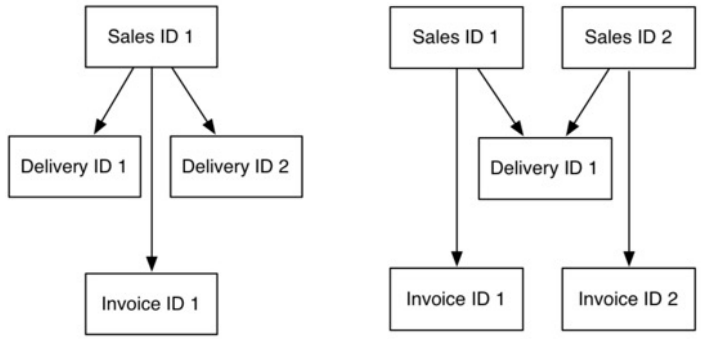
Different Case IDs in different parts of the process

1. Sales order: traced by Sales order ID
2. Delivery: traced by Delivery ID
3. Invoicing: traced by Invoicing ID

"flattening reality" (like putting a 3D-world in a 2D-picture).

- You need to choose which perspective you want to take on your process.

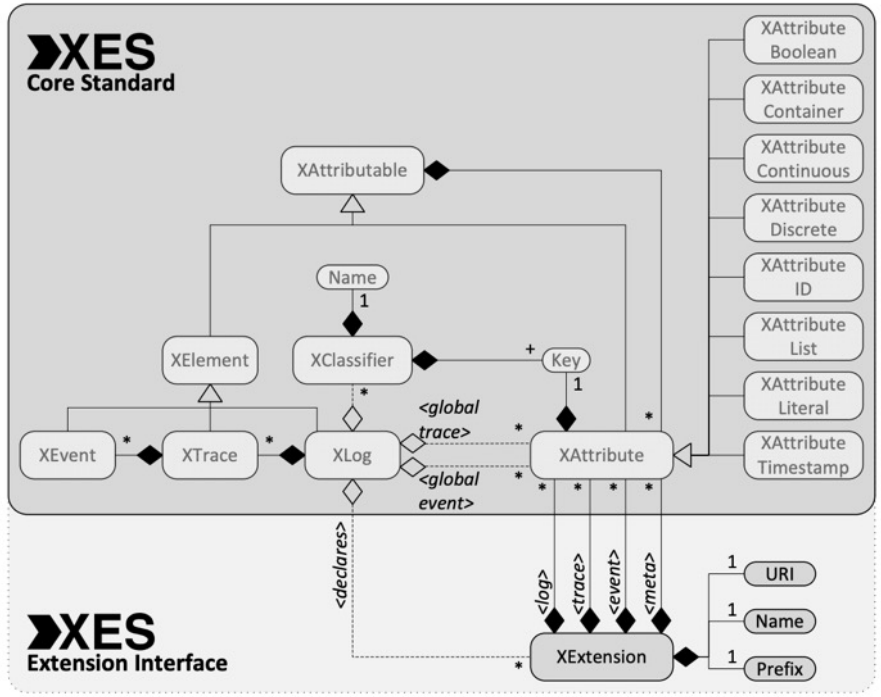
N:M Relationships



XES (eXtensible Event Stream)

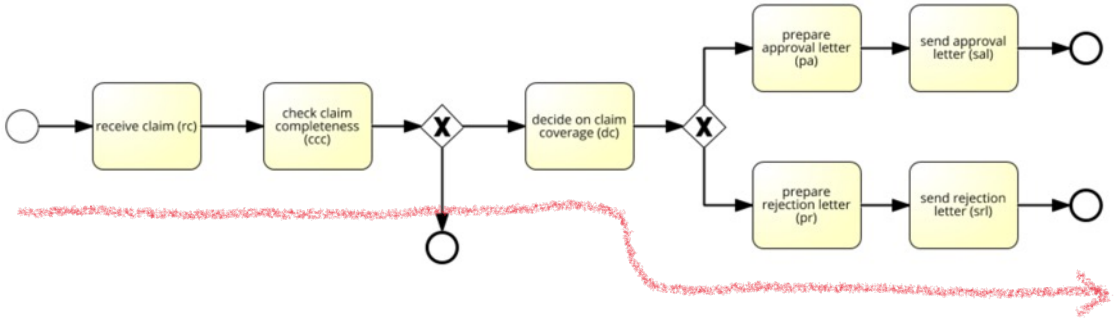
Official IEEE Standard

- Driven by IEEE Task Force on Process Mining
- XML syntax and OpenXES library
- Supported by most of the existing tools
- Conversion from other formats (CSV) is easy

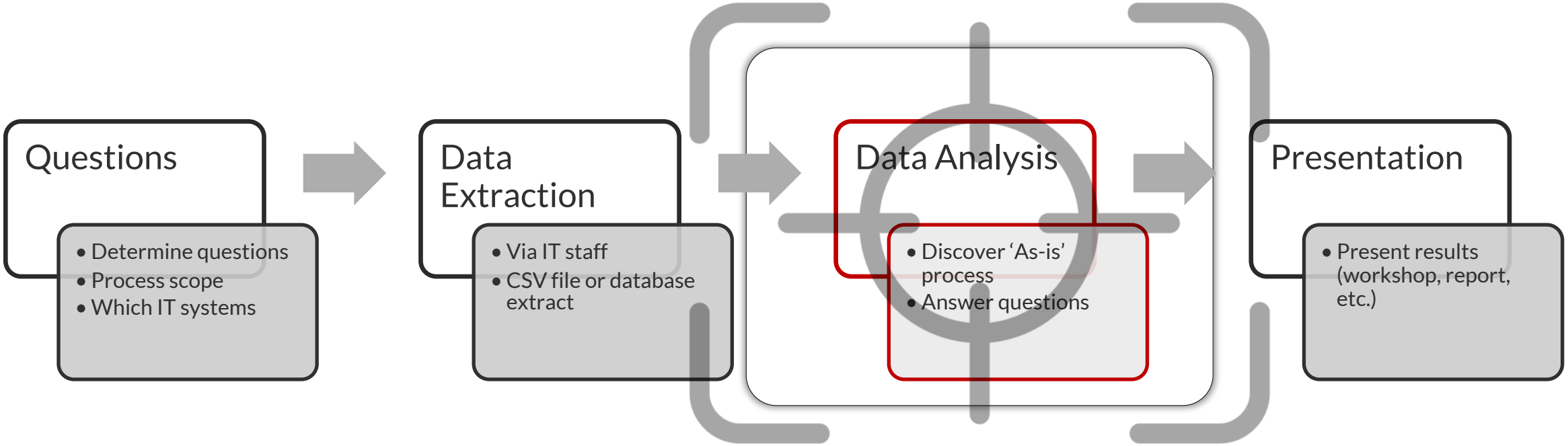



```
<?xml version="1.0" encoding="UTF-8" ?>
<!-- This file has been generated with the OpenXES library. -->
<!-- OpenXES is available from http://www.openxes.org/ -->
<log xes.version="1.0" xes.features="nested-attributes" openxes.version="1.0RC7">
    <string key="concept:name" value="Process_04483xw"/>
    <trace>
        <string key="concept:name" value="1241"/>
        <event>
            <string key="org:resource" value="Clerk_#3940"/>
            <string key="concept:name" value="receive claim"/>
            <date key="time:timestamp" value="2020-12-18T14:01:13+01:00"/>
        </event>
        <event>
            <string key="org:resource" value="Clerk_#3940"/>
            <string key="concept:name" value="check claim completeness"/>
            <date key="time:timestamp" value="2020-12-18T14:05:29+01:00"/>
        </event>
        <event>
            <string key="org:resource" value="Clerk_#3819"/>
            <string key="concept:name" value="decide on claim coverage"/>
            <date key="time:timestamp" value="2020-12-18T14:16:45+01:00"/>
        </event>
        <event>
            <string key="org:resource" value="Clerk_#1173"/>
            <string key="concept:name" value="prepare rejection letter"/>
            <date key="time:timestamp" value="2020-12-18T14:25:08+01:00"/>
        </event>
        <event>
            <string key="org:resource" value="Clerk_#1173"/>
            <string key="concept:name" value="send rejection letter"/>
            <date key="time:timestamp" value="2020-12-18T14:25:08+01:00"/>
        </event>
    </trace>
```

(Weske, 2021)

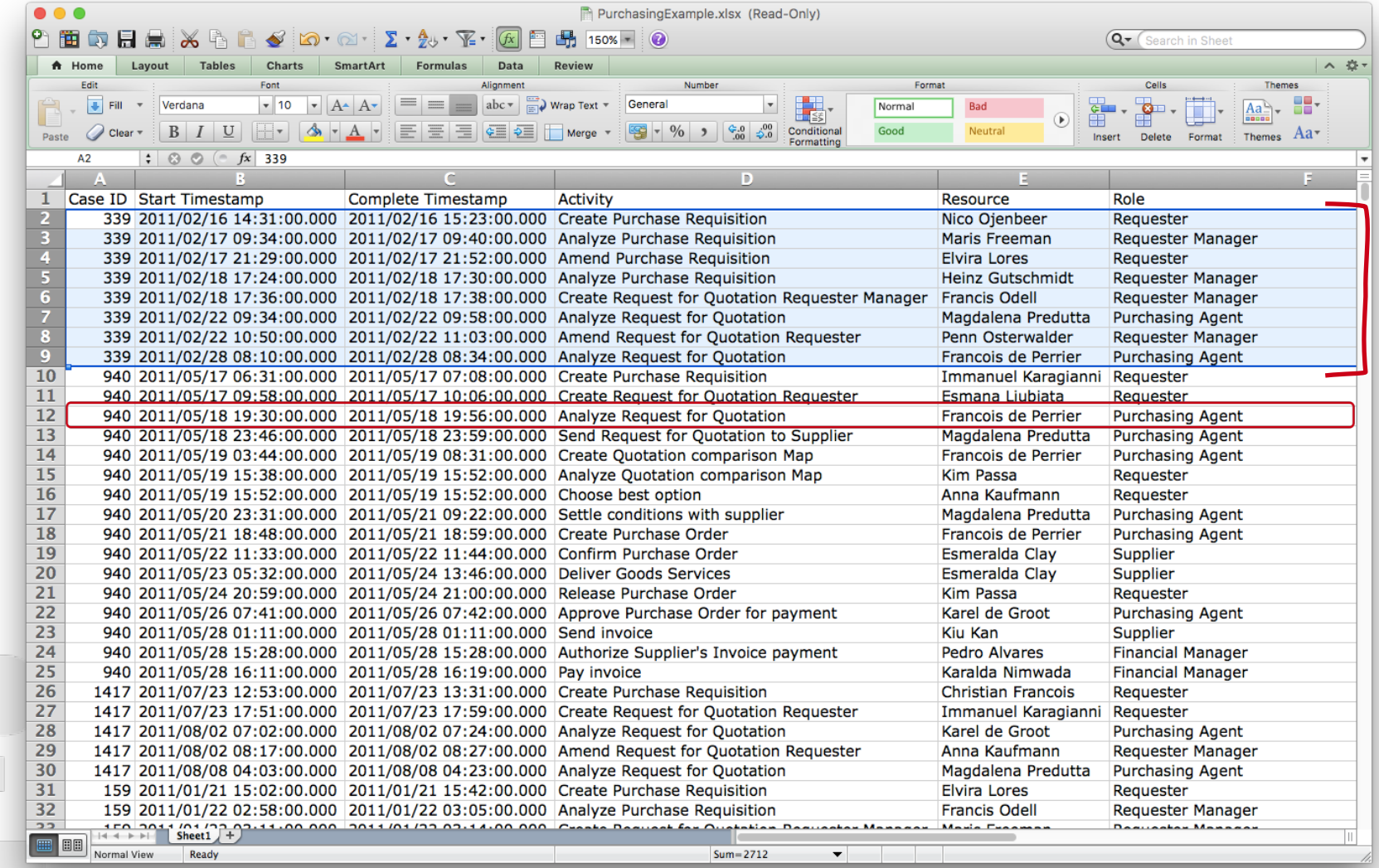


Roadmap – Data Analysis



1. Inspect Data
2. Import Data
3. Inspect Process
4. Inspect Statistics
5. Inspects Cases
6. Filter on performance
7. Visualize Bottlenecks
8. Animate Process
9. Compliance Check
10. Organizational View

1. Inspect Data



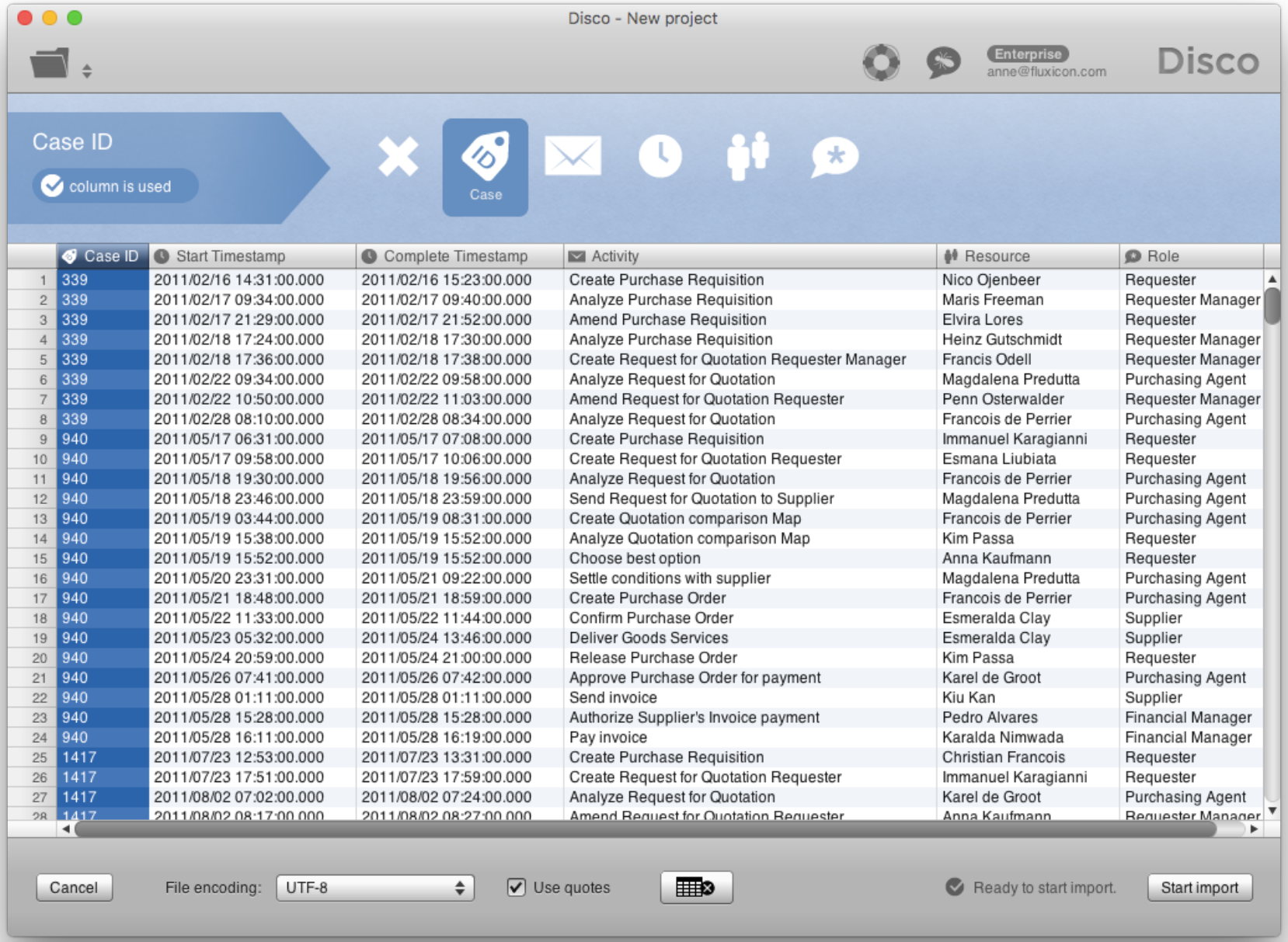
	A	B	C	D	E	F
1	Case ID	Start Timestamp	Complete Timestamp	Activity	Resource	Role
2	339	2011/02/16 14:31:00.000	2011/02/16 15:23:00.000	Create Purchase Requisition	Nico Ojenbeer	Requester
3	339	2011/02/17 09:34:00.000	2011/02/17 09:40:00.000	Analyze Purchase Requisition	Maris Freeman	Requester Manager
4	339	2011/02/17 21:29:00.000	2011/02/17 21:52:00.000	Amend Purchase Requisition	Elvira Lores	Requester
5	339	2011/02/18 17:24:00.000	2011/02/18 17:30:00.000	Analyze Purchase Requisition	Heinz Gutschmidt	Requester Manager
6	339	2011/02/18 17:36:00.000	2011/02/18 17:38:00.000	Create Request for Quotation Requester Manager	Francis Odell	Requester Manager
7	339	2011/02/22 09:34:00.000	2011/02/22 09:58:00.000	Analyze Request for Quotation	Magdalena Predutta	Purchasing Agent
8	339	2011/02/22 10:50:00.000	2011/02/22 11:03:00.000	Amend Request for Quotation Requester	Penn Osterwalder	Requester Manager
9	339	2011/02/28 08:10:00.000	2011/02/28 08:34:00.000	Analyze Request for Quotation	Francois de Perrier	Purchasing Agent
10	940	2011/05/17 06:31:00.000	2011/05/17 07:08:00.000	Create Purchase Requisition	Immanuel Karagianni	Requester
11	940	2011/05/17 09:58:00.000	2011/05/17 10:06:00.000	Create Request for Quotation Requester	Esmana Liubiata	Requester
12	940	2011/05/18 19:30:00.000	2011/05/18 19:56:00.000	Analyze Request for Quotation	Francois de Perrier	Purchasing Agent
13	940	2011/05/18 23:46:00.000	2011/05/18 23:59:00.000	Send Request for Quotation to Supplier	Magdalena Predutta	Purchasing Agent
14	940	2011/05/19 03:44:00.000	2011/05/19 08:31:00.000	Create Quotation comparison Map	Francois de Perrier	Purchasing Agent
15	940	2011/05/19 15:38:00.000	2011/05/19 15:52:00.000	Analyze Quotation comparison Map	Kim Passa	Requester
16	940	2011/05/19 15:52:00.000	2011/05/19 15:52:00.000	Choose best option	Anna Kaufmann	Requester
17	940	2011/05/20 23:31:00.000	2011/05/21 09:22:00.000	Settle conditions with supplier	Magdalena Predutta	Purchasing Agent
18	940	2011/05/21 18:48:00.000	2011/05/21 18:59:00.000	Create Purchase Order	Francois de Perrier	Purchasing Agent
19	940	2011/05/22 11:33:00.000	2011/05/22 11:44:00.000	Confirm Purchase Order	Esmeralda Clay	Supplier
20	940	2011/05/23 05:32:00.000	2011/05/24 13:46:00.000	Deliver Goods Services	Esmeralda Clay	Supplier
21	940	2011/05/24 20:59:00.000	2011/05/24 21:00:00.000	Release Purchase Order	Kim Passa	Requester
22	940	2011/05/26 07:41:00.000	2011/05/26 07:42:00.000	Approve Purchase Order for payment	Karel de Groot	Purchasing Agent
23	940	2011/05/28 01:11:00.000	2011/05/28 01:11:00.000	Send invoice	Kiu Kan	Supplier
24	940	2011/05/28 15:28:00.000	2011/05/28 15:28:00.000	Authorize Supplier's Invoice payment	Pedro Alvares	Financial Manager
25	940	2011/05/28 16:11:00.000	2011/05/28 16:19:00.000	Pay invoice	Karalda Nimwada	Financial Manager
26	1417	2011/07/23 12:53:00.000	2011/07/23 13:31:00.000	Create Purchase Requisition	Christian Francois	Requester
27	1417	2011/07/23 17:51:00.000	2011/07/23 17:59:00.000	Create Request for Quotation Requester	Immanuel Karagianni	Requester
28	1417	2011/08/02 07:02:00.000	2011/08/02 07:24:00.000	Analyze Request for Quotation	Karel de Groot	Purchasing Agent
29	1417	2011/08/02 08:17:00.000	2011/08/02 08:27:00.000	Amend Request for Quotation Requester	Anna Kaufmann	Requester Manager
30	1417	2011/08/08 04:03:00.000	2011/08/08 04:23:00.000	Analyze Request for Quotation	Magdalena Predutta	Purchasing Agent
31	159	2011/01/21 15:02:00.000	2011/01/21 15:42:00.000	Create Purchase Requisition	Elvira Lores	Requester
32	159	2011/01/22 02:58:00.000	2011/01/22 03:05:00.000	Analyze Purchase Requisition	Francis Odell	Requester Manager

Case 1

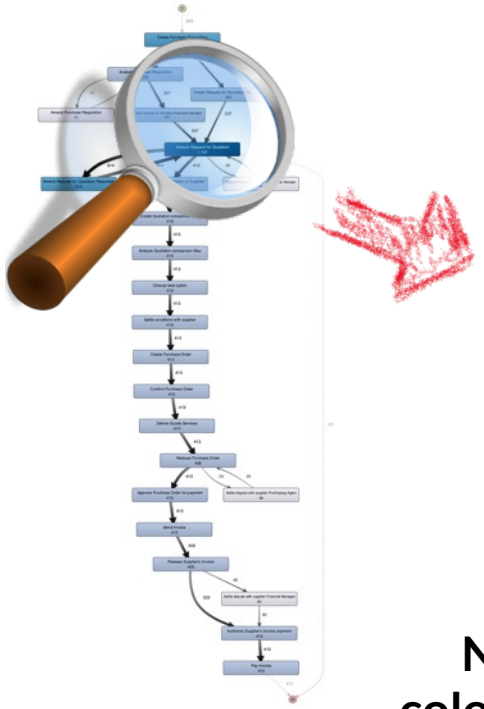
Event

2. Import Data

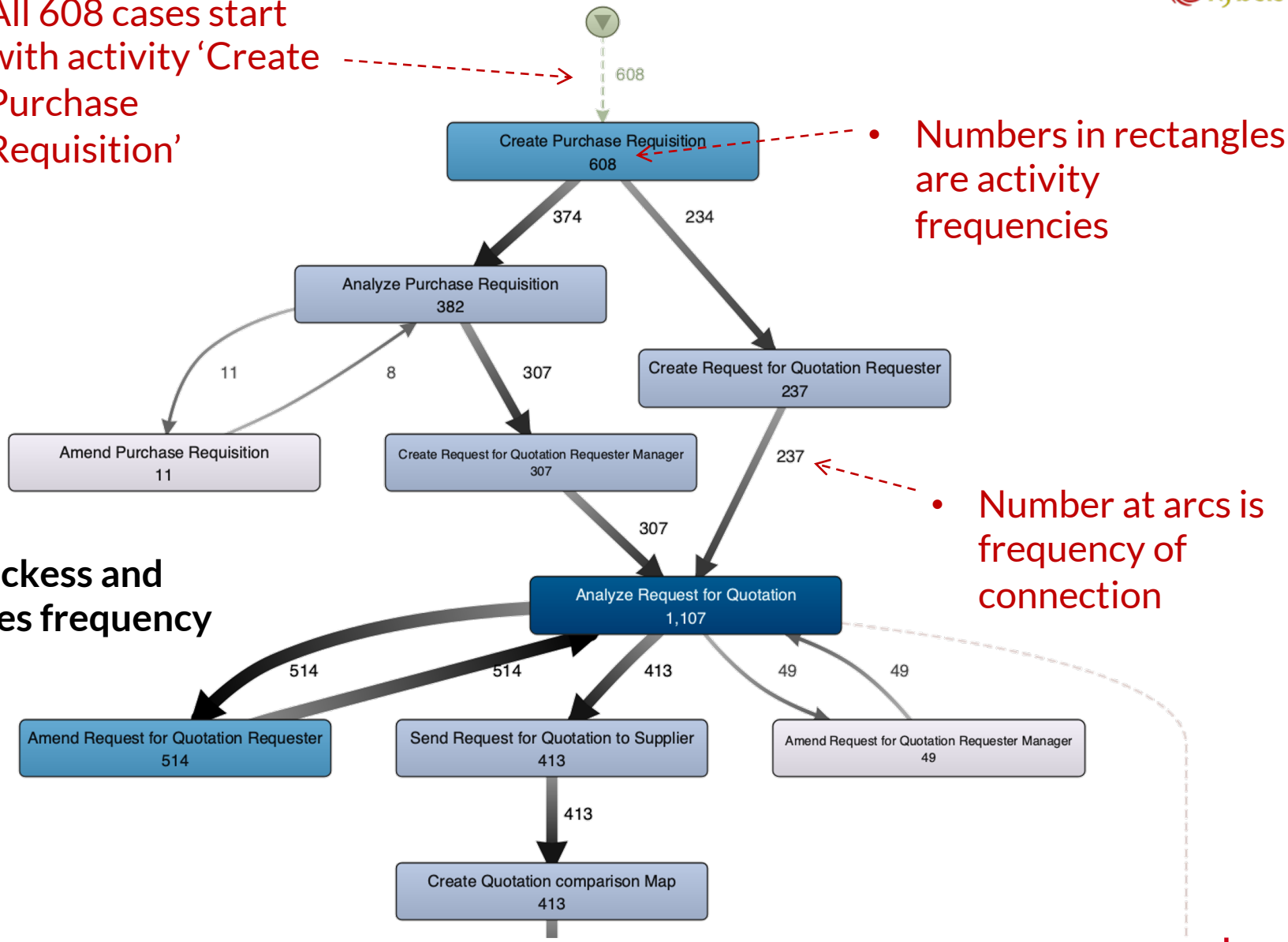
Check the guess made by the tool



3. Inspect Process



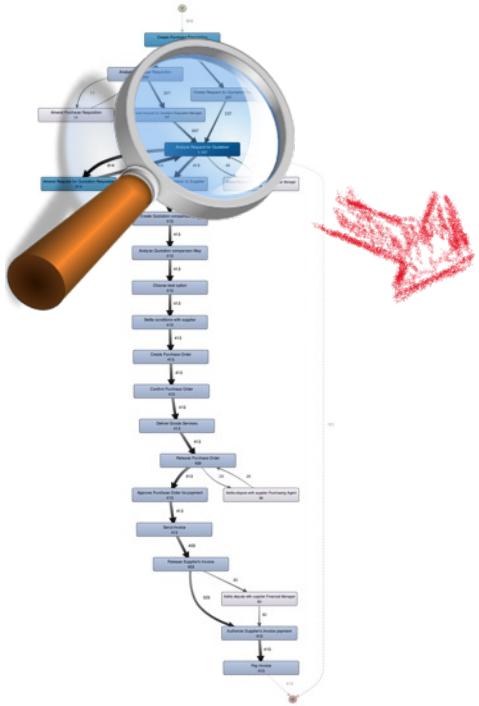
- All 608 cases start with activity ‘Create Purchase Requisition’



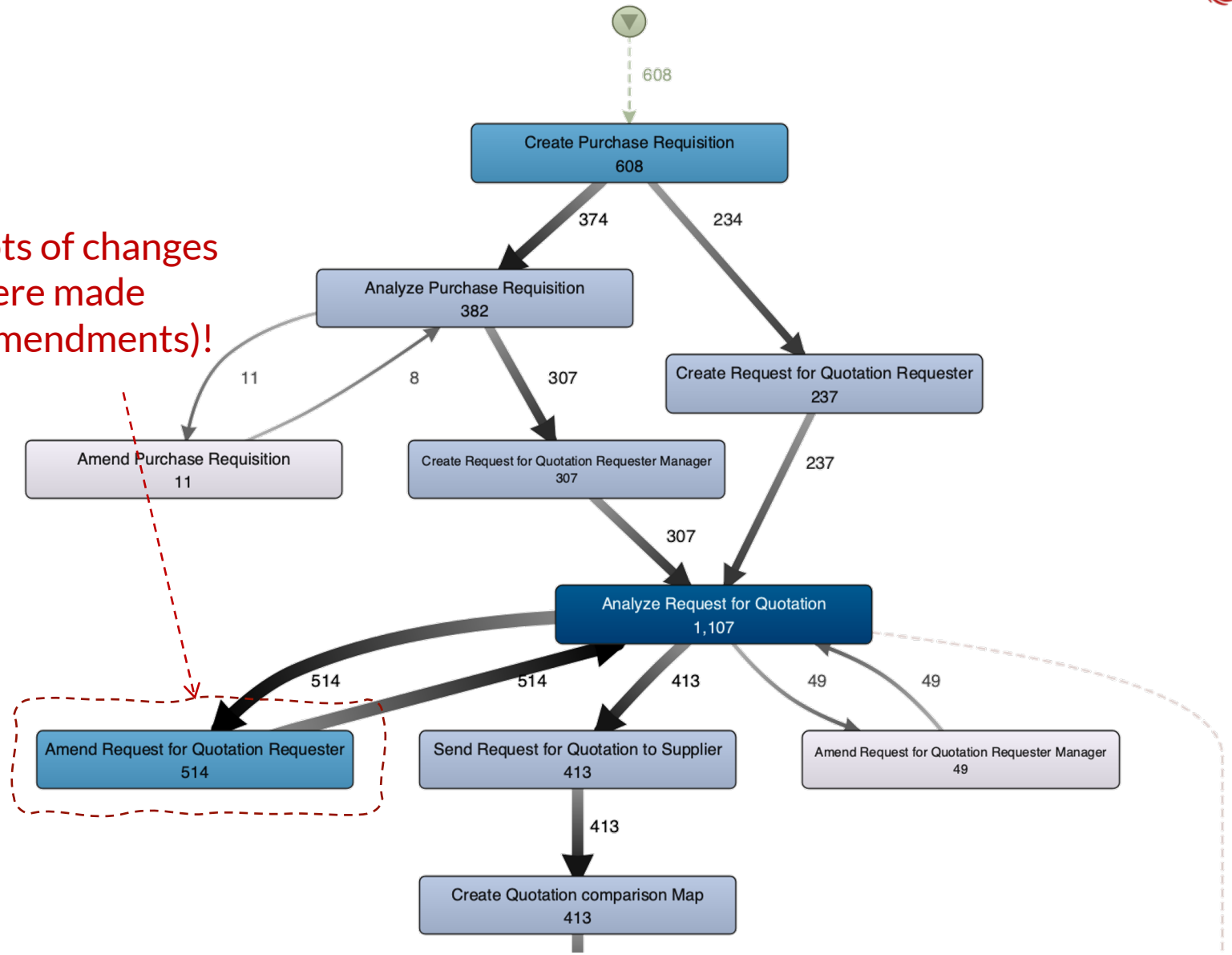
- Numbers in rectangles are activity frequencies
- Number at arcs is frequency of connection

Numbers, thickess and coloring indicates frequency

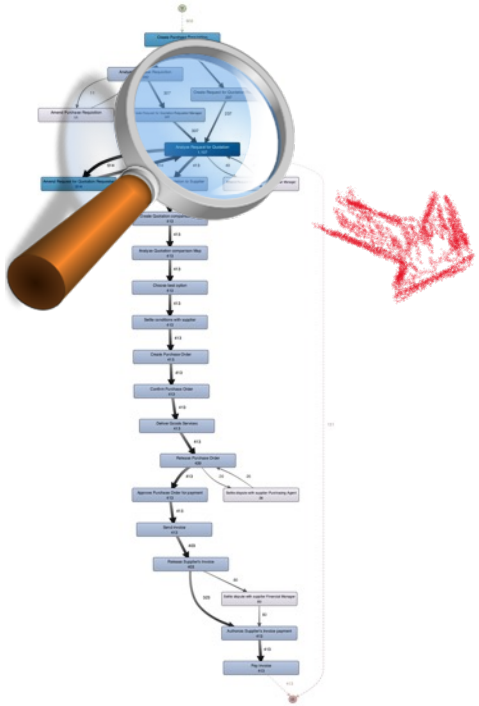
3. Inspect Process



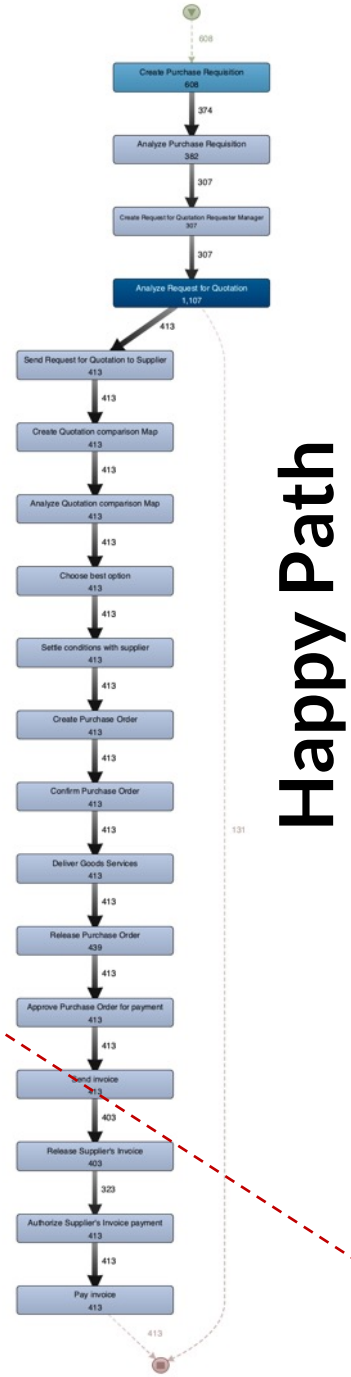
- Lots of changes were made (amendments)!



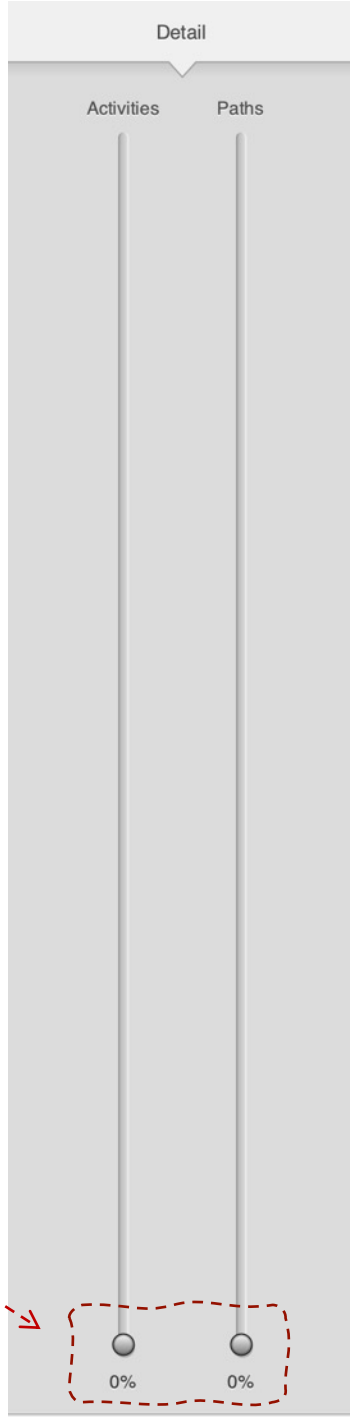
3. Inspect Process



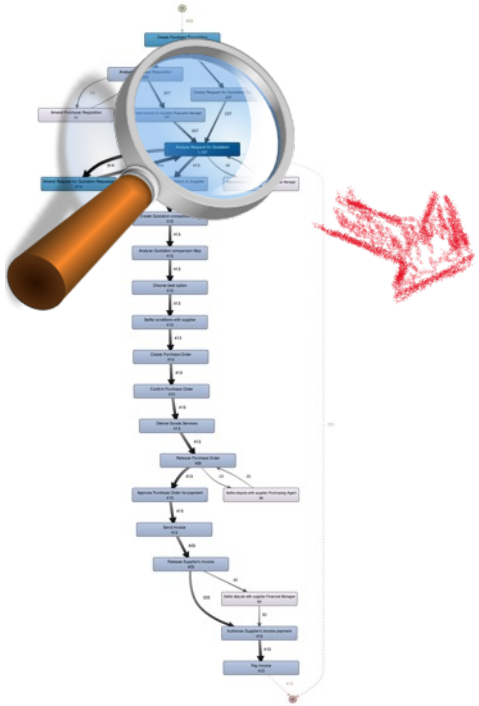
- At the lowest point only the activities of the most frequent variant (Happy Path) are shown

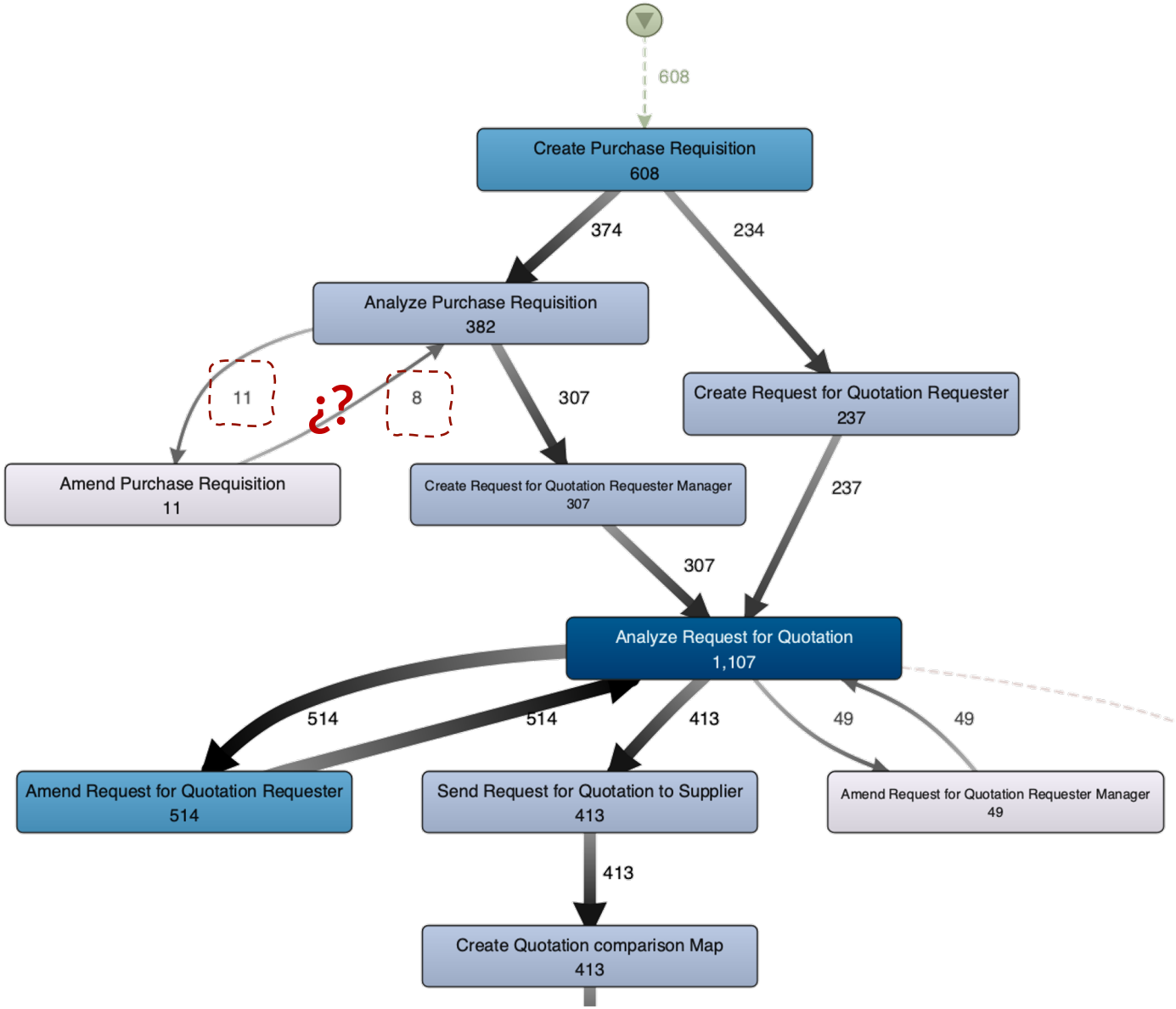


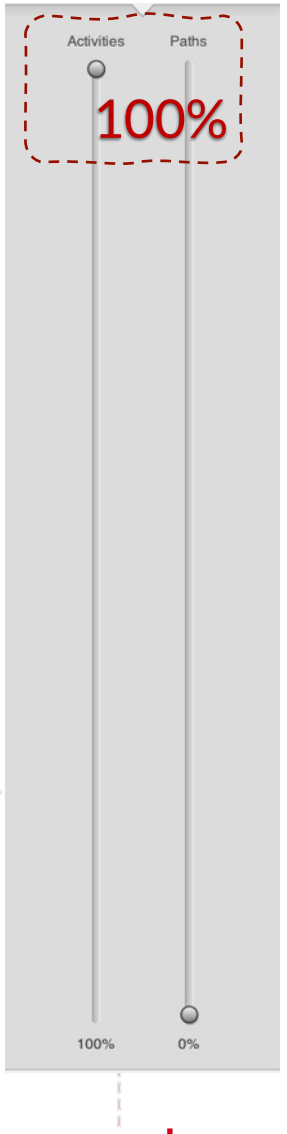
Happy Path



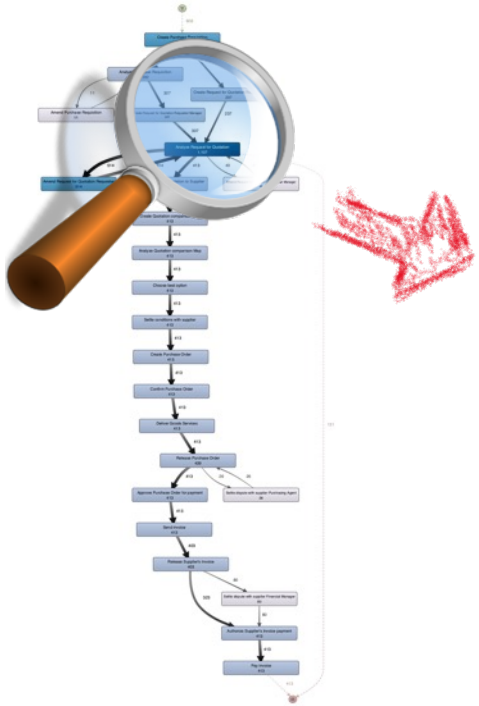
3. Inspect Process

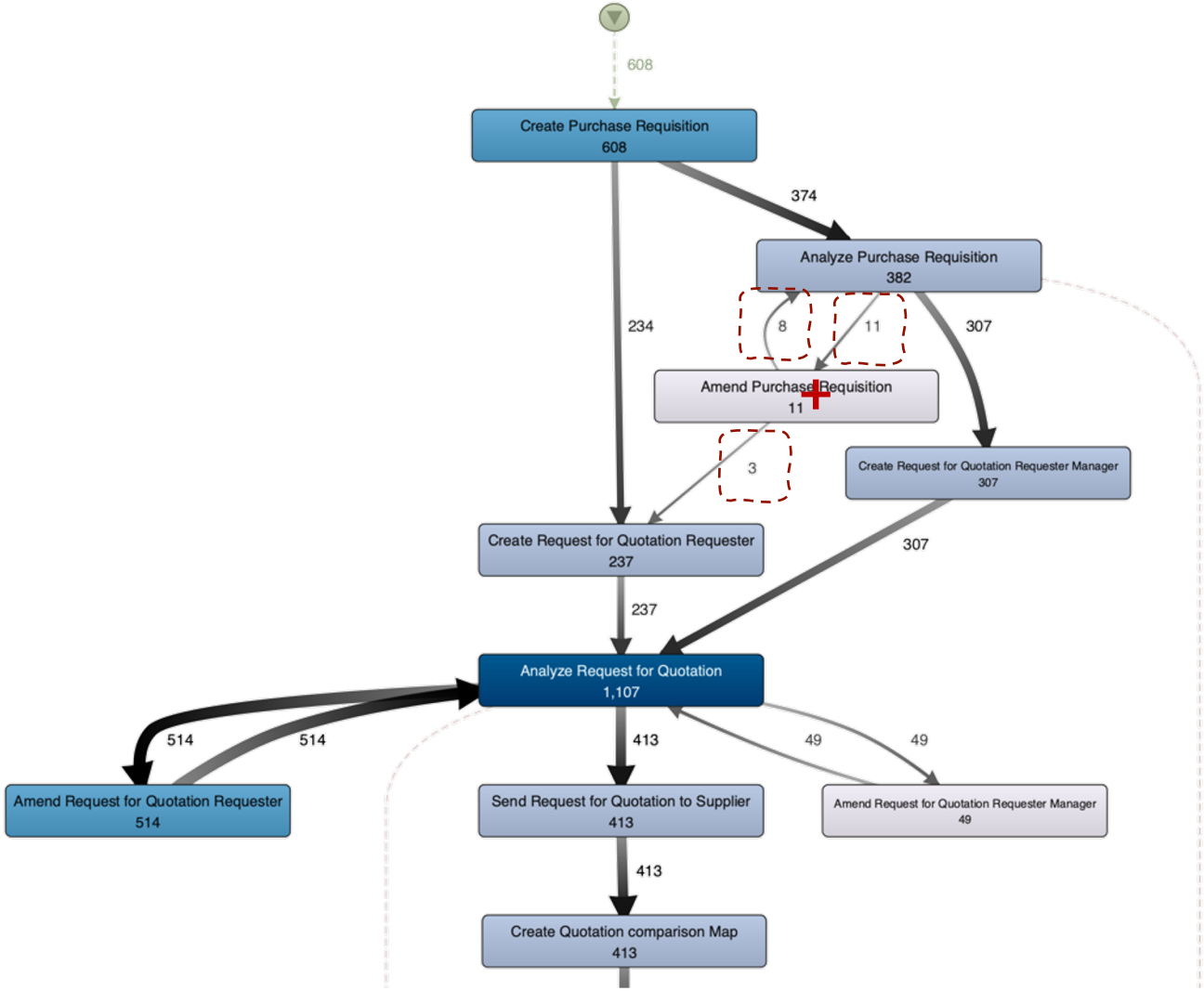


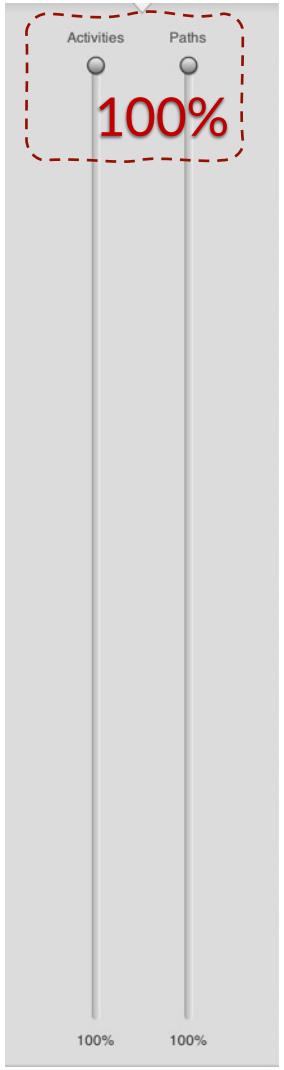




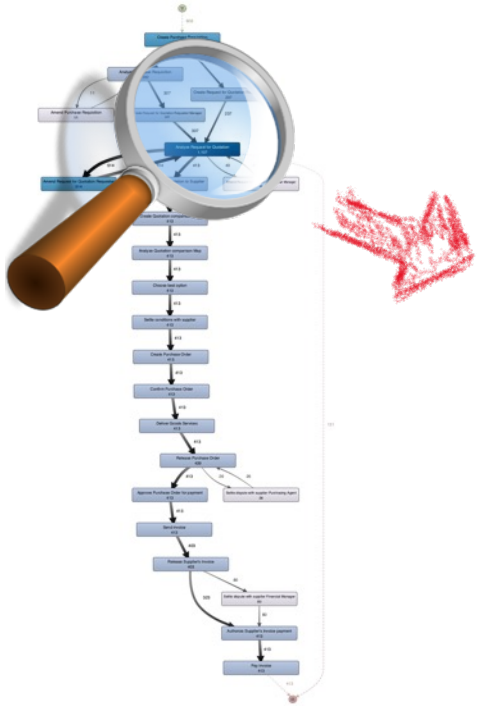
3. Inspect Process





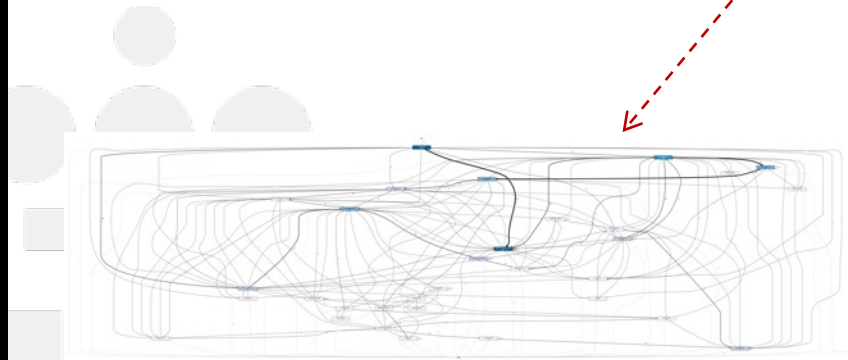


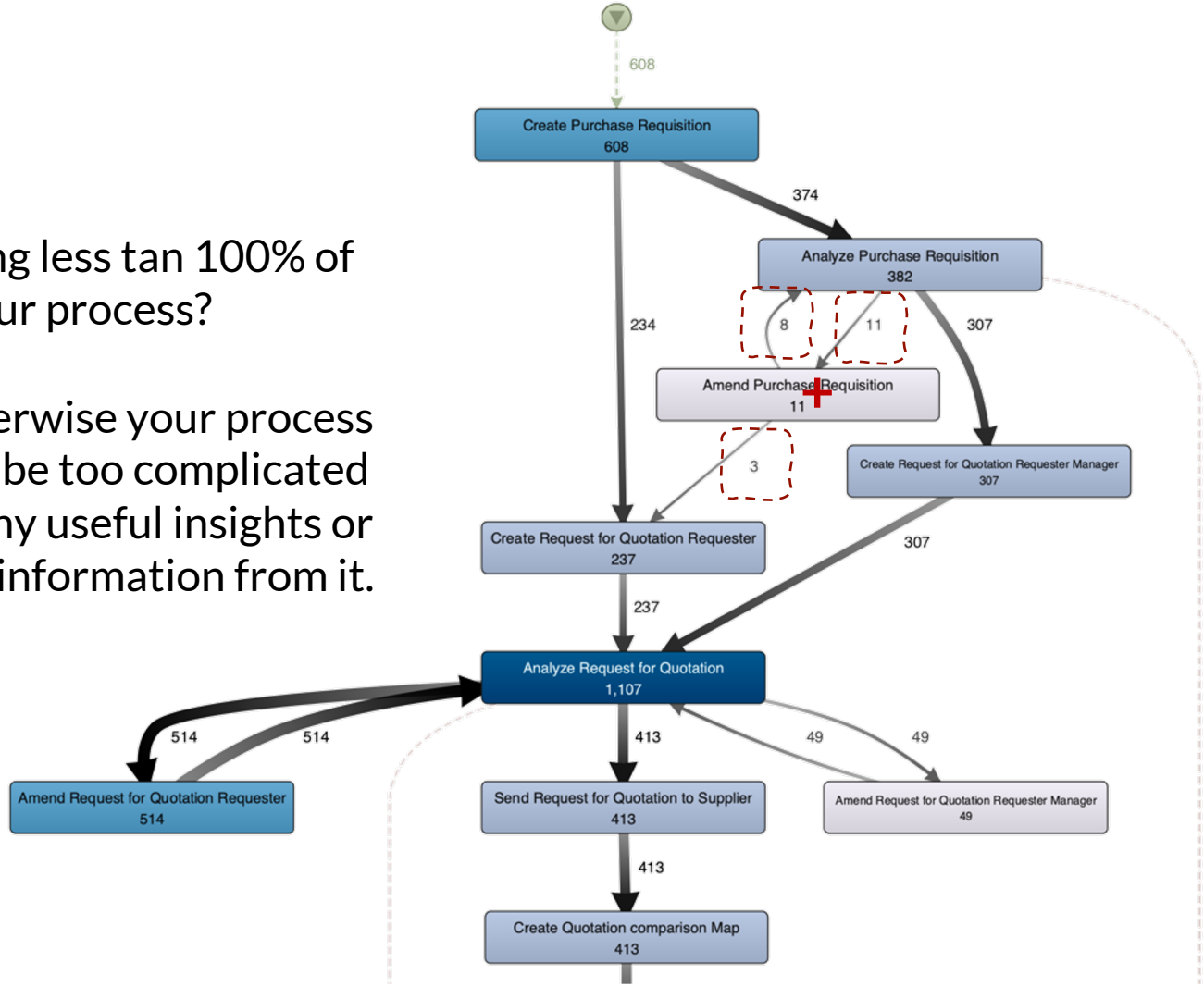
3. Inspect Process

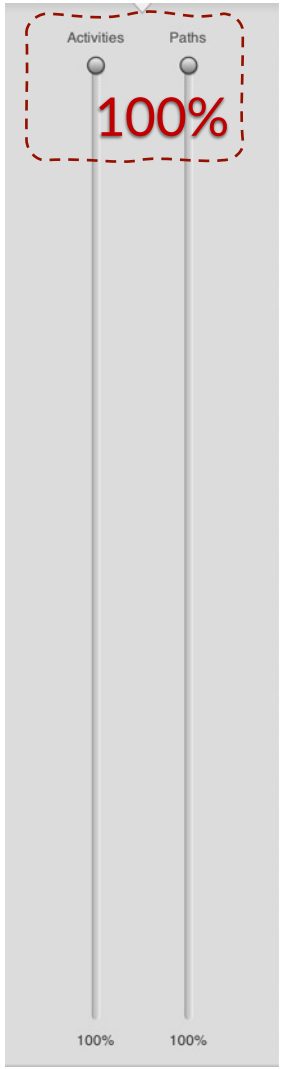


Why seeing less tan 100% of your process?

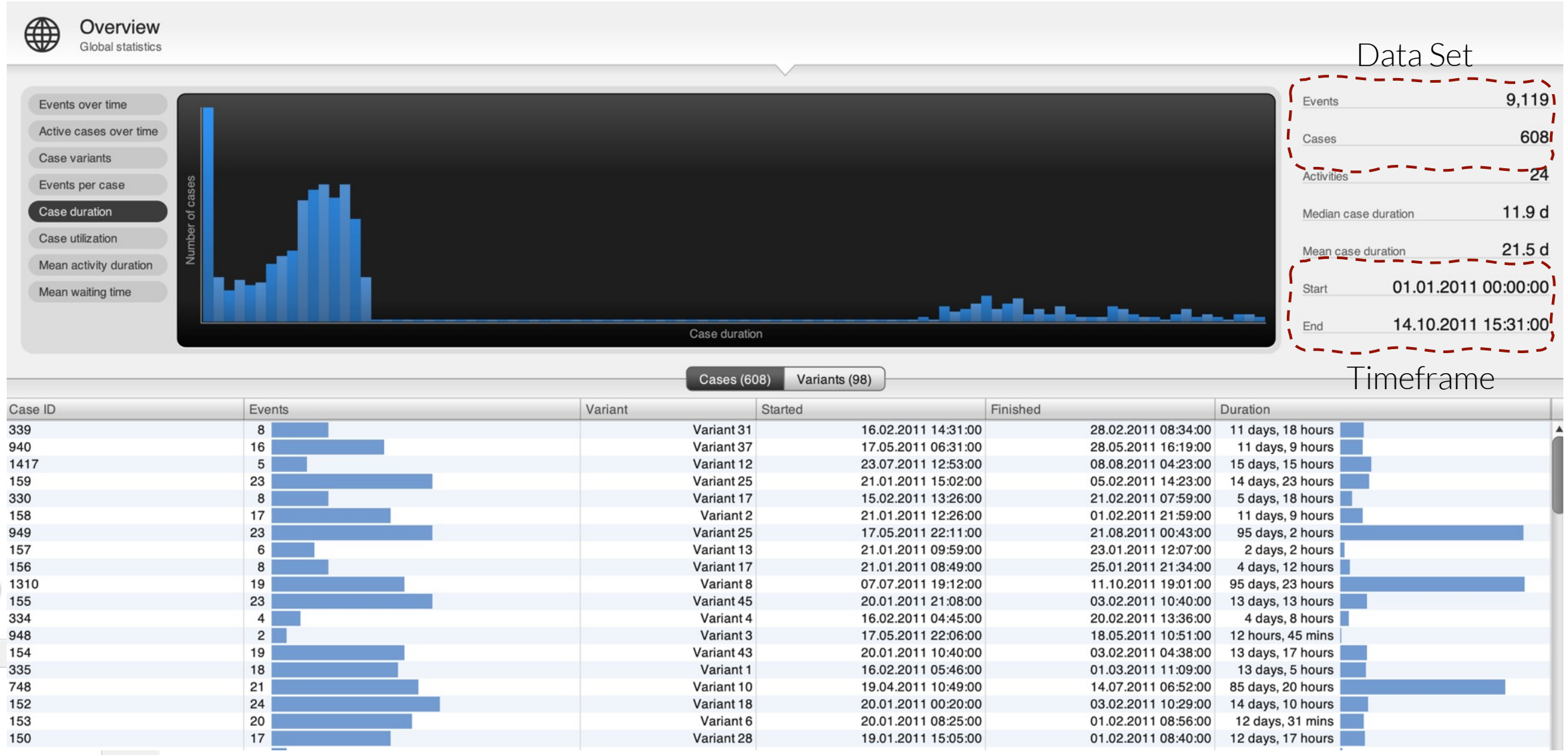
‘Cause otherwise your process map might be too complicated to derive any useful insights or actionable information from it.





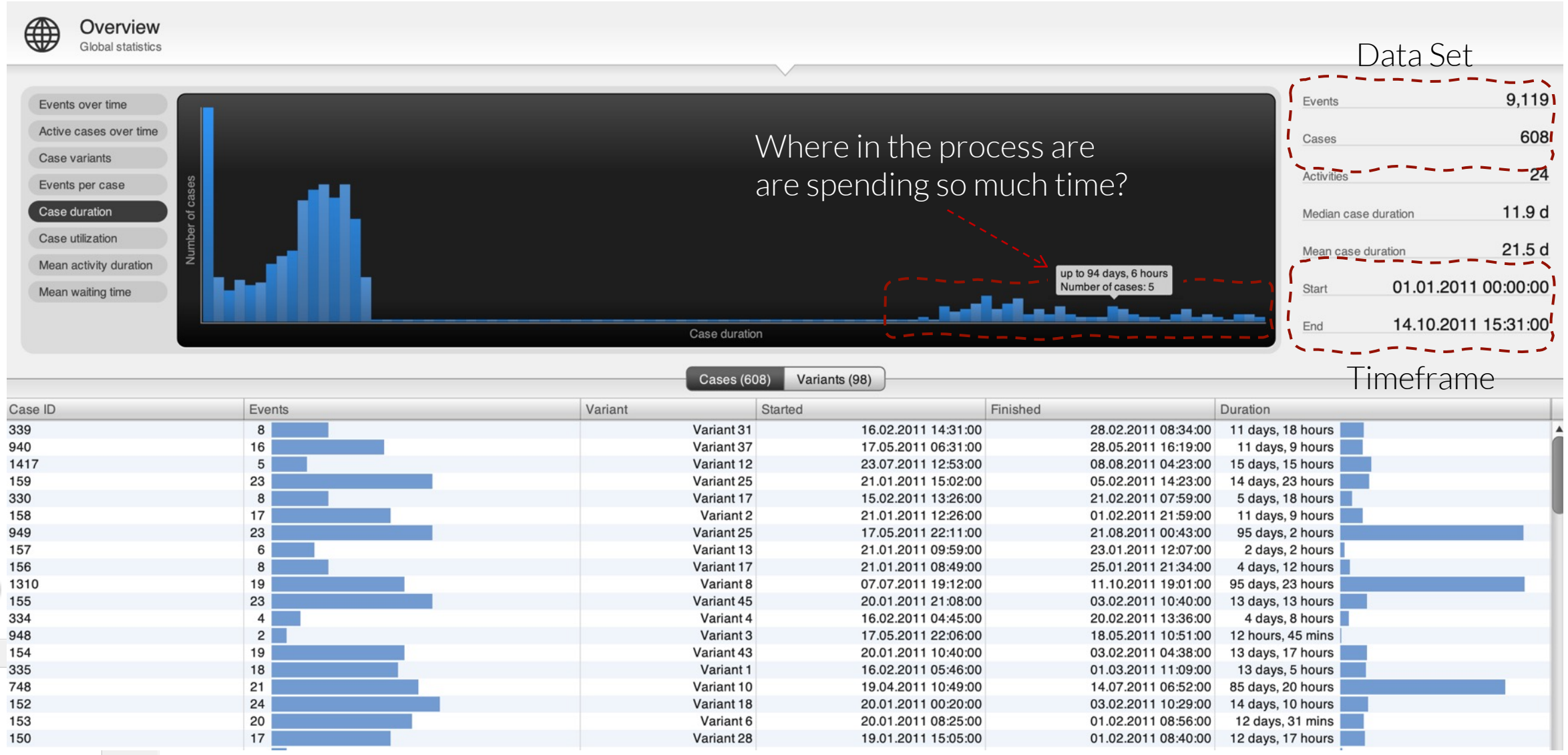


4. Inspect Statistics



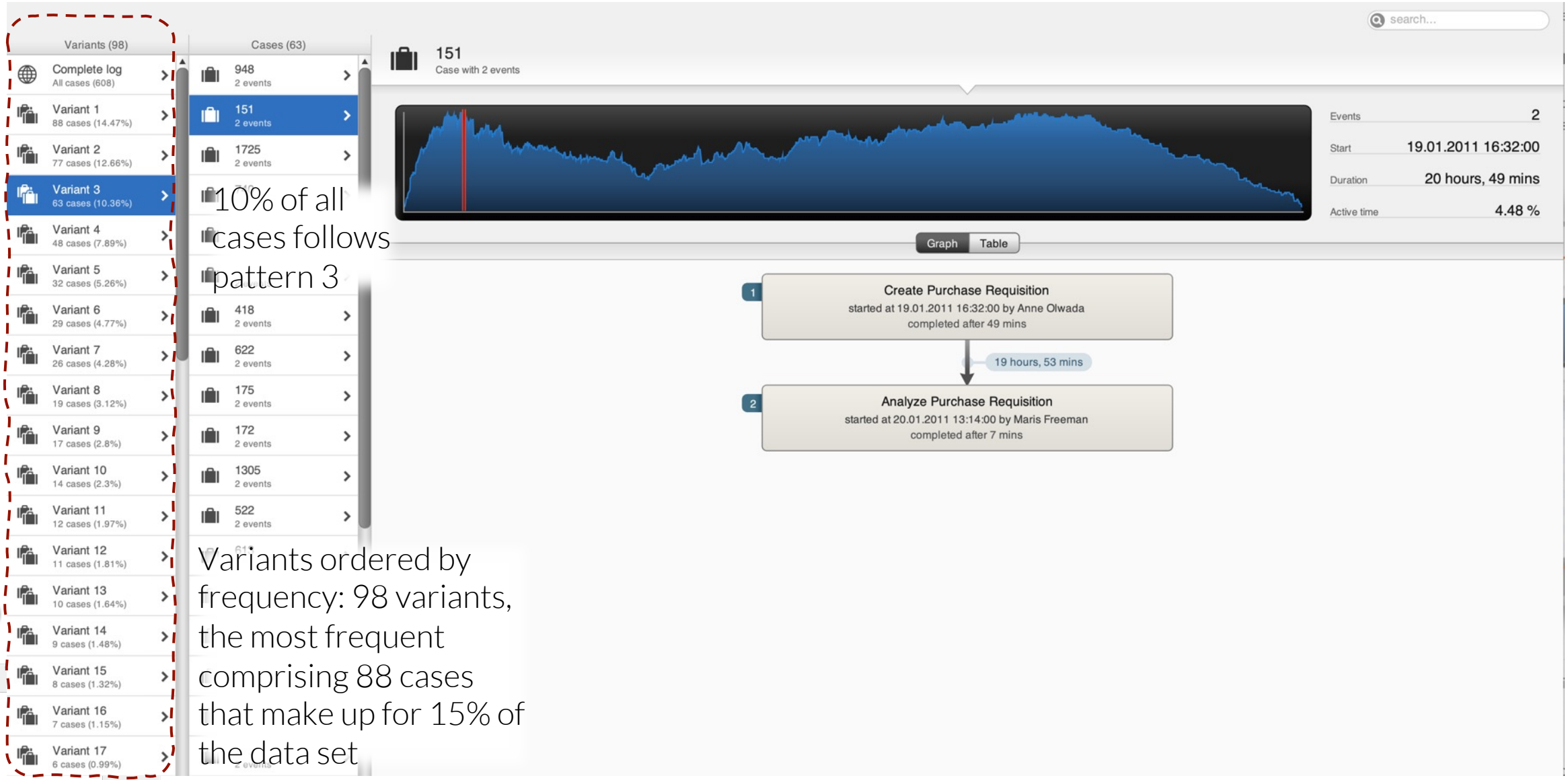
Case ID	Events	Variant	Started	Finished	Duration
339	8	Variant 31	16.02.2011 14:31:00	28.02.2011 08:34:00	11 days, 18 hours
940	16	Variant 37	17.05.2011 06:31:00	28.05.2011 16:19:00	11 days, 9 hours
1417	5	Variant 12	23.07.2011 12:53:00	08.08.2011 04:23:00	15 days, 15 hours
159	23	Variant 25	21.01.2011 15:02:00	05.02.2011 14:23:00	14 days, 23 hours
330	8	Variant 17	15.02.2011 13:26:00	21.02.2011 07:59:00	5 days, 18 hours
158	17	Variant 2	21.01.2011 12:26:00	01.02.2011 21:59:00	11 days, 9 hours
949	23	Variant 25	17.05.2011 22:11:00	21.08.2011 00:43:00	95 days, 2 hours
157	6	Variant 13	21.01.2011 09:59:00	23.01.2011 12:07:00	2 days, 2 hours
156	8	Variant 17	21.01.2011 08:49:00	25.01.2011 21:34:00	4 days, 12 hours
1310	19	Variant 8	07.07.2011 19:12:00	11.10.2011 19:01:00	95 days, 23 hours
155	23	Variant 45	20.01.2011 21:08:00	03.02.2011 10:40:00	13 days, 13 hours
334	4	Variant 4	16.02.2011 04:45:00	20.02.2011 13:36:00	4 days, 8 hours
948	2	Variant 3	17.05.2011 22:06:00	18.05.2011 10:51:00	12 hours, 45 mins
154	19	Variant 43	20.01.2011 10:40:00	03.02.2011 04:38:00	13 days, 17 hours
335	18	Variant 1	16.02.2011 05:46:00	01.03.2011 11:09:00	13 days, 5 hours
748	21	Variant 10	19.04.2011 10:49:00	14.07.2011 06:52:00	85 days, 20 hours
152	24	Variant 18	20.01.2011 00:20:00	03.02.2011 10:29:00	14 days, 10 hours
153	20	Variant 6	20.01.2011 08:25:00	01.02.2011 08:56:00	12 days, 31 mins
150	17	Variant 28	19.01.2011 15:05:00	01.02.2011 08:40:00	12 days, 17 hours

4. Inspect Statistics



Case ID	Events	Variant	Started	Finished	Duration
339	8	Variant 31	16.02.2011 14:31:00	28.02.2011 08:34:00	11 days, 18 hours
940	16	Variant 37	17.05.2011 06:31:00	28.05.2011 16:19:00	11 days, 9 hours
1417	5	Variant 12	23.07.2011 12:53:00	08.08.2011 04:23:00	15 days, 15 hours
159	23	Variant 25	21.01.2011 15:02:00	05.02.2011 14:23:00	14 days, 23 hours
330	8	Variant 17	15.02.2011 13:26:00	21.02.2011 07:59:00	5 days, 18 hours
158	17	Variant 2	21.01.2011 12:26:00	01.02.2011 21:59:00	11 days, 9 hours
949	23	Variant 25	17.05.2011 22:11:00	21.08.2011 00:43:00	95 days, 2 hours
157	6	Variant 13	21.01.2011 09:59:00	23.01.2011 12:07:00	2 days, 2 hours
156	8	Variant 17	21.01.2011 08:49:00	25.01.2011 21:34:00	4 days, 12 hours
1310	19	Variant 8	07.07.2011 19:12:00	11.10.2011 19:01:00	95 days, 23 hours
155	23	Variant 45	20.01.2011 21:08:00	03.02.2011 10:40:00	13 days, 13 hours
334	4	Variant 4	16.02.2011 04:45:00	20.02.2011 13:36:00	4 days, 8 hours
948	2	Variant 3	17.05.2011 22:06:00	18.05.2011 10:51:00	12 hours, 45 mins
154	19	Variant 43	20.01.2011 10:40:00	03.02.2011 04:38:00	13 days, 17 hours
335	18	Variant 1	16.02.2011 05:46:00	01.03.2011 11:09:00	13 days, 5 hours
748	21	Variant 10	19.04.2011 10:49:00	14.07.2011 06:52:00	85 days, 20 hours
152	24	Variant 18	20.01.2011 00:20:00	03.02.2011 10:29:00	14 days, 10 hours
153	20	Variant 6	20.01.2011 08:25:00	01.02.2011 08:56:00	12 days, 31 mins
150	17	Variant 28	19.01.2011 15:05:00	01.02.2011 08:40:00	12 days, 17 hours

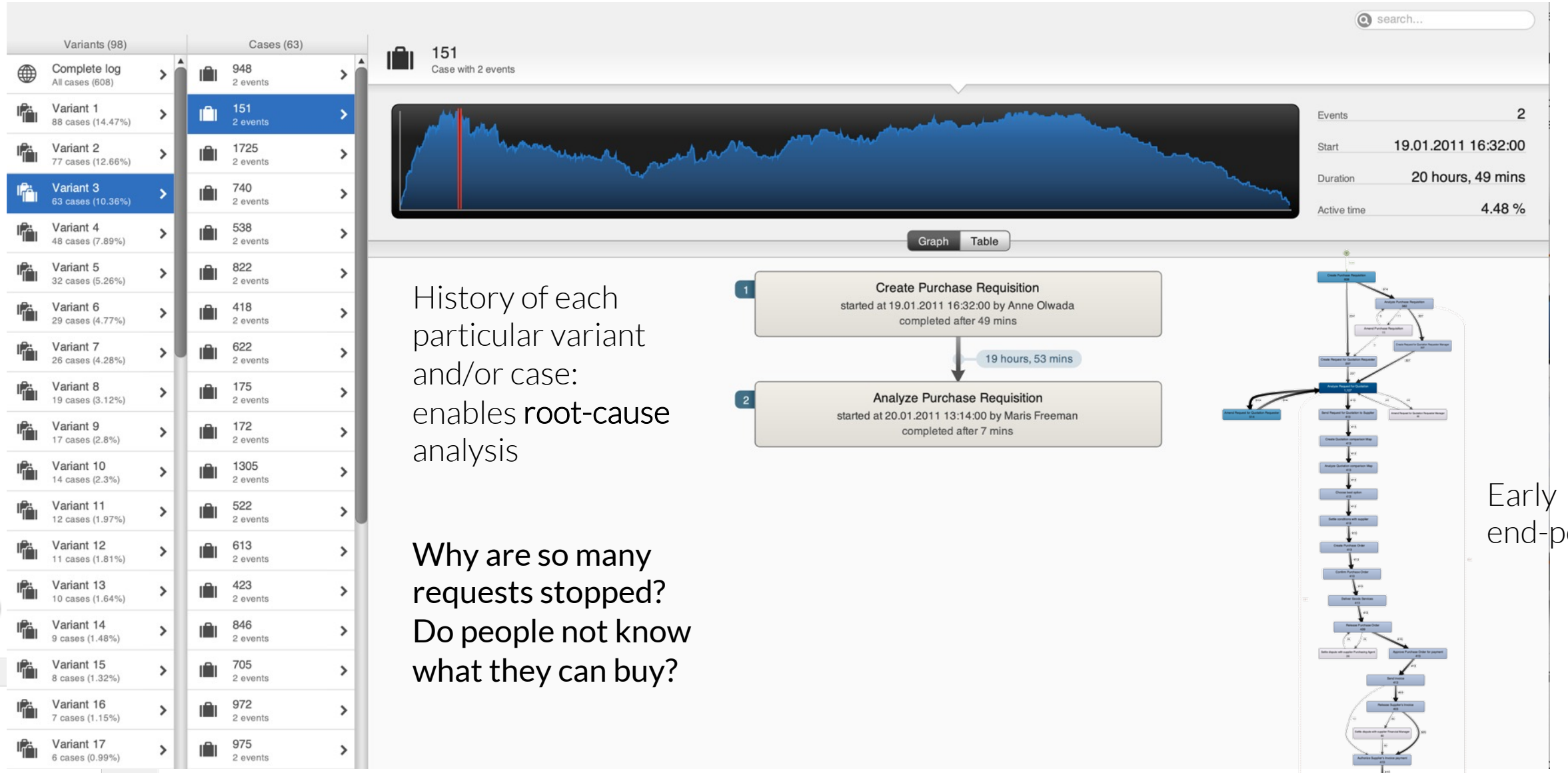
5. Inspect Cases



10% of all cases follows pattern 3

Variants ordered by frequency: 98 variants, the most frequent comprising 88 cases that make up for 15% of the data set

5. Inspect Cases



History of each particular variant and/or case: enables **root-cause** analysis

Why are so many requests stopped? Do people not know what they can buy?

Early end-point

Results so far

1. How does the process actually look like?
   - Objective process map discovered
   - Lots of amendments and stopped requests:

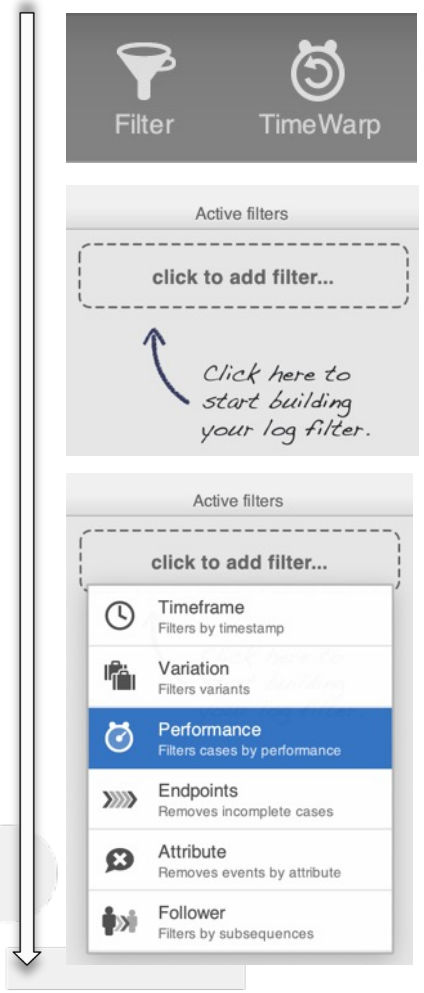
     Update of purchasing guidelines needed

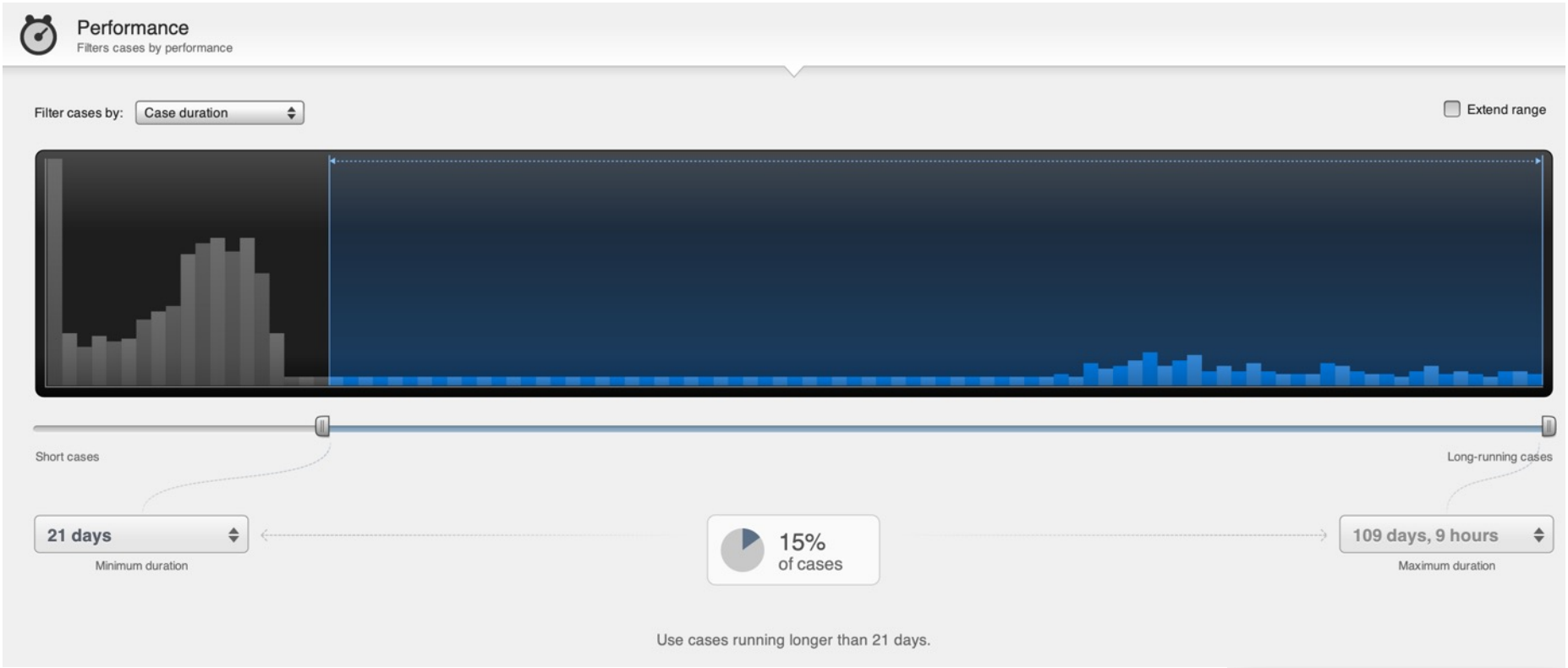
2. Do we meet the performance targets?
   - Not by all (some take longer than 21 days):

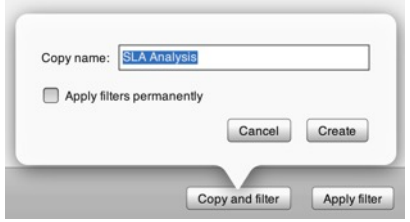
     Where in the process do we lose the time? → Next

3. Are there deviations from the prescribed process?

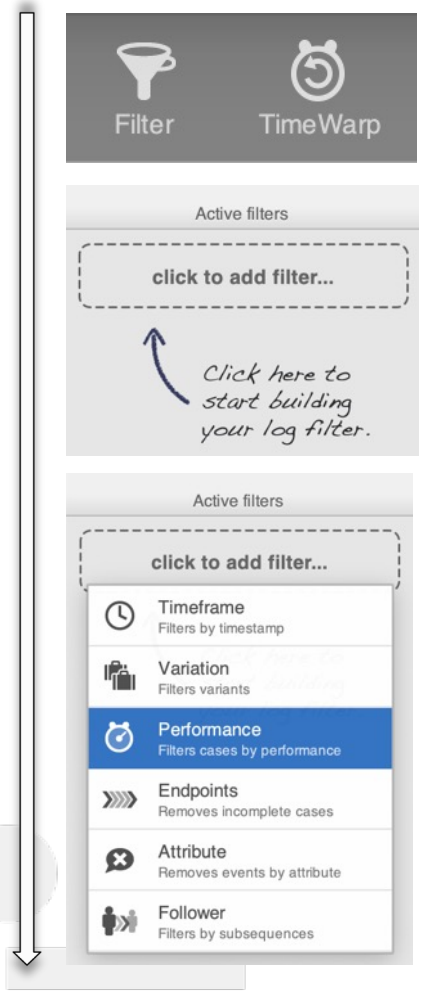
6. Filter on performance



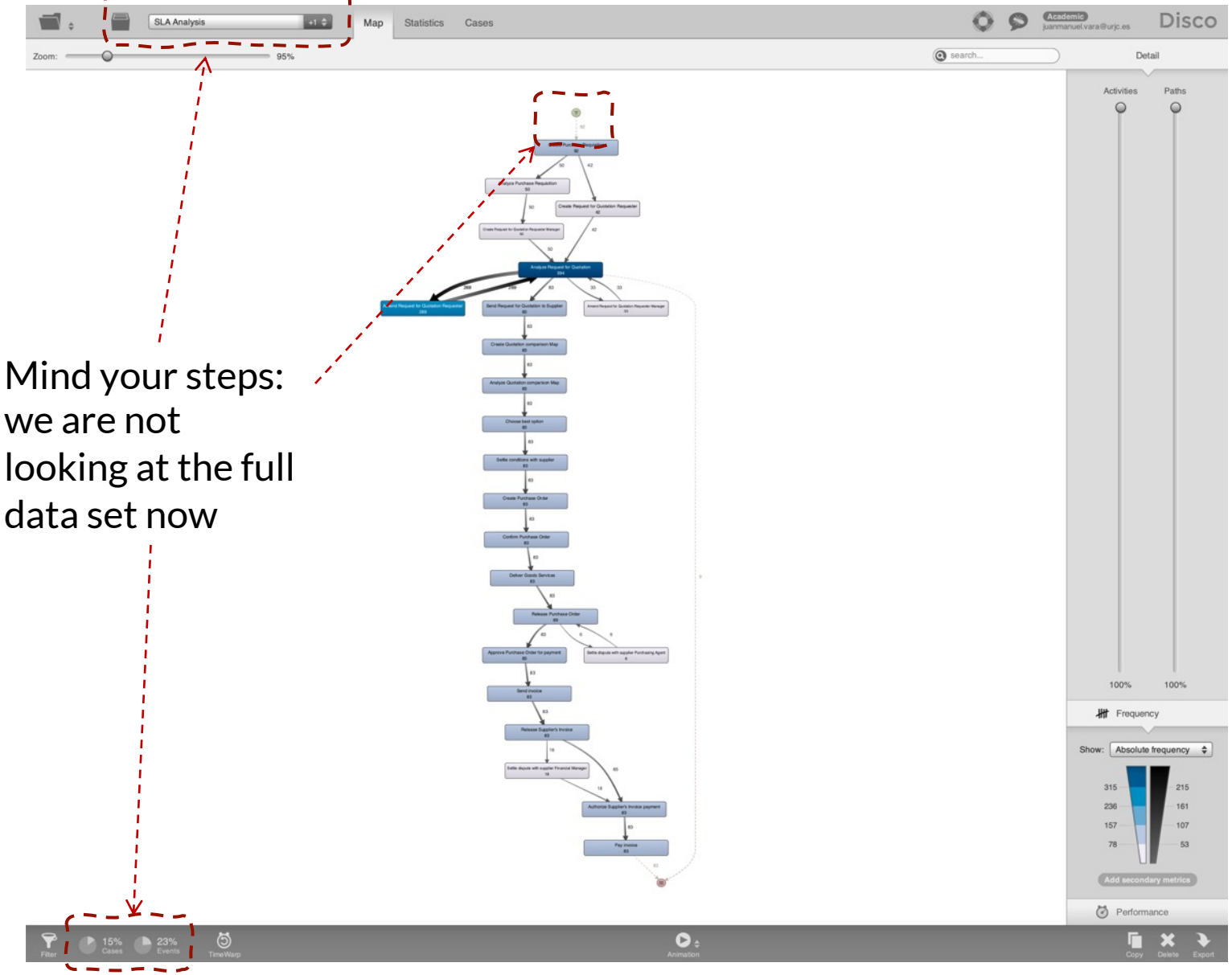




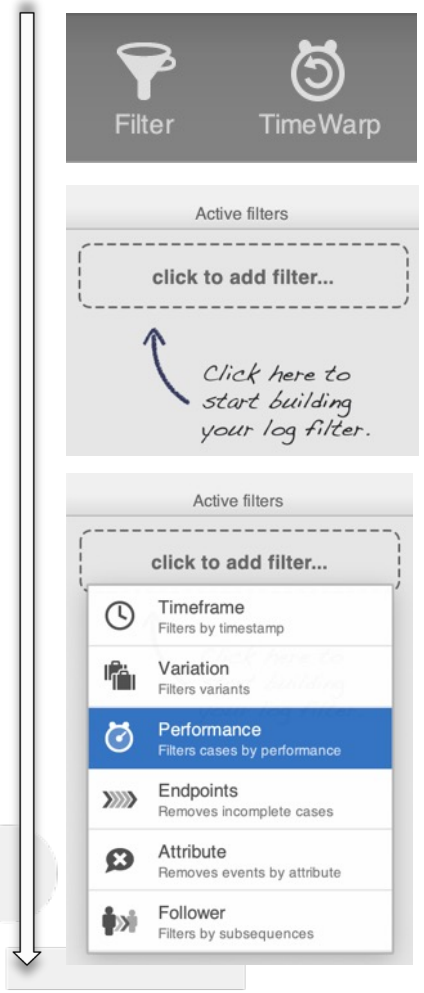
6. Filter on performance



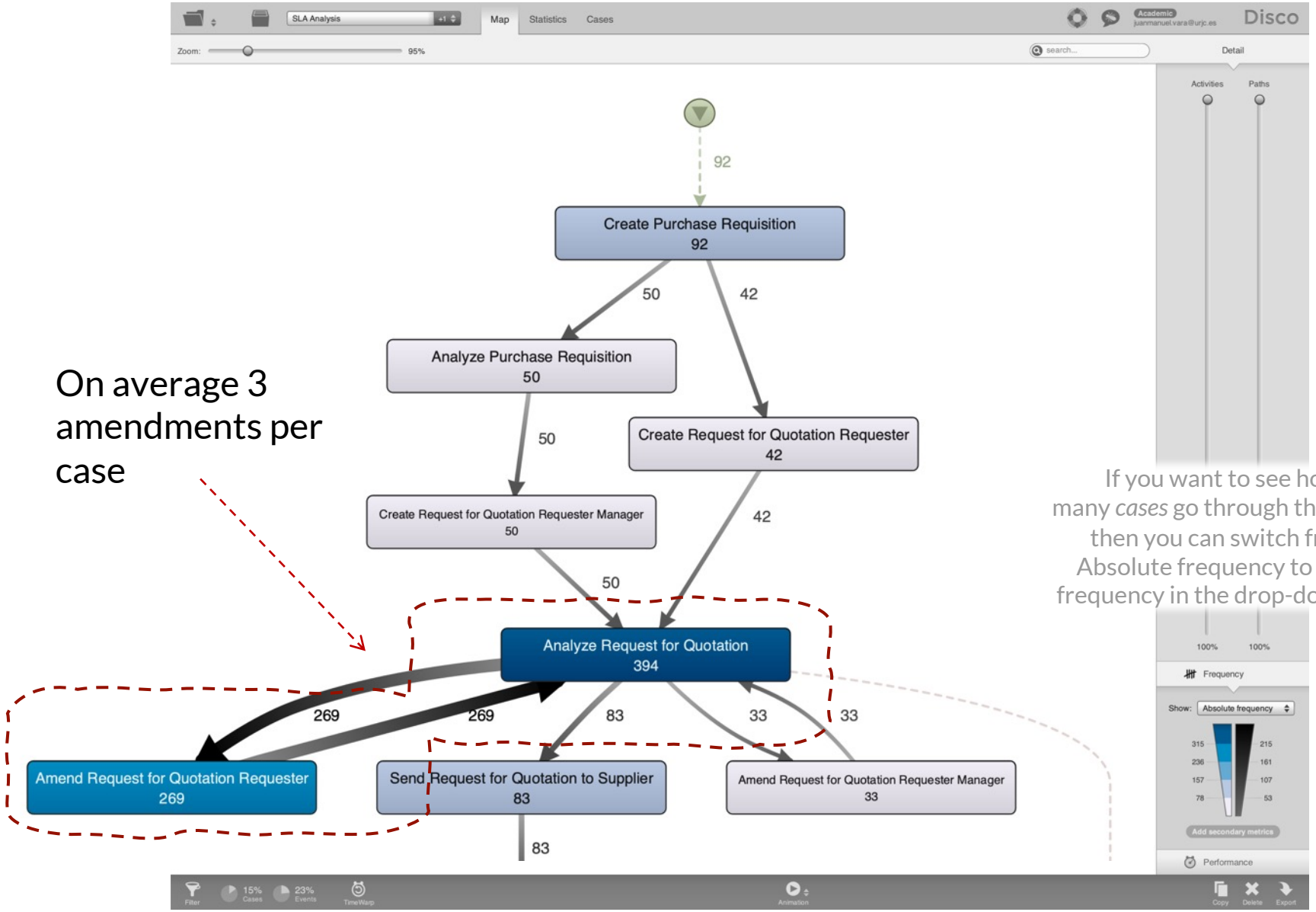
Mind your steps: we are not looking at the full data set now



6. Filter on performance

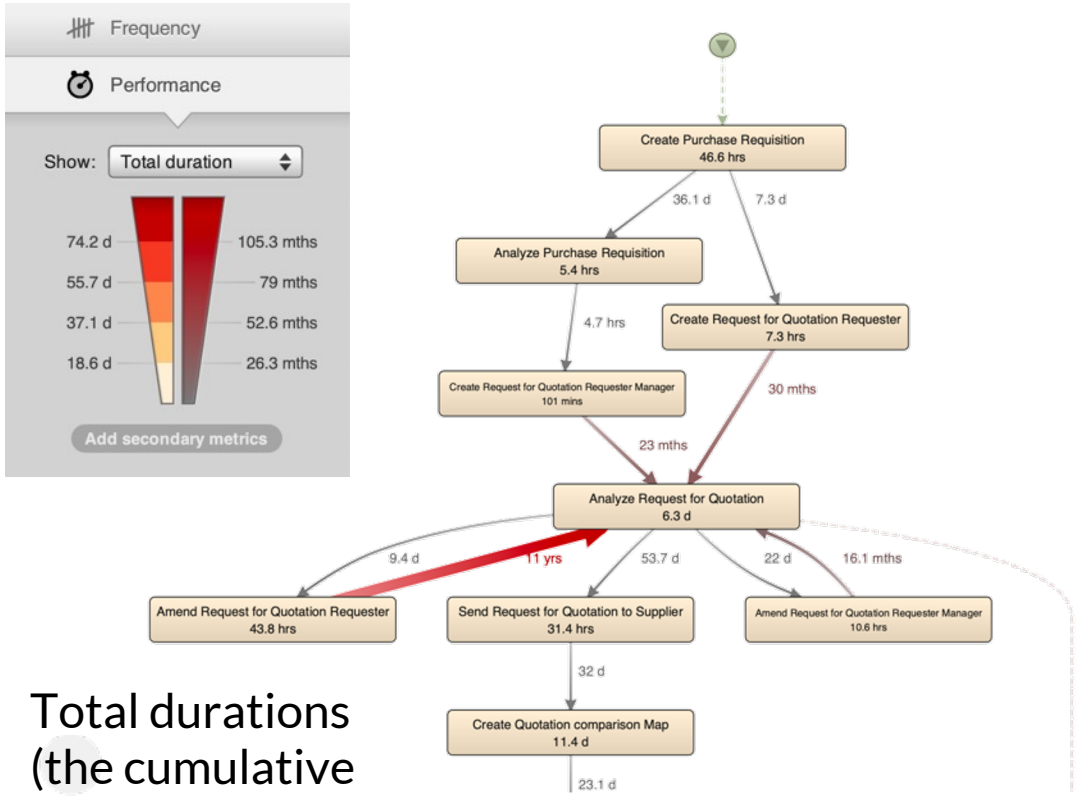


On average 3 amendments per case

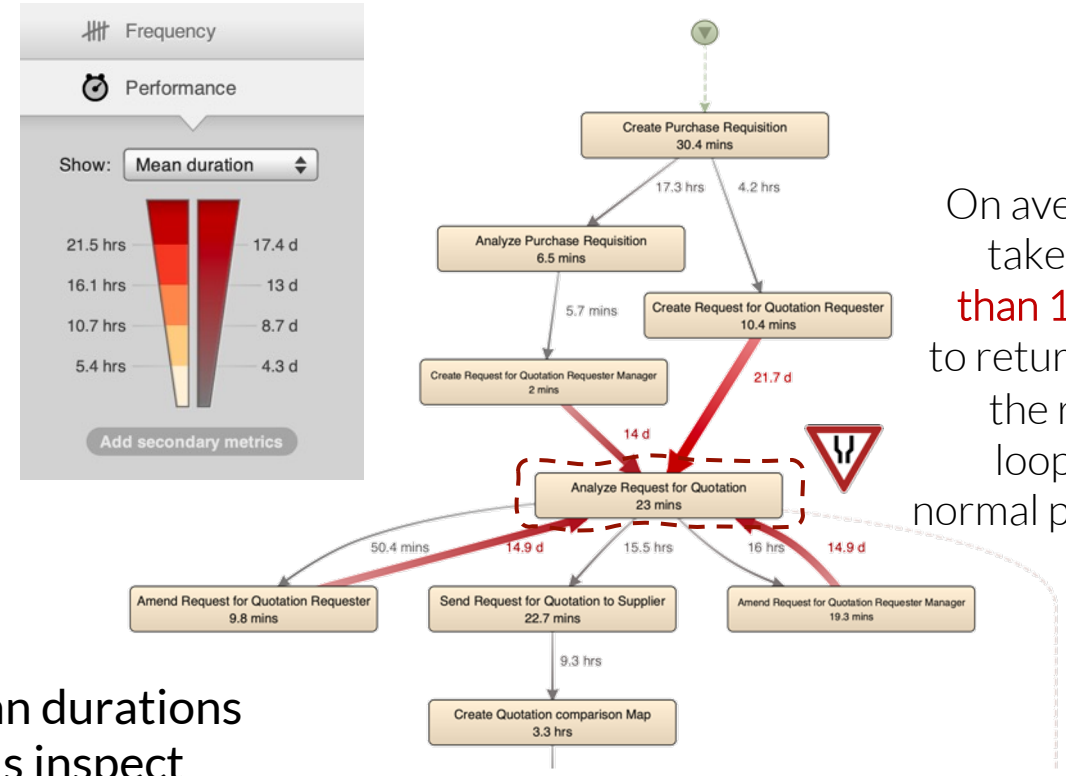


If you want to see how many *cases* go through that loop, then you can switch from Absolute frequency to Case frequency in the drop-down list

7. Visualize bottlenecks



Total durations (the cumulative delays over all cases) shows the high-impact areas



Mean durations let us inspect average delays

On average it takes more than 14 days to return from the rework loop to the normal process

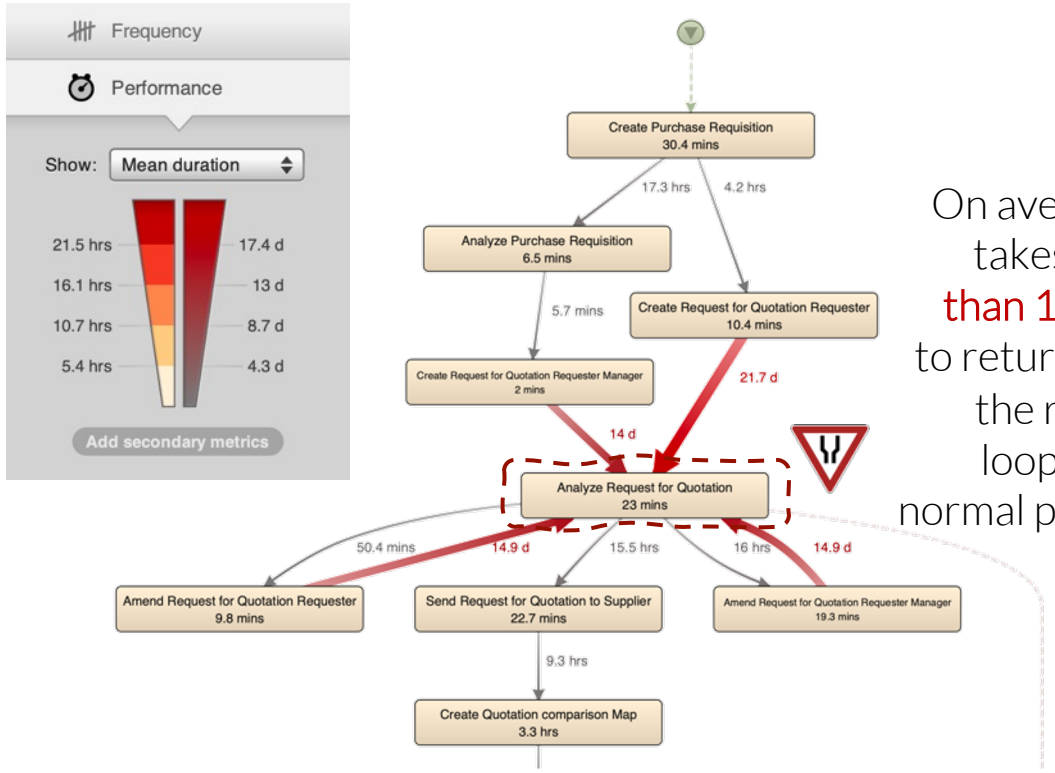
7. Visualize bottlenecks

Process mining show us *where* we have problems

- Data-based analysis allow us to objectively know where we should focus
  - Waiting times are often magnitudes higher than execution times
  - Focus is not on making people work faster but to organize the process in smarter way

Process mining helps to *communicate* our findings

- Change initiatives are hard
- Processes are complex to understand
- Charts and statistics are ofetn too abstract
- Visual representations might help to engage stakeholders



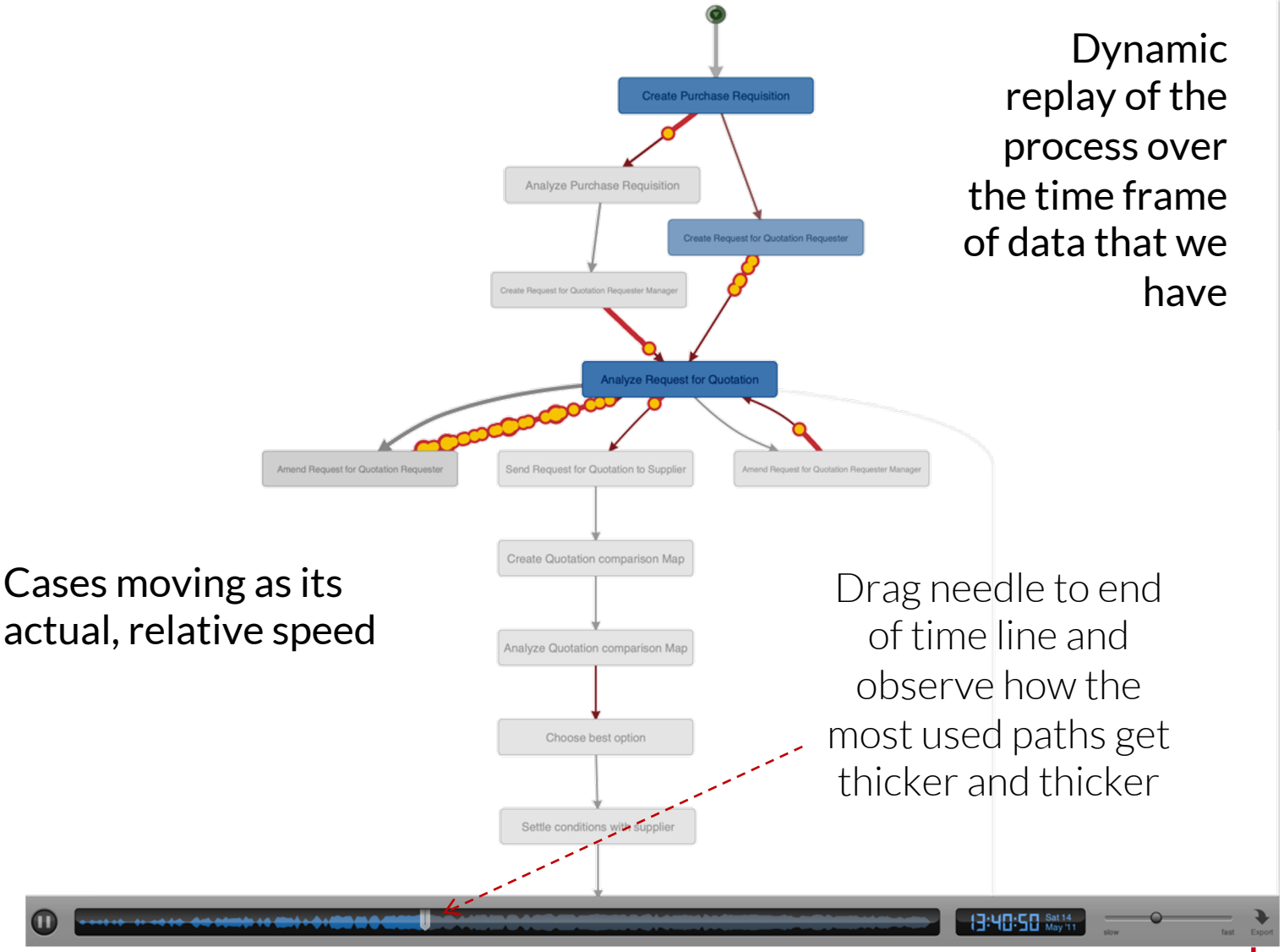
On average it takes more than 14 days to return from the rework loop to the normal process

Interactive analysis workshops to take advantage from domain experts

8. Animate Process

Animation can be extremely helpful for communication

- Can make the discovered bottlenecks really tangible for people and "bring them to life"



Dynamic replay of the process over the time frame of data that we have

Cases moving as its actual, relative speed

Drag needle to end of time line and observe how the most used paths get thicker and thicker

Results so far

1. How does the process actually look like?
   - Objective process map discovered
   - Lots of amendments and stopped requests:

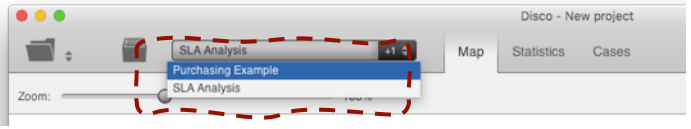
     Update of purchasing guidelines needed

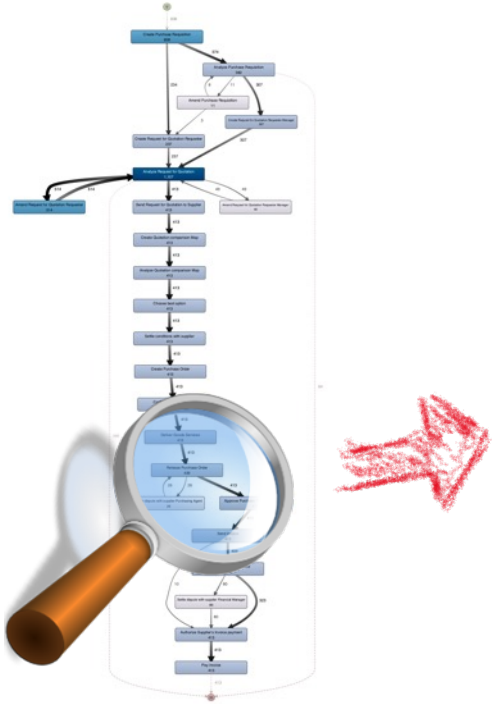
2. Do we meet the performance targets?
   - Not by all (some take longer than 21 days)
   - The 'Analyze Request for Quotation' activity is a huge bottleneck:

     Process change is needed

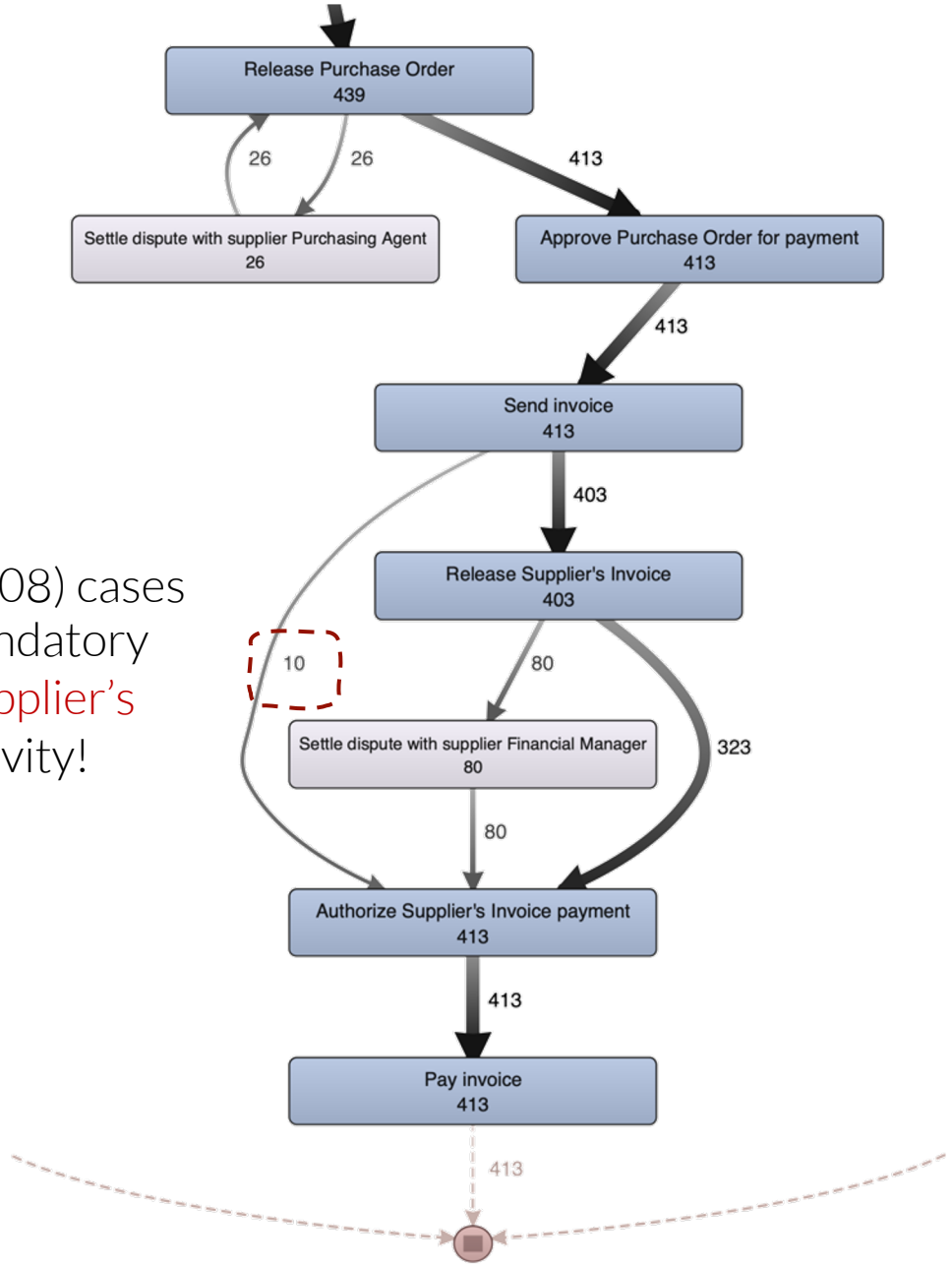
3. Are there deviations from the prescribed process? → Next

9. Compliance Check





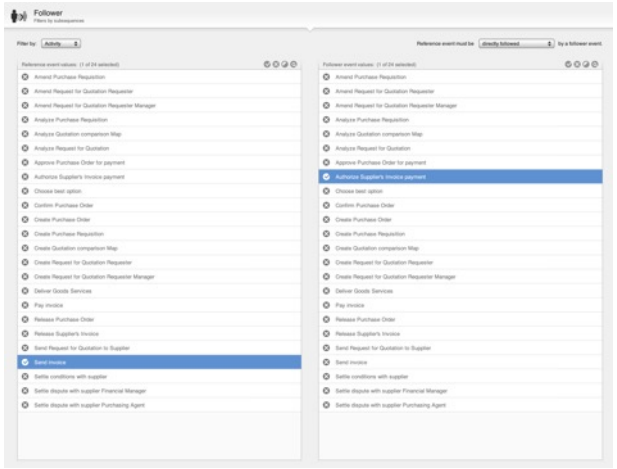
10 (out of 608) cases skip the mandatory 'Release Supplier's Invoice' activity!



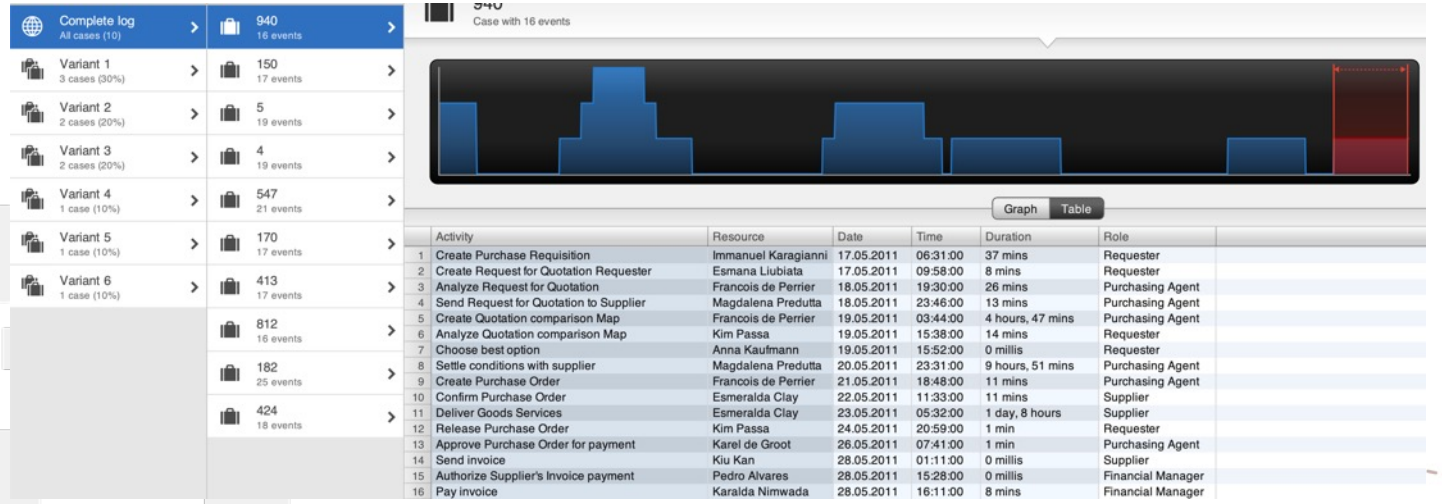
9. Compliance Check

Actionable result:

- Change the operational system to prevent the violation
- Provide targeted training



We can filter the data set to focus on the cases that go through a particular path



	Activity	Resource	Date	Time	Duration	Role
1	Create Purchase Requisition	Immanuel Karagianni	17.05.2011	06:31:00	37 mins	Requester
2	Create Request for Quotation Requester	Esmana Liubiata	17.05.2011	09:58:00	8 mins	Requester
3	Analyze Request for Quotation	Francois de Perrier	18.05.2011	19:30:00	26 mins	Purchasing Agent
4	Send Request for Quotation to Supplier	Magdalena Predutta	18.05.2011	23:46:00	13 mins	Purchasing Agent
5	Create Quotation comparison Map	Francois de Perrier	19.05.2011	03:44:00	4 hours, 47 mins	Purchasing Agent
6	Analyze Quotation comparison Map	Kim Passa	19.05.2011	15:38:00	14 mins	Requester
7	Choose best option	Anna Kaufmann	19.05.2011	15:52:00	0 millis	Requester
8	Settle conditions with supplier	Magdalena Predutta	20.05.2011	23:31:00	9 hours, 51 mins	Purchasing Agent
9	Create Purchase Order	Francois de Perrier	21.05.2011	18:48:00	11 mins	Purchasing Agent
10	Confirm Purchase Order	Esmeralda Clay	22.05.2011	11:33:00	11 mins	Supplier
11	Deliver Goods Services	Esmeralda Clay	23.05.2011	05:32:00	1 day, 8 hours	Supplier
12	Release Purchase Order	Kim Passa	24.05.2011	20:59:00	1 min	Requester
13	Approve Purchase Order for payment	Karel de Groot	26.05.2011	07:41:00	1 min	Purchasing Agent
14	Send invoice	Kiu Kan	28.05.2011	01:11:00	0 millis	Supplier
15	Authorize Supplier's Invoice payment	Pedro Alvares	28.05.2011	15:28:00	0 millis	Financial Manager
16	Pay invoice	Karalda Nimwada	28.05.2011	16:11:00	8 mins	Financial Manager

Results so far

1. How does the process actually look like?
   - Objective process map discovered
   - Lots of amendments and stopped requests:

     Update of purchasing guidelines needed

2. Do we meet the performance targets?
   - Not by all (some take longer than 21 days)
   - The ‘Analyze Request for Quotation’ activity is a huge bottleneck:

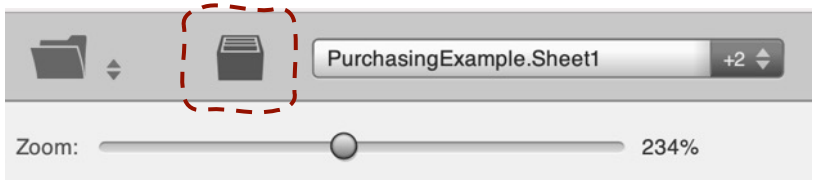
     Process change is needed

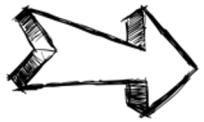
3. Are there deviations from the prescribed process?
   - Yes, training or system change needed

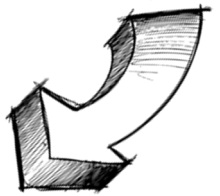
10. Organizational View

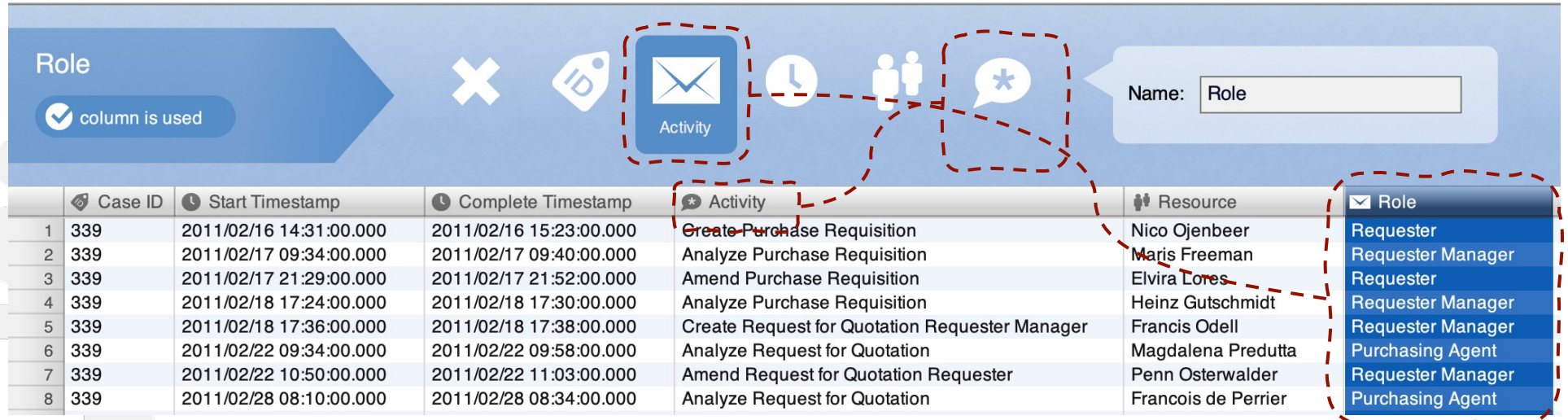
Alternative views on the data

- Re-configure the data import









	Case ID	Start Timestamp	Complete Timestamp	Activity	Resource	Role
1	339	2011/02/16 14:31:00.000	2011/02/16 15:23:00.000	Create Purchase Requisition	Nico Ojenbeer	Requester
2	339	2011/02/17 09:34:00.000	2011/02/17 09:40:00.000	Analyze Purchase Requisition	Maris Freeman	Requester Manager
3	339	2011/02/17 21:29:00.000	2011/02/17 21:52:00.000	Amend Purchase Requisition	Elvira Lores	Requester
4	339	2011/02/18 17:24:00.000	2011/02/18 17:30:00.000	Analyze Purchase Requisition	Heinz Gutschmidt	Requester Manager
5	339	2011/02/18 17:36:00.000	2011/02/18 17:38:00.000	Create Request for Quotation Requester Manager	Francis Odell	Requester Manager
6	339	2011/02/22 09:34:00.000	2011/02/22 09:58:00.000	Analyze Request for Quotation	Magdalena Predutta	Purchasing Agent
7	339	2011/02/22 10:50:00.000	2011/02/22 11:03:00.000	Amend Request for Quotation Requester	Penn Osterwalder	Requester Manager
8	339	2011/02/28 08:10:00.000	2011/02/28 08:34:00.000	Analyze Request for Quotation	Francois de Perrier	Purchasing Agent

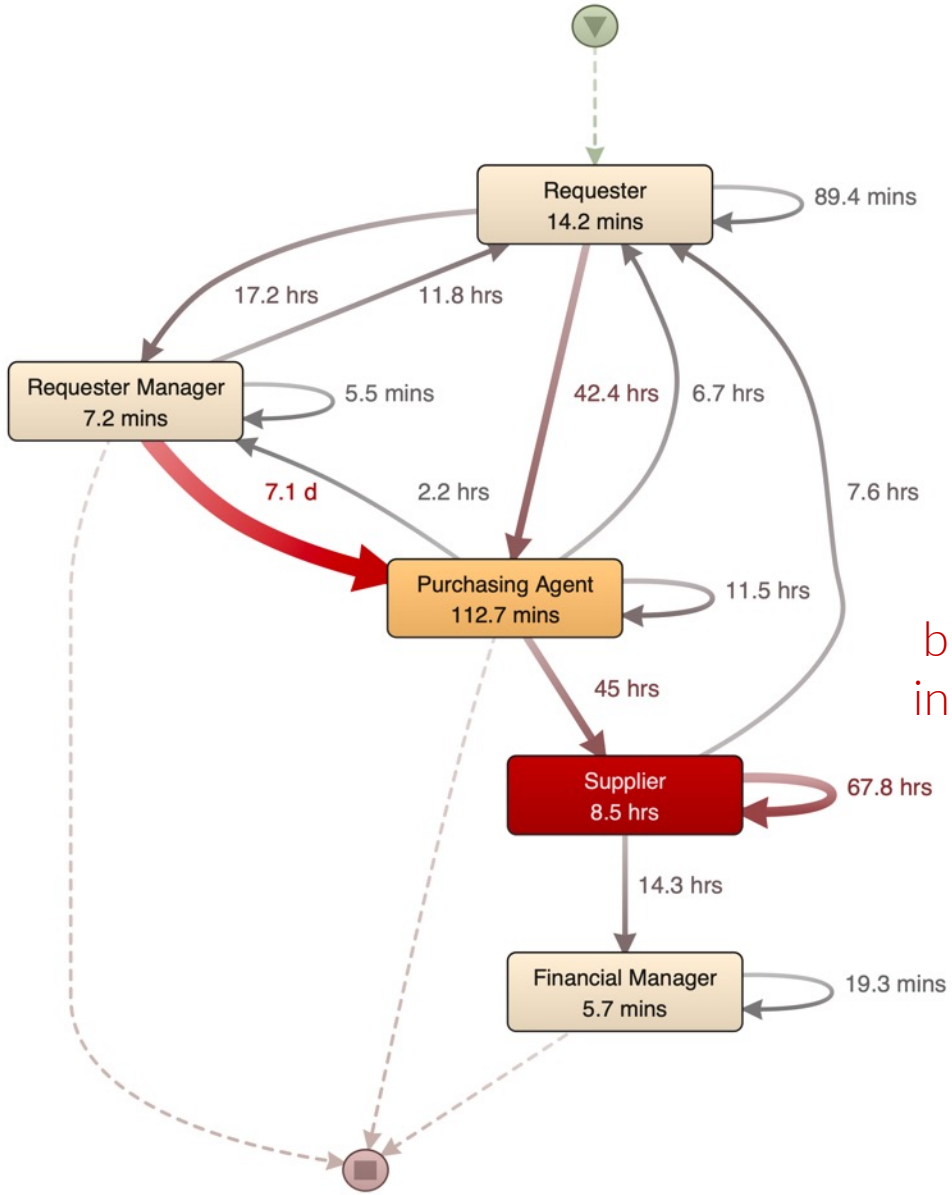
10. Organizational View

Different perspective

- We can take many different perspectives, even based on the **same data set**

We now don't see the activity sequences displayed but rather the hand-over of work between different functions, or departments

- Inefficiencies can often be found at the borders of organizational units

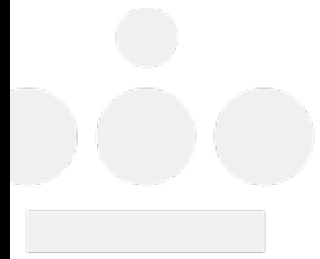


Clearly, the Purchasing agents are causing the biggest delays in the process!

Take-away

A few things you can learn

- Real processes are often more complex than you might expect
  - Manual process analysis get the *Sunny day scenario*
  - Process Mining get a hold of all the exceptions
- There is not one *right* model
- Process mining is not about mining a data set to create one process model
  - Explore the process by translating questions into filters
  - Interactive activity needing from domain knowledge to interpret findings



The hype

TECH

How three friends turned a college project into a $2.5 billion software unicorn

CNBC

PUBLISHED THU, NOV 21 2019•8:01 AM EST | UPDATED THU, NOV 21 2019•9:56 AM EST

Microsoft brings new process mining features to Power Automate

TC

IBM to acquire myInvenio with an eye on AI-enabled automation

by R. Dallon Adams in Innovation
on April 15, 2021, 5:00 AM PST

With the deal, IBM looks to offer organizations a suite of capabilities such as robotic process automation, document processing and process mining.

TECHNOLOGY, MEDIA & TELECOM - INNOVATION APRIL 1, 2021 / 6:07 AM / UPDATED A MONTH AGO

German process mining startup Celonis teams up with IBM and Red Hat

REUTERS

Tooling landscape

ProM

- Open-source, extensible framework
- More than 1500 plugins (#29 in 2004!)

- Lower the threshold for process mining.
- Inability to discover concurrency well.
- Focus on performance analysis rather than conformance checking and precise models.

Disco

- Low threshold
- Easy to use yet scalable

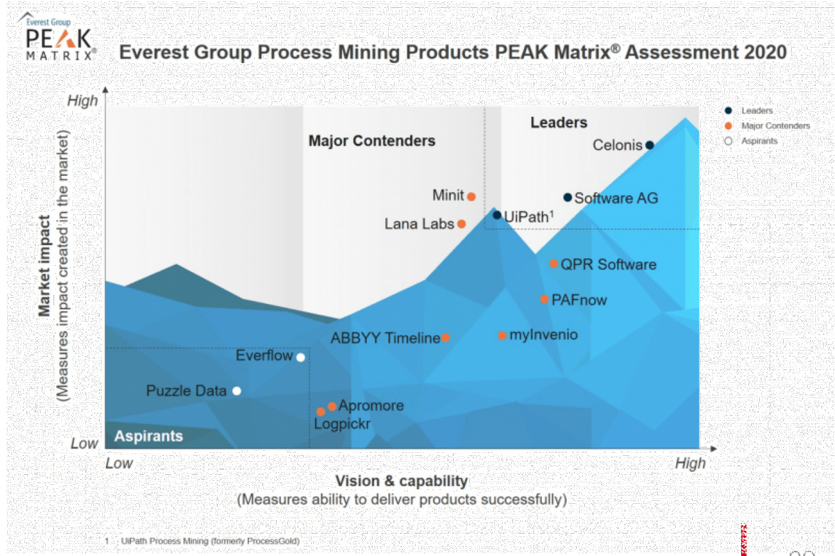
Disco by Fluxicon

CELONIS

- Cloud-based + BI capabilities
- Strong connection with SAP

celonis

Over 35 process mining vendors by the end of 2020

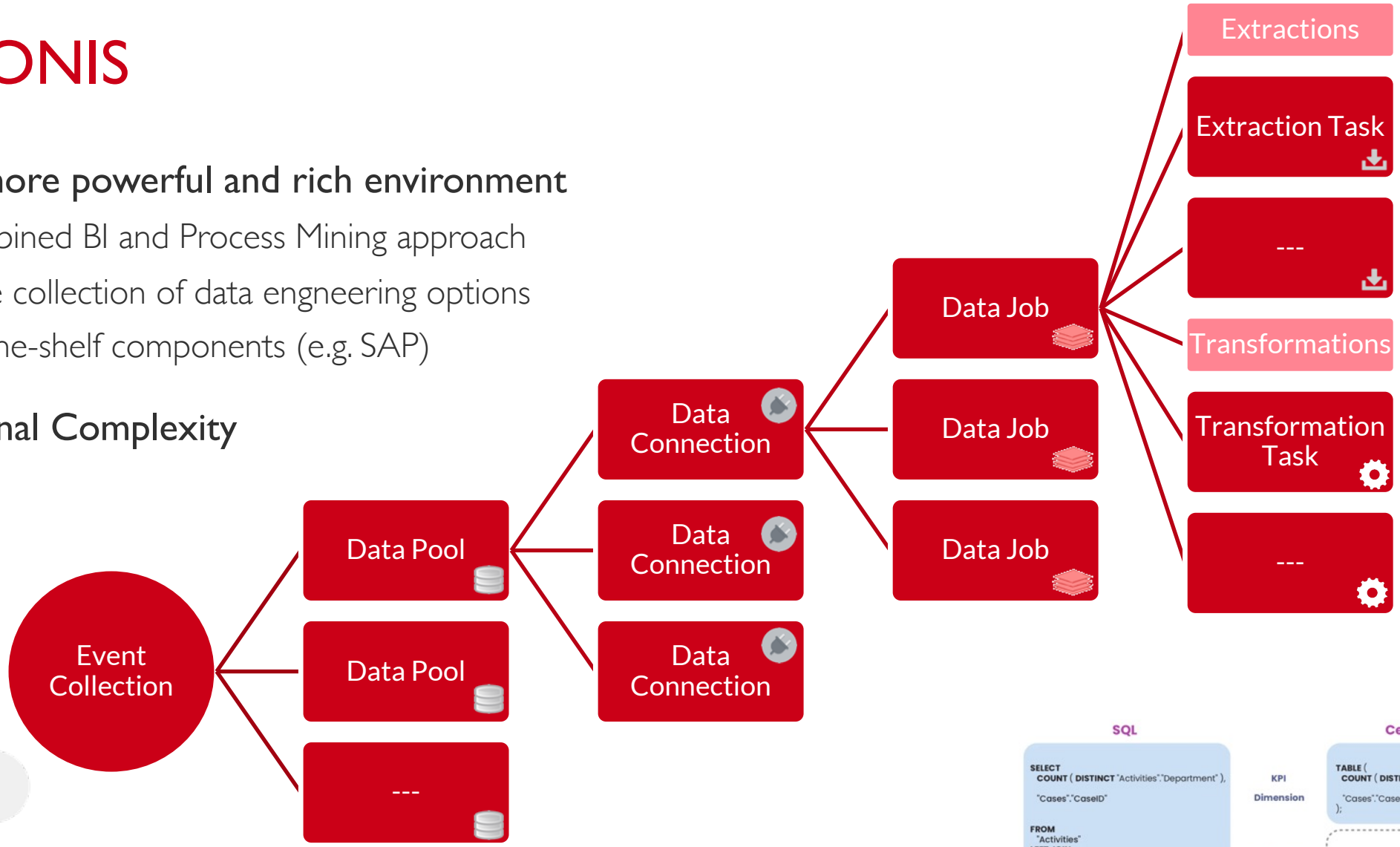


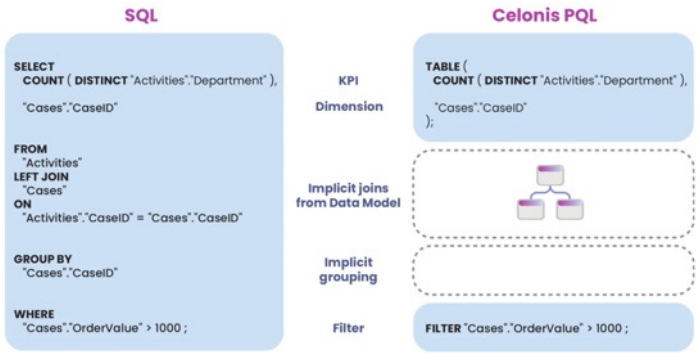
CELONIS

Much more powerful and rich environment

- Combined BI and Process Mining approach
- Wide collection of data engneering options
- Off-the-shelf components (e.g. SAP)

Additional Complexity





Data Engineer Tasks

- Create new Data Pool in Event Collection
- Create Data Connections to your source systems
- Create a Data Job and set up an Extraction Task to extract all relevant tables
- Set up the Transformation Tasks
  - Create Activity Table
  - Insert activities into the empty Activity Table
    - Specify object
    - Select Informacion
    - Insert Activity
- Create a data model and extend it by adding master and data tables
- Apply name mapping

Connect

Extract

Transform

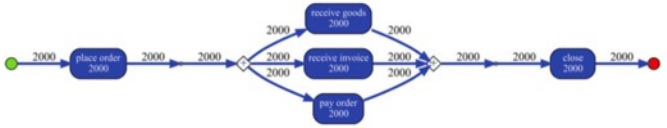
Set up data model

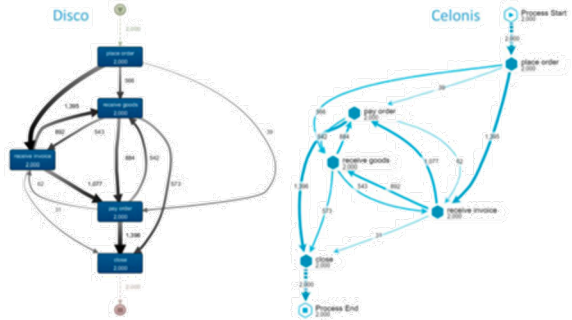
Challenges & Open Issues

Directly-Followed Graphs Issues

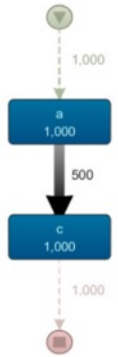
Existing tools start from DFGs for Discovery

- DFGs cannot express choice & concurrency
  - This may lead to Spaghetti-like DFGs with loops





- Filtering DFGs using frequency-based thresholds may provide misleading results



- Performance information mapped onto DFGs can be misleading
  - The average time reported between two activities is conditional

- Limited support for conformance checking
  - Rule-based approach being the least bad solution

Simplicity & Performance
(over accuracy)

(Van der Aalst, 2019)

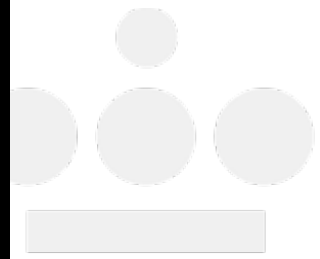
Hurdles for widespread adoption

Data quality

- 80% of the efforts and time are spent on locating, selecting, extracting and transforming the data
- Time needed to apply process mining is short (say 20%) once the data is available in the right format

People

- Stakeholders are unaware of process mining capabilities
- Hard to distinguish from ML and AI
- May reveal mismanagement and complicance problems



(Gartner, 2020)

Trends (mainly Academic)

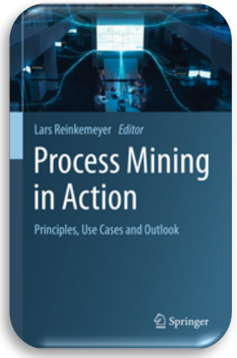
Digital Twins of an Organization (DTO)

"A dynamic software model of any organization, that relies on operational and/or other data to understand how an organization operationalizes its business model, connects with its current state, responds to changes, deploys resources and delivers expected customer value."

(Kerremans, 2018)

Process Discovery as the starting point

- Uptake in conformance checking & performance analysis techniques
- Data Mining & ML to find root causes

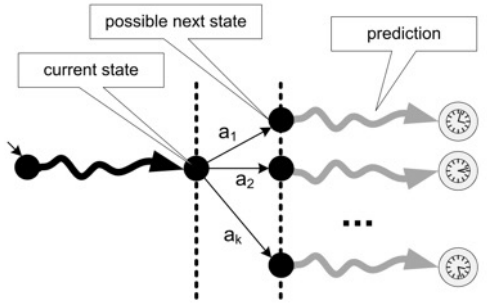


None of the Use Cases compiled in (Reinkenmeyer, 2020) dealing with conformance checking

Trends (mainly Academic)

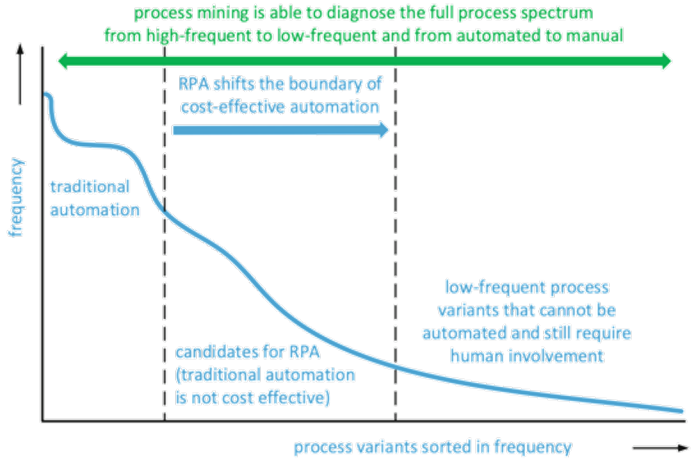
Forward looking

- Backward-looking can be used to fundamentally improve processes
- Provides little support for the day-to-day management of process
- Need to be able to analyze cases that are still running



Call to action

- Process Mining to identify candidates for RPA and monitor SW robots
- RPM to translate Process Mining diagnostics into management actions



(van der Aalst, 2020)

The Challenge

The need for **Business Process Hygiene**

There exists so much information that people are looking for new ways to leverage their data

From silos to corporate process mining

- The need for a digital workforce
- Mindset change
  - *"do I have to use data"* → *"can I afford not to use available data"*

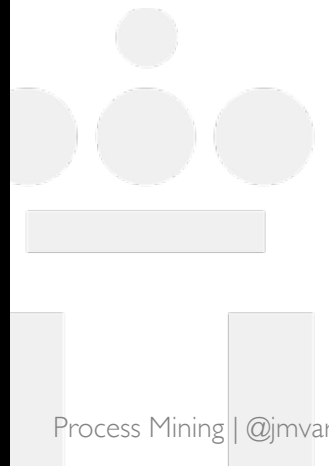
We are probably paving the way

- Spreadsheets for process analysts
- 35+ Process Mining tools
- More than 135K following the Process Mining MOOC

(van der Aalst, 2019; Reinkenmeyer, 2020b)

Personal reflection

Academia Basic Research	Algorithms	Discovery
		Conformance Checking
		Predictive process monitoring
		Object-centric Process Mining
	Data Engineering	Automate ETL
		Deal with uncertain and continuous event data
Industry Applied Research	New application fields	University Processes Portfolio
		Service Design
		Blockchain
	New applications in explored fields	Gather new insights
		Ad-hoc methods & techniques



References

- Bridgewater, A. (2021). A Rising Tide, High Times For Low-Code. *Forbes,* Jan 27, 2021.
- Celonis.(2020). *Process Mining for Experts Certificate.*
- Davenport, T. H., & Spanyi, A. (2019). What process mining is, and why companies should do it. *Harvard Business Review, 97*(2).
- Fluxicon. (2020). *Hands-on Tutorial.* https://fluxicon.com/book/read/tutorial/
- Gartner. (2020). Interview with Prof.dr.ir. Wil van der Aalst. In *2020 Gartner Market Guide for Process Mining.* Research Note G00733123
- Kerremans, M. (2018). Market Guide for Technologies Supporting a DTO. *Gartner Inc.*
- Jans, M. (2017). *From relational database to valuable event logs for process mining purposes: a procedure.* Technical report, Hasselt University.
- Reinkemeyer, L. (2020). *Process Mining in Action*. Cham: Springer International Publishing.
- Reinkemeyer, L. (2020b). Business View: Towards a Digital Enabled Organization. In *Process Mining in Action* (pp. 197-206). Springer, Cham.
- Rozinat, A., Günther, C. (2018). *Disco, Process Mining for Professionals*. : https://fluxicon.com/

References

- van der Aalst, W. (2016). *Process Mining: Data Science in Action*. Springer, Berlin, Heidleberg.
- van der Aalst, W. (2018). (MOOC) ***Process Mining: Data science in Action***. Coursera: https://www.coursera.org/learn/process-mining
- van der Aalst, W. (2019). *Business Process Hygiene: No Business Case Needed.* https://bit.ly/3fg2KpR
- van der Aalst, W. (2020). Academic view: Development of the process mining discipline. In *Process Mining in Action* (pp. 181-196). Springer, Cham.
- Weske, M. (2021). Process Mining: An openHPI course. *Practical Aspects Simulating Event Logs.*