Data Science × Business Process Mgt =

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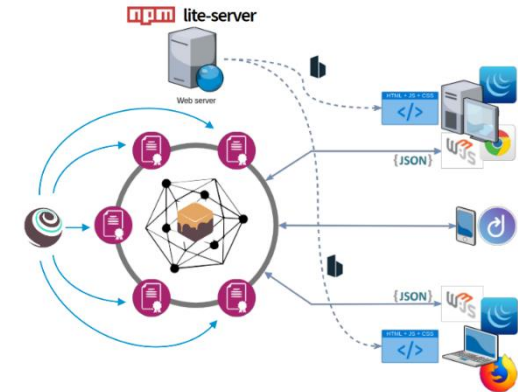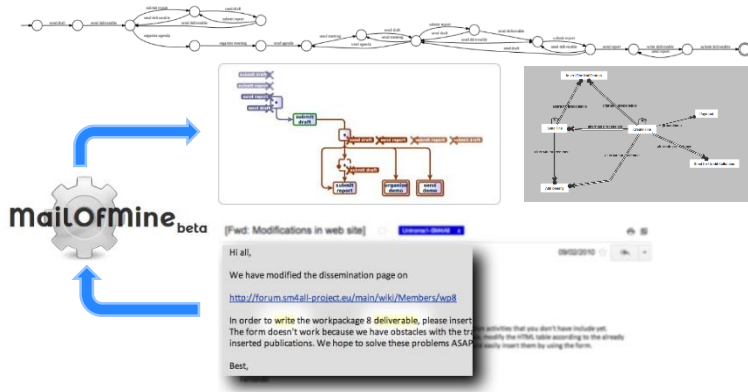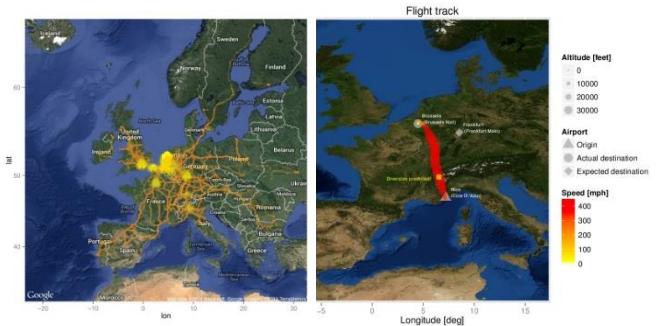
Dr. Claudio Di Ciccio
http://diciccio.net/
claudio.di.ciccio@wu.ac.at

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

RISE_BPM

Claudio Di Ciccio /klˈaːudio di ˈtʃiːttʃo/

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

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

A process mining perspective

Business process mgt.

Organisation, modelling, execution, and tracking of processes

Half empty or half full?

Which is more fundamental: processes or things?

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

Processes are into dynamics

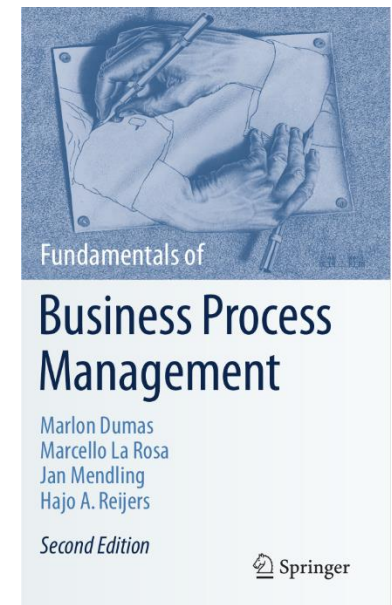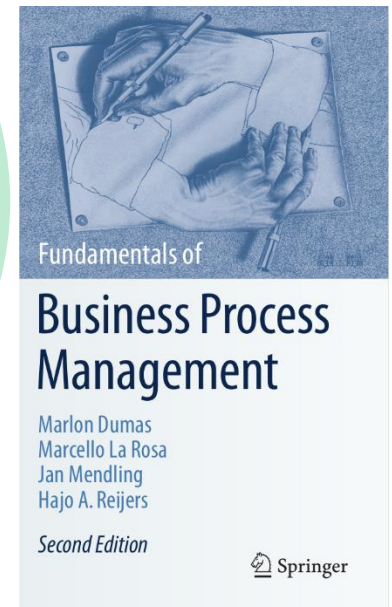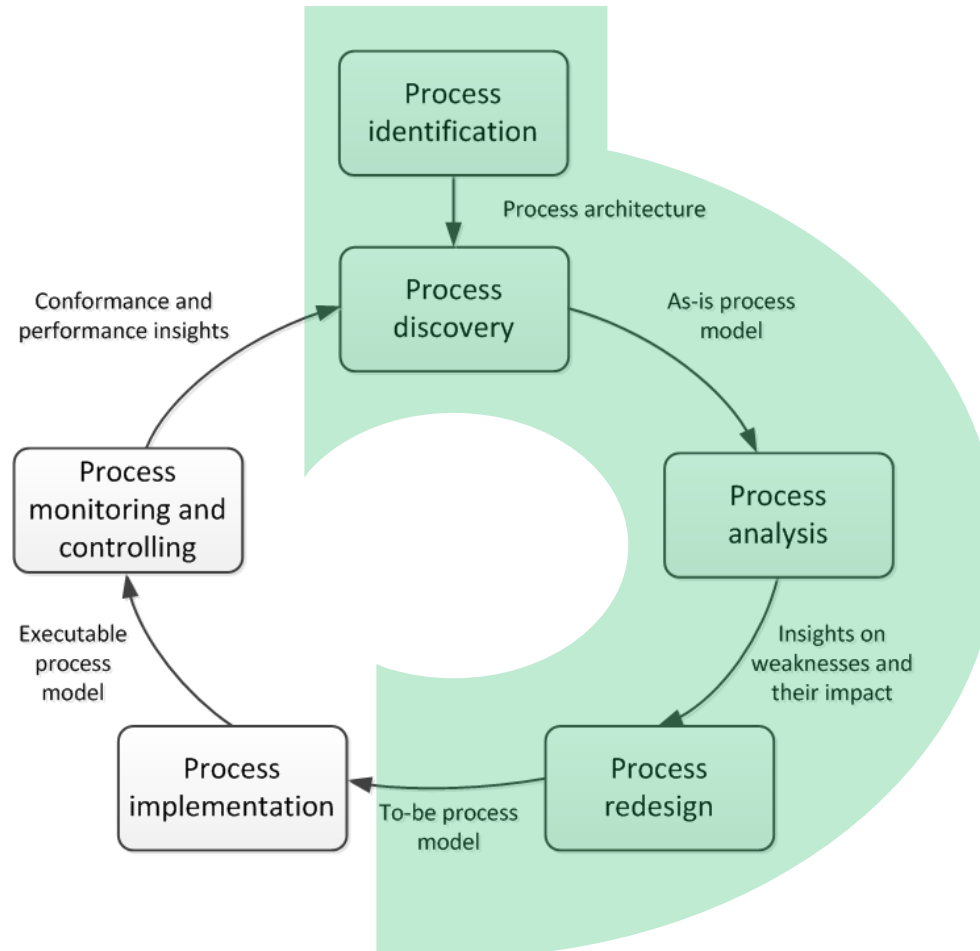# Relevance of business processes

- 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]

Fundamentals of

Business Process Management

Marlon Dumas
Marcello La Rosa
Jan Mendling
Hajo A. Reijers

Second Edition

Springer

BPM lifecycle and positioning

Processes and division of labour

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 straights 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.
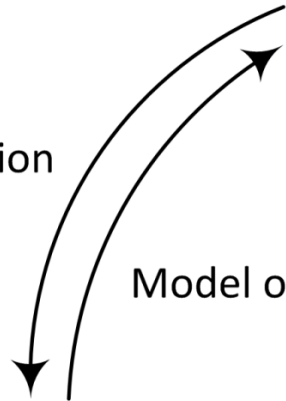
Process models

© Stiftung Deutsches Technikmuseum Berlin
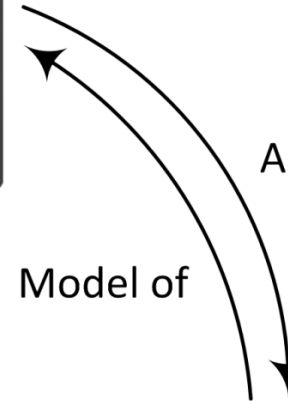
BPM lifecycle and positioning

Workflow

PROCESS MINING

IT Support

IT Support
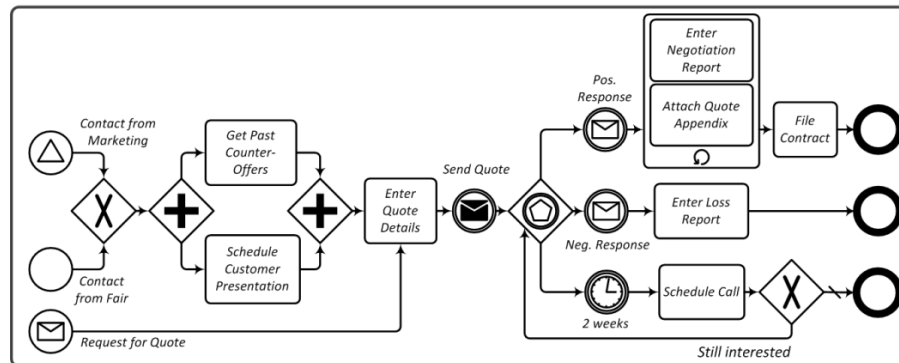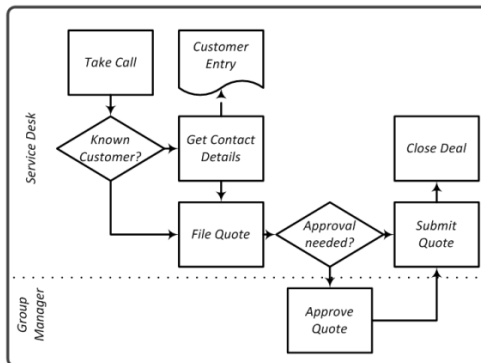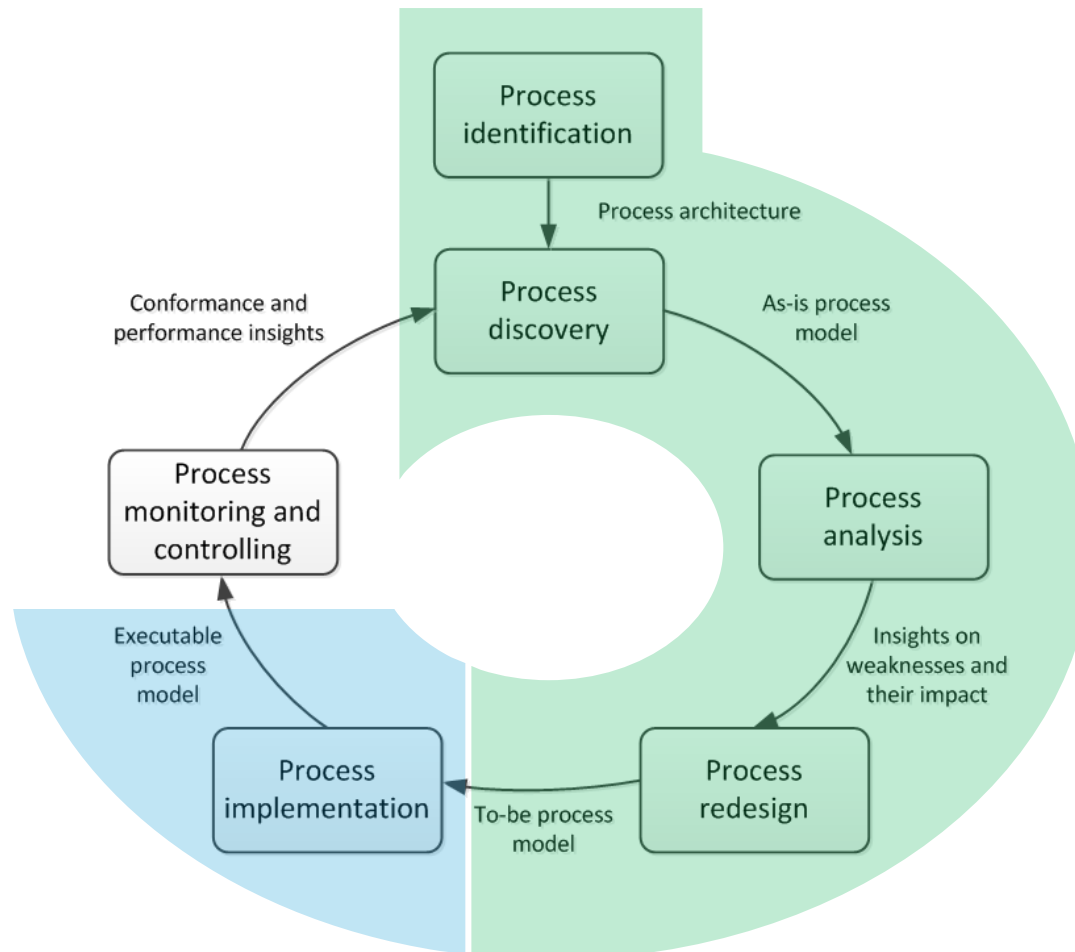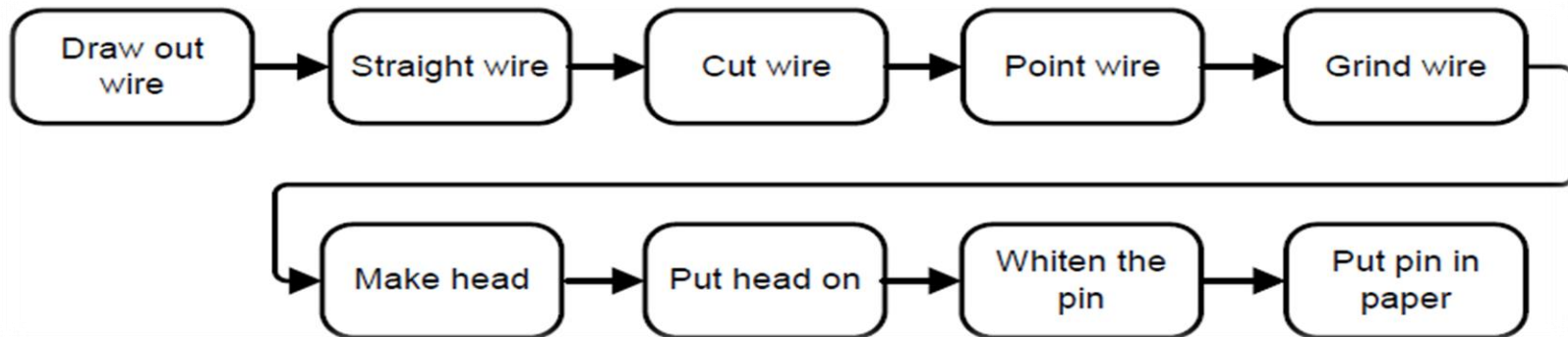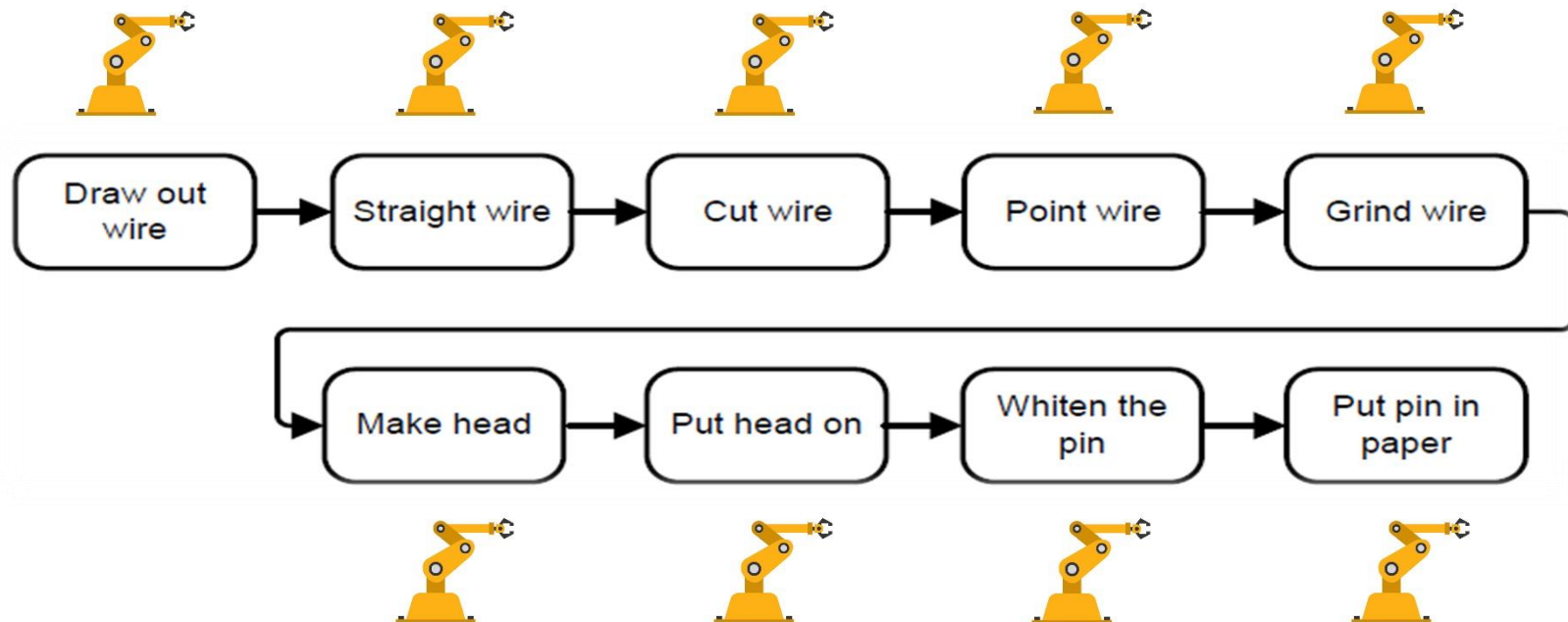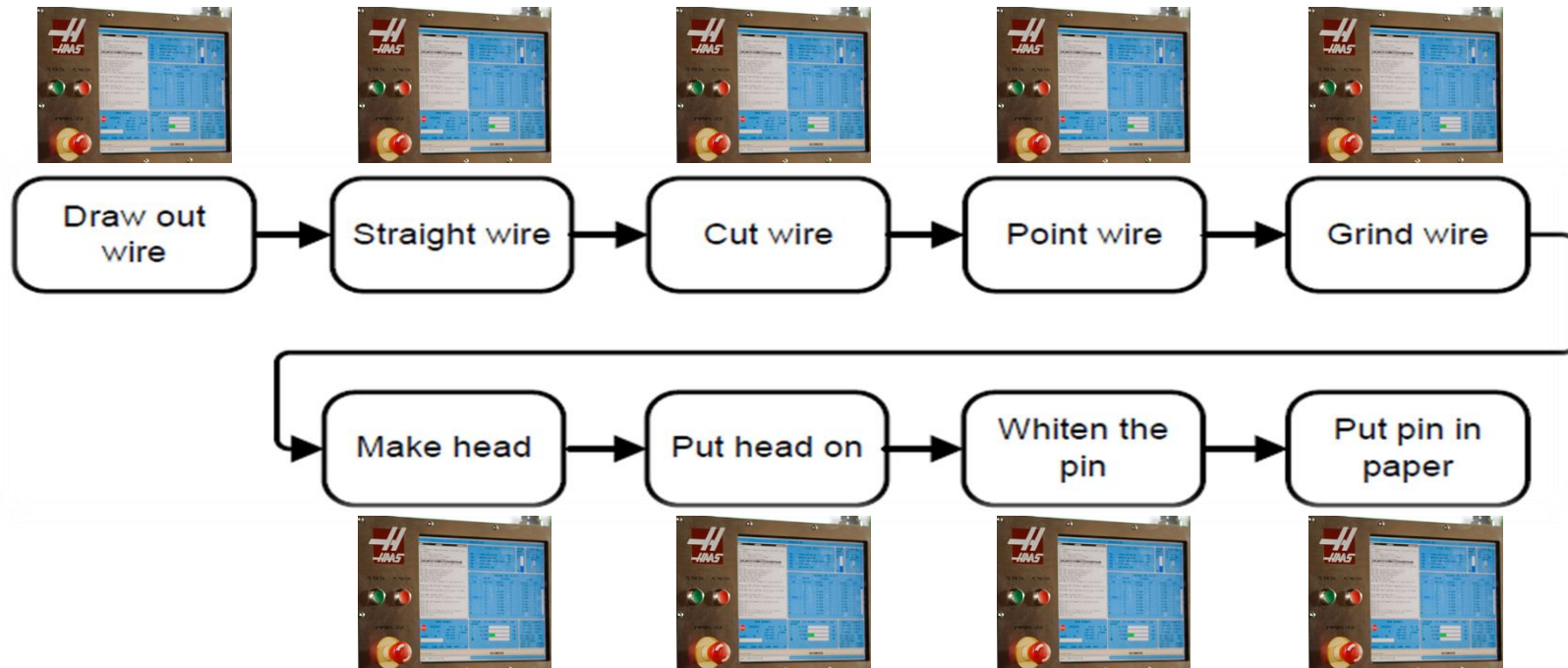
The event log

Execution model

[Boarding pass check passed]

Proceed to security check

[In front of security check, Hand luggage on screening belt]

Security screening not passed

Pass security screening

Pass luggage screening

[Beyond security screening, Hand luggage on screening belt]

[In front of security check, Hand luggage after monitor]

Pass luggage screening

Pass security screening

Security screening not passed

[Beyond security screening, Hand luggage after monitor]
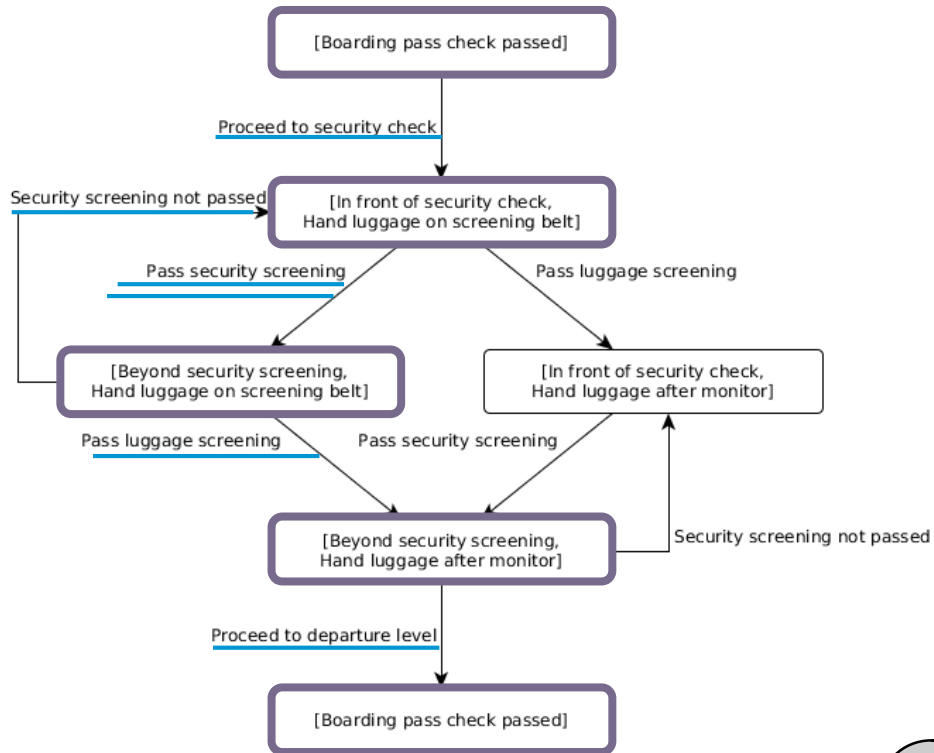
Proceed to departure level
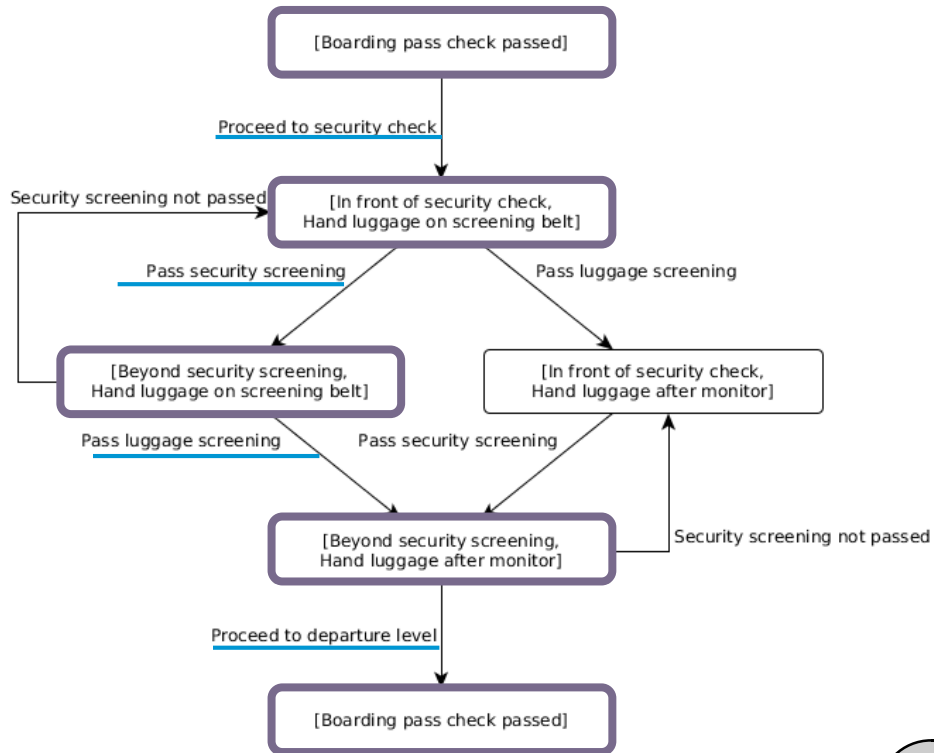
[Boarding pass check passed]

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
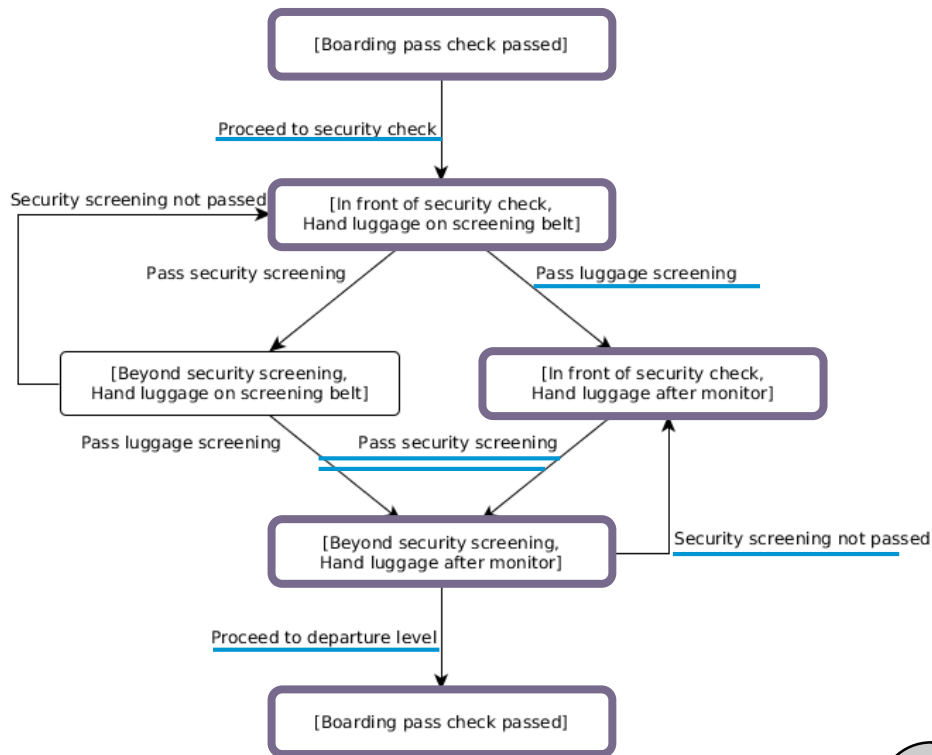
The event log

- 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
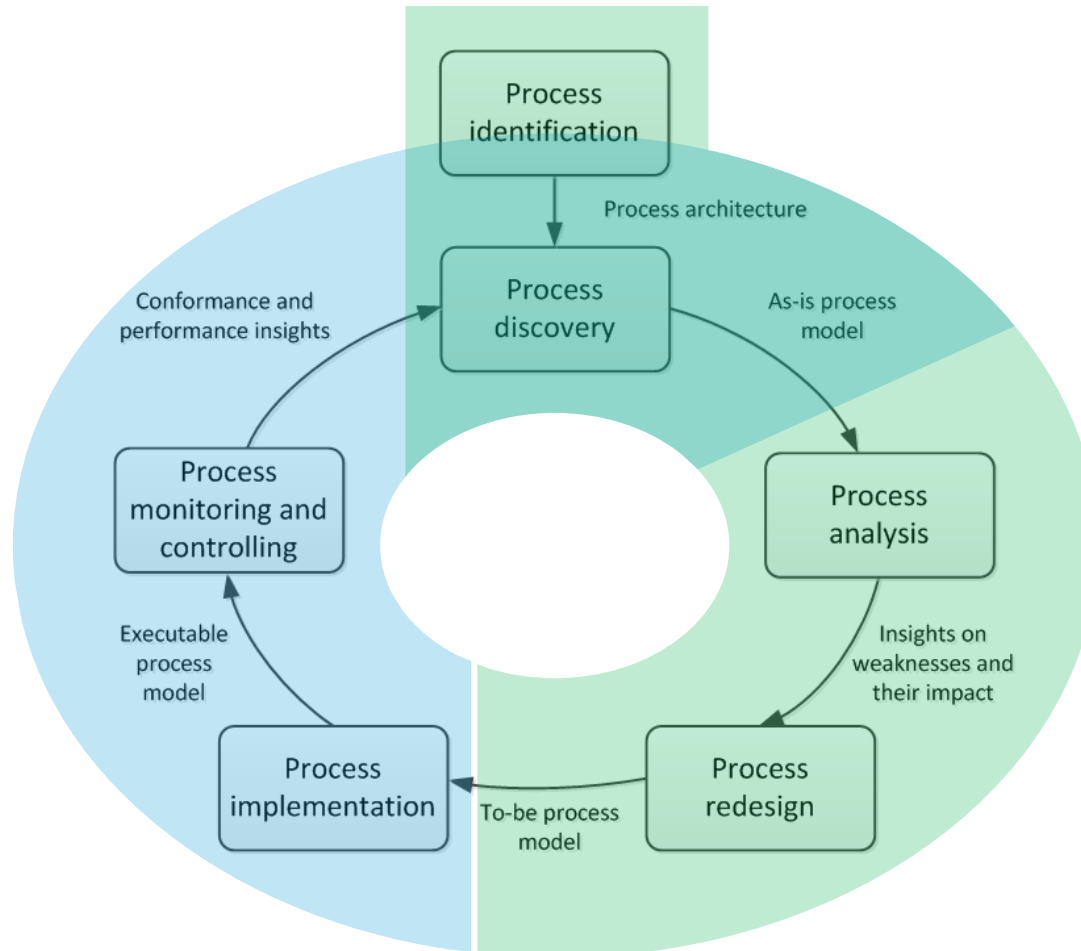
The event log

- 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

PROCESS IMPLEMENTATION

The event log

[Boarding pass check passed]

Proceed to security check

Security screening not passed → [In front of security check, Hand luggage on screening belt]

Pass security screening   Pass luggage screening

?

[Beyond security screening, Hand luggage on screening belt]   [In front of security check, Hand luggage after monitor]

Pass luggage screening   Pass security screening

[Beyond security screening, Hand luggage after monitor]   Security screening not passed

Proceed to departure level

[Boarding pass check passed]

- 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
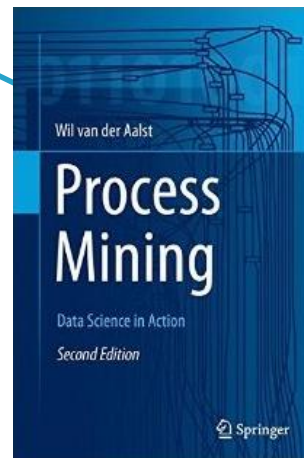
From data to action

Process mining

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

Process mining at large

Business Process → Execution & Tracing → Event log

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

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
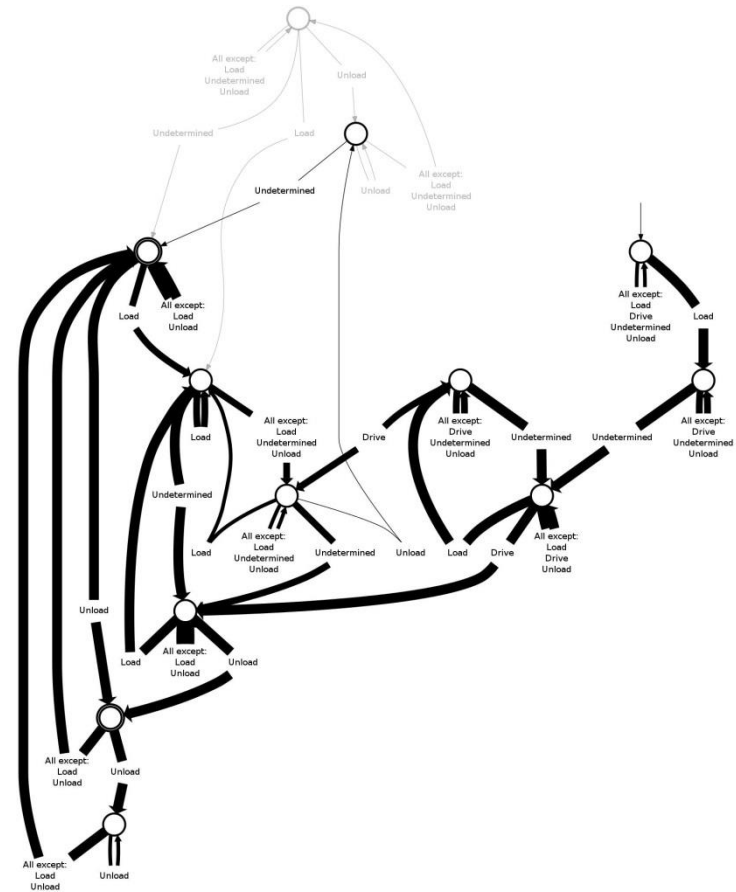
Play in, play out, replay
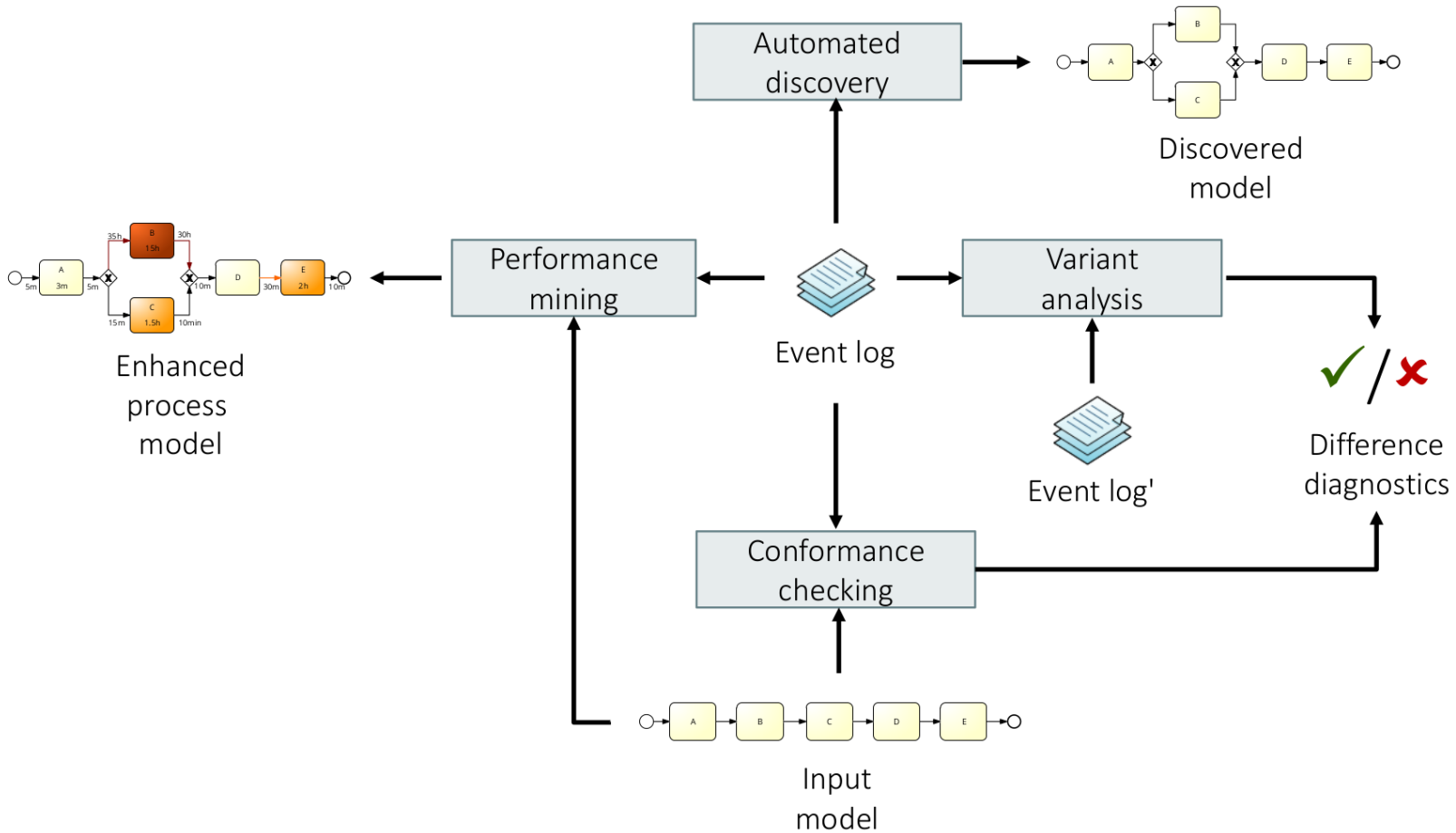

Play in	Play out	Replay
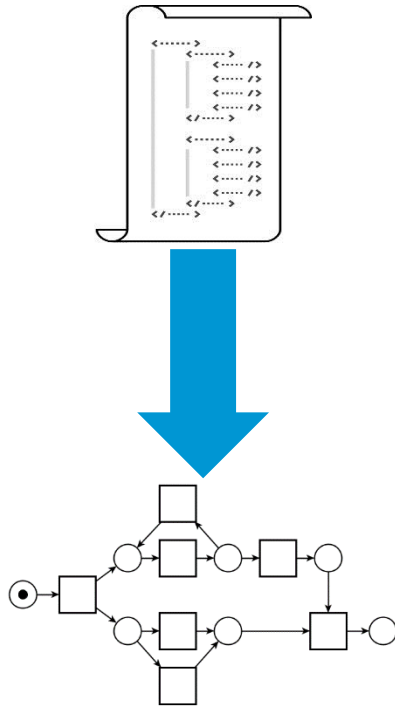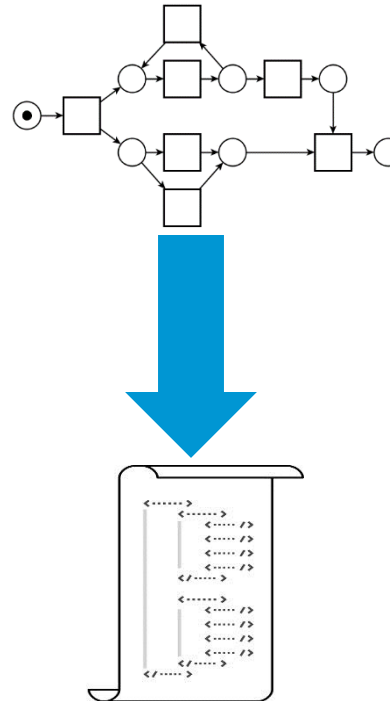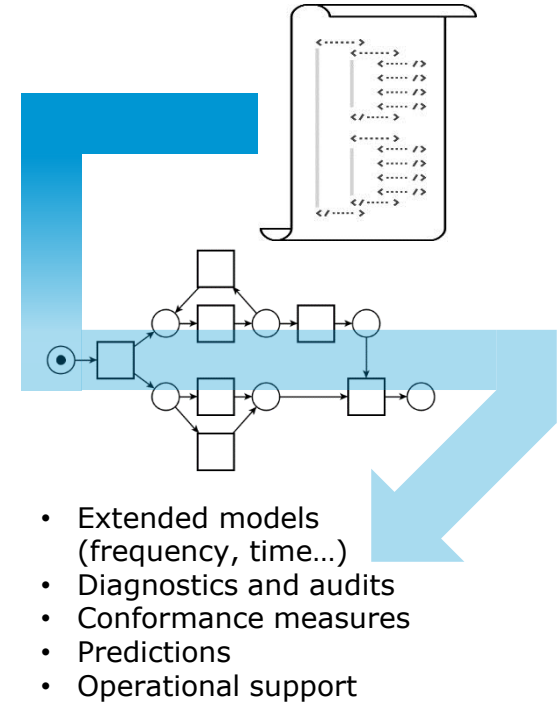
Replay bullets:
- Extended models (frequency, time…)
- Diagnostics and audits
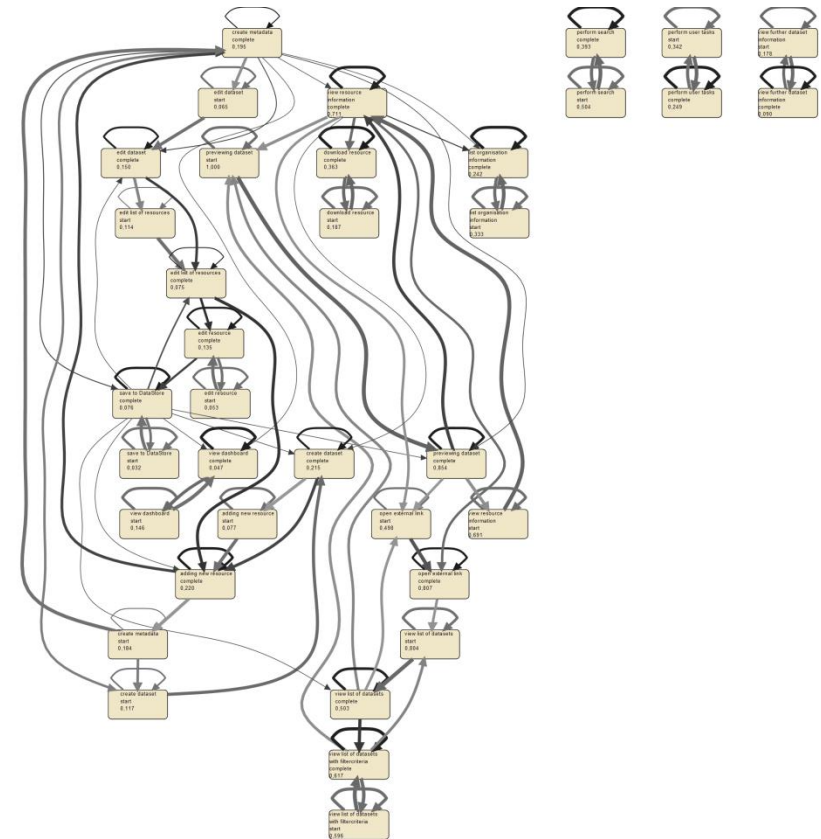- Conformance measures
- Predictions
- Operational support

Discovery	Execution / Simulation	Analytics

Process mining techniques

Automated process discovery

Conformance checking

From conformance checking to process enhancement

Process analytics

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
…	…	…	…	…	…

case id   activity name   timestamp   resource   other data

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
…	…	…	…	…	…

case id | activity name | timestamp | resource | other data

Event logs v. data at hand

https://api.flightstats.com/flex/weather/rest/v1/json/all/LAX?codeType=fs
</weatherUrl>
</airport>
</airport>
- <airport>
<fs>JFK</fs>
<iata>JFK</iata>
<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>1</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>
</airport>
</airports>
</appendix>
- <flightTracks>
- <flightTrack>
<flightId>305231585</flightId>
<carrierFsCode>AA</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>
<bearing>229.53397527577076</bearing>
<heading>228.69263064295163</heading>
- <positions>
- <position>
<lon>-73.7300033569336</lon>
<lat>40.68000030517578</lat>
<speedMph>137</speedMph>
<altitudeFt>800</altitudeFt>
<source>ASDI</source>
<date>2013-08-08T09:14:56.000Z</date>
</position>
- <position>
<lon>-73.69999694824219</lon>
<lat>40.70000076293945</lat>
<speedMph>143</speedMph>
<altitudeFt>1400</altitudeFt>
<source>ASDI</source>
<date>2013-08-08T09:13:56.000Z</date>
</position>
- <position>
<lon>-73.69999694824219</lon>
<lat>40.70000076293945</lat>
<speedMph>157</speedMph>
<altitudeFt>1500</altitudeFt>
<source>ASDI</source>
<date>2013-08-08T09:13:56.000Z</date>
</position>
- <position>
<lon>-73.66999816894531</lon>
<lat>40.72999954223633</lat>
<speedMph>163</speedMph>
<altitudeFt>2000</altitudeFt>
<source>ASDI</source>
<date>2013-08-08T09:12:55.000Z</date>
</position>

The ETL procedure (preventing the GIGO effect)

Correlation

Scoping

Snapshots

Granularity

...

- 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)

Correlation

Scoping

Snapshots

Granularity

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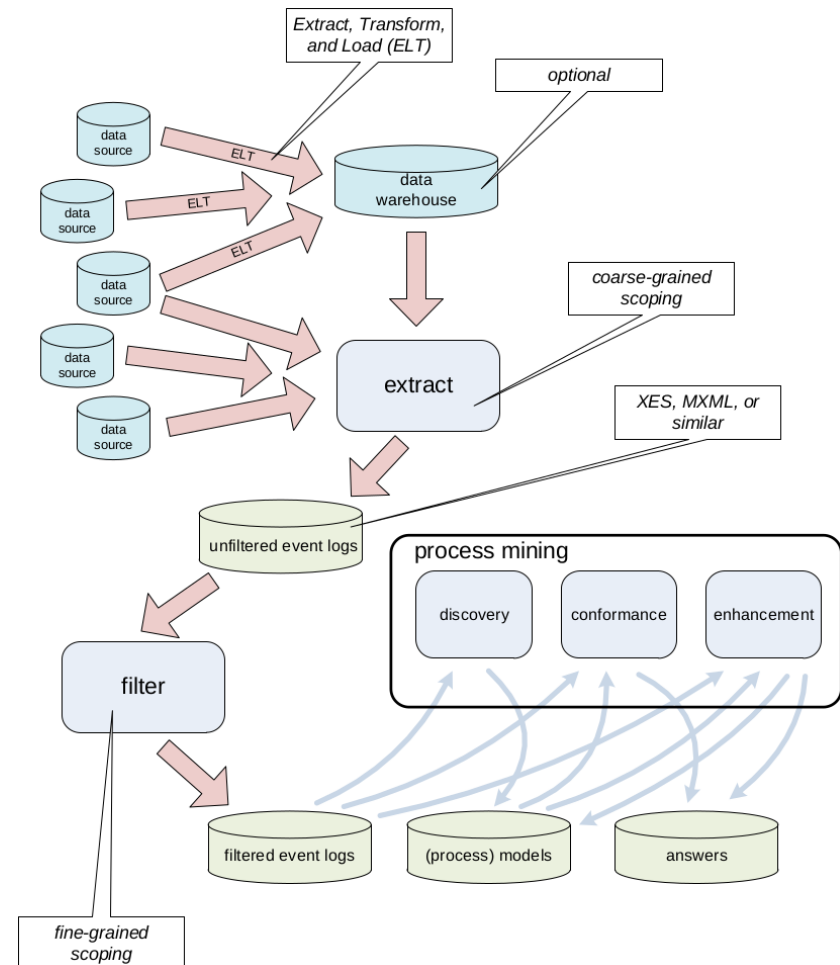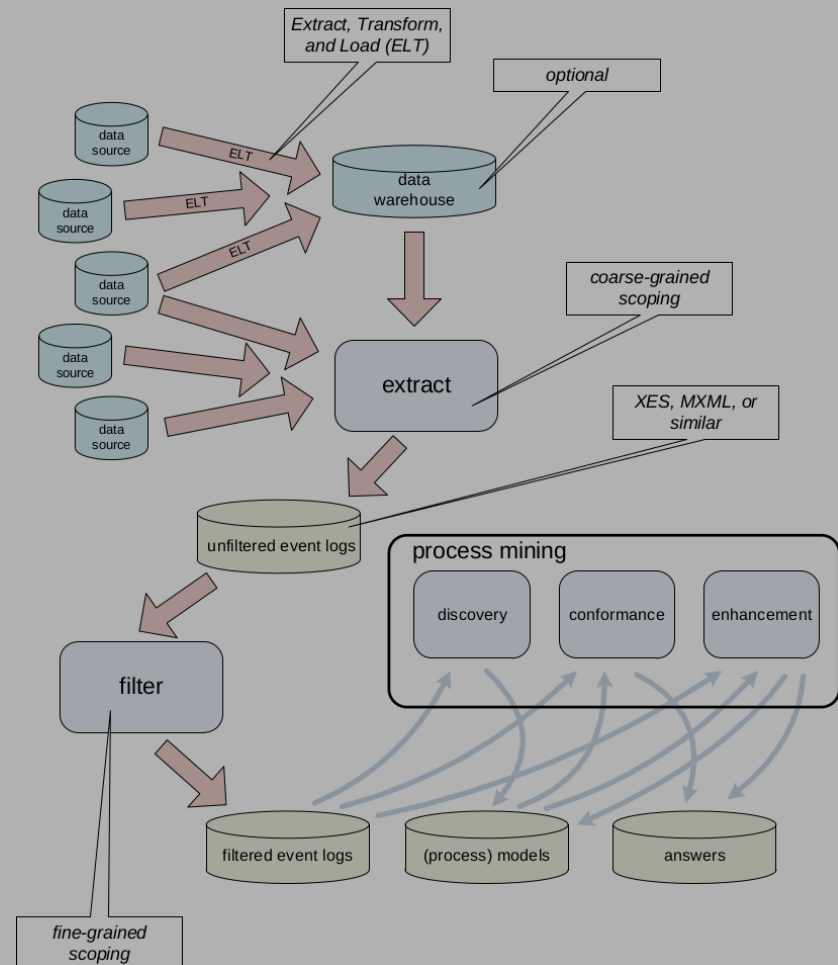
- Extract, Transform, and Load (ETL)
  1. Extract data from outside sources
  2. Transform it to fit operational needs
  3. Load it into target systems

☠ Timestamps

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

- http://www.xes-standard.org/

- Adopted by the IEEE Task Force on Process Mining
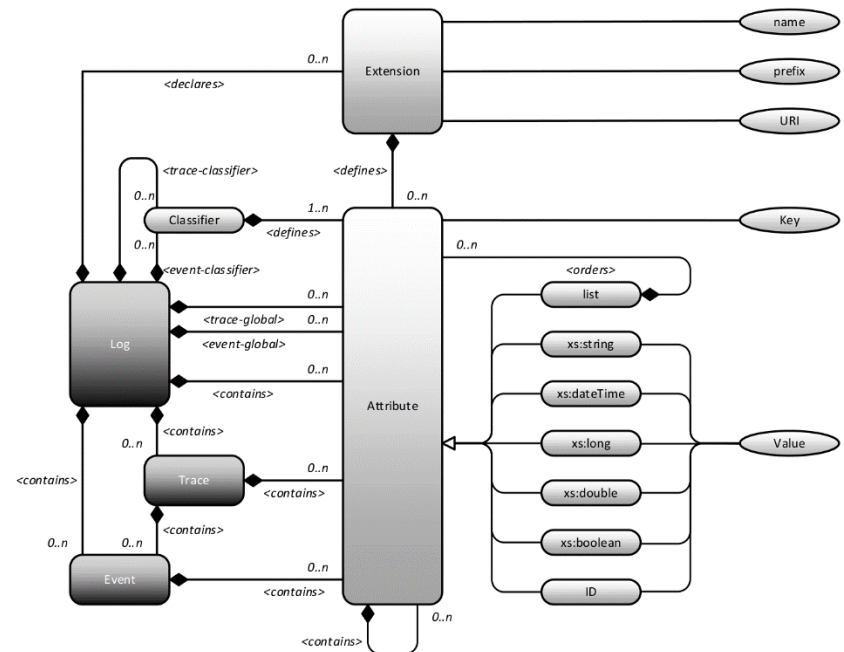
Everyday challenges in a process miner life

Assessment of the outcome quality

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

a
Receive
application form

l
Close application

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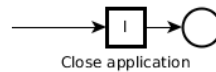$$\langle a, b, c, d, b, c, d, b, e, g, f, h, j, l \rangle^{10},$$
$$\langle a, b, e, g, f, h, k, l \rangle^{60},$$
$$\langle a, b, e, f, g, h, j, l \rangle^{20},$$
$$\langle a, b, e, g, f, h, j, l \rangle^{20}$$
$$]$$

Fitness: ☑
　　All traces can be replayed
Generalisation: ☑
　　More unforeseen traces can be replayed too

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

Precision: ☒
　　Basically any trace starting with a and ending with l can be replayed

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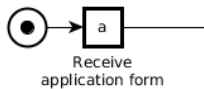$$\langle a, b, c, d, b, c, d, b, e, g, f, h, j, l \rangle^{10},$$
$$\langle a, b, e, g, f, h, k, l \rangle^{60},$$
$$\langle a, b, e, f, g, h, j, l \rangle^{20},$$
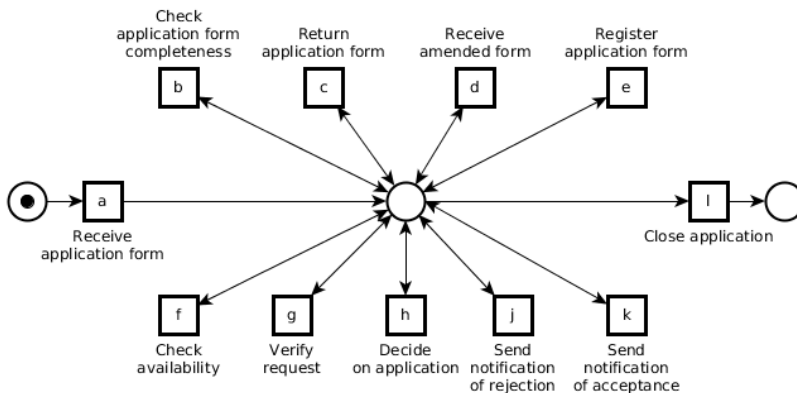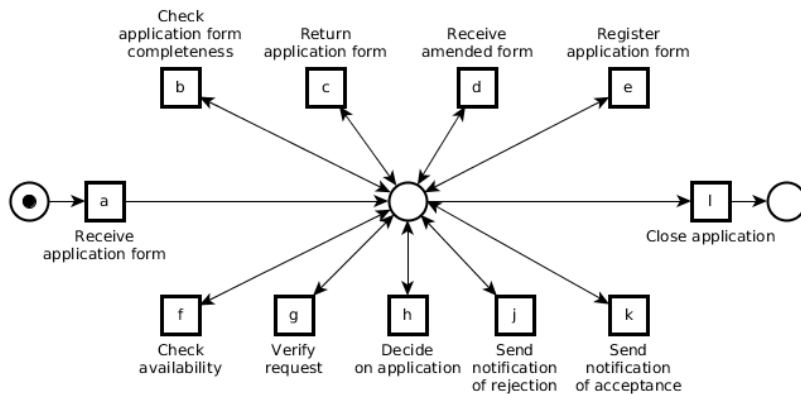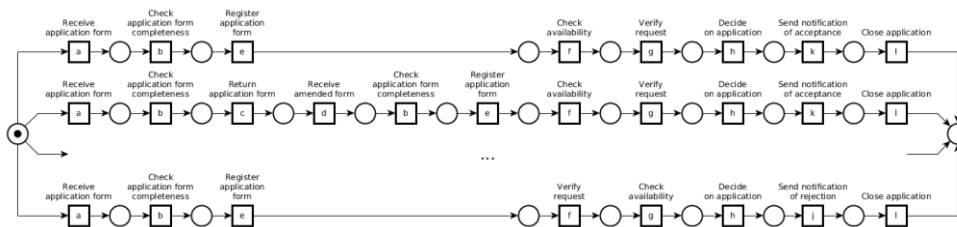$$\langle a, b, e, g, f, h, j, l \rangle^{20}$$
$$]$$

Flower model

Underfitting

Check application form completeness — b
Return application form — c
Receive amended form — d
Register application form — e

a — Receive application form
l — Close application

Check availability — f
Verify request — g
Decide on application — h
Send notification of rejection — j
Send notification of acceptance — k

Precision: ☒
Basically any trace starting with a and ending with l can be replayed

Source: https://www.aliexpress.com/
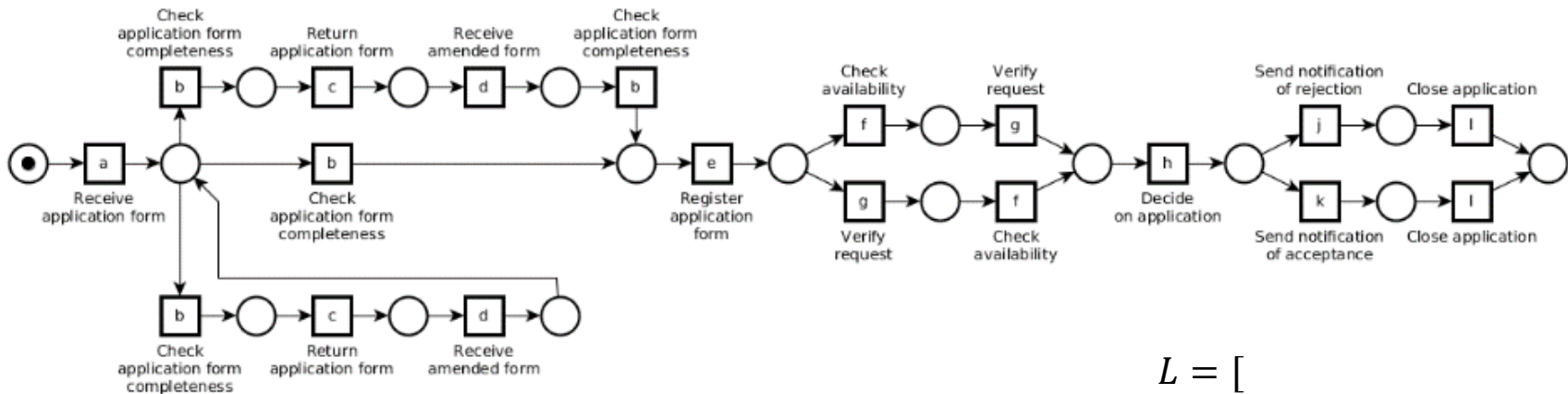
Overfitting

Generalisation: ☒
          Nothing outside the log is accepted

Fitness: ☑
Precision: ☑
Generalisation: ☑

Simplicity: ☒

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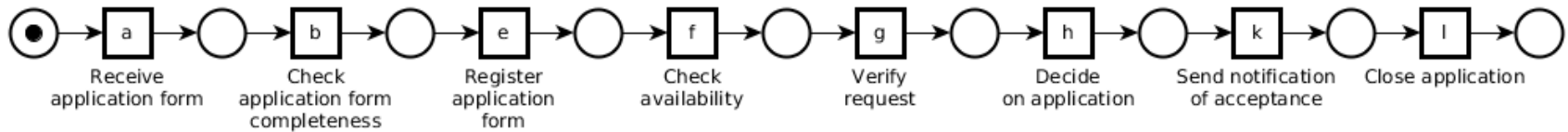$$\langle a, b, c, d, b, c, d, b, e, g, f, h, j, l \rangle^{10},$$
$$\langle a, b, e, g, f, h, k, l \rangle^{60},$$
$$\langle a, b, e, f, g, h, j, l \rangle^{20},$$
$$\langle a, b, e, g, f, h, j, l \rangle^{20}$$
$$]$$

Most frequent trace(s) as model

Receive application form (a), Check application form completeness (b), Register application form (e), Check availability (f), Verify request (g), Decide on application (h), Send notification of acceptance (k), Close application (l)
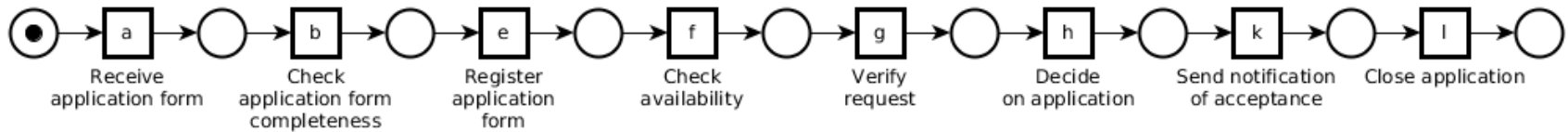
Simplicity: ☑

OK, all the other the traces?

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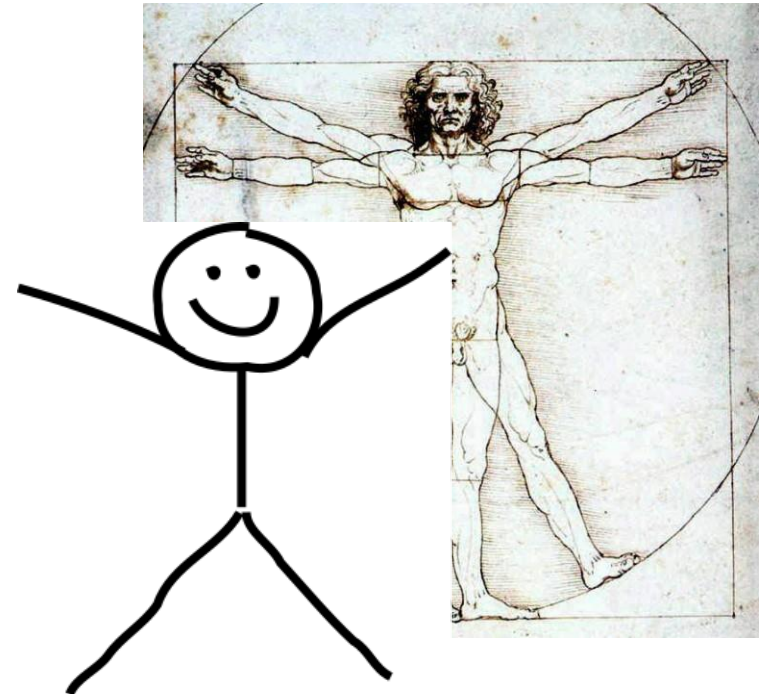$$\langle a, b, e, g, f, h, j, l \rangle^{20}$$
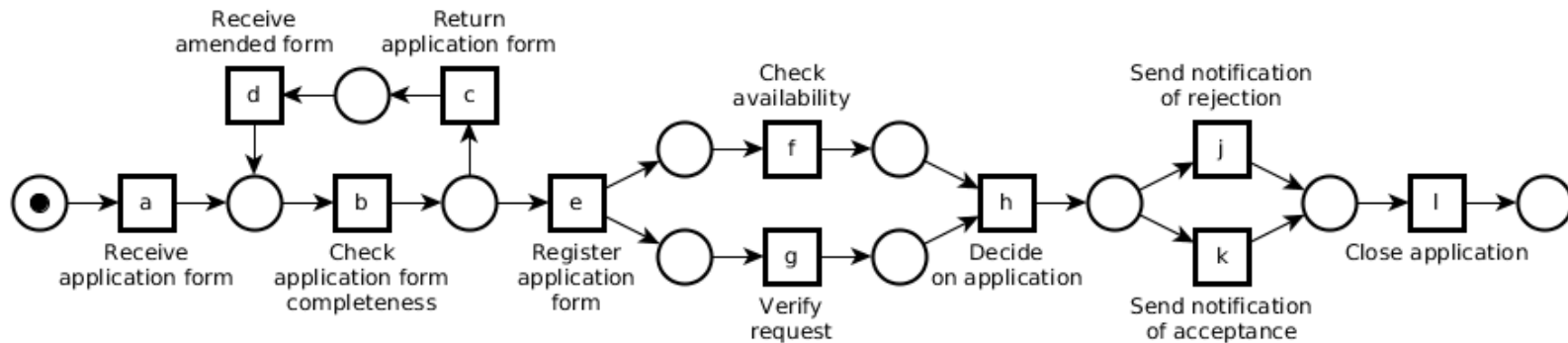$$]$$

Most frequent trace(s) as model

Receive application form | Check application form completeness | Register application form | Check availability | Verify request | Decide on application | Send notification of acceptance | Close application

Simplicity: ☑

OK, all the rest?

A reasonable model

Fitness: ☑
Precision: ☑
Generalisation: ☑
Simplicity: ☑

The objective

Before
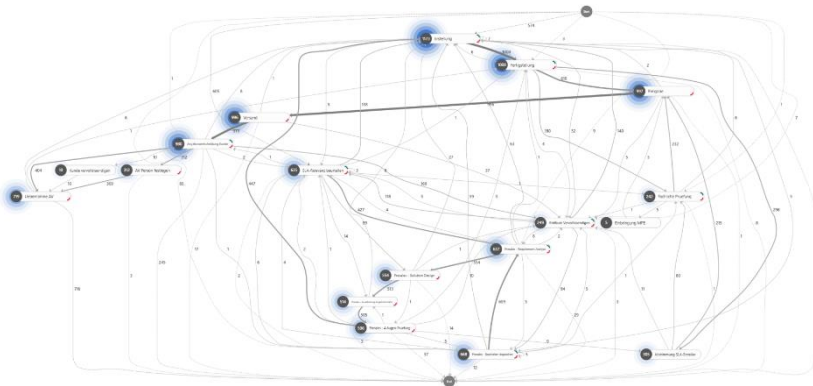
After
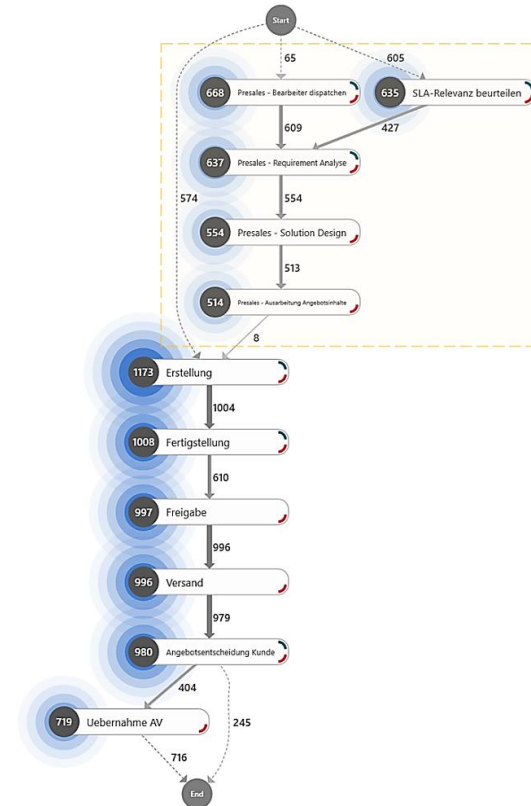
Research

Process mining

Ongoing challenges

Process mining research institutes
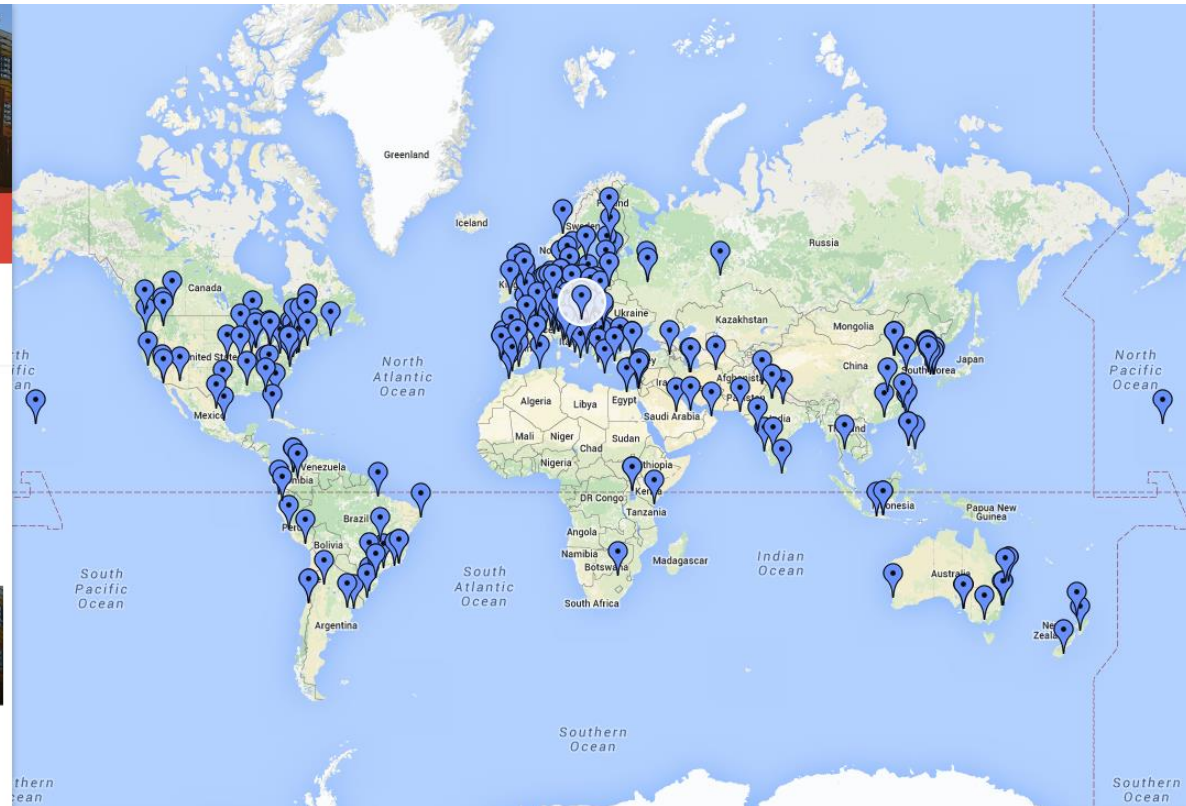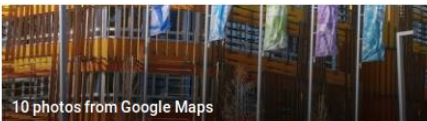
Process mining at WU Vienna
https://wu.ac.at/infobiz

- 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