

Data Science × Business Process Mgt =

Process Mining

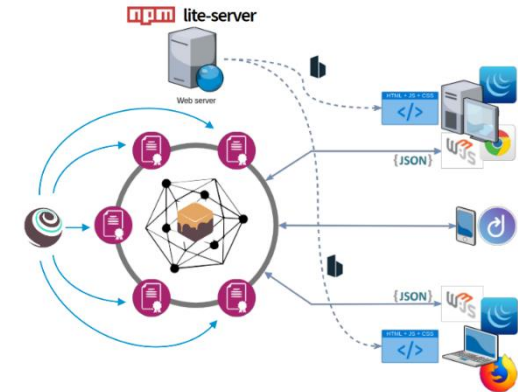
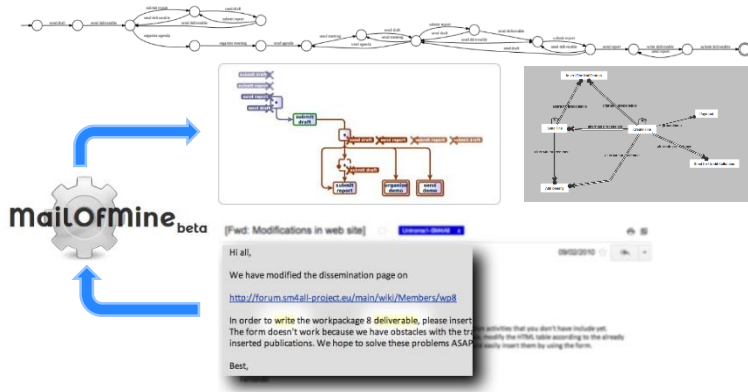
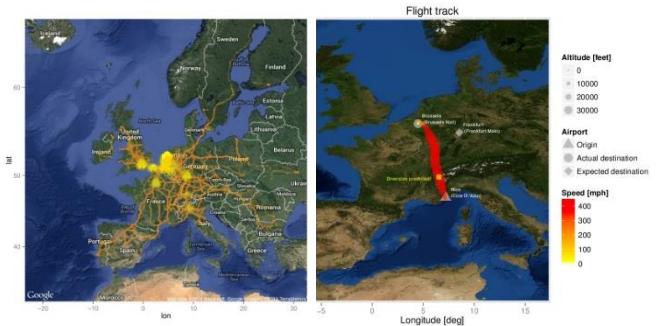
Dr. Claudio Di Ciccio
<http://diciccio.net/>
claudio.di.ciccio@wu.ac.at



Claudio Di Ciccio /kl'audio di 'tʃi:ttʃo/



Assistant professor
 Ph.D. in Computer Science
 Main research interests:
 process mining,
 declarative processes,
 cryptoeconomics,
 service-oriented arch's



Key facts



- Est. 1898 (Imperial Export Academy)
- 1919: The University of World Trade
- 1975: WU, Vienna University of Economics and Business
- 2013: New campus
- 2015: Triple accreditation (1% worldwide)

23,545 **Students enrolled (2017)**

2,300 **Staff & faculty members**

1,000 **Ph.D students**

35,000 **Campus WU area (m²)**

13th/100 **FT ranking
(Masters in Mgt.) 2018**

51-100 **QS Ranking
Business & Mgt. Studies**

58th/100 **Executive MBA Ranking
2017**





WU

WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

Institute for Information Business
<https://wu.ac.at/infobiz>



A process mining perspective

Business process mgt.

WU

WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

Organisation, modelling, execution, and tracking of processes



Half empty or half full?

<http://aeon.co/ideas/which-is-more-fundamental-processes-or-things>

Which is more fundamental: processes or things?



Neither half-full nor half-empty. *Courtesy Wikipedia*

Processes are into dynamics

Thanks to Dr. Mieke Jans for offering her hand

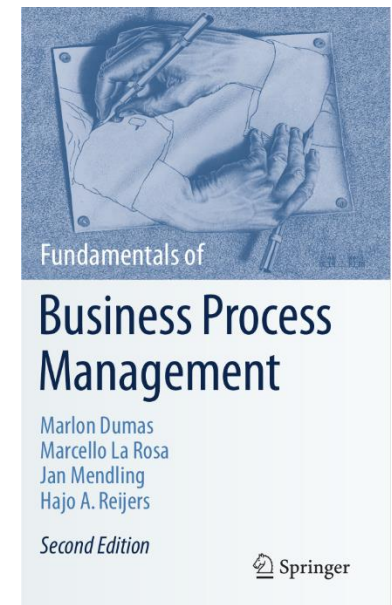


Relevance of business processes

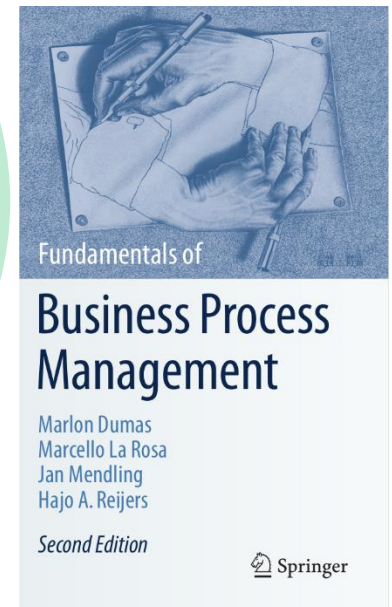
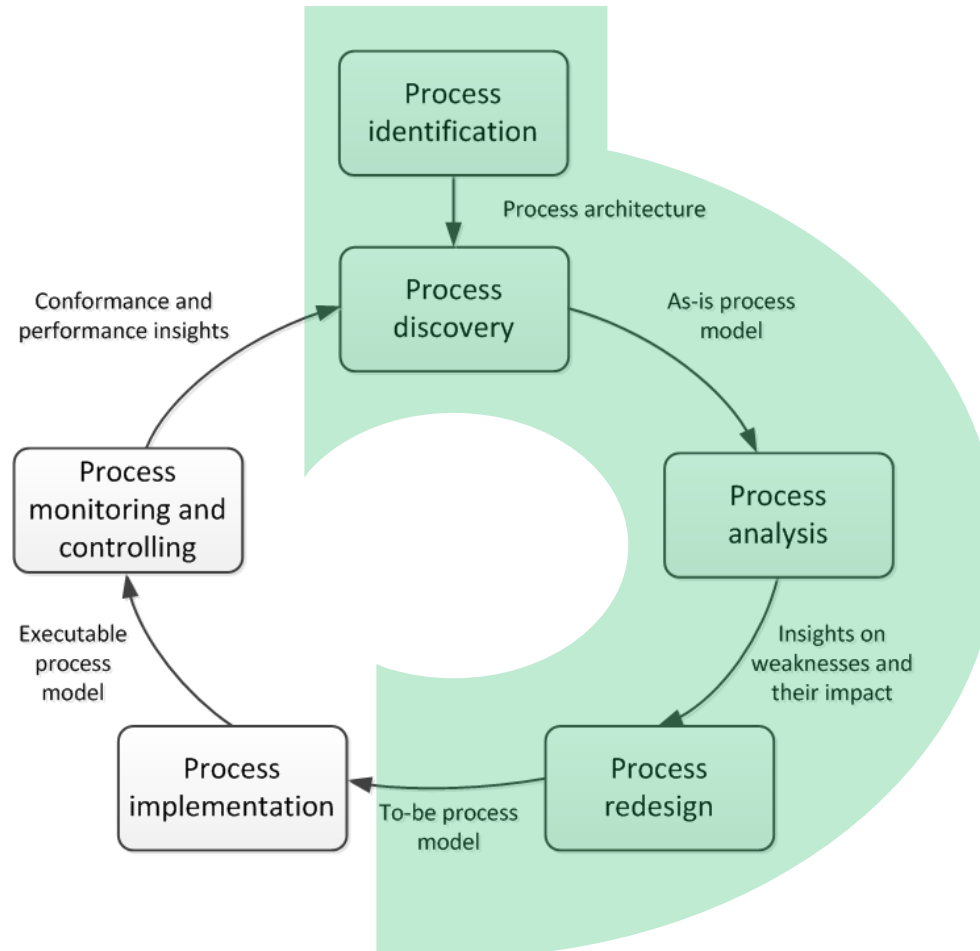
Dumas et al., 2018

- Business processes are everywhere
 - Products and services are provided by **activities**
 - Execution of activities requires **coordination**
 - **Success** of this coordination influences costs, time, and quality of products and services

- A noticeable definition:
 - "a set of **logically related tasks** performed to **achieve** a defined business **outcome** for a particular customer or market"
[Davenport 1992]



BPM lifecycle and positioning



Processes and division of labour

Smith, 1776

To take an example, the trade of a pin-maker: But in the way in which this business is now carried on, it is divided into a number of branches:

- One man draws out the wire;
- another straightens it;
- a third cuts it;
- a fourth points it;
- a fifth grinds it at the top for receiving the head;
- to make the head requires three operations;
- to put it on is a peculiar business;
- to whiten the pins is another;
- to put them into the paper;

and the important business of making a pin is, in this manner, divided into about eighteen distinct operations.



Process models

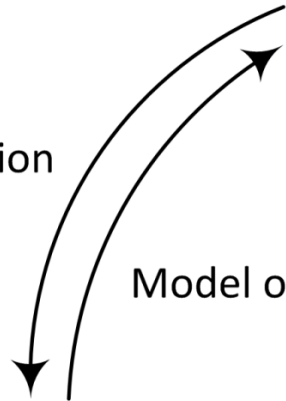
Slide courtesy of Prof. Jan Mendling

Original



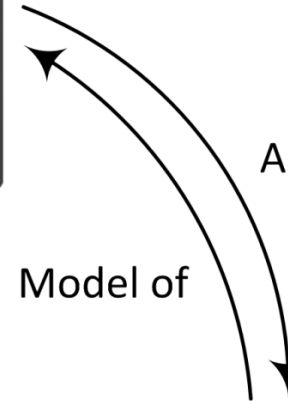
© Stiftung Deutsches Technikmuseum Berlin

Abstraction



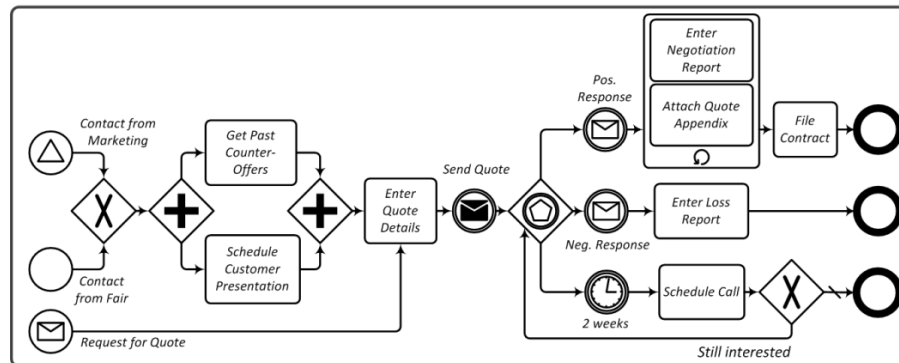
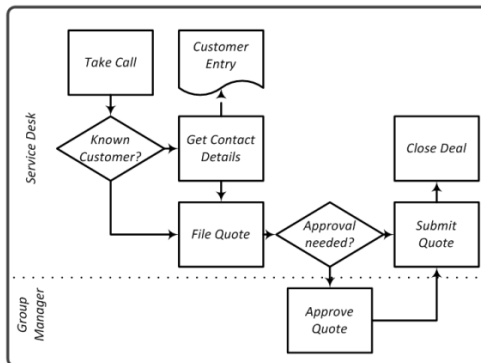
Model of

Abstraction

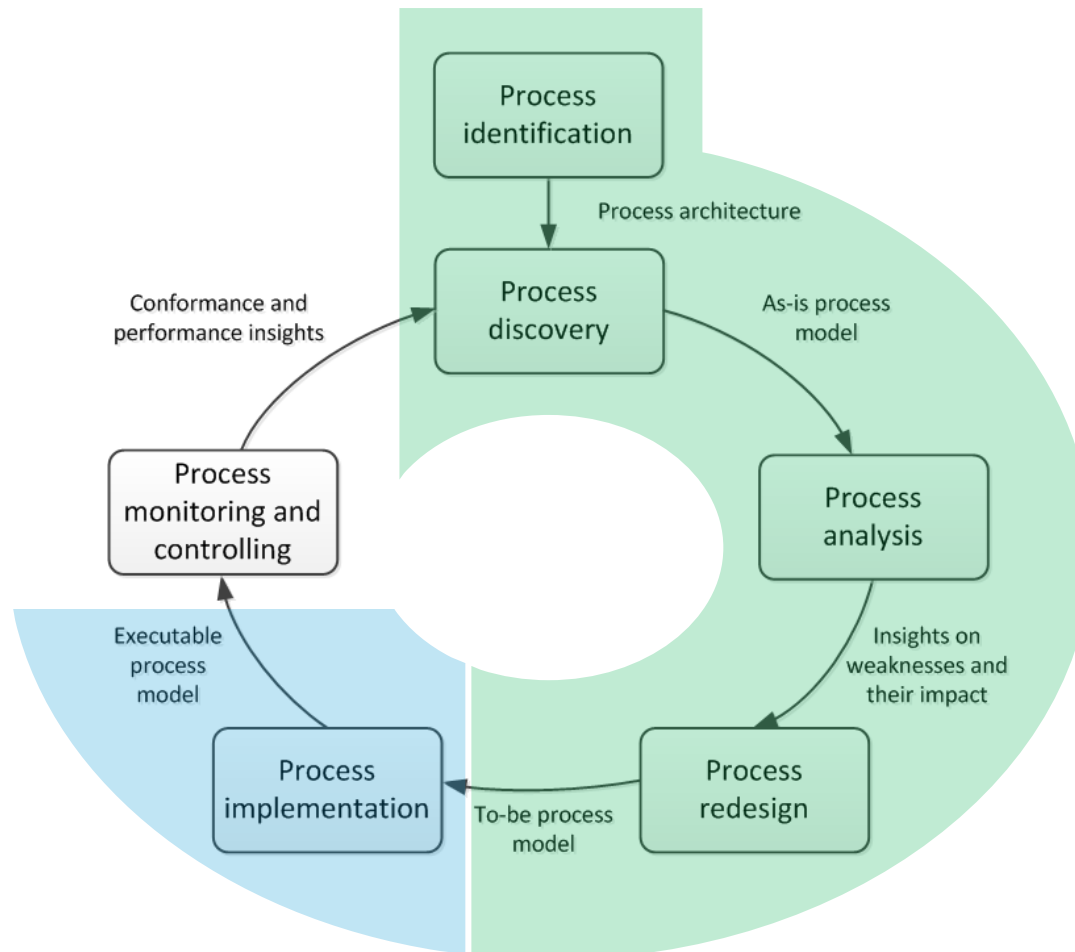


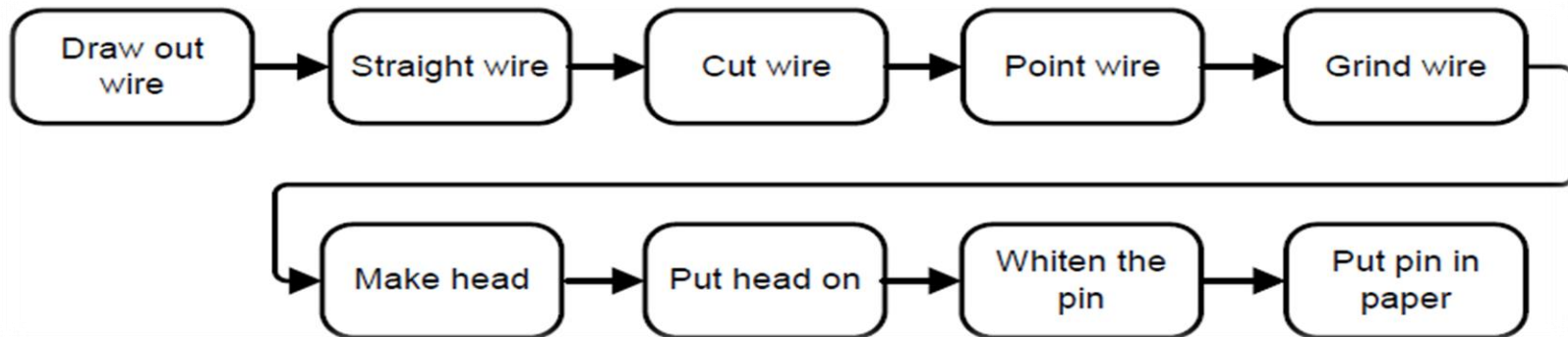
Model of

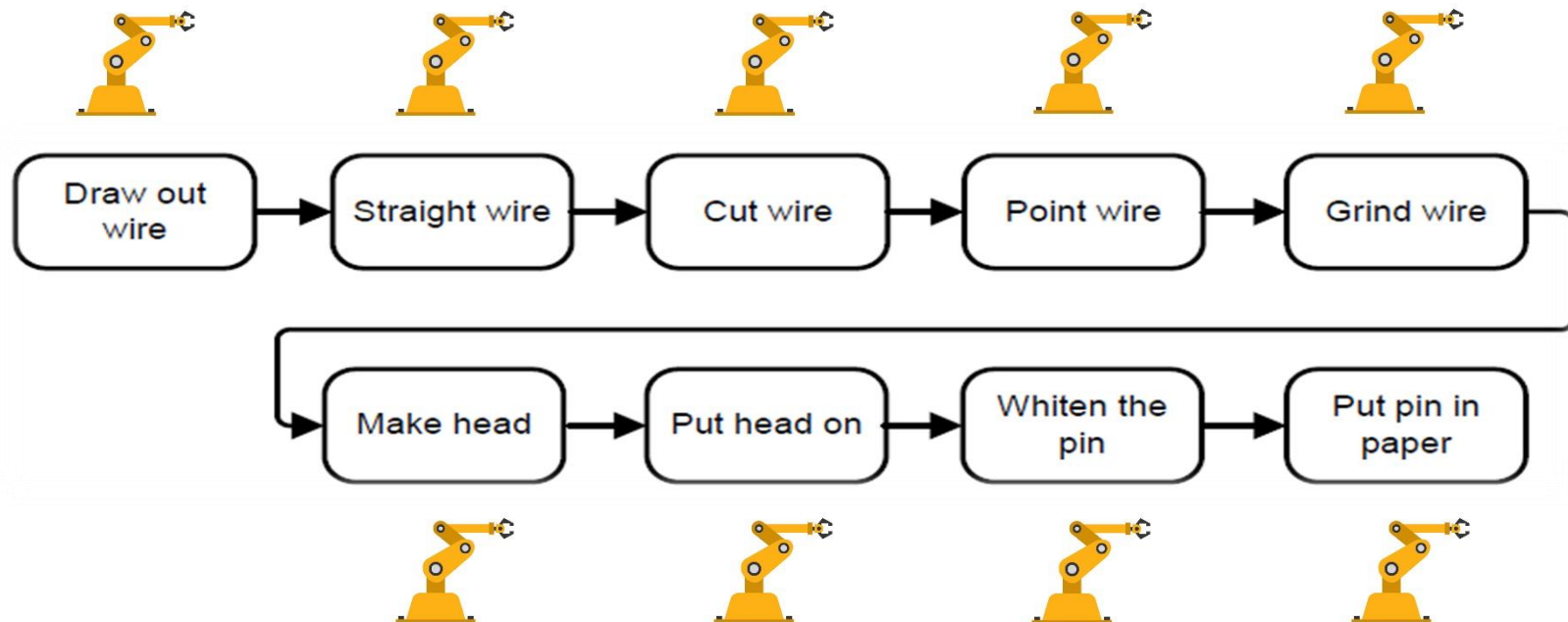
Model Level



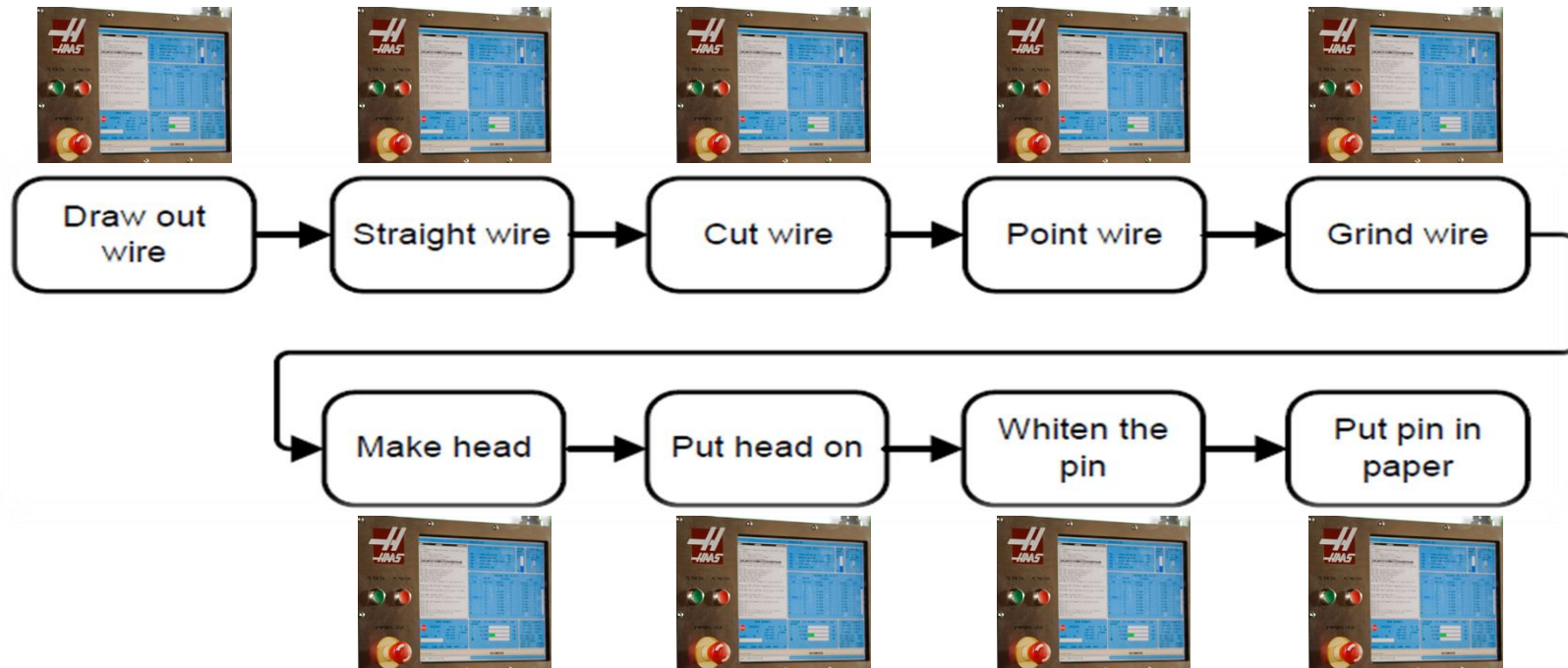
BPM lifecycle and positioning





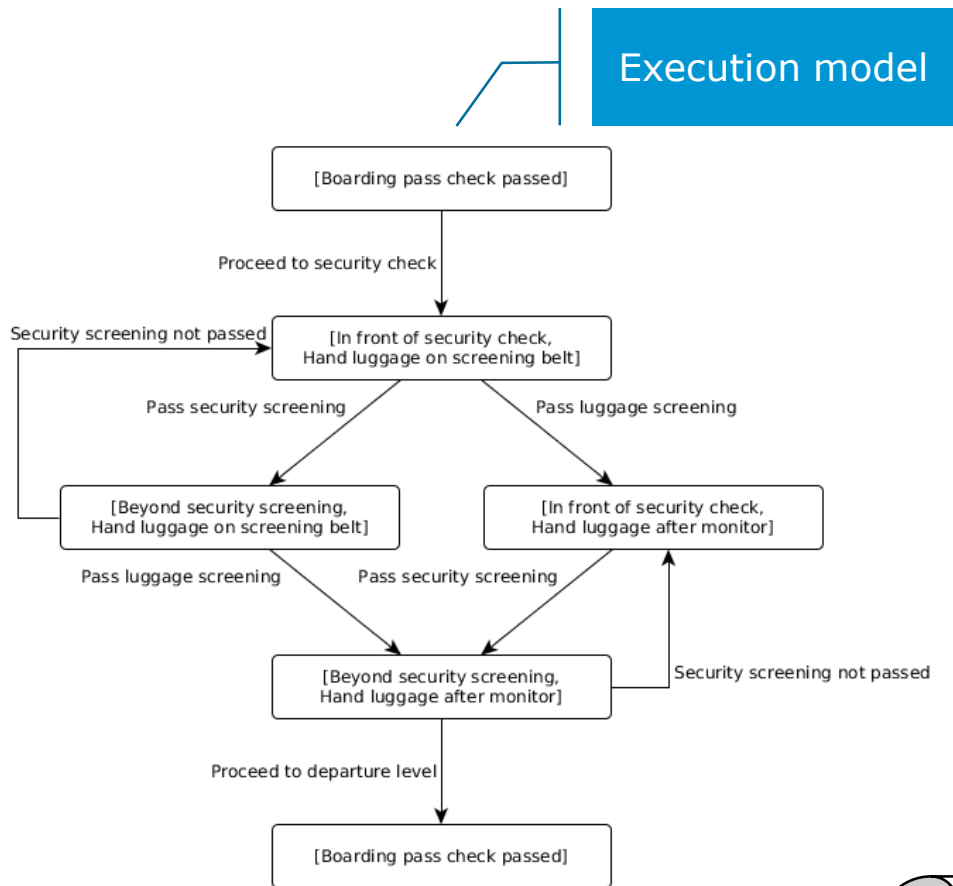


Slide courtesy of Prof. Jan Mendling

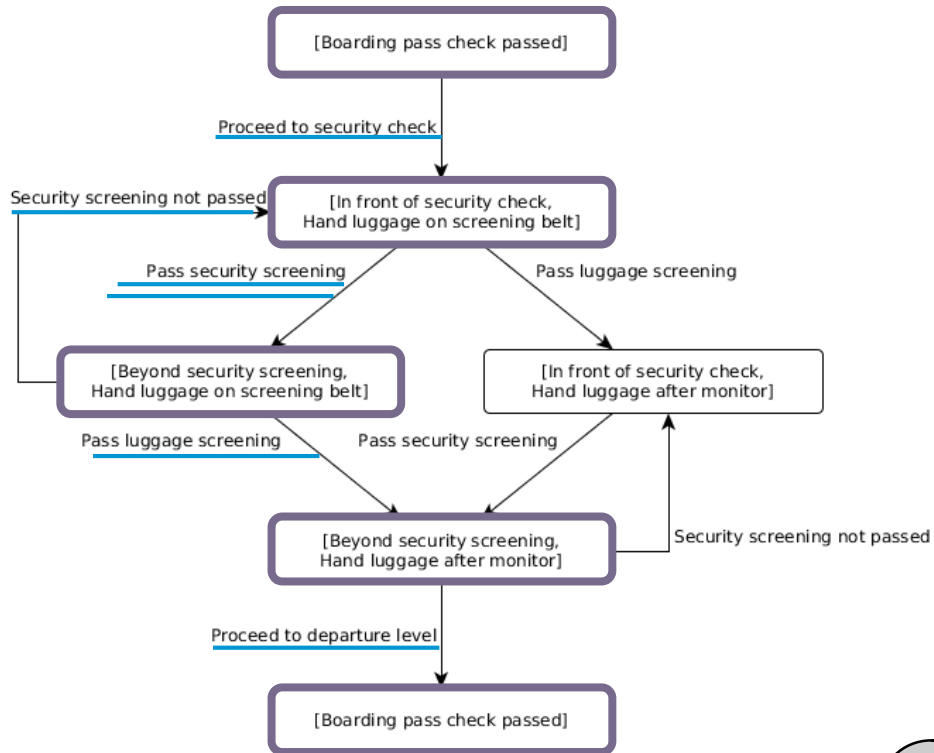


The event log

Execution model



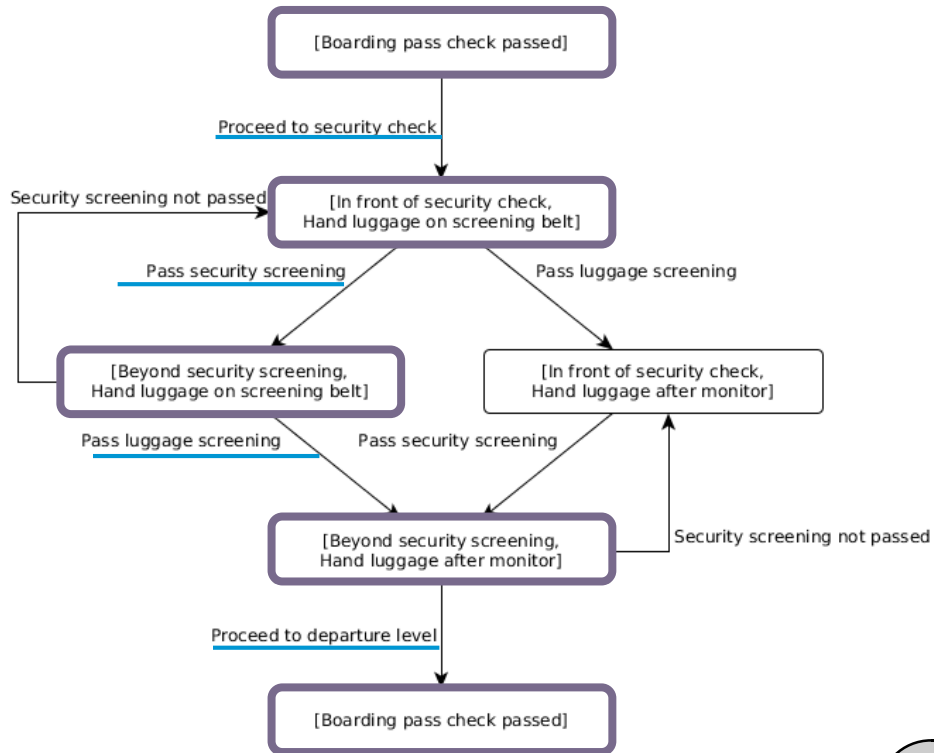
The event log



Case #1

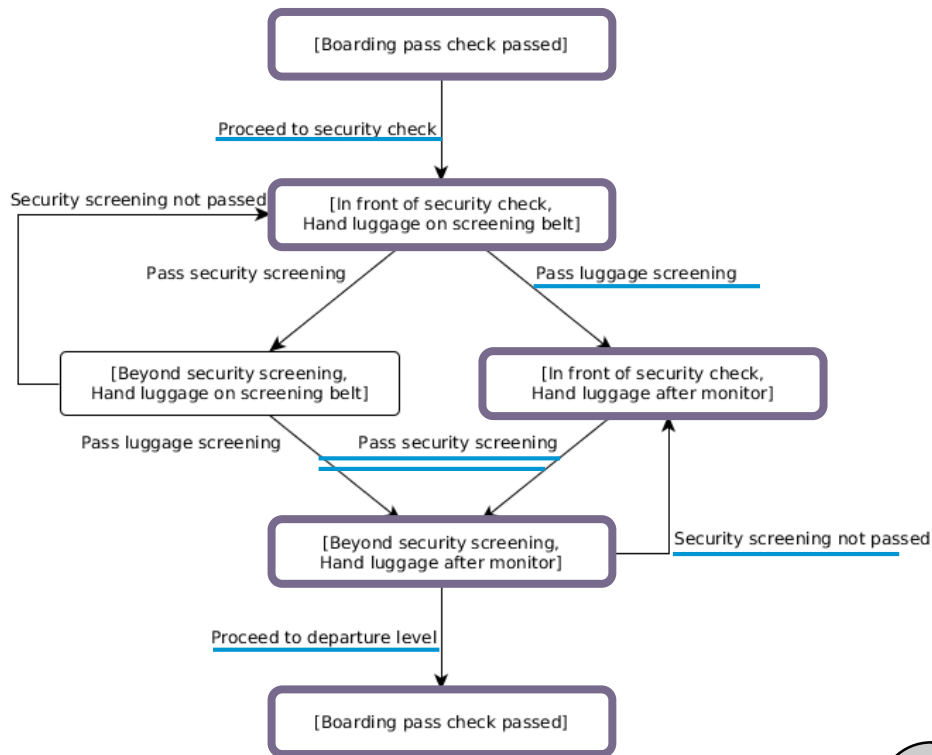
- Proceed to security check
- Pass security screening
- Security screening not passed
- Pass security screening
- Pass luggage screening
- Proceed to departure level

The event log



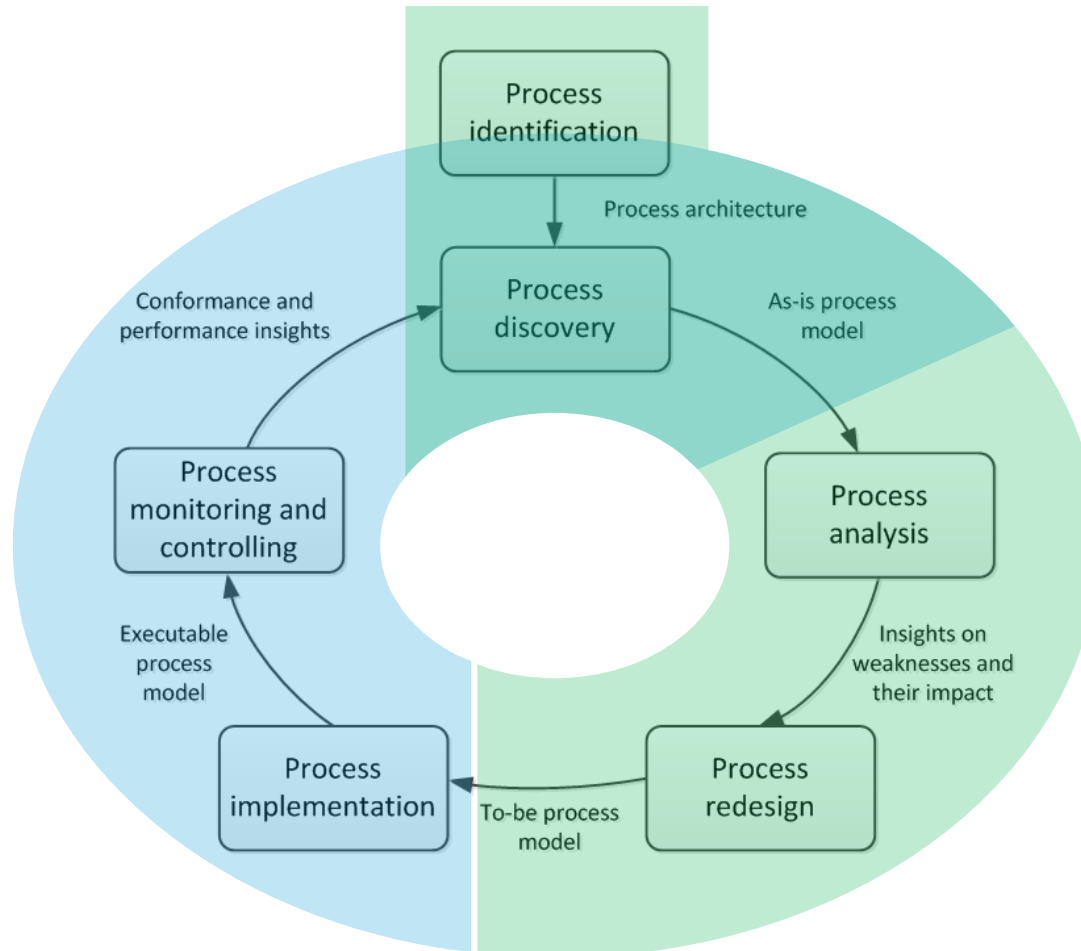
- Case #1
 - Proceed to security check
 - Pass security screening
 - Security screening not passed
 - Pass security screening
 - Pass luggage screening
 - Proceed to departure level
- Case #2
 - Proceed to security check
 - Pass security screening
 - Pass luggage screening
 - Proceed to departure level

The event log



- Case #1
 - Proceed to security check
 - Pass security screening
 - Security screening not passed
 - Pass security screening
 - Pass luggage screening
 - Proceed to departure level
- Case #2
 - Proceed to security check
 - Pass security screening
 - Pass luggage screening
 - Proceed to departure level
- Case #3
 - Proceed to security check
 - Pass luggage screening
 - Pass security screening
 - Security screening not passed
 - Pass security screening
 - Proceed to departure level

BPM lifecycle and positioning



The event log



- Case #1
 - Proceed to security check
 - Pass security screening
 - Security screening not passed
 - Pass security screening
 - Pass luggage screening
 - Proceed to departure level
- Case #2
 - Proceed to security check
 - Pass security screening
 - Pass luggage screening
 - Proceed to departure level
- Case #3
 - Proceed to security check
 - Pass luggage screening
 - Pass security screening
 - Security screening not passed
 - Pass security screening
 - Proceed to departure level

From data to action

Process mining

WU

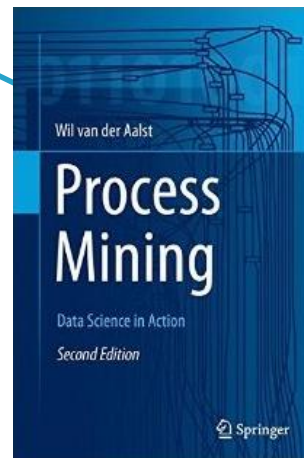
WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

Discovery, conformance, analytics,
enhancement of processes based on data



Process mining at large

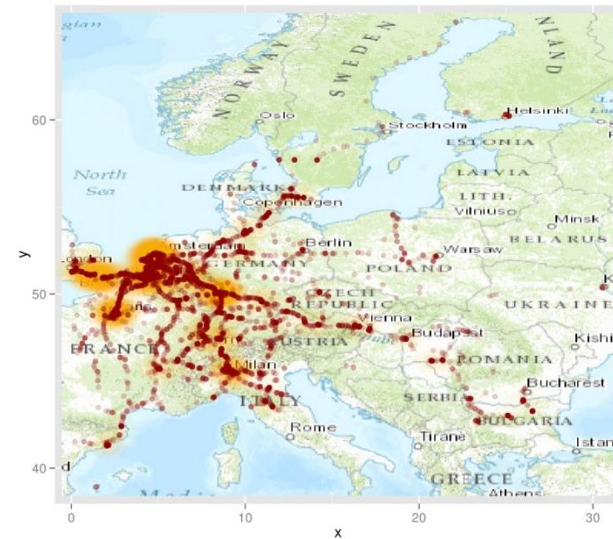
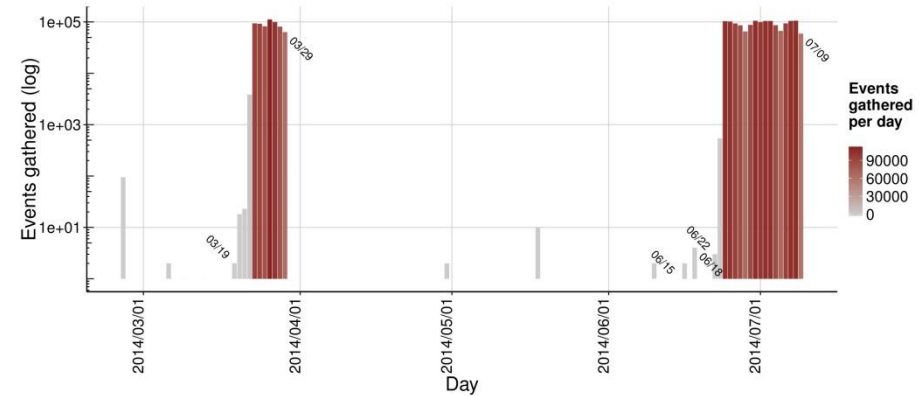
van der Aalst, 2016



Process mining is not (only) a static analysis of data

<http://getservice-project.eu/>

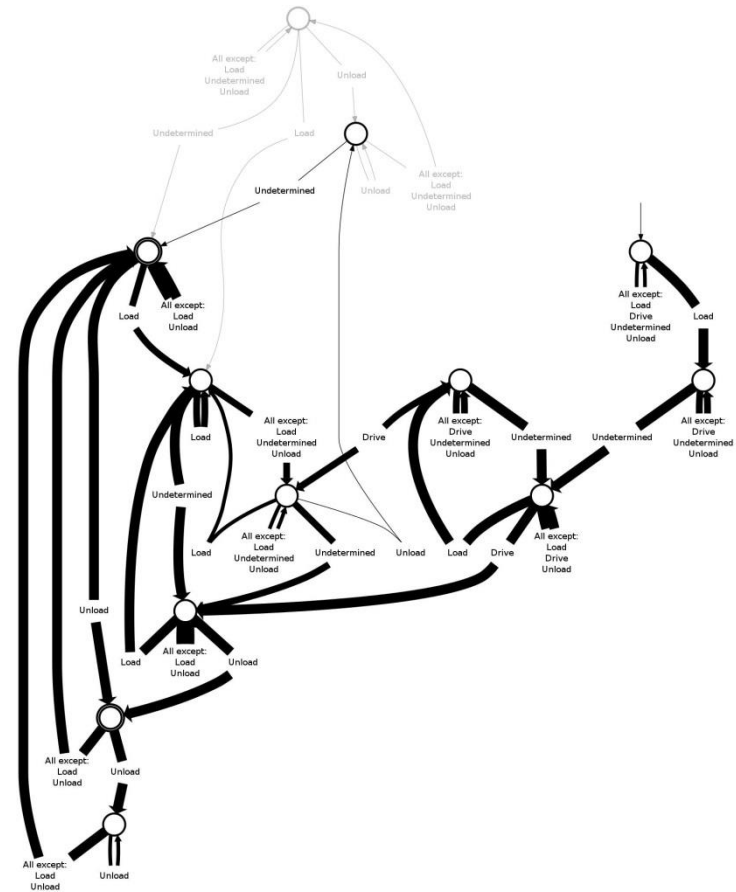
	A	B	C	D	E
1	truck	latitude	longitude	activity	datetime
2	1849	51.9468	4.03339	Start of Drive	01-05-2013T04:55:00Z
3	1849	51.9461	4.03387	Driving times state event	01-05-2013T04:56:00Z
4	1849	51.9461	4.03387	Driving times state event	01-05-2013T04:56:00Z
5	5708	51.9509	4.4427	Driving times driving violation	01-05-2013T04:56:00Z
6	5708	51.9517	4.4429	End of Drive	01-05-2013T04:57:00Z
7	5708	51.9517	4.4429	Start of UN	01-05-2013T04:57:00Z
8	5708	51.9517	4.4429	Driving times state event	01-05-2013T05:00:00Z
9	7262	51.9491	4.44093	Basic record	01-05-2013T05:00:00Z
10	7262	51.9501	4.44347	Navigation destination reached	01-05-2013T05:00:00Z
11	1849	51.9426	4.0355	Basic record	01-05-2013T05:01:00Z
12	7262	51.9514	4.44241	End of Drive	01-05-2013T05:01:00Z
13	7262	51.9514	4.44241	Contact OFF	01-05-2013T05:01:00Z
14	7262	51.9514	4.44241	Start of UN	01-05-2013T05:01:00Z
15	7262	51.9514	4.44241	Contact ON	01-05-2013T05:01:00Z
16	7207	51.8733	4.57282	Basic record	01-05-2013T05:02:00Z
17	7262	51.9514	4.44241	Cancellation of UN	01-05-2013T05:03:00Z
18	7262	51.9514	4.44241	Start of Drive	01-05-2013T05:03:00Z
19	7262	51.9516	4.44281	End of Drive	01-05-2013T05:03:00Z
20	7262	51.9516	4.44281	Start of UN	01-05-2013T05:03:00Z
21	7262	51.9516	4.44281	Contact OFF	01-05-2013T05:03:00Z
22	7262	51.9516	4.44281	Contact ON	01-05-2013T05:03:00Z
23	7262	51.9516	4.44283	Cancellation of UN	01-05-2013T05:04:00Z
24	7262	51.9516	4.44283	Start of Drive	01-05-2013T05:04:00Z
25	5708	51.9517	4.44328	Driving times state event	01-05-2013T05:04:00Z
26	7262	51.9517	4.44293	End of Drive	01-05-2013T05:04:00Z
27	7262	51.9517	4.44293	Start of UN	01-05-2013T05:04:00Z
28	7262	51.9517	4.44293	Contact OFF	01-05-2013T05:04:00Z
29	7262	51.9517	4.44293	Contact ON	01-05-2013T05:04:00Z
30	7262	51.9517	4.44293	Contact OFF	01-05-2013T05:04:00Z
31	1849	51.9277	4.03795	Basic record	01-05-2013T05:06:00Z
32	7262	51.9517	4.44293	Driving times state event	01-05-2013T05:06:00Z
33	7207	51.9269	4.53878	Basic record	01-05-2013T05:07:00Z
34	1849	51.9319	4.10066	Start of speed limit violation	01-05-2013T05:09:00Z
35	5708	51.9517	4.44328	Basic record	01-05-2013T05:09:00Z
36	1179	51.8874	4.56323	Basic record	01-05-2013T05:10:00Z
37	5708	51.9517	4.44328	Driving times state event	01-05-2013T05:10:00Z
38	5708	51.9518	4.44356	Driving times state event	01-05-2013T05:11:00Z
39	1849	51.935	4.15209	End of speed limit violation	01-05-2013T05:12:00Z



Process mining is (mainly) about dynamics

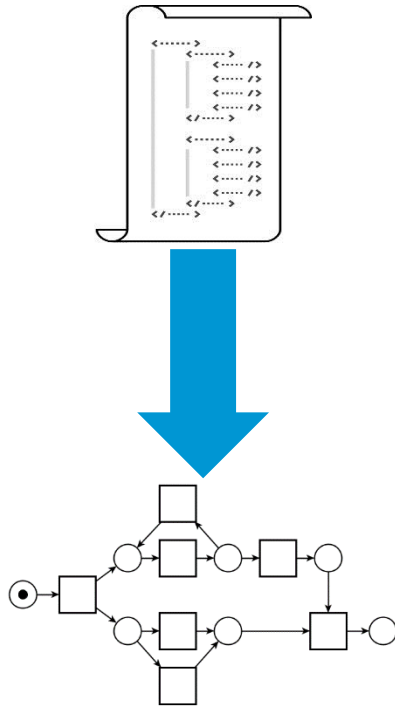
<https://github.com/cdc08x/MINERful>

	A	B	C	D	E
1	truck	latitude	longitude	activity	datetime
2	1849	51.9468	4.03339	Start of Drive	01-05-2013T04:55:00Z
3	1849	51.9461	4.03387	Driving times state event	01-05-2013T04:56:00Z
4	1849	51.9461	4.03387	Driving times state event	01-05-2013T04:56:00Z
5	5708	51.9509	4.4427	Driving times driving violation	01-05-2013T04:56:00Z
6	5708	51.9517	4.4429	End of Drive	01-05-2013T04:57:00Z
7	5708	51.9517	4.4429	Start of UN	01-05-2013T04:57:00Z
8	5708	51.9517	4.4429	Driving times state event	01-05-2013T05:00:00Z
9	7262	51.9491	4.44093	Basic record	01-05-2013T05:00:00Z
10	7262	51.9501	4.44347	Navigation destination reached	01-05-2013T05:00:00Z
11	1849	51.9426	4.0355	Basic record	01-05-2013T05:01:00Z
12	7262	51.9514	4.44241	End of Drive	01-05-2013T05:01:00Z
13	7262	51.9514	4.44241	Contact OFF	01-05-2013T05:01:00Z
14	7262	51.9514	4.44241	Start of UN	01-05-2013T05:01:00Z
15	7262	51.9514	4.44241	Contact ON	01-05-2013T05:01:00Z
16	7207	51.8733	4.57282	Basic record	01-05-2013T05:02:00Z
17	7262	51.9514	4.44241	Cancellation of UN	01-05-2013T05:03:00Z
18	7262	51.9514	4.44241	Start of Drive	01-05-2013T05:03:00Z
19	7262	51.9516	4.44281	End of Drive	01-05-2013T05:03:00Z
20	7262	51.9516	4.44281	Start of UN	01-05-2013T05:03:00Z
21	7262	51.9516	4.44281	Contact OFF	01-05-2013T05:03:00Z
22	7262	51.9516	4.44281	Contact ON	01-05-2013T05:03:00Z
23	7262	51.9516	4.44283	Cancellation of UN	01-05-2013T05:04:00Z
24	7262	51.9516	4.44283	Start of Drive	01-05-2013T05:04:00Z
25	5708	51.9517	4.44328	Driving times state event	01-05-2013T05:04:00Z
26	7262	51.9517	4.44293	End of Drive	01-05-2013T05:04:00Z
27	7262	51.9517	4.44293	Start of UN	01-05-2013T05:04:00Z
28	7262	51.9517	4.44293	Contact OFF	01-05-2013T05:04:00Z
29	7262	51.9517	4.44293	Contact ON	01-05-2013T05:04:00Z
30	7262	51.9517	4.44293	Contact OFF	01-05-2013T05:04:00Z
31	1849	51.9277	4.03795	Basic record	01-05-2013T05:06:00Z
32	7262	51.9517	4.44293	Driving times state event	01-05-2013T05:06:00Z
33	7207	51.9269	4.53878	Basic record	01-05-2013T05:07:00Z
34	1849	51.9319	4.10066	Start of speed limit violation	01-05-2013T05:09:00Z
35	5708	51.9517	4.44328	Basic record	01-05-2013T05:09:00Z
36	1179	51.8874	4.56323	Basic record	01-05-2013T05:10:00Z
37	5708	51.9517	4.44328	Driving times state event	01-05-2013T05:10:00Z
38	5708	51.9518	4.44356	Driving times state event	01-05-2013T05:11:00Z
39	1849	51.935	4.15209	End of speed limit violation	01-05-2013T05:12:00Z



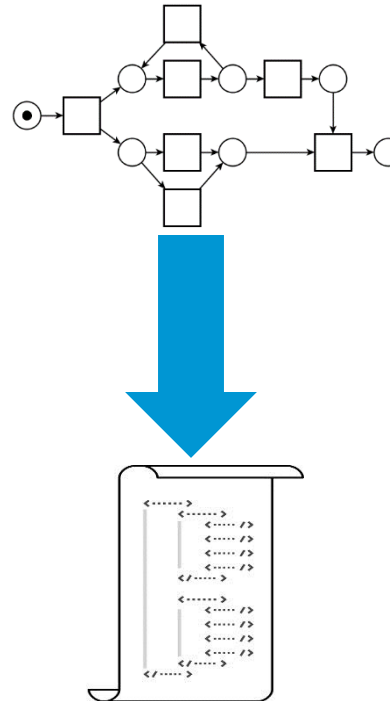
Play in, play out, replay

Play in



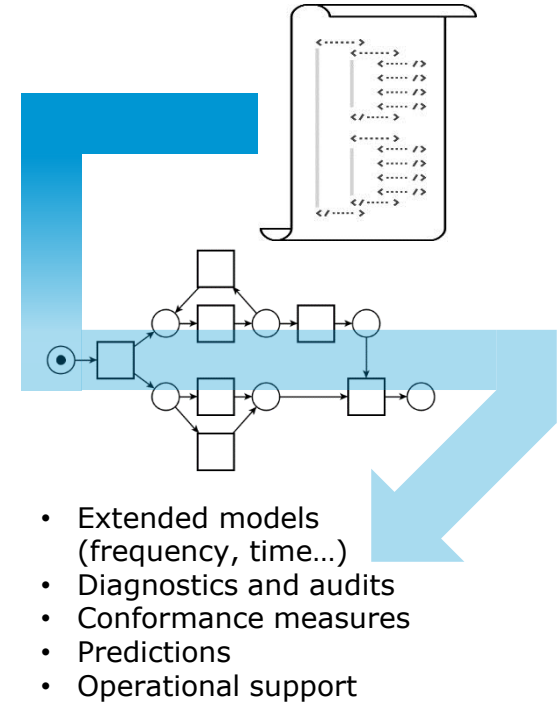
Discovery

Play out



Execution / Simulation

Replay

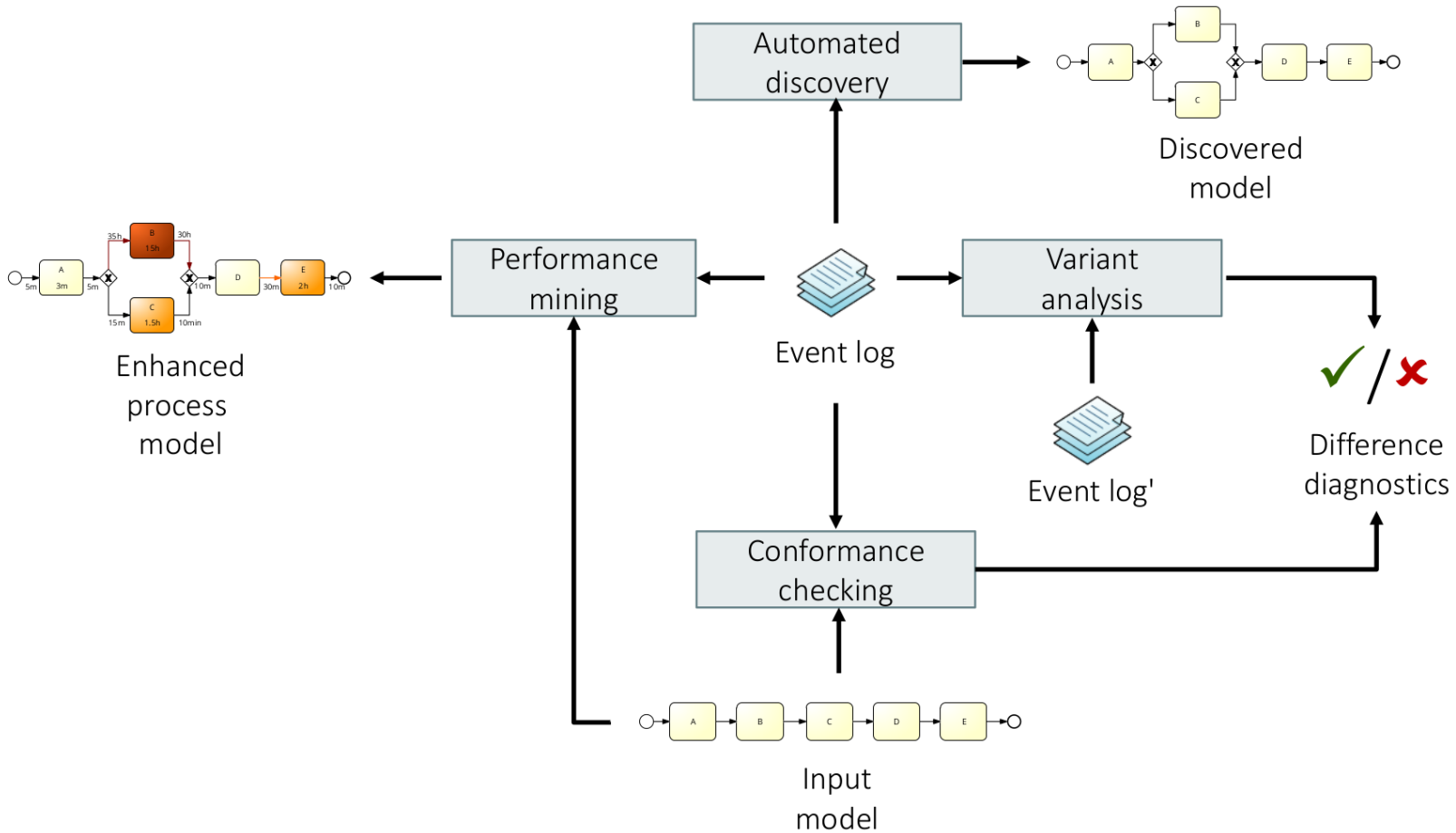


- Extended models (frequency, time...)
- Diagnostics and audits
- Conformance measures
- Predictions
- Operational support

Analytics

Process mining techniques

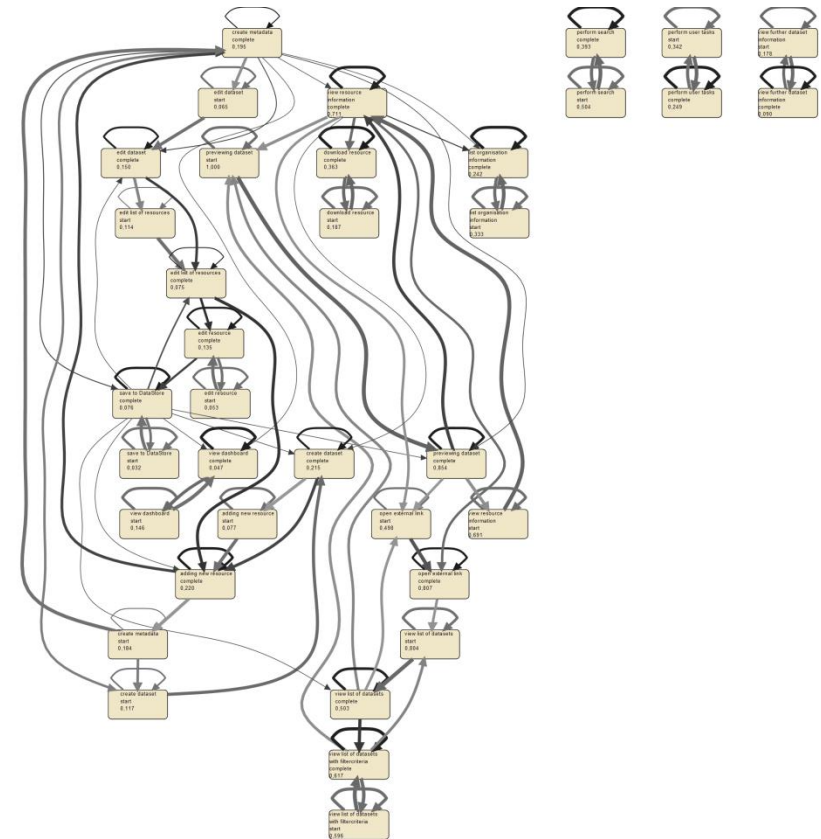
Dumas et al., 2018



Automated process discovery

<http://www.promtools.org/>

```
<log xes.version="1.0" xes.features="nested-attributes" openxes.version="1.0RC7">  
<string key="concept:name" value="Klassifikation_v4.csv"/>  
<trace>  
<string key="visitIP" value="178.114.0.0"/>  
<string key="concept:name" value="77095"/>  
<string key="visitorId" value="e8a0cfadda07194d"/>  
<event>  
<string key="concept:instance" value="1"/>  
<string key="lifecycle:transition" value="start"/>  
<string key="concept:name" value="outlink"/>  
<date key="time:timestamp" value="2014-07-08T11:09:43.000+02:00"/>  
</event>  
<event>  
<string key="siteSearchKeyword"/>  
<string key="country" value="Österreich"/>  
<string key="pageTitle"/>  
<string key="visitorType" value="returning"/>  
<string key="visitDurationPretty" value="2 Minuten 45s"/>  
<string key="customVariables"/>  
<string key="concept:name" value="outlink"/>  
<string key="visitServerHour" value="9"/>  
<string key="providerName" value="unbekannt"/>  
<string key="longitude" value="13.333000"/>  
<string key="searches" value="0"/>  
<string key="referrerTypeName" value="Direkte Zugriffe"/>  
<string key="visitorTypeIcon" value="plugins/Morpheus/images/returningVisitor.gif"/>  
<string key="pageId" value="184467"/>  
<string key="serverTimePrettyFirstAction" value="11:12:17"/>  
<string key="providerUrl"/>  
<date key="time:timestamp" value="2014-07-08T11:12:22.000+02:00"/>  
<string key="serverDate" value="08.07.2014"/>  
<string key="generationTime"/>  
<string key="lastActionTimestamp" value="1404810901"/>  
<string key="visitEcommerceStatusIcon"/>  
<string key="referrerSearchEngineUrl"/>  
<string key="visitLocalTime" value="11:12:04"/>  
<string key="region"/>  
<string key="actions" value="3"/>  
<string key="lastActionDateTime" value="08.07.2014 09:15"/>  
<string key="continent" value="Europa"/>  
<string key="visitDuration" value="165"/>  
<int key="daysSinceLastVisit" value="1"/>  
<string key="pageIdAction" value="13130"/>  
<string key="city"/>  
<string key="latitude" value="47.333000"/>  
<string key="icon" value="plugins/Morpheus/images/link.gif"/>  
<string key="serverDatePrettyFirstAction" value="Dienstag, 8. Juli 2014"/>  
<string key="referrerKeyword"/>  
<int key="daysSinceFirstVisit" value="1"/>  
<string key="serverTimePretty" value="08.7.2014 11:12:22/11:15:01"/>  
<string key="regionCode"/>  
<string key="visitCount" value="2"/>  
<string key="referrerType" value="direct"/>  
<string key="countryFlag" value="plugins/BeerCountry/images/flag/at.png"/>
```



Conformance checking

<http://www.promtools.org/>

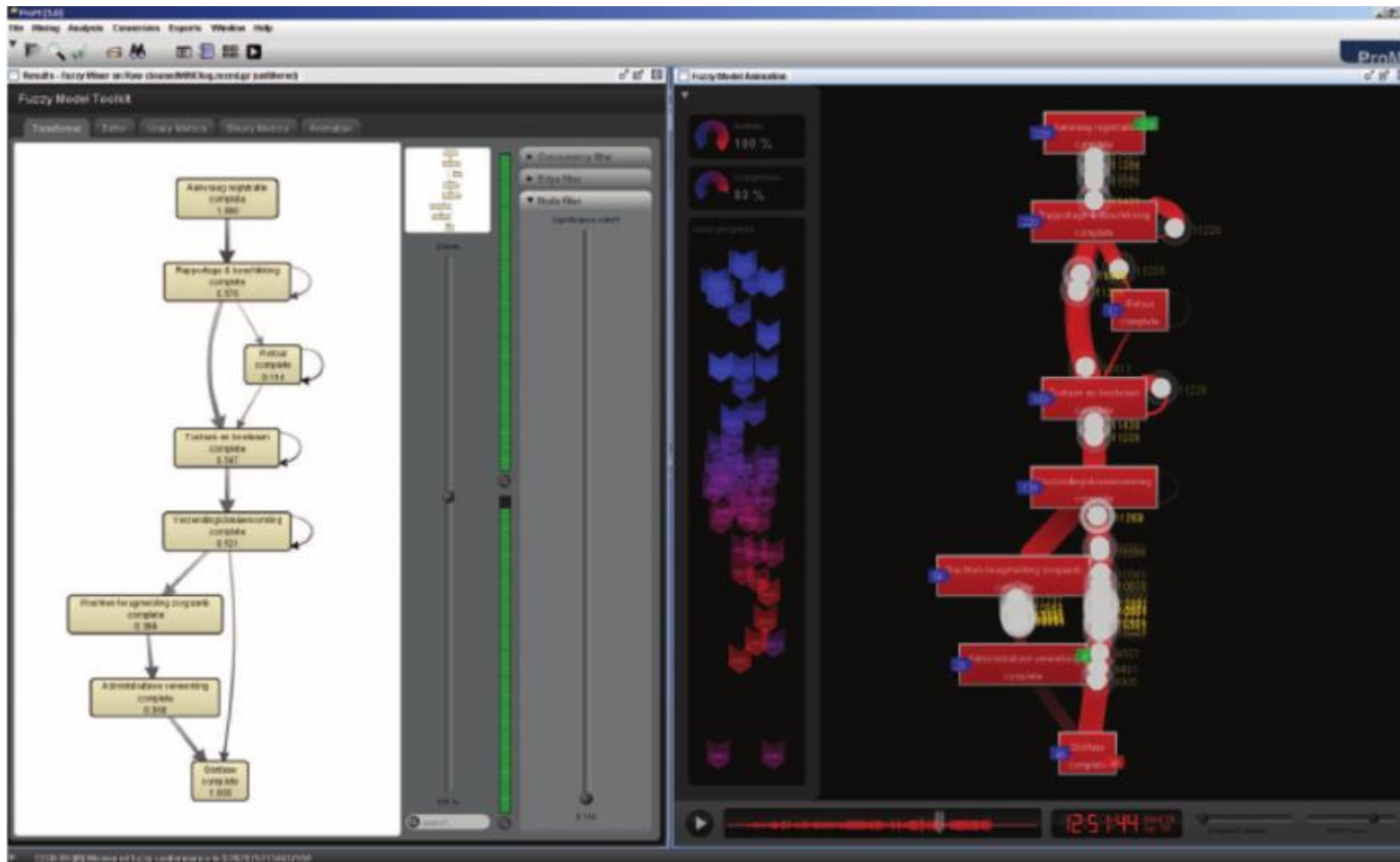
The screenshot displays the ProM 6 interface for Log-Model Alignments. The main window shows a list of traces with their respective fitness and precision values. A legend on the right explains the color coding for alignment steps: green for perfect alignment, purple for missing events, and yellow for wrong events. An 'Inspector' window is open, showing a table of alignment statistics for a selected element.

Property	Value
#Move log-model (total)	168
#Move log-model (in 100% fitting traces)	142
#Traces where move log-model occur	168
#Move model only (in all traces)	40
#Traces where move model only occur	40

Below the main window, the 'Details of the Alignment for Trace 1000181554' are shown, featuring a sequence of colored blocks representing different process steps. A Petri net diagram on the right side of the interface illustrates the underlying process flow with transitions and places.

From conformance checking to process enhancement

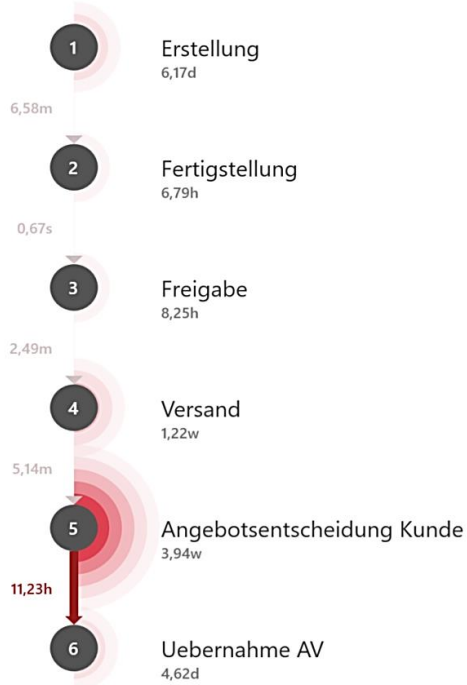
<http://www.promtools.org/>



Process analytics

Slide courtesy of Matej Puchovsky

Variant 1



absolute case coverage

14% 180 / 1265

absolute event coverage

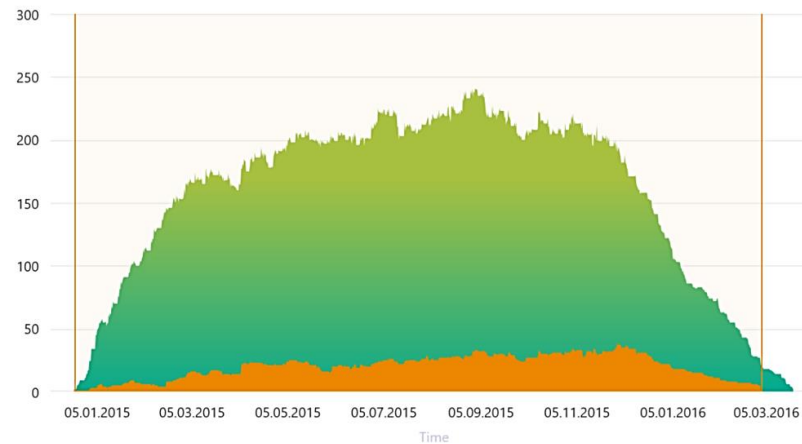
9% 1080 / 12577

filtered case coverage

14% 180 / 1265

filtered event coverage

9% 1080 / 12577



First variant timestamp	Last variant timestamp	Total duration	Total active time	Total waiting time
05.01.2015 11:34:00	17.03.2016 12:56:00	23y 8M 15h 30m	23y 5M 7d 14h 37m	2M 24d 53m

	Minimum	Mean	Maximum
Duration (cases)	51m	1M 17d 29m 10s	9M 13d 7h 17m
Active time (cases)	0	1M 16d 13h 52s 333ms	9M 13d 7h 17m
Utilization (cases)	0,00 %	99,44 %	100,00 %
Waiting time (cases)	0	11h 28m 17s 666ms	2M 24d 48m

Everyday challenges in a process miner life



Data treatment

Starting point for process mining: Event data

Every row is an event (here: an exam attempt)

student name	course name	exam date	mark
Peter Jones	Business Information systems	16-1-2014	8
Sandy Scott	Business Information systems	16-1-2014	5
Bridget White	Business Information systems	16-1-2014	9
John Anderson	Business Information systems	16-1-2014	8
Sandy Scott	BPM Systems	17-1-2014	7
Bridget White	BPM Systems	17-1-2014	8
Sandy Scott	Process Mining	20-1-2014	5
Bridget White	Process Mining	20-1-2014	9
John Anderson	Process Mining	20-1-2014	8
...

case id

activity name

timestamp

other data

Another event log: order handling

order n.	activity	timestamp	user	product	qty.
9901	register order	22-1-2014@09.15	Sara Jones	iPhone5S	1
9902	register order	22-1-2014@09.18	Sara Jones	iPhone5S	2
9903	register order	22-1-2014@09.27	Sara Jones	iPhone4S	1
9901	check stock	22-1-2014@09.49	Pete Scott	iPhone5S	1
9901	ship order	22-1-2014@10.11	Sue Fox	iPhone5S	1
9903	check stock	22-1-2014@10.34	Pete Scott	iPhone4S	1
9901	handle payment	22-1-2014@10.41	Carol Hope	iPhone5S	1
9902	check stock	22-1-2014@10.57	Pete Scott	iPhone5S	2
9902	cancel order	22-1-2014@11.08	Carol Hope	iPhone5S	2
...



Another event log: patient treatment

patient	activity	timestamp	doctor	age	cost
5781	make X-ray	23-1-2014@10.30	Dr. Jones	45	70.00
5541	blood test	23-1-2014@10.18	Dr. Scott	61	40.00
5833	blood test	23-1-2014@10.27	Dr. Scott	24	40.00
5781	blood test	23-1-2014@10.49	Dr. Scott	45	40.00
5781	CT scan	23-1-2014@11.10	Dr. Fox	45	1200.00
5833	surgery	23-1-2014@12.34	Dr. Scott	24	2300.00
5781	handle payment	23-1-2014@12.41	Carol Hope	45	0.00
5541	put in plaster cast	23-1-2014@13.57	Dr. Jones	61	440.00
5541	put in plaster cast	23-1-2014@13.08	Dr. Jones	61	440.00
...



Event logs v. data at hand

LoadingStart	LoadingEnd	Reference (Shipment)	Remarks (Shipment)	Department (Shipment/Booking)	Department (Shipment/Booking)	BookingNr (Shipment/Booking)	ShipmentNr (Shipment)	ShipmentRouteNr
25/03/2014 12:00	25/03/2014 13:00	20140321500062				1000	100001065568	100001065568
26/03/2014 10:00	26/03/2014 11:00	20140324500037				1000	100001065576	100001065576
24/03/2014 15:00	24/03/2014 16:00	AMST048022				1000	100001079143	100001079143
31/03/2014 07:30	31/03/2014 08:30		1135917			1000	100001081155	100001081155
31/03/2014 08:30	31/03/2014 09:30		1135918			1000	100001081157	100001081157
25/03/2014 10:00	25/03/2014 11:00	2 drivers				1000	100001081473	100001081473
24/03/2014 10:18	24/03/2014 13:54	2966/DBFL				1000	100001082425	100001082425
24/03/2014 22:00	24/03/2014 23:00	15 Wines-Garone				1000	100001082327	100001082327
26/03/2014 21:00	26/03/2014 22:00	Italy				1000	100001082343	100001082343
26/03/2014 21:00	26/03/2014 22:00	Italy				1000	100001082342	100001082342
29/03/2014 09:00	29/03/2014 10:00	Italy				1000	100001082346	100001082346
29/03/2014 09:00	29/03/2014 10:00	Italy				1000	100001082348	100001082348
29/03/2014 09:00	29/03/2014 10:00	Italy				1000	100001082349	100001082349
25/03/2014 10:00	25/03/2014 11:00	LX6406				1000	100001082404	100001082405
26/03/2014 03:00	26/03/2014 04:00	LX6407 CHECK				1000	100001082404	100001082406
29/03/2014 18:00	29/03/2014 19:00	Drison's/Sotr				1000	100001082422	100001082422
29/03/2014 18:00	29/03/2014 19:00	Drison's/Sotr				1000	100001082423	100001082423
25/03/2014 17:00	25/03/2014 18:00	Gateway FRA				1000	100001082424	100001082424
24/03/2014 17:00	28/03/2014 18:00	Gateway FRA				1000	100001082425	100001082425
24/03/2014 21:46	24/03/2014 22:46	240314 KLB434				1000	100001082615	100001082615
25/03/2014 07:00	25/03/2014 08:00	84-1403-023531				1000	100001082710	100001082710
25/03/2014 08:00	25/03/2014 09:00	84-1403-023532				1000	100001082716	100001082716
22/03/2014 13:16	22/03/2014 15:30	210314 CK1113/3A				1000	100001082731	100001082732
24/03/2014 10:59	24/03/2014 12:44		1135914			1000	100001082869	100001082869
24/03/2014 22:29	24/03/2014 23:29	240314 KLB818				1000	100001082883	100001082883
24/03/2014 07:34	24/03/2014 08:00	240314 KLB424				1000	100001082884	100001082884
25/03/2014 08:15	25/03/2014 09:15	LP 101247				1000	100001082942	100001082943
26/03/2014 04:06	26/03/2014 05:06	LP 101247				1000	100001082942	100001082945
24/03/2014 09:16	24/03/2014 12:29		33793			1000	100001082954	100001082954
25/03/2014 10:00	25/03/2014 11:00		33817			1000	100001082956	100001082956
25/03/2014 11:00	25/03/2014 12:00		33819			1000	100001082957	100001082957
24/03/2014 13:05	24/03/2014 14:05	D0000103322				1000	100001083077	100001083077
24/03/2014 14:59	24/03/2014 15:59	D0000103321				1000	100001083072	100001083071
24/03/2014 08:00	24/03/2014 11:45		1133477			1000	100001083093	100001083093
24/03/2014 09:52	24/03/2014 10:02		1135867			1000	100001083144	100001083147
24/03/2014 10:23	24/03/2014 11:30		1135867			1000	100001083144	100001083148
24/03/2014 19:31	24/03/2014 20:31	240314 KLB552				1000	100001083149	100001083149

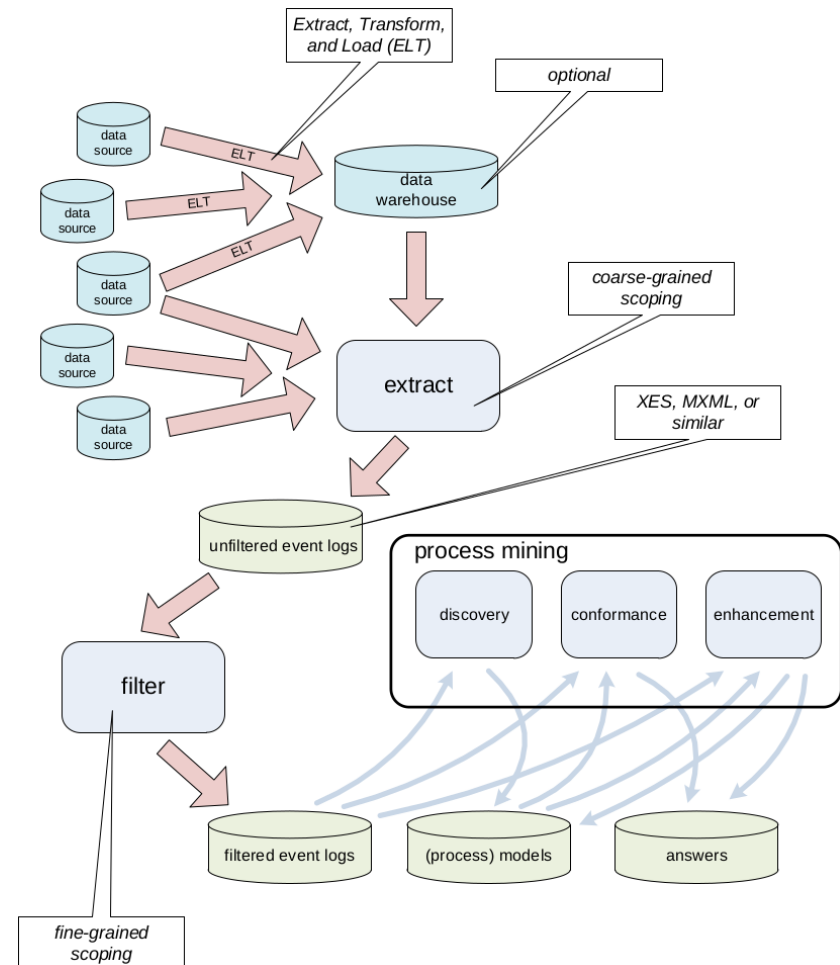
```
https://api.flightstats.com/flex/weather/rest/v1/json/all/LAX?codeType=fs
</weatherUrl>
<airport>
- <airport>
<fs>JFK</fs>
<ata>JFK</ata>
<icao>KJFK</icao>
<faa>JFK</faa>
<name>John F. Kennedy International Airport</name>
<street1>JFK Airport</street1>
<city>New York</city>
<cityCode>NYC</cityCode>
<stateCode>NY</stateCode>
<postalCode>11430</postalCode>
<countryCode>US</countryCode>
<countryName>United States</countryName>
<regionName>North America</regionName>
<timeZoneRegionName>America/New_York</timeZoneRegionName>
<weatherZone>NYZ178</weatherZone>
<localTime>2013-08-09T04:00:32.489</localTime>
<utcOffsetHours>-4.0</utcOffsetHours>
<latitude>40.642335</latitude>
<longitude>-73.78817</longitude>
<elevationFeet>13</elevationFeet>
<classification>I</classification>
<active>true</active>
- <delayIndexUrl>
https://api.flightstats.com/flex/delayindex/rest/v1/json/airports/JFK?codeType=fs
</delayIndexUrl>
<weatherUrl>
https://api.flightstats.com/flex/weather/rest/v1/json/all/JFK?codeType=fs
</weatherUrl>
<airports>
</airports>
<appendix>
- <flightTracks>
- <flightTrack>
<flightId>305231585</flightId>
<carrierCode>AA</carrierCode>
<carrierFsCode>
<flightNumber>10</flightNumber>
<callsign>AAL10</callsign>
<departureAirportFsCode>LAX</departureAirportFsCode>
<arrivalAirportFsCode>JFK</arrivalAirportFsCode>
- <departureDate>
<dateLocal>2013-08-07T21:30:00.000</dateLocal>
<dateUtc>2013-08-08T04:30:00.000Z</dateUtc>
</departureDate>
<equipment>B762</equipment>
<heading>229.5397527577076</heading>
<heading>228.69263064295163</heading>
- <positions>
- <position>
<lon>-73.7300033569336</lon>
<lat>40.68000030517578</lat>
<speedMph>137</speedMph>
<altitudeFeet>800</altitudeFeet>
<source>ASDI</source>
<date>2013-08-08T09:14:56.000Z</date>
</position>
<position>
<lon>-73.69999694824219</lon>
<lat>40.70000076293945</lat>
<speedMph>157</speedMph>
<altitudeFeet>1500</altitudeFeet>
<source>ASDI</source>
<date>2013-08-08T09:13:56.000Z</date>
</position>
<position>
<lon>-73.69999694824219</lon>
<lat>40.70000076293945</lat>
<speedMph>163</speedMph>
<altitudeFeet>2000</altitudeFeet>
<source>ASDI</source>
<date>2013-08-08T09:12:55.000Z</date>
</position>
</positions>
```

The ETL procedure (preventing the GIGO effect)

van der Aalst, 2016



- Extract, Transform, and Load (ETL)
 1. Extract data from outside sources
 2. Transform it to fit operational needs
 3. Load it into target systems



The ETL procedure (preventing the GIGO effect)

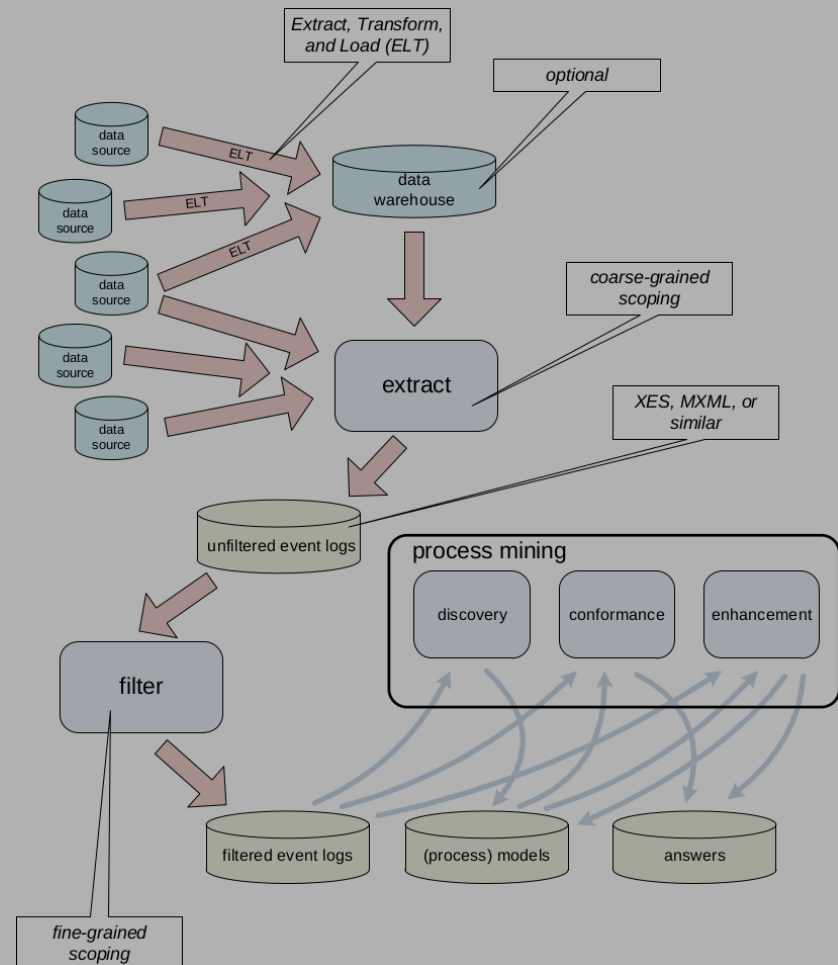
van der Aalst, 2016



- Extract, Transform, and Load (ETL)
 1. Extract data from outside sources
 2. Transform it to fit operational needs
 3. Load it into target systems

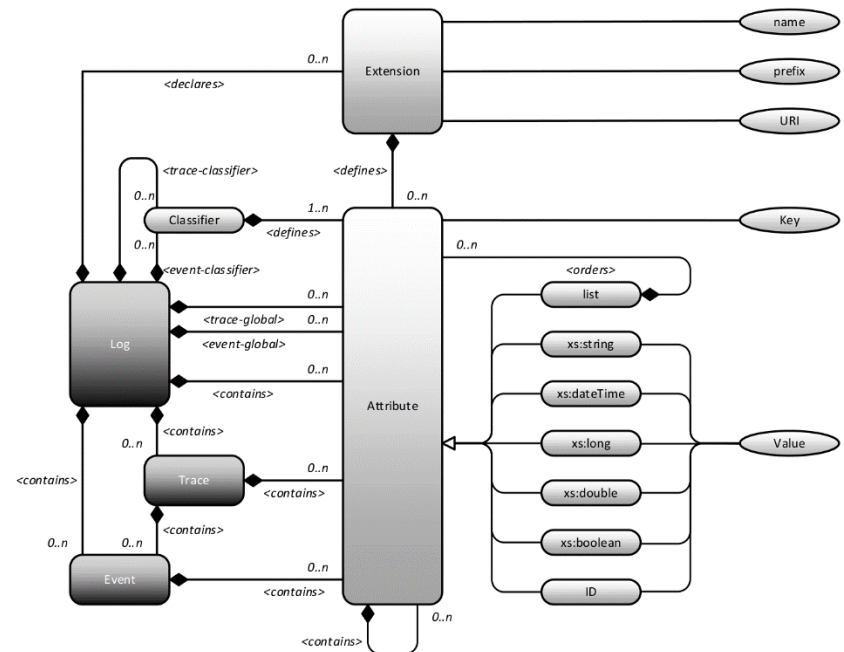


Timestamps



IEEE Standard 1849-2016 for eXtensible Event Stream (XES)

- <http://www.xes-standard.org/>
- Adopted by the IEEE Task Force on Process Mining



Everyday challenges in a process miner life



Assessment of the outcome quality

An example event log

Ext. name	Short name
Receive application form	a
Check application form completeness	b
Return application form	c
Receive amended form	d
Register application form	e
Check availability	f
Verify request	g
Decide on application	h
Send notification of rejection	j
Send notification of acceptance	k
Close application	l

$$L = [$$

$$\langle a, b, e, f, g, h, k, l \rangle^{70},$$

$$\langle a, b, c, d, b, e, f, g, h, k, l \rangle^{50},$$

$$\langle a, b, c, d, b, e, g, f, h, k, l \rangle^{30},$$

$$\langle a, b, c, d, b, c, d, b, e, g, f, h, j, l \rangle^{10},$$

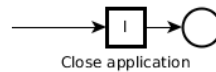
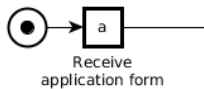
$$\langle a, b, e, g, f, h, k, l \rangle^{60},$$

$$\langle a, b, e, f, g, h, j, l \rangle^{20},$$

$$\langle a, b, e, g, f, h, j, l \rangle^{20}$$

$$]$$

Flower model


$$L = [$$
$$\langle a, b, e, f, g, h, k, l \rangle^{70},$$
$$\langle a, b, c, d, b, e, f, g, h, k, l \rangle^{50},$$
$$\langle a, b, c, d, b, e, g, f, h, k, l \rangle^{30},$$
$$\langle a, b, c, d, b, c, d, b, e, g, f, h, j, l \rangle^{10},$$
$$\langle a, b, e, g, f, h, k, l \rangle^{60},$$
$$\langle a, b, e, f, g, h, j, l \rangle^{20},$$
$$\langle a, b, e, g, f, h, j, l \rangle^{20}$$
$$]$$

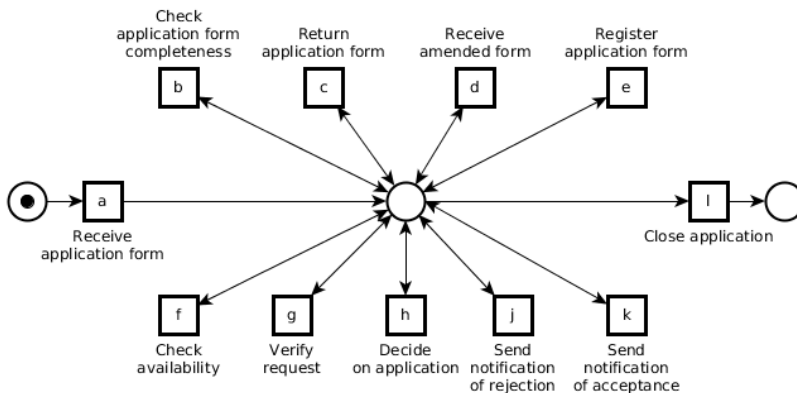
Flower model

Fitness:

All traces can be replayed

Generalisation:

More unforeseen traces can be replayed too



Precision:

Basically any trace starting with *a* and ending with *l* can be replayed

$$L = [$$

$$\langle a, b, e, f, g, h, k, l \rangle^{70},$$

$$\langle a, b, c, d, b, e, f, g, h, k, l \rangle^{50},$$

$$\langle a, b, c, d, b, e, g, f, h, k, l \rangle^{30},$$

$$\langle a, b, c, d, b, c, d, b, e, g, f, h, j, l \rangle^{10},$$

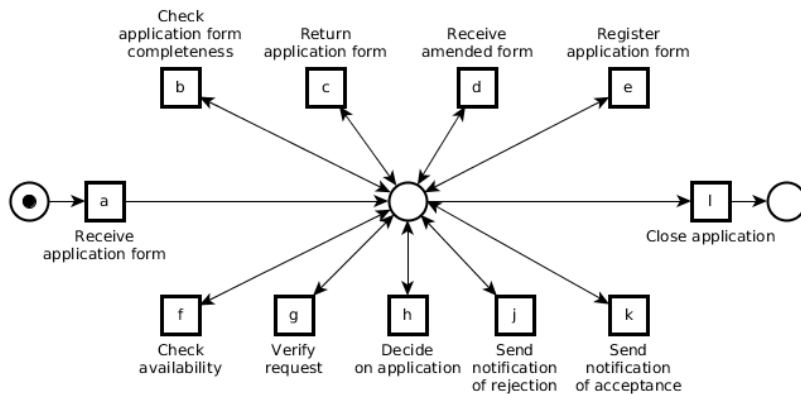
$$\langle a, b, e, g, f, h, k, l \rangle^{60},$$

$$\langle a, b, e, f, g, h, j, l \rangle^{20},$$

$$\langle a, b, e, g, f, h, j, l \rangle^{20}$$

$$]$$

Underfitting



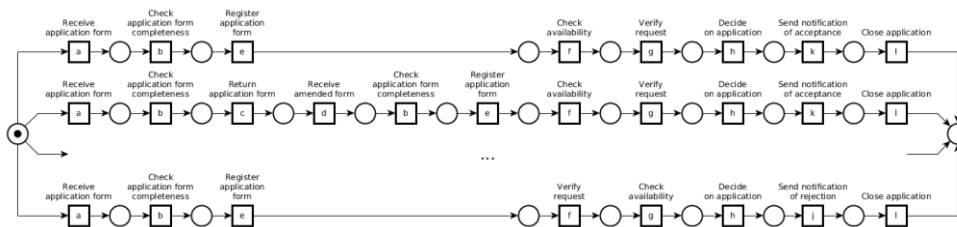
Precision:

Basically any trace starting with *a* and ending with *l* can be replayed



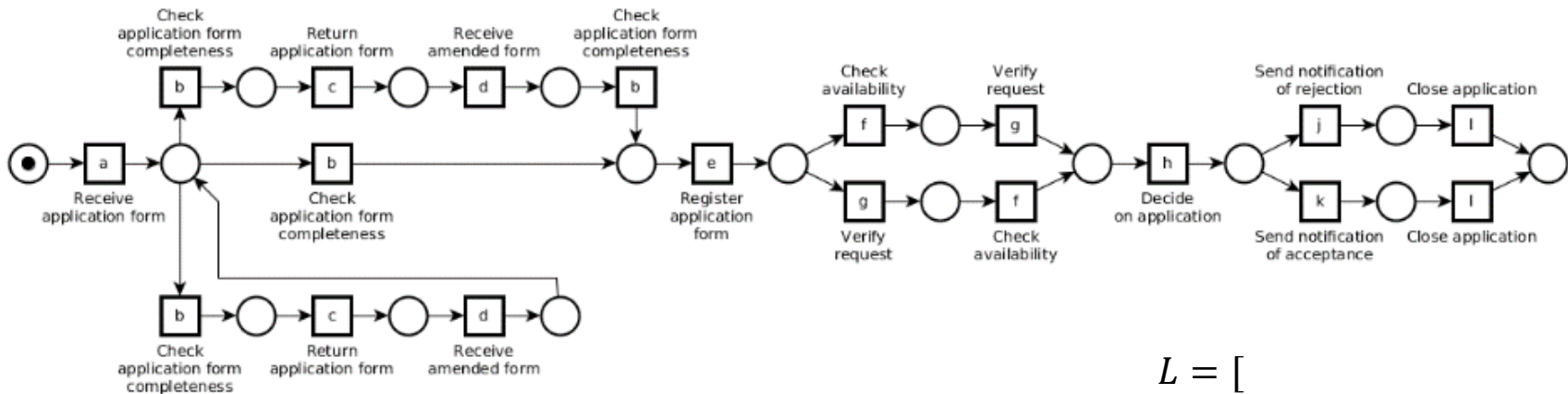
Source: <https://www.aliexpress.com/>

Overfitting



Generalisation: ☒
Nothing outside the log is accepted

Finding the balance



Fitness:
Precision:
Generalisation:

Simplicity:

$$L = [$$

$$\langle a, b, e, f, g, h, k, l \rangle^{70},$$

$$\langle a, b, c, d, b, e, f, g, h, k, l \rangle^{50},$$

$$\langle a, b, c, d, b, e, g, f, h, k, l \rangle^{30},$$

$$\langle a, b, c, d, b, c, d, b, e, g, f, h, j, l \rangle^{10},$$

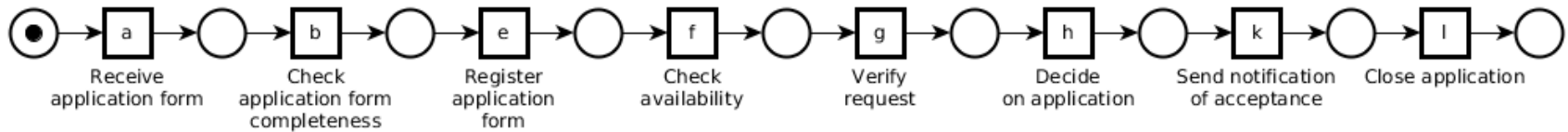
$$\langle a, b, e, g, f, h, k, l \rangle^{60},$$

$$\langle a, b, e, f, g, h, j, l \rangle^{20},$$

$$\langle a, b, e, g, f, h, j, l \rangle^{20}$$

$$]$$

Most frequent trace(s) as model



Simplicity:

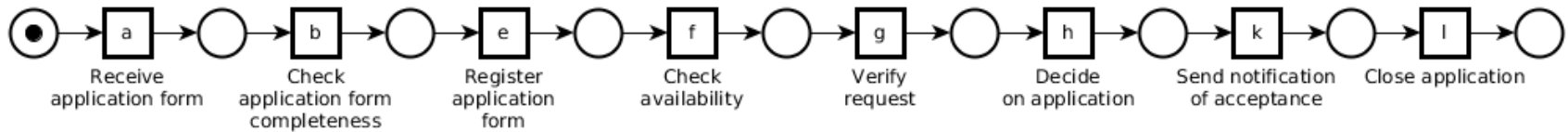
OK, all the other the traces?

$$L = [$$

- $\langle a, b, e, f, g, h, k, l \rangle^{70},$
- $\langle a, b, c, d, b, e, f, g, h, k, l \rangle^{50},$
- $\langle a, b, c, d, b, e, g, f, h, k, l \rangle^{30},$
- $\langle a, b, c, d, b, c, d, b, e, g, f, h, j, l \rangle^{10},$
- $\langle a, b, e, g, f, h, k, l \rangle^{60},$
- $\langle a, b, e, f, g, h, j, l \rangle^{20},$
- $\langle a, b, e, g, f, h, j, l \rangle^{20}$

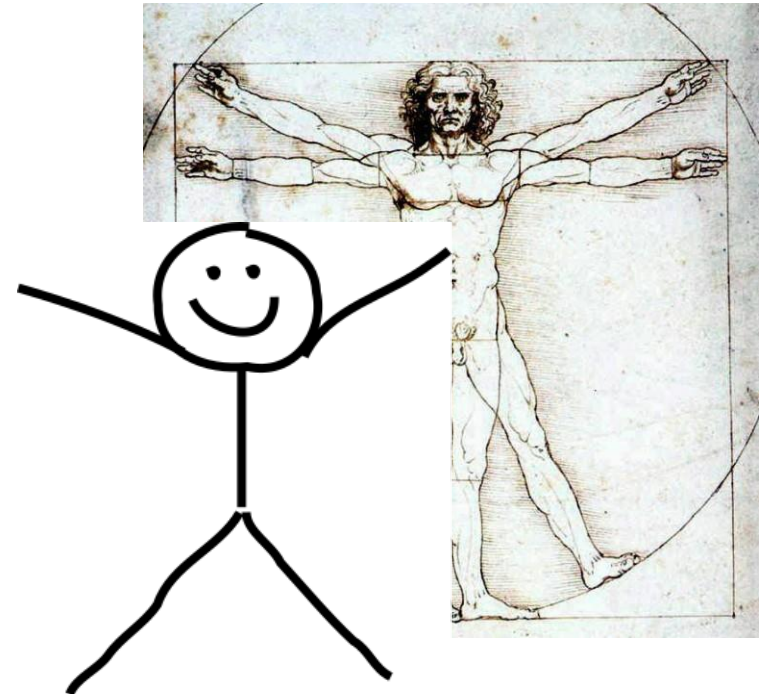
$$]$$

Most frequent trace(s) as model



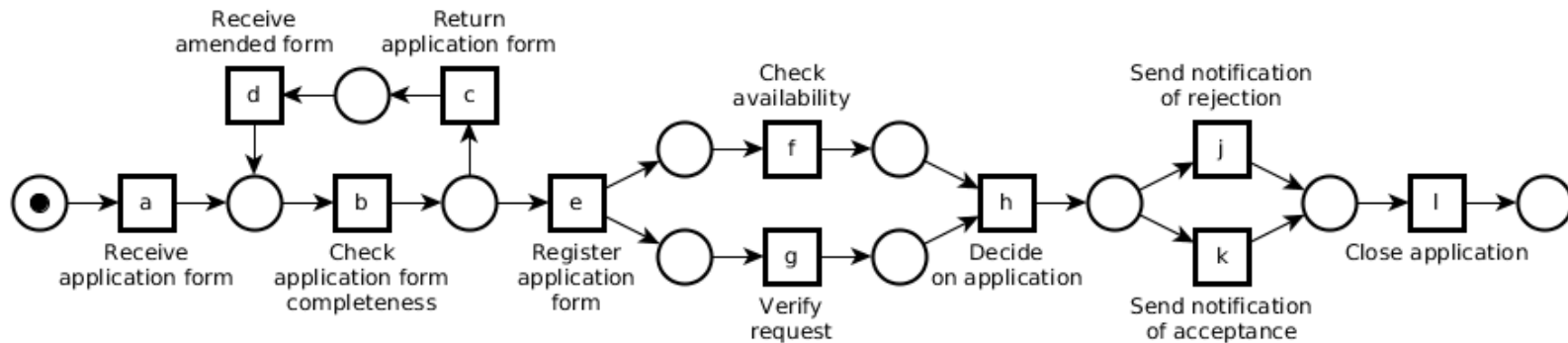
Simplicity:

OK, all the rest?



A reasonable model

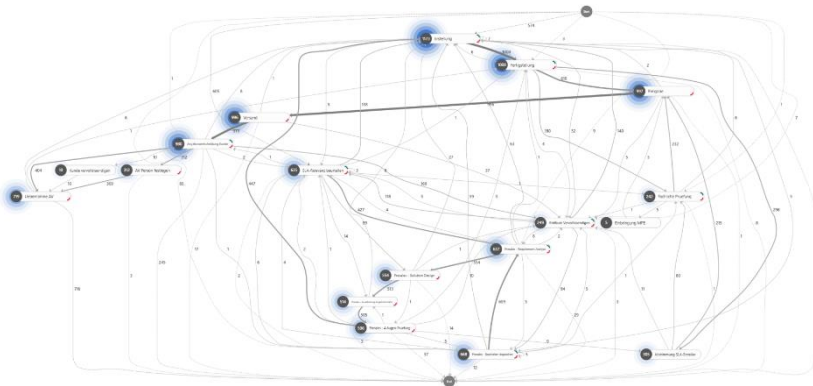
Fitness:
Precision:
Generalisation:
Simplicity:



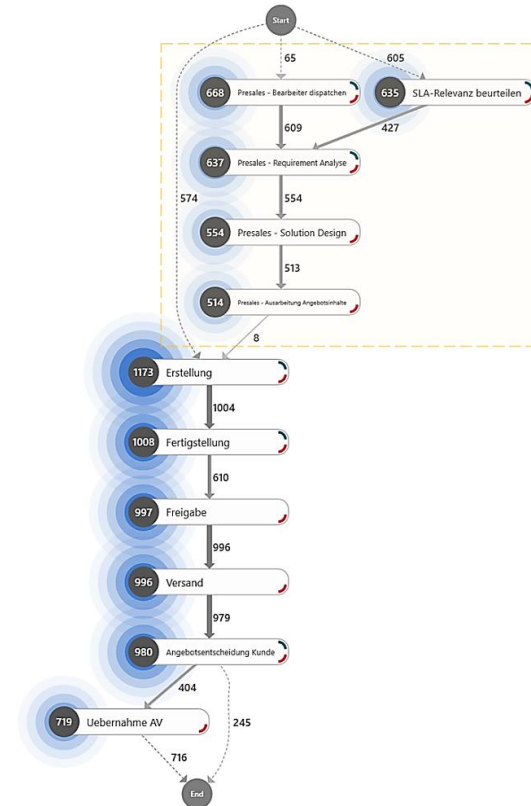
The objective

Slide courtesy of Matej Puchovsky

Before



After



Research

Process mining

WU

WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

Ongoing challenges



Process mining research institutes



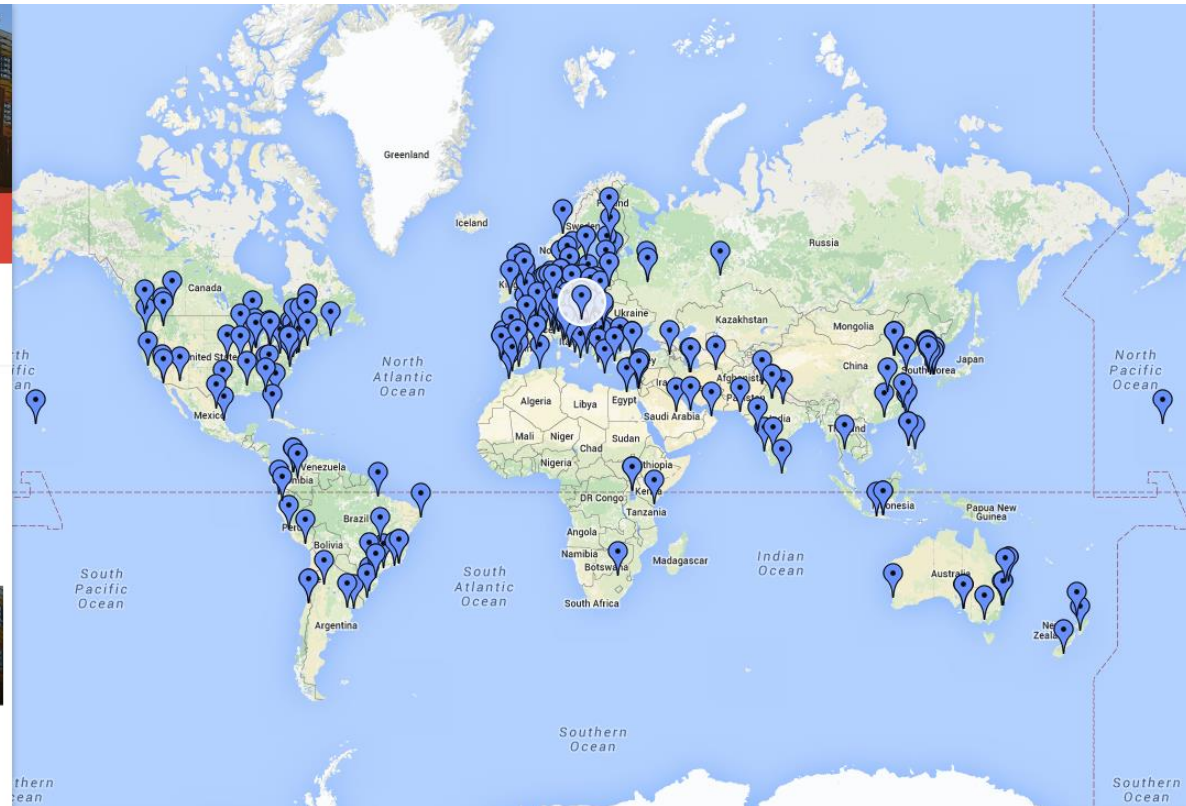
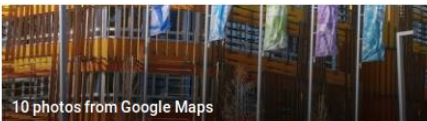
← Vienna University of Economics a...

name
Vienna University of Economics and Business

Details from Google Maps

Welthandelsplatz 1 1, 1020 Wien, Austria
+43 1 313360
www.wu.ac.at
4.4 ★★★★★

[View in Google Maps](#)



Process mining at WU Vienna

<https://wu.ac.at/infobiz>

- Enabling process mining on the blockchain
- Enabling process mining on data lakes
- Enabling process mining on software projects
- Predictive analytics with process mining
- Discovering business rules via process mining
- Informing resource management with process mining
- Explaining development and change with process mining
- ... and more



Data Science × Business Process Mgt =

Process Mining

Dr. Claudio Di Ciccio
<http://diciccio.net/>
claudio.di.ciccio@wu.ac.at

