Extraction et Gestion des Connaissances, EGC-2023

Data-aware Processes and their Executions

what's in for Knowledge Representation and Graphs

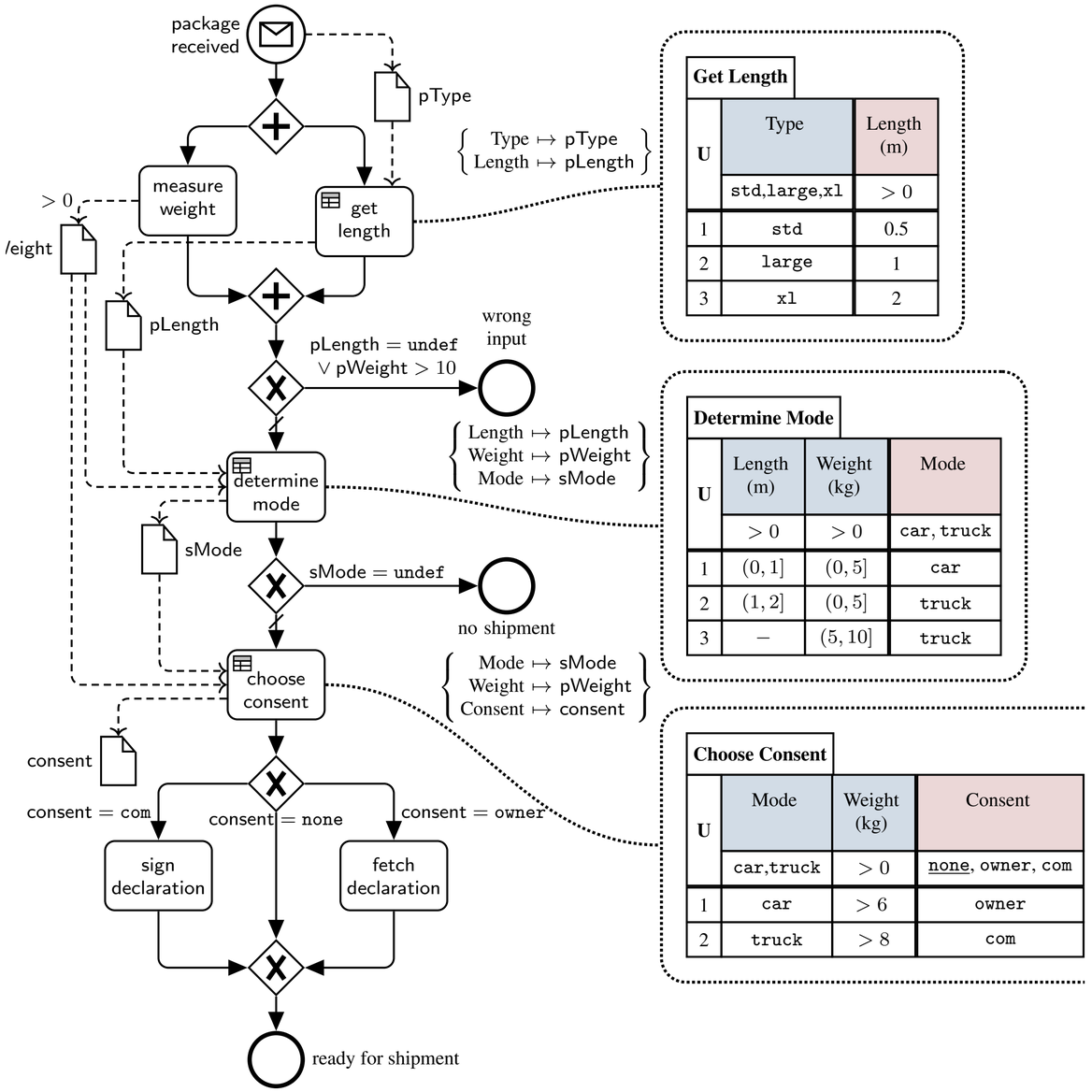
Chiara Ghidini, Fondazione Bruno Kessler

The agenda

- Work-Processes and Process Mining
- …. Why did I get intersted in that?
- **Challenges!**
  - **Conceptual** & **Semantic modeling and analysis** of process executions
  - **Event Knowledge Graphs**
  - **Extraction of Process Knowledge Graphs from text**

Work-processes

how to model them



Work process

A set of **logically related tasks** performed to achieve a **defined business outcome** for a particular customer or market.

(Davenport, 1992)

A **collection of activities** that **take** one or more kinds of **input** and **create an output** that is of **value to the customer**.

(Hammer & Champy, 1993)

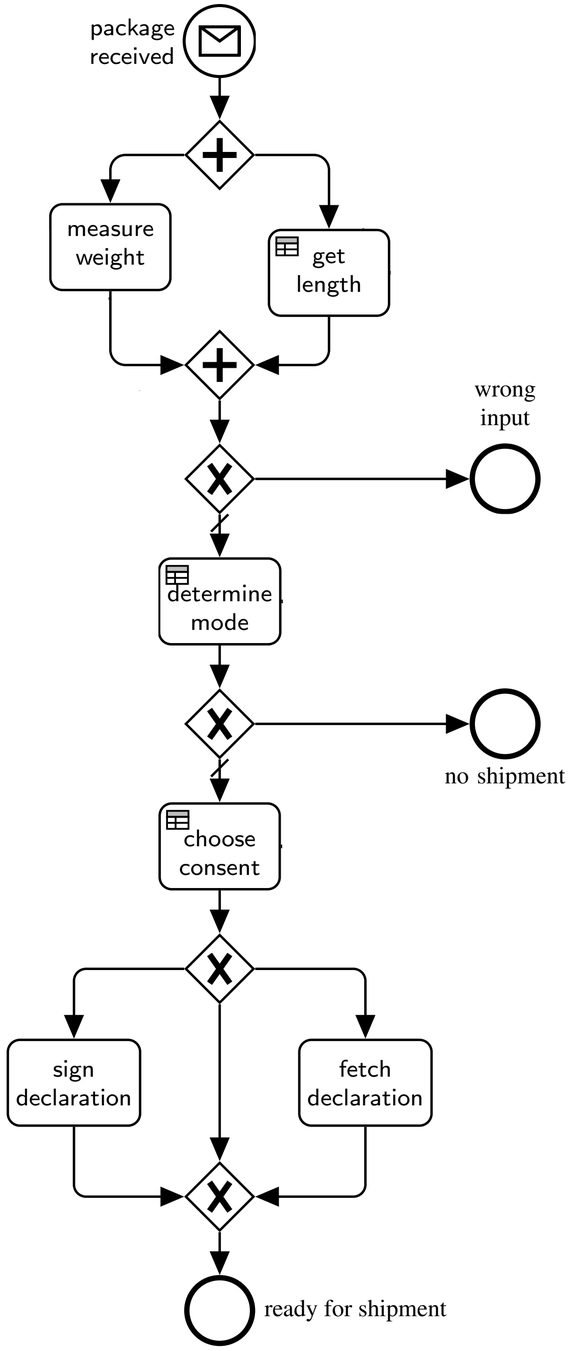
A **set of activities** performed **in coordination** in an **organizational** and **technical environment**. These activities **jointly realize a business goal**.

(Weske, 2011)

credits to Marco Montali

Different perspectives

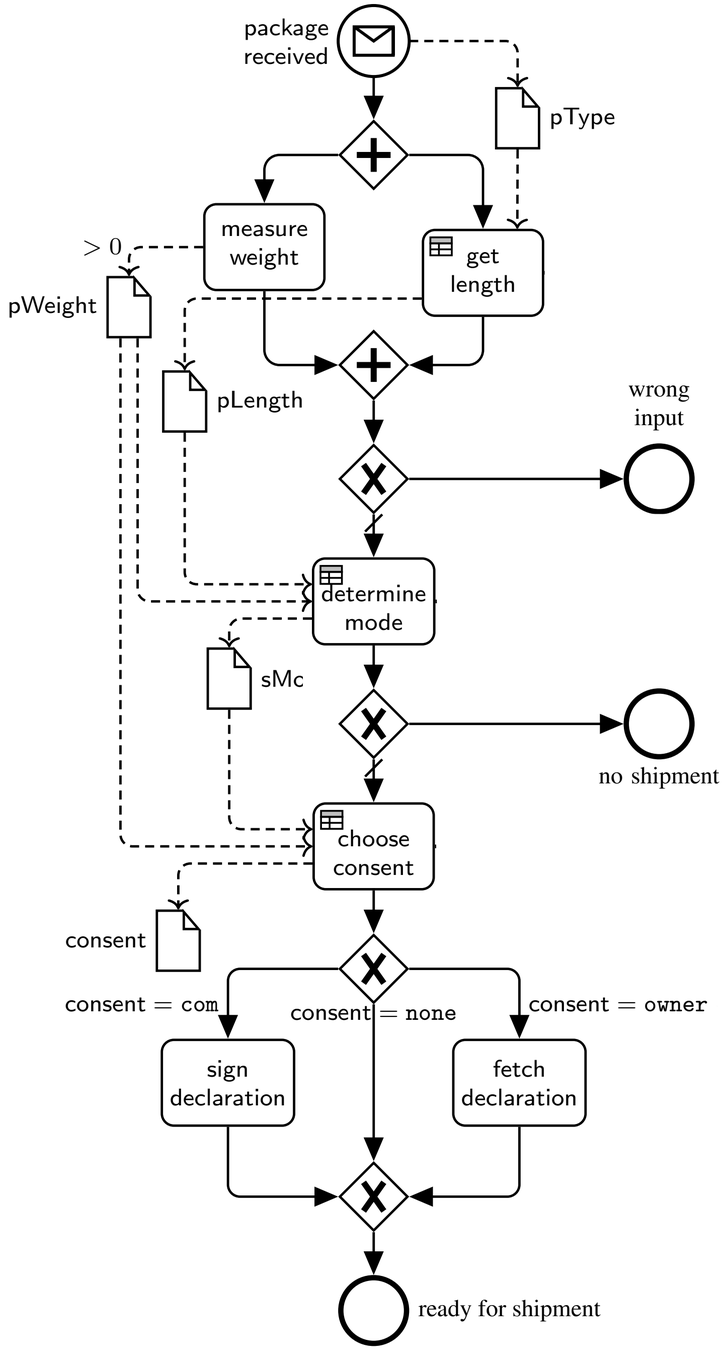
- Control-flow



credits to Marco Montali

Different languages for different perspectives

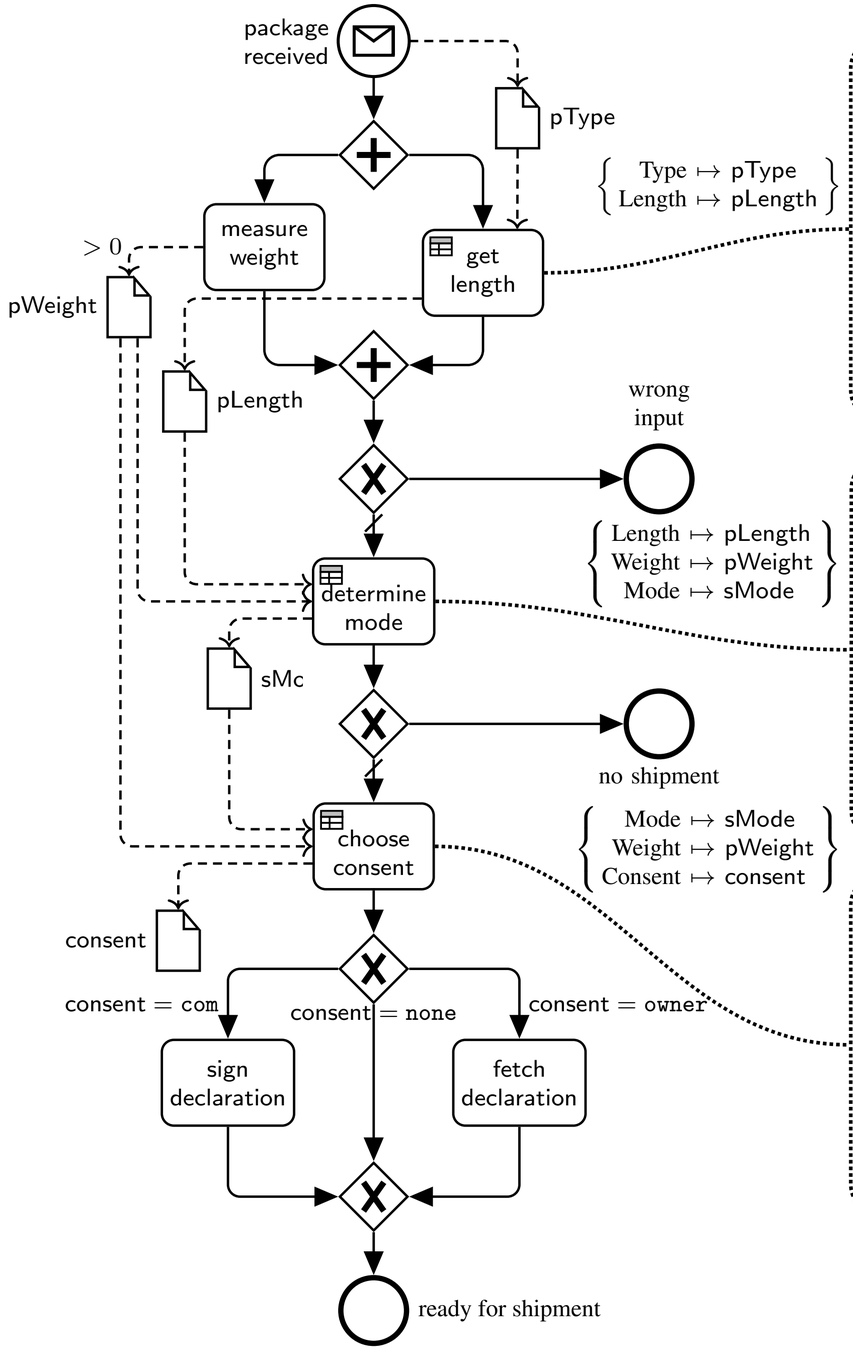
- Control-flow
- Data



credits to Marco Montali

Different languages for different perspectives

- Control-flow
- Data
- Decisions



Get Length

U	Type	Length (m)
	std,large,xl	> 0
1	std	0.5
2	large	1
3	xl	2

Determine Mode

U	Length (m)	Weight (kg)	Mode
	> 0	> 0	car, truck
1	$(0, 1]$	$(0, 5]$	car
2	$(1, 2]$	$(0, 5]$	truck
3	–	$(5, 10]$	truck

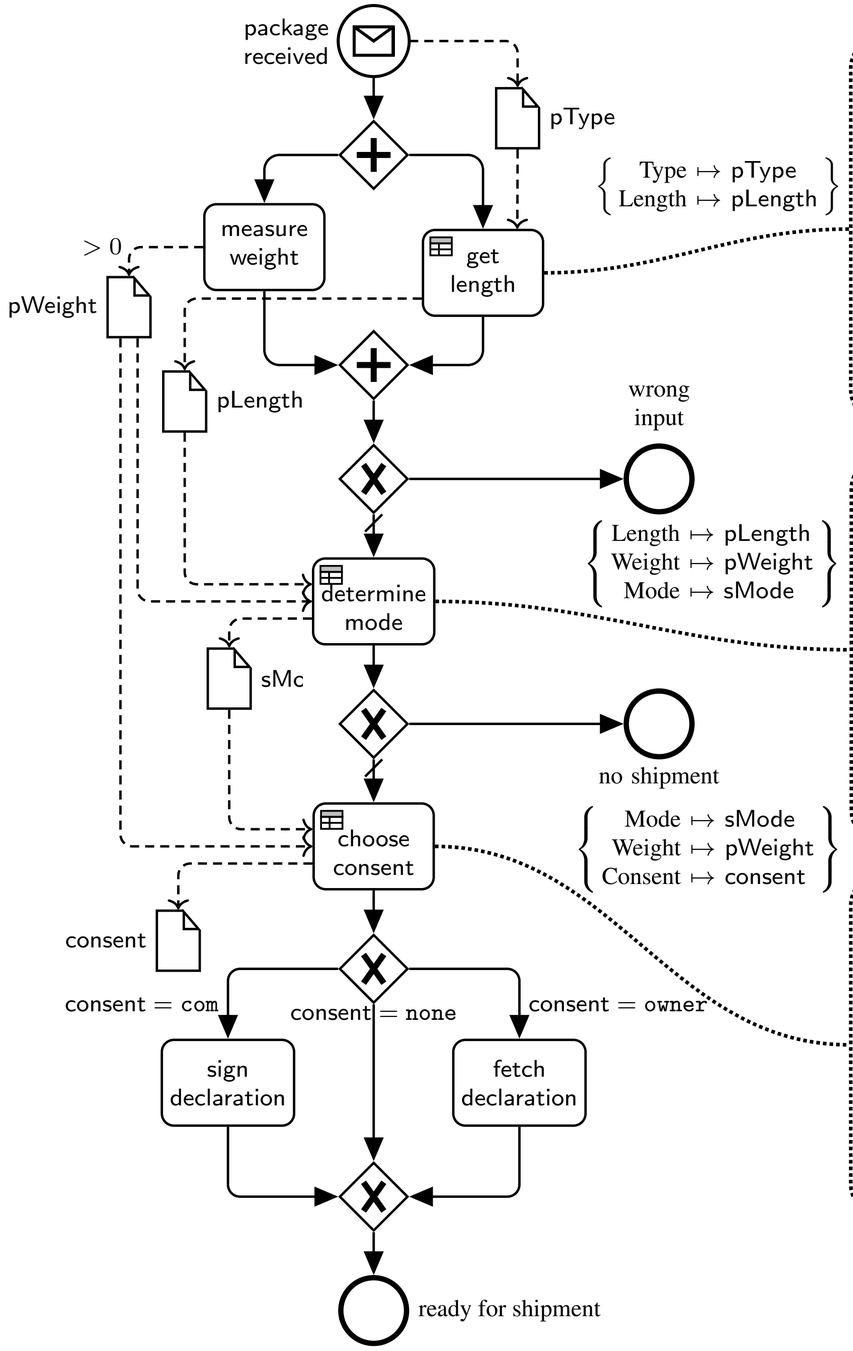
Choose Consent

U	Mode	Weight (kg)	Consent
	car,truck	> 0	none, owner, com
1	car	> 6	owner
2	truck	> 8	com

credits to Marco Montali

Different languages for different perspectives

- Control-flow
- Data
- Decisions
- Resources
- Time
- …



Get Length

U	Type	Length (m)
	std,large,xl	> 0
1	std	0.5
2	large	1
3	xl	2

Determine Mode

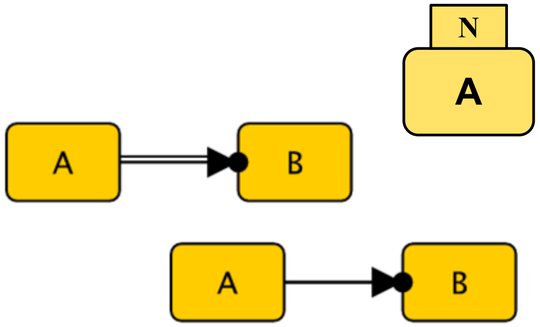
U	Length (m)	Weight (kg)	Mode
	> 0	> 0	car, truck
1	$(0, 1]$	$(0, 5]$	car
2	$(1, 2]$	$(0, 5]$	truck
3	–	$(5, 10]$	truck

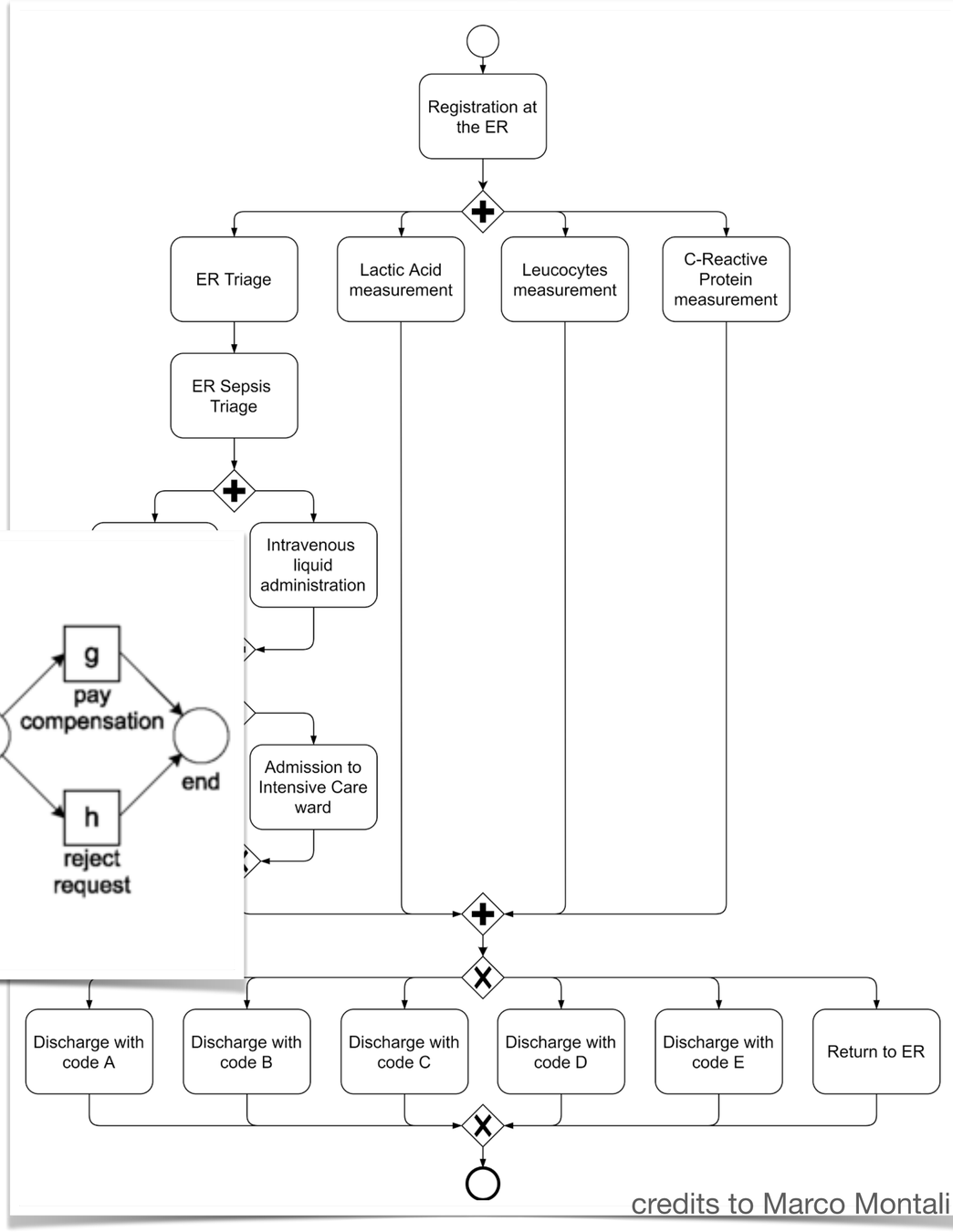
Choose Consent

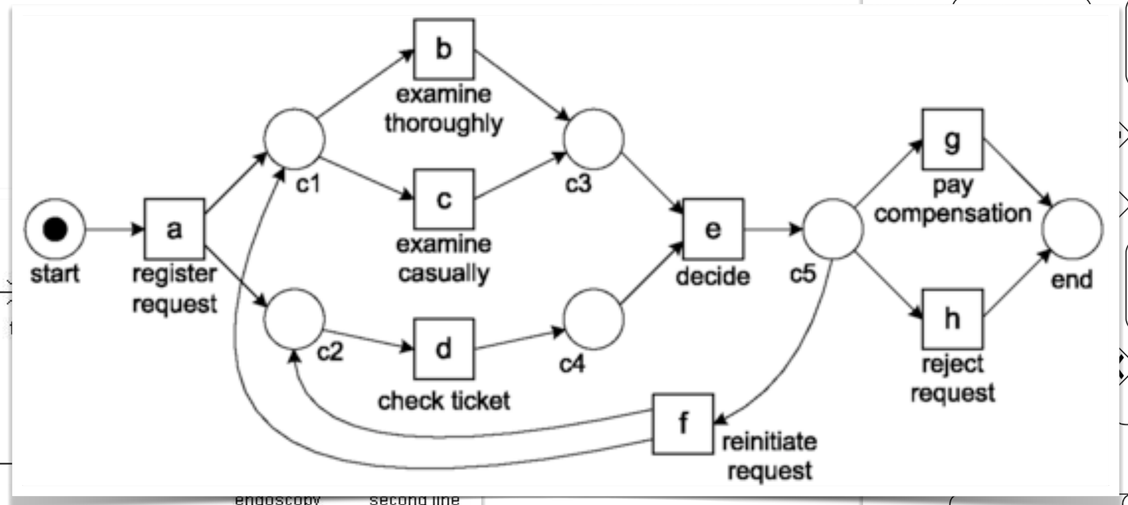
U	Mode	Weight (kg)	Consent
	car,truck	> 0	none, owner, com
1	car	> 6	owner
2	truck	> 8	com

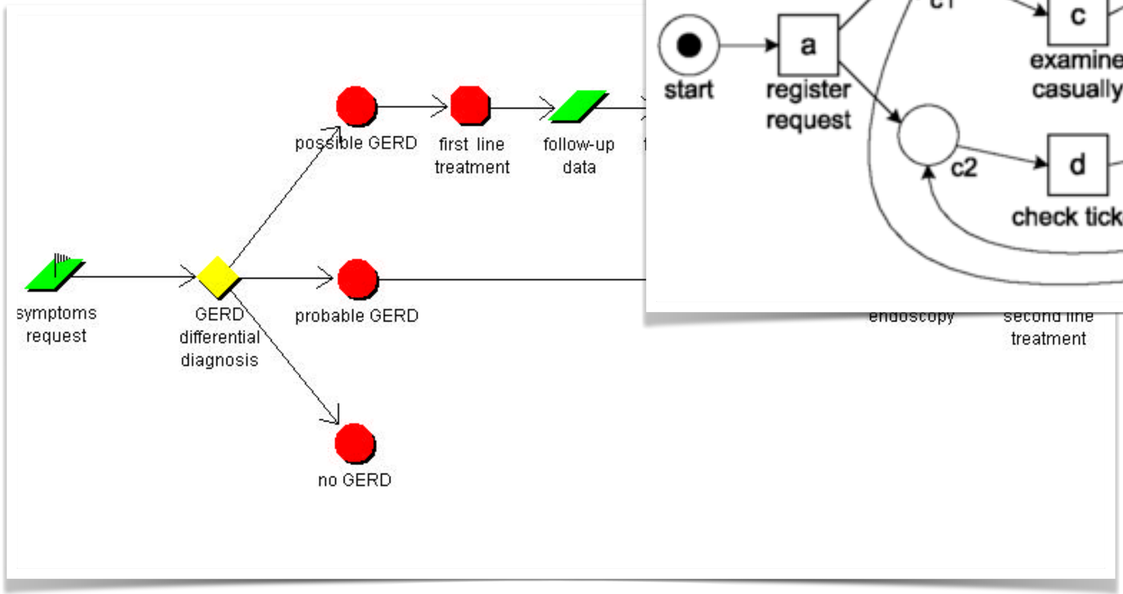
credits to Marco Montali

Importance of Models









credits to Marco Montali

Why maintaining process models?

Documentation

Design-time support

- What-if analysis, teaching
- Simulation
- Verification

Runtime support: enactment and orchestration

credits to Marco Montali

Why maintaining process models?

Documentation

Design-time support

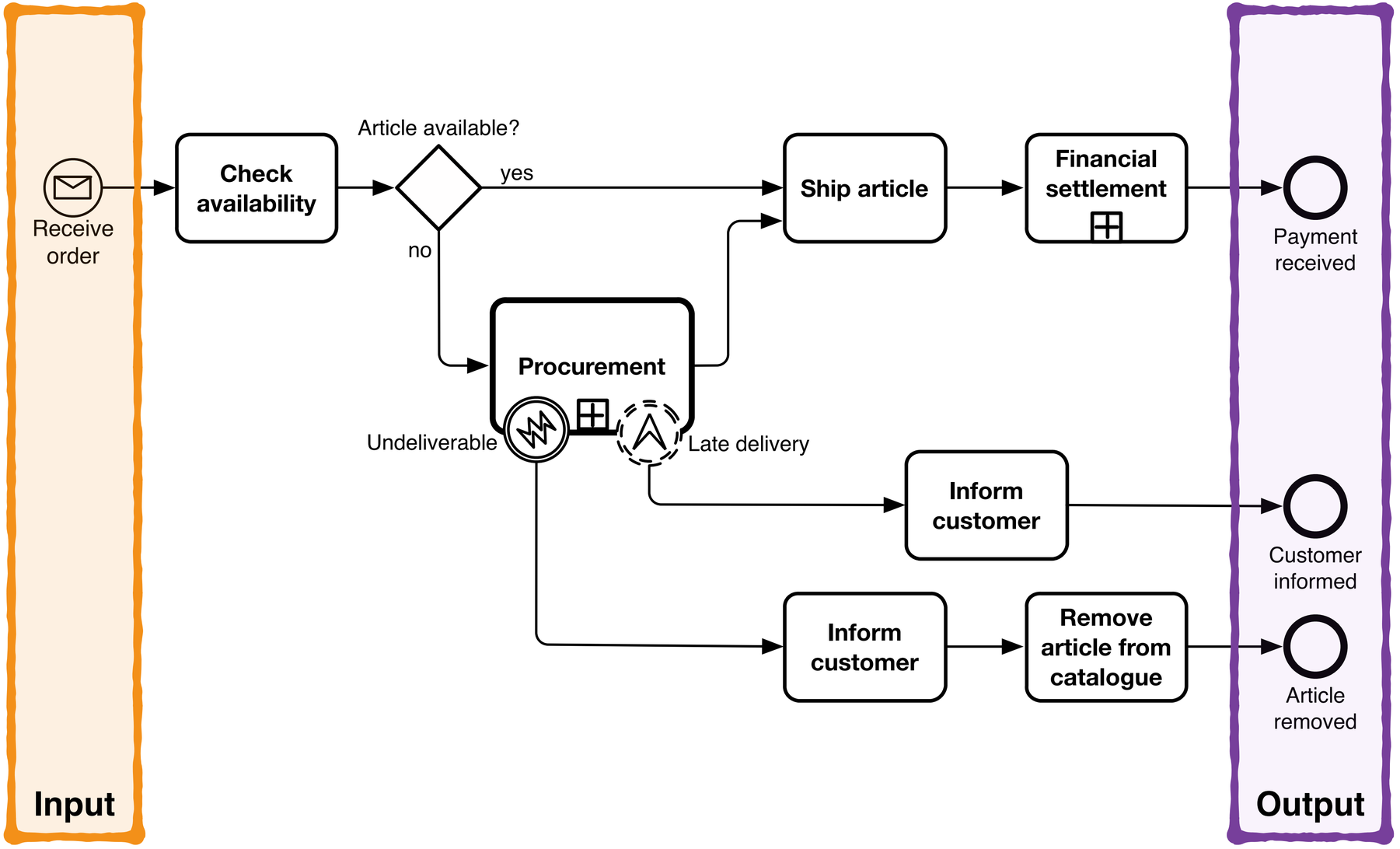
- What-if analysis, teaching
- Simulation
- Verification

Runtime support: enactment and orchestration

Only useful if they accurately represent reality!

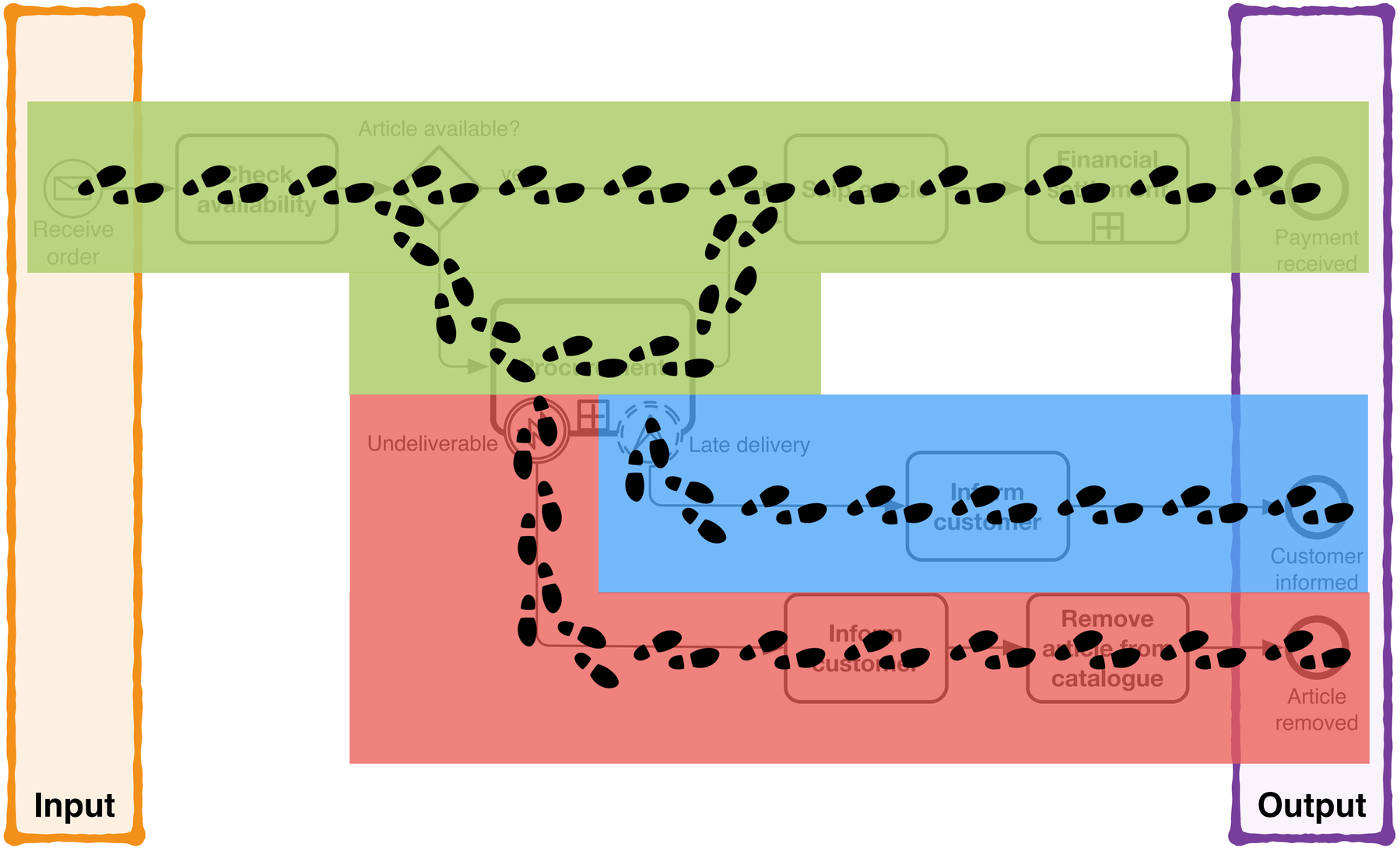
credits to Marco Montali

Work-process modelling = modelling of pathways



credits to Marco Montali

(Business) process modelling = modelling of pathways

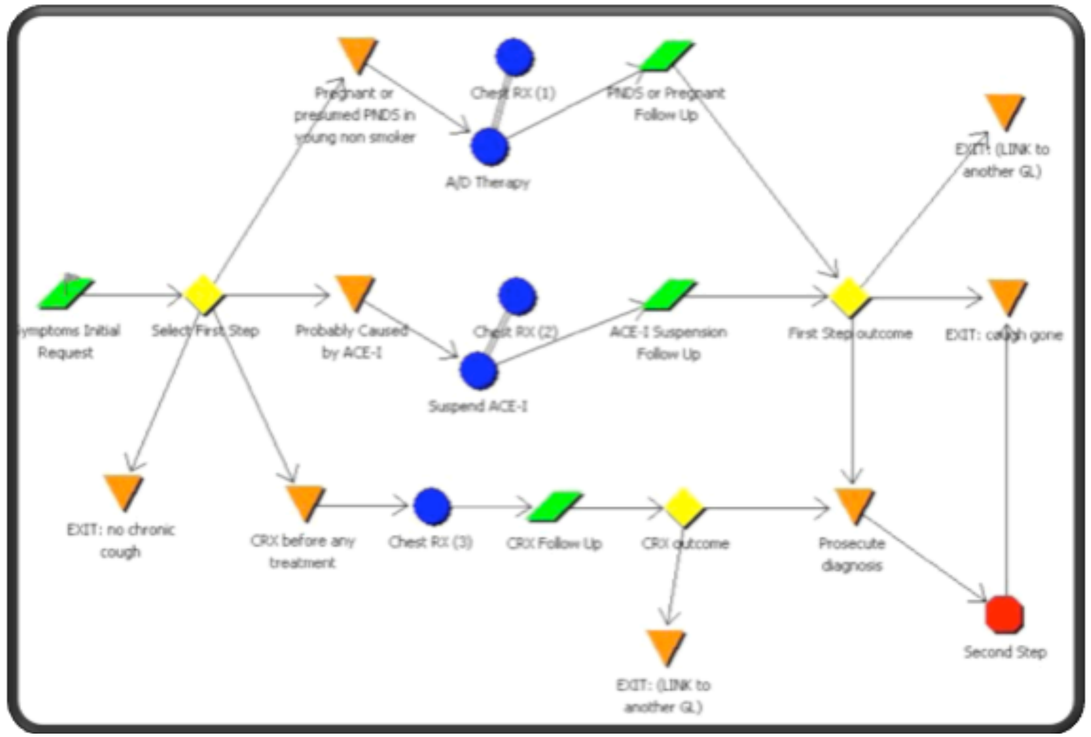


credits to Marco Montali

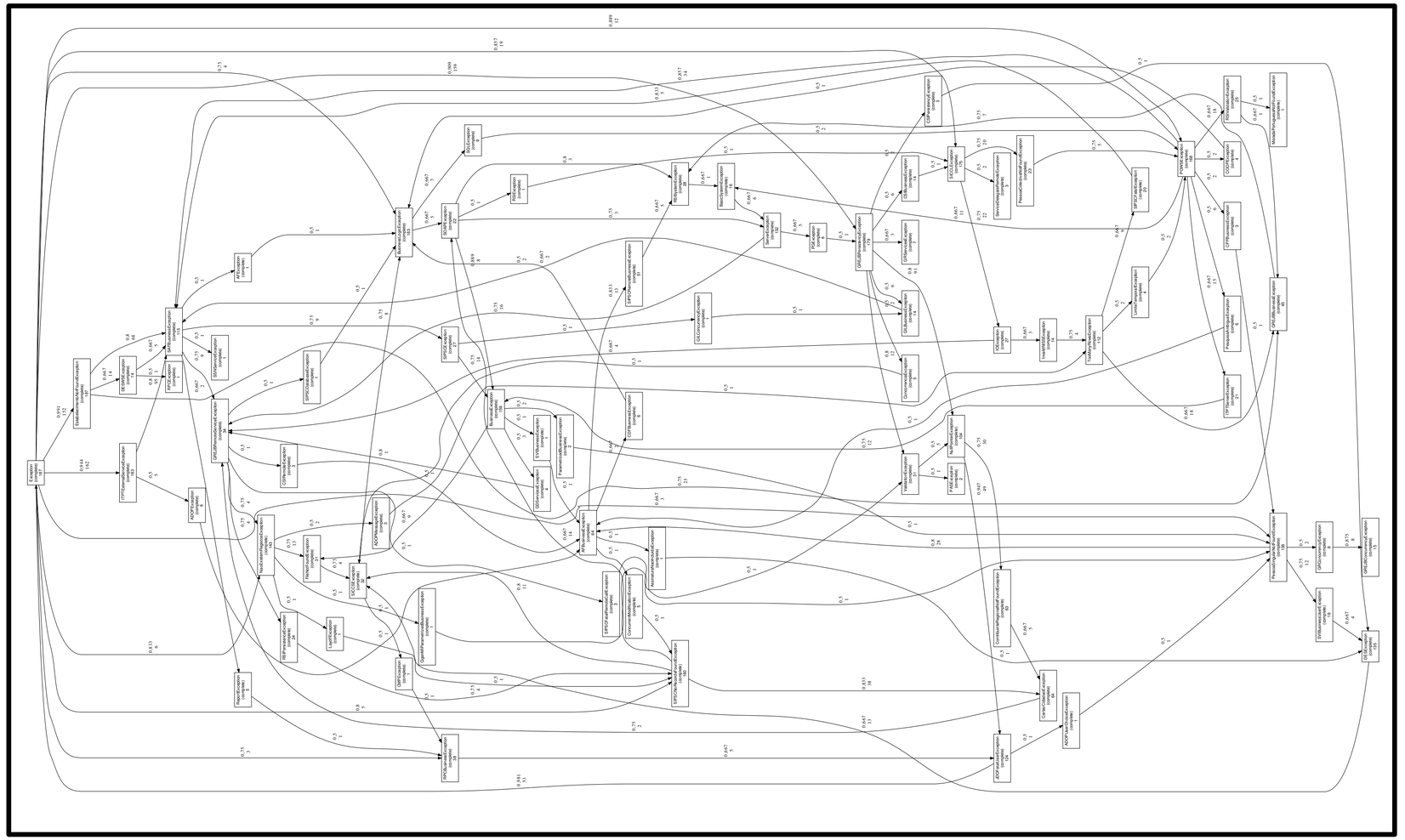

The Model


The reality

Model of Clinical Pathways

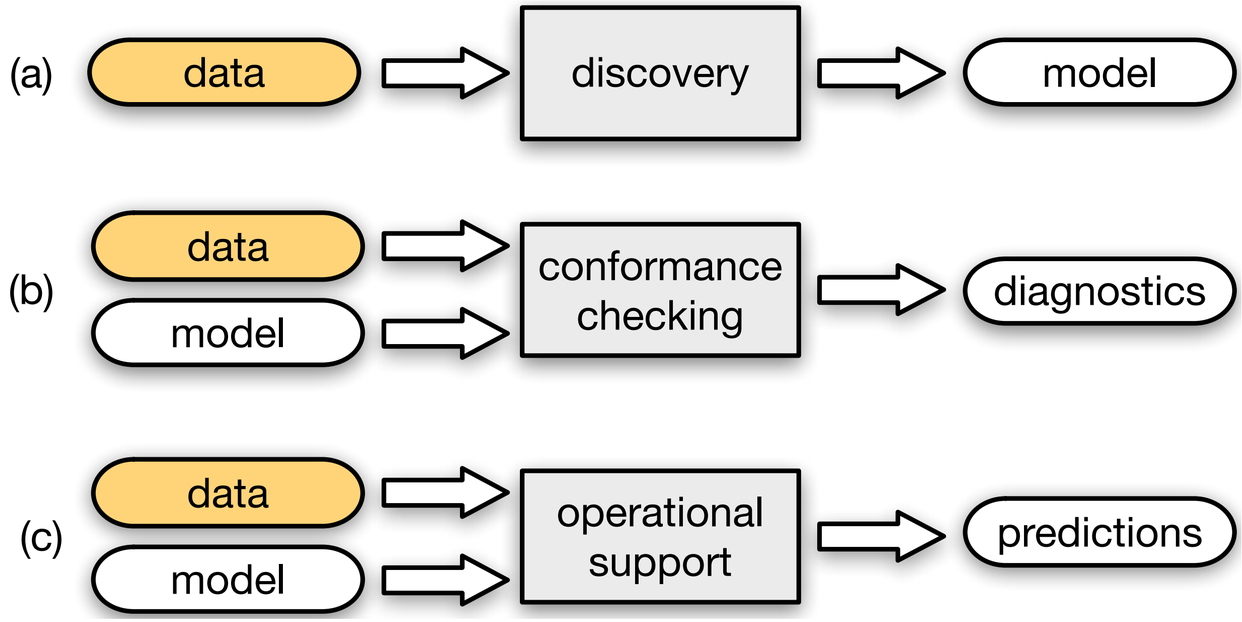


Real Pathways



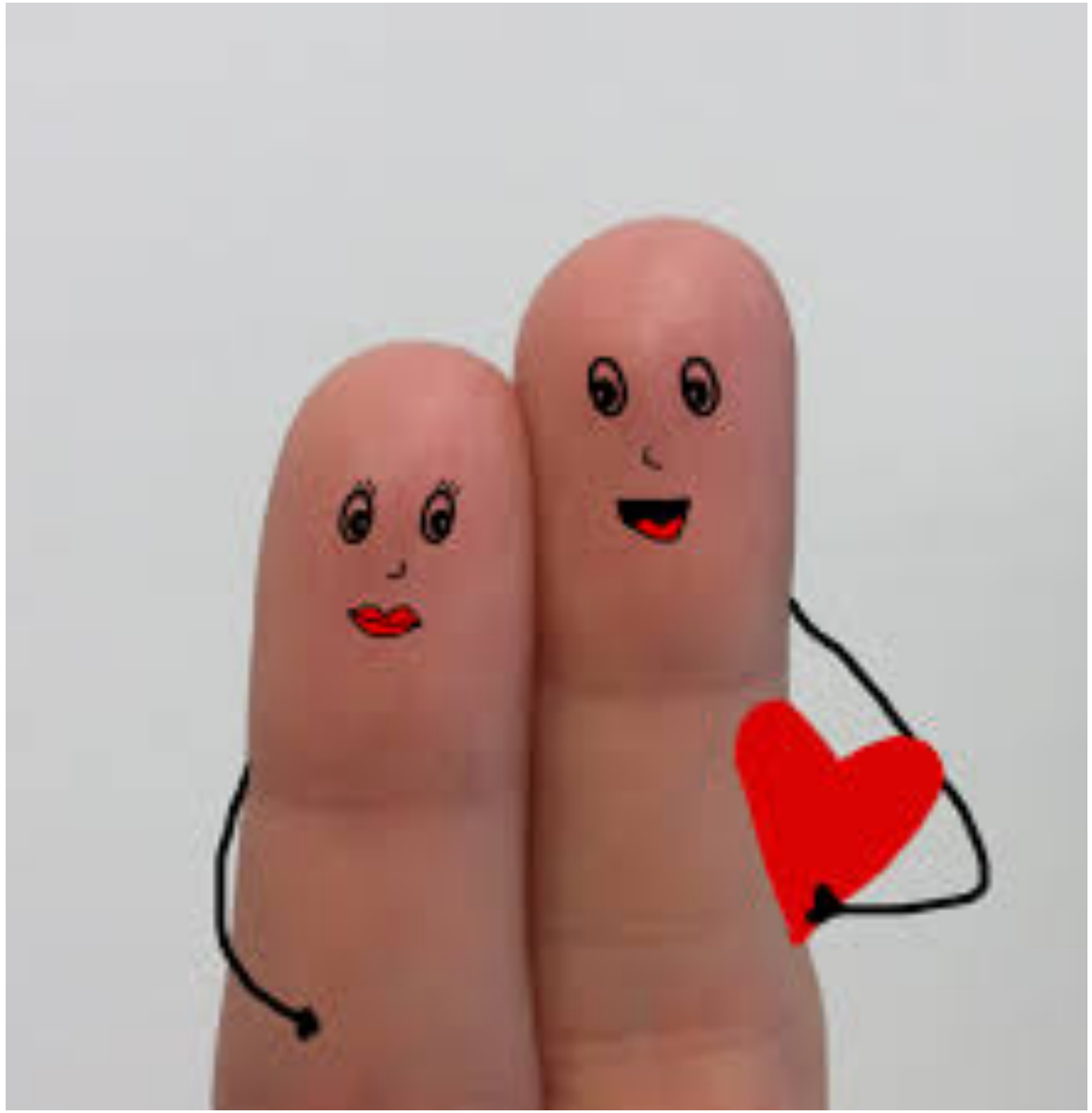
Bridging the gap

Process Mining



Process Mining

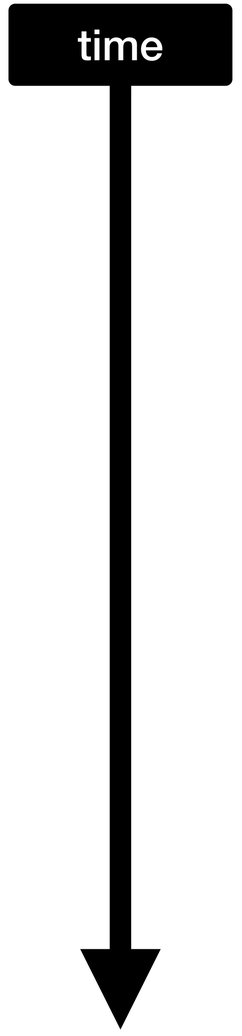
a happy relationship between models and data



Event log

https://ceur-ws.org/Vol-1859/bpmds-08-paper.pdf

Case id	Activity	Timestamp	Transaction type	Resource	…
…	…	…	…	…	
253	ER Triage	04–13-2021 11:33:50	complete	Nurse 1	…
255	Release A	04–13-2021 11:35:05	complete	Physician 02	…
259	Lactic Acid	04–13-2021 11:38:55	complete	Nurse 4	…
254	Leucocytes	04–13-2021 11:41:23	complete	Nurse 5	…
256	Lactic Acid	04–13-2021 11:52:35	complete	Nurse 4	…
257	ER Triage	04–13-2021 11:53:16	complete	Nurse 7	…
258	ER Registration	04–13-2021 11:54:47	complete	Nurse 8	…
253	Admission NC	04–13-2021 11:55:26	complete	Physician 02	…
259	Admission IC	04–13-2021 11:58:30	complete	Physician 03	…
260	CRP	04–13-2021 12:01:12	complete	Nurse 07	…
261	Release B	04–13-2021 12:02:00	complete	Physician 03	…
253	IV Liquid	04–13-2021	complete	Nurse 2	…



Event log

https://ceur-ws.org/Vol-1859/bpmds-08-paper.pdf

subject/instance | what | when | step | who | other data attributes

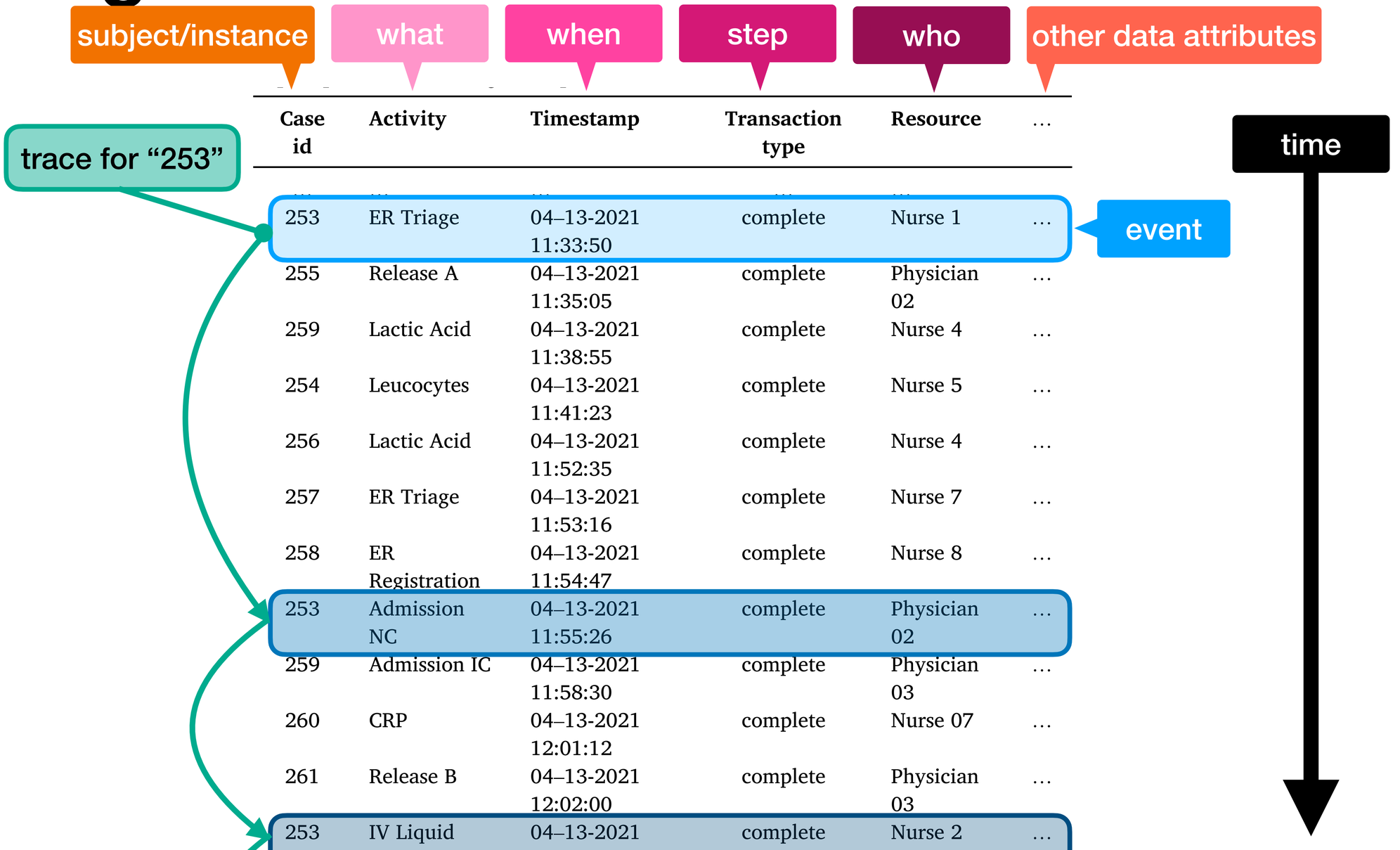
Case id	Activity	Timestamp	Transaction type	Resource	…
253	ER Triage	04–13-2021 11:33:50	complete	Nurse 1	…
255	Release A	04–13-2021 11:35:05	complete	Physician 02	…
259	Lactic Acid	04–13-2021 11:38:55	complete	Nurse 4	…
254	Leucocytes	04–13-2021 11:41:23	complete	Nurse 5	…
256	Lactic Acid	04–13-2021 11:52:35	complete	Nurse 4	…
257	ER Triage	04–13-2021 11:53:16	complete	Nurse 7	…
258	ER Registration	04–13-2021 11:54:47	complete	Nurse 8	…
253	Admission NC	04–13-2021 11:55:26	complete	Physician 02	…
259	Admission IC	04–13-2021 11:58:30	complete	Physician 03	…
260	CRP	04–13-2021 12:01:12	complete	Nurse 07	…
261	Release B	04–13-2021 12:02:00	complete	Physician 03	…
253	IV Liquid	04–13-2021	complete	Nurse 2	…

event

time

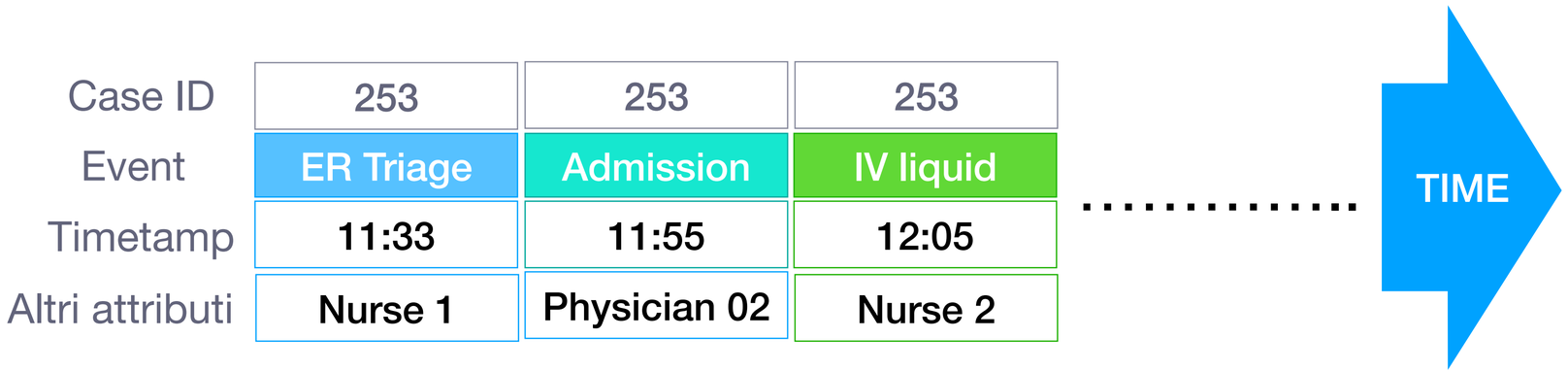
Event log

https://ceur-ws.org/Vol-1859/bpmds-08-paper.pdf



Case id	Activity	Timestamp	Transaction type	Resource	…
253	ER Triage	04–13-2021 11:33:50	complete	Nurse 1	…
255	Release A	04–13-2021 11:35:05	complete	Physician 02	…
259	Lactic Acid	04–13-2021 11:38:55	complete	Nurse 4	…
254	Leucocytes	04–13-2021 11:41:23	complete	Nurse 5	…
256	Lactic Acid	04–13-2021 11:52:35	complete	Nurse 4	…
257	ER Triage	04–13-2021 11:53:16	complete	Nurse 7	…
258	ER Registration	04–13-2021 11:54:47	complete	Nurse 8	…
253	Admission NC	04–13-2021 11:55:26	complete	Physician 02	…
259	Admission IC	04–13-2021 11:58:30	complete	Physician 03	…
260	CRP	04–13-2021 12:01:12	complete	Nurse 07	…
261	Release B	04–13-2021 12:02:00	complete	Physician 03	…
253	IV Liquid	04–13-2021	complete	Nurse 2	…

What we produce: …. a story

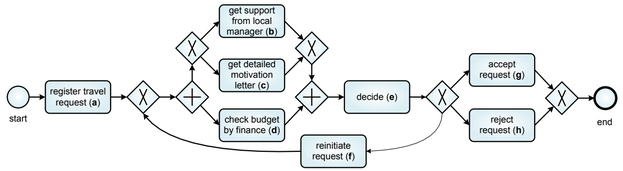


Starting point: **execution trace**

Event Log: set of execution traces

Discovery

Case	Activity	Timestamp	Resource
432	register travel request (a)	18-3-2014:9.15	John
432	get support from local manager (b)	18-3-2014:9.25	Mary
432	check budget by finance (d)	19-3-2014:8.55	John
432	decide (e)	19-3-2014:9.36	Sue
432	accept request (g)	19-3-2014:9.48	Mary



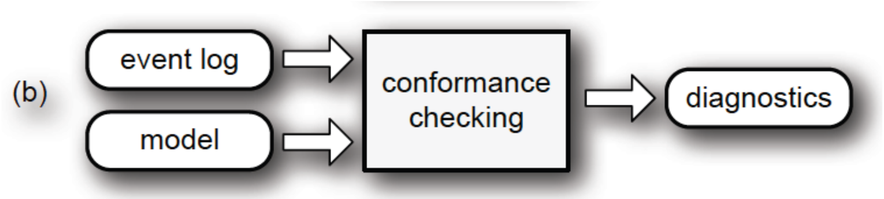
Awareness of **expected** and **unexpected** paths

credits to W.M.P. van der Aalst

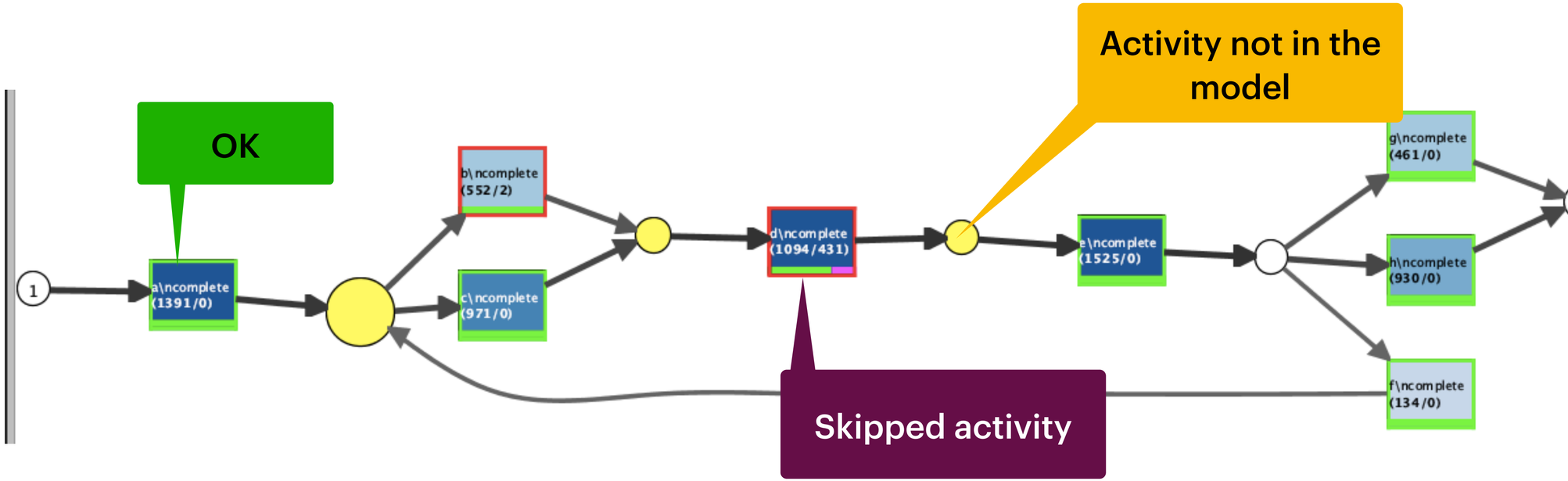
Conformance checking

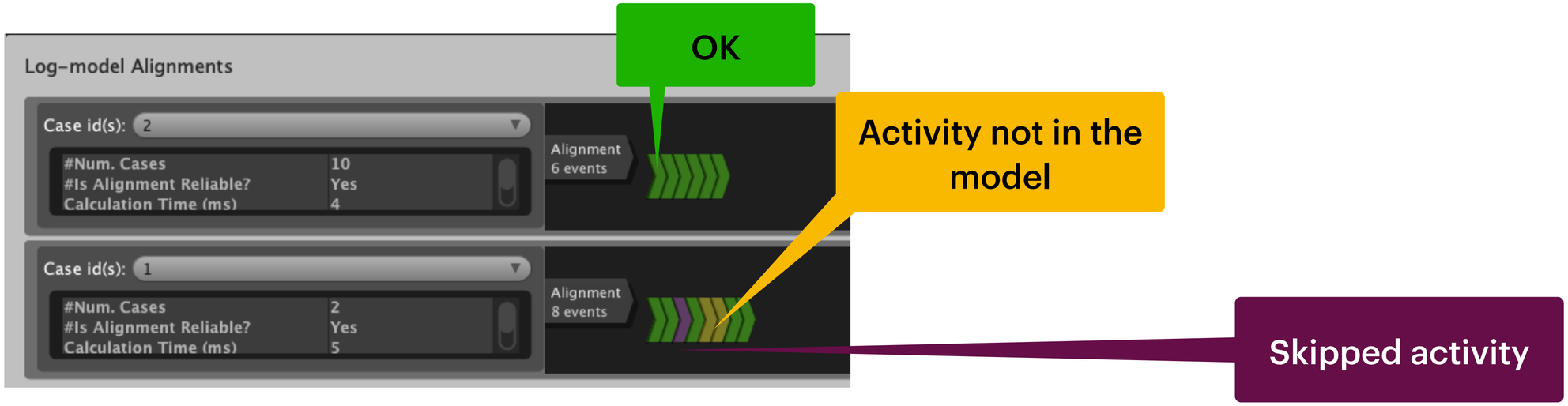
How many (and which) traces follow the **prescriptive** model?

- consciousness
- discussion
- corrective measures



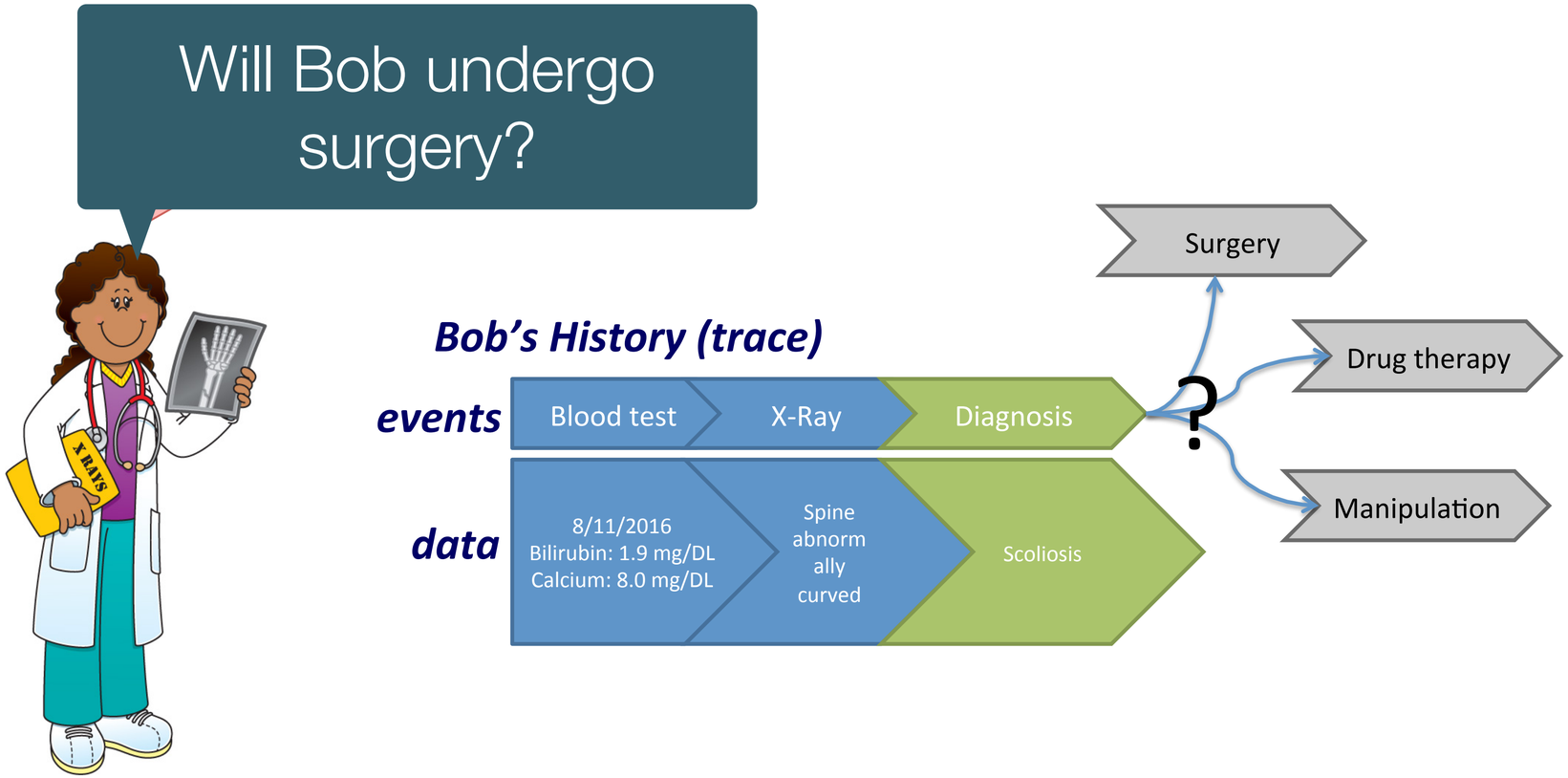
What can I do?


OK
Activity not in the model
Skipped activity
1
a\ncomplete (1391/0)
b\ncomplete (552/2)
c\ncomplete (971/0)
d\ncomplete (1094/431)
e\ncomplete (1525/0)
g\ncomplete (461/0)
h\ncomplete (930/0)
f\ncomplete (134/0)


Log–model Alignments
OK
Activity not in the model
Skipped activity
Case id(s): 2
#Num. Cases 10
#Is Alignment Reliable? Yes
Calculation Time (ms) 4
Alignment
6 events
Case id(s): 1
#Num. Cases 2
#Is Alignment Reliable? Yes
Calculation Time (ms) 5
Alignment
8 events

Operational Support

An example: predictions!



Chiara Di Francescomarino, Chiara Ghidini:
Predictive Process Monitoring. Process Mining Handbook 2022: 320-346

Tools! (Some)

Open source

Apromore

ProM
process mining workbench

Commercial vendors

ARIS software AG

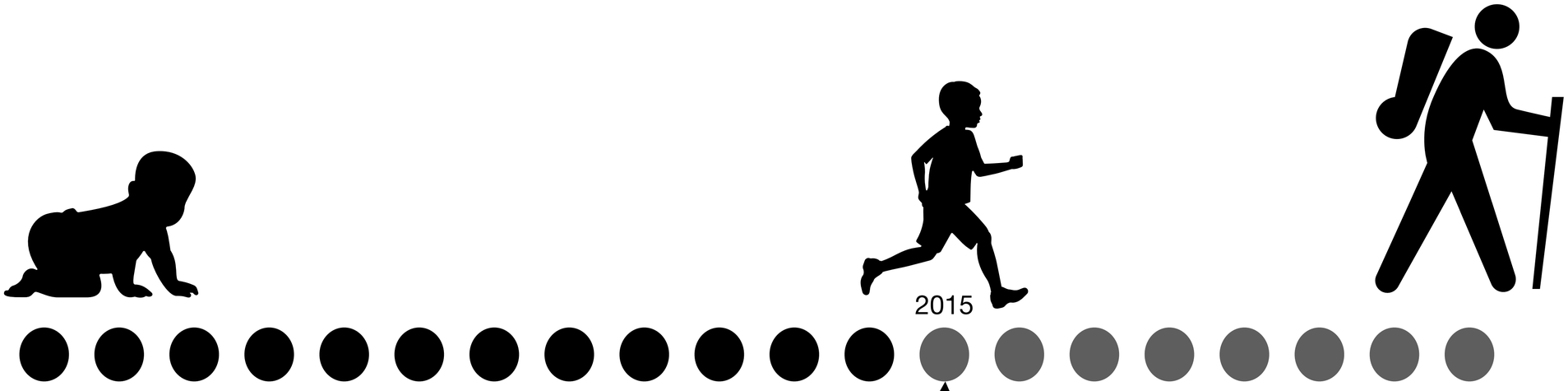
BIPOD SIAV

celonis

Disco by Fluxicon

EverFlow ICARO TECH

Why did I get intersted in Process Mining?

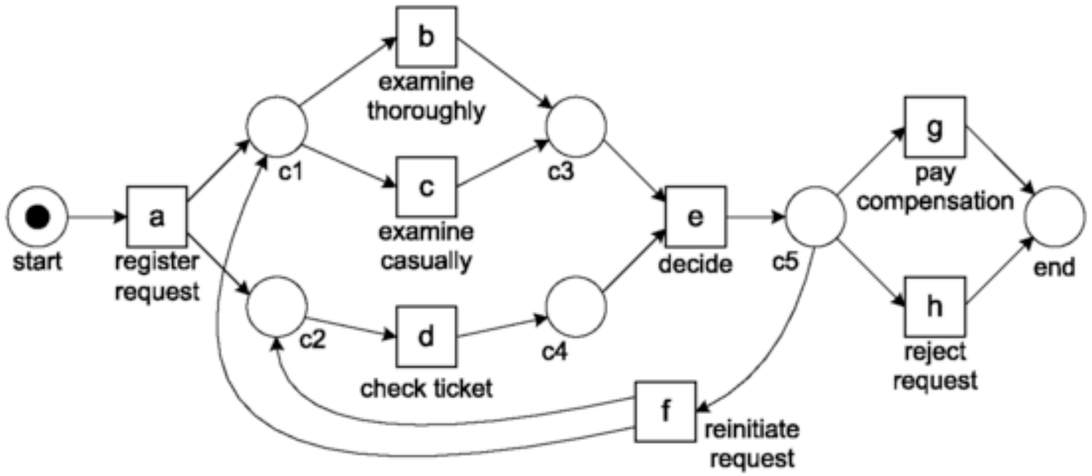


Semantic Web
Description Logics
Temporal Logic
Contextual Logics

Me!

Process Model Discovery

Process
Model



Execution
Traces

Discovery

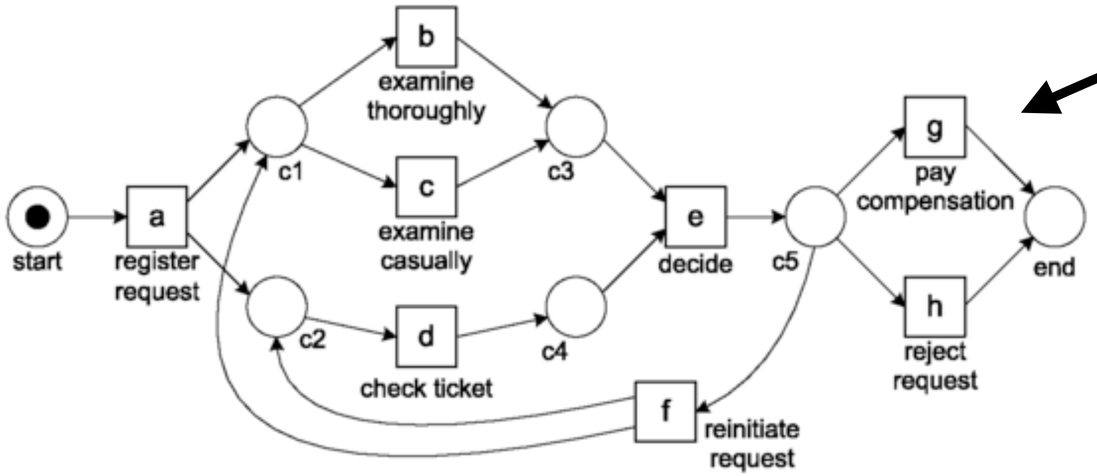
— Models & Data #1 —

in BPM data have full meaning when paired with a conceptual / formal model

Process Model Discovery

Two views of the same world

Process
Model



Execution
Traces

Discovery

and Conformance Checking

— Models & Data #2 —

in BPM conceptual / formal models have full meaning when paired with data

Discovery on multidimensional data

- Discovery of **data conditions**

  Massimiliano de Leoni, Wil M. P. van der Aalst: **Data-aware process mining**: discovering decisions in processes using alignments. SAC 2013: 1454-1461

- Discovery of **multi-entity process** models

  Dirk Fahland: Process Mining over Multiple Behavioral Dimensions with **Event Knowledge Graphs**. Process Mining Handbook 2022: 274-319

- Discover **action-response-effect** patterns

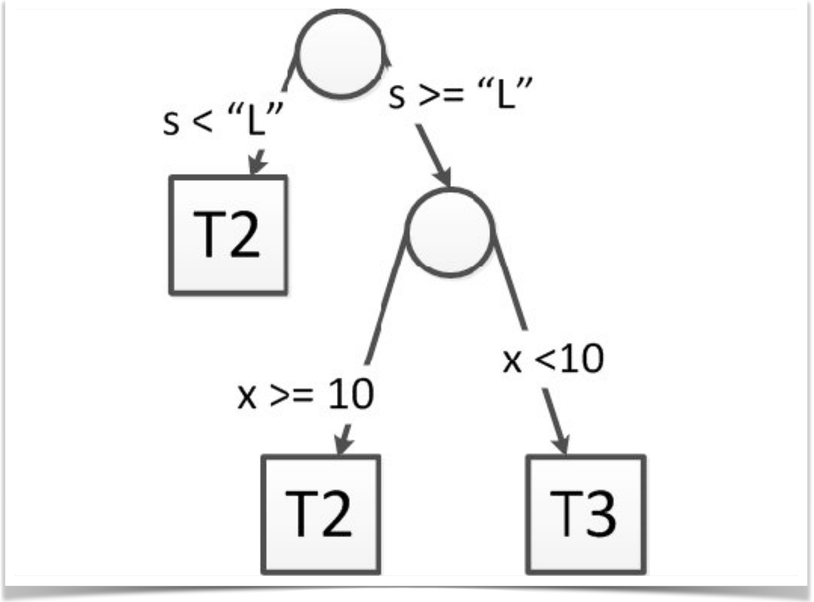
  Jelmer Jan Koorn, Xixi Lu, Henrik Leopold, Hajo A. Reijers: From **action to response to effect**: Mining statistical relations in work processes. Inf. Syst. 109: 102035 (2022)

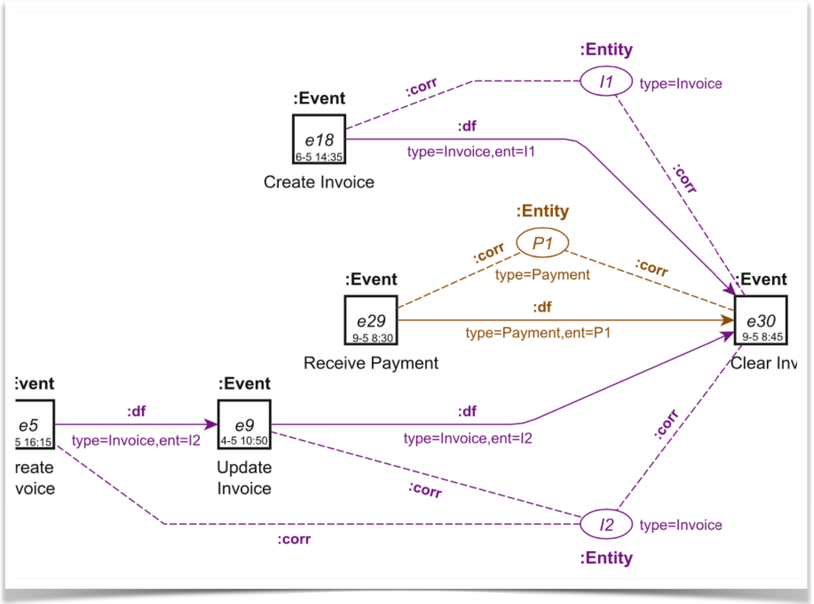
- Discovery of **probabilistic** models

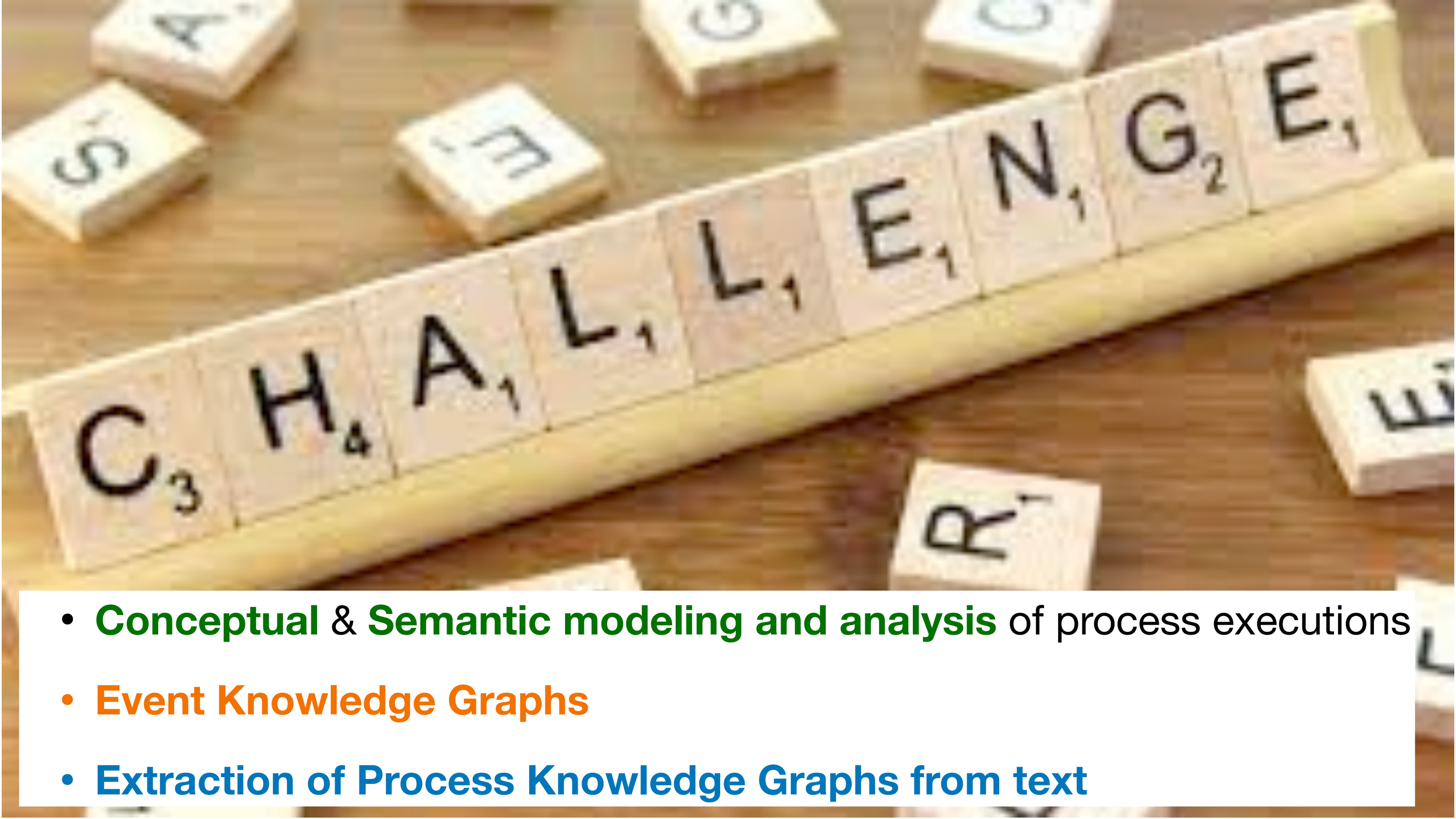
  Anti Alman, Fabrizio Maria Maggi, Marco Montali, Rafael Peñaloza: Probabilistic declarative process mining. Inf. Syst. 109: 102033 (2022)

- Discovery from **text**

  Patrizio Bellan, Mauro Dragoni, Chiara Ghidini: Extracting Business Process Entities and Relations from Text Using Pre-trained Language Models and In-Context Learning. EDOC 2022: 182-199







- **Conceptual** & **Semantic modeling and analysis** of process executions
- **Event Knowledge Graphs**
- **Extraction of Process Knowledge Graphs from text**

Discovery on multidimensional d

Data need conceptual models and conceptual m

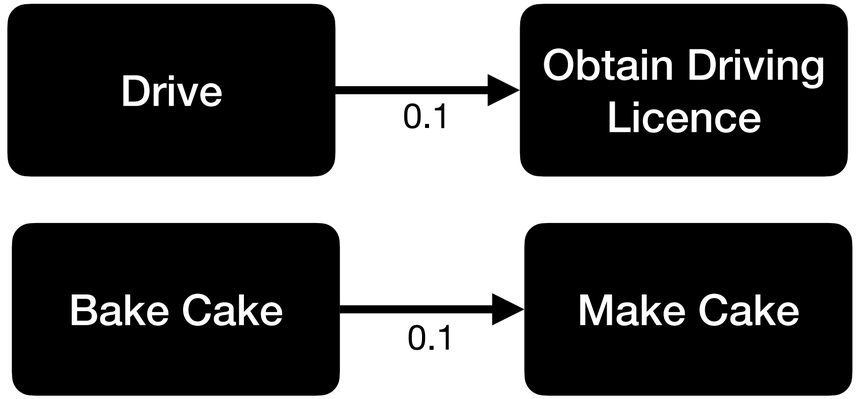
The more we discover the more we need to understand about the conceptual nature of data.

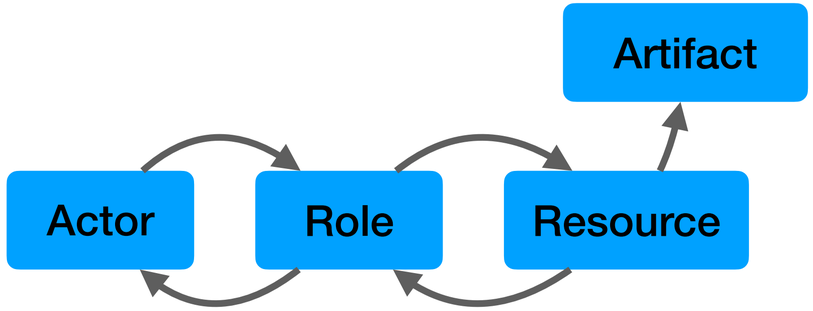
Motivations: WHY?

Kinds of dependences

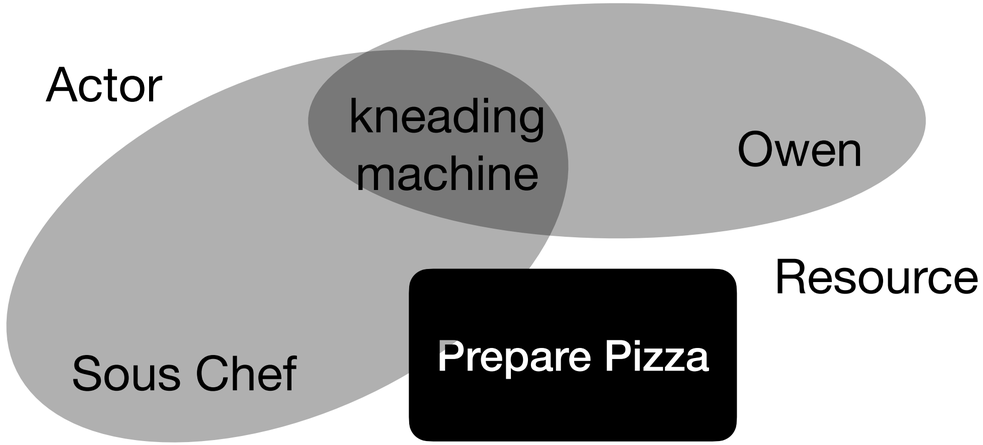
Let ***x*** and ***y*** two activities:	"Law of nature"	Goal of the process	Norms
Historical dependence "The occurrence of ***y*** presupposes that another activity ***x*** occurred ***in the past***"	BAKE PIZZA → DELIVER PIZZA	MAKE DIAGNOSIS → PROPOSE TREATMENT	LOGIN → PURCHASE GOODS
Causal dependence "***x*** causes ***y***"	SEND MESSAGE → RECEIVE MESSAGE	ORDER PIZZA DELIVERY → DELIVER PIZZA	FIRST USE SOFTWARE → EVALUATE TERMS AND CONDITIONS
Goal-based co-occurrence "The occurrence of both, ***x*** and ***y***, is necessary for the **satisfaction of *G***"		MAKE PAYMENT ↔ DELIVER PIZZA	

Greta Adamo, Chiara Di Francescomarino, Chiara Ghidini, Fabrizio Maria Maggi: Beyond arrows in process models: A user study on activity dependences and their rationales. Inf. Syst. 100: 101762 (2021)

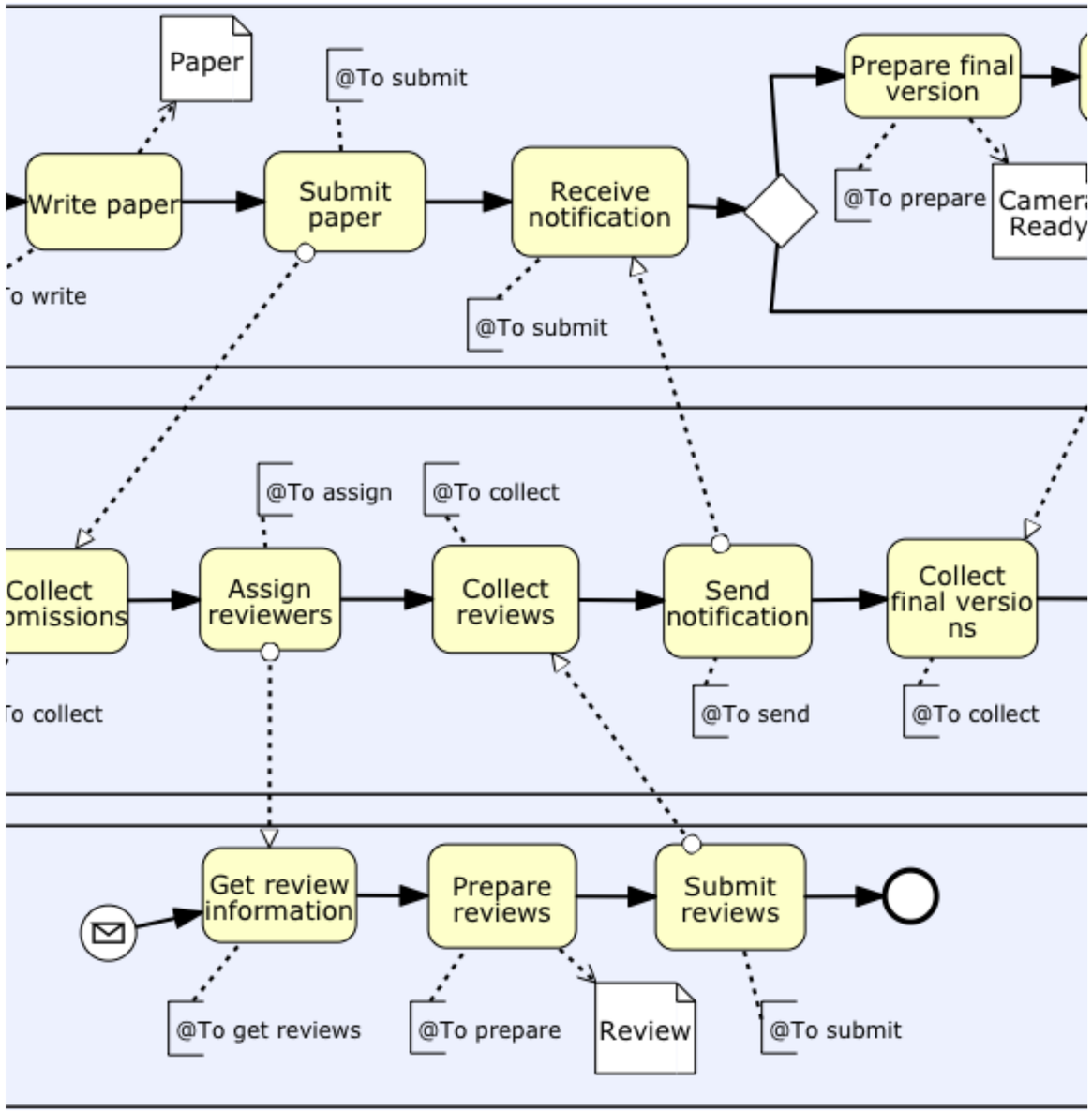




Greta Adamo, Chiara Di Francescomarino, Chiara Ghidini: Digging into Business Process Meta-models: A First Ontological Analysis. CAiSE 2020: 384-400

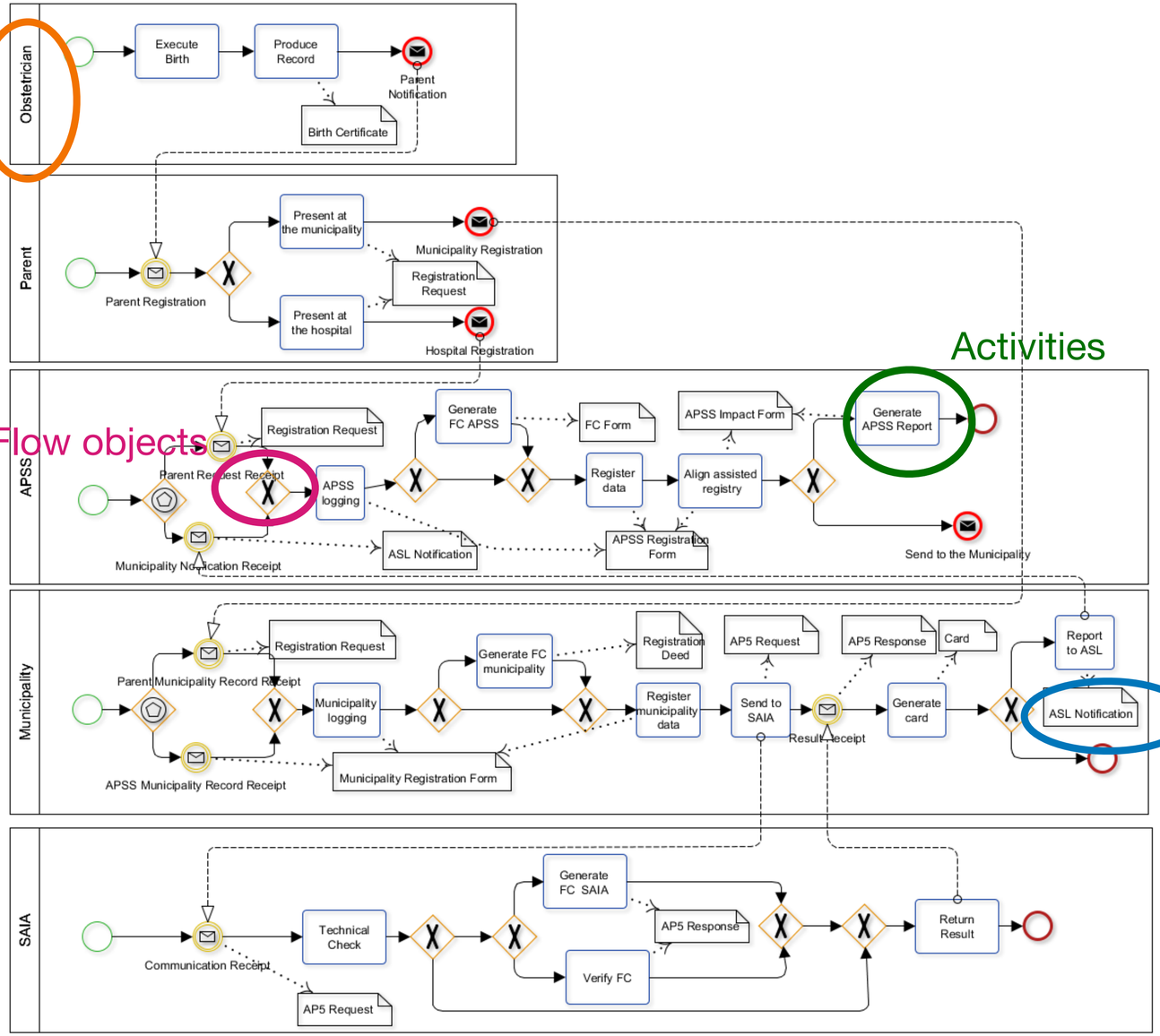


Semantic-based Process Data Analysis



An example: Birth Management Process

Roles

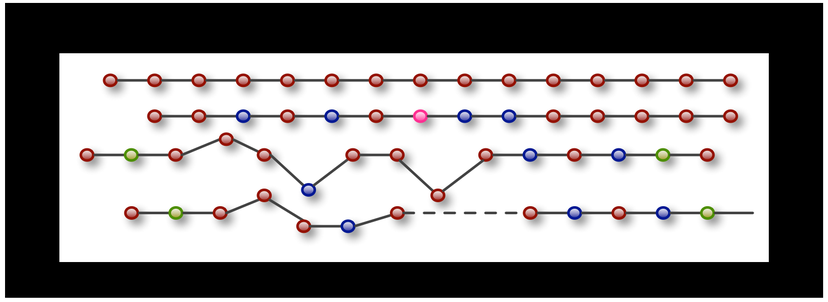


Activities

Flow objects

Documents

Execution Traces



Example analysis

Query	Description	P	K	T	Inference
Q.1	Average time per process execution spent by a specific municipality.			X	
Q.2	Total number of **Registration Request documents filled from January, 1st, 2014.**		X	X	
Q.3	**Percentage of times** in which the **flow** followed is the one which **passes first through the APSS pool and then through the municipality one.**	X		X	
Q.4	**Number of cases and average time** spent by **each public office** involved in the birth management procedure for **executing optional activities**.	X	X	X	X
Q.5	Number of times in which the municipality sends to SAIA a request without FiscalCode.	X	X	X	X
Q.6	Last event of trace TraceID.			X	
Q.7	Average time spent by trace TraceID.			X	
Q.8	Does the trace TraceID pass through the activity labeled with “PresentAtTheHospital”?	X		X	

Process Analysis

Extracts **analytical knowledge** about the **performances** of a business process starting from collected **process execution data**

Three Challenges

- **Challenge 1**: Combining **three** different dimensions.
  - D1: the **procedural** dimension (P)
  - D2: the **domain** of interest (K)
  - D3: the **execution** dimension (T)
- **Challenge 2**: Semantic Reasoning
- **Challenge 3**: Scalability

Semantic Process Analysis

A possible solution!

Employs **Semantic Web techniques** that leverage the explicit formalization of the semantics of a business process and the data it manipulates

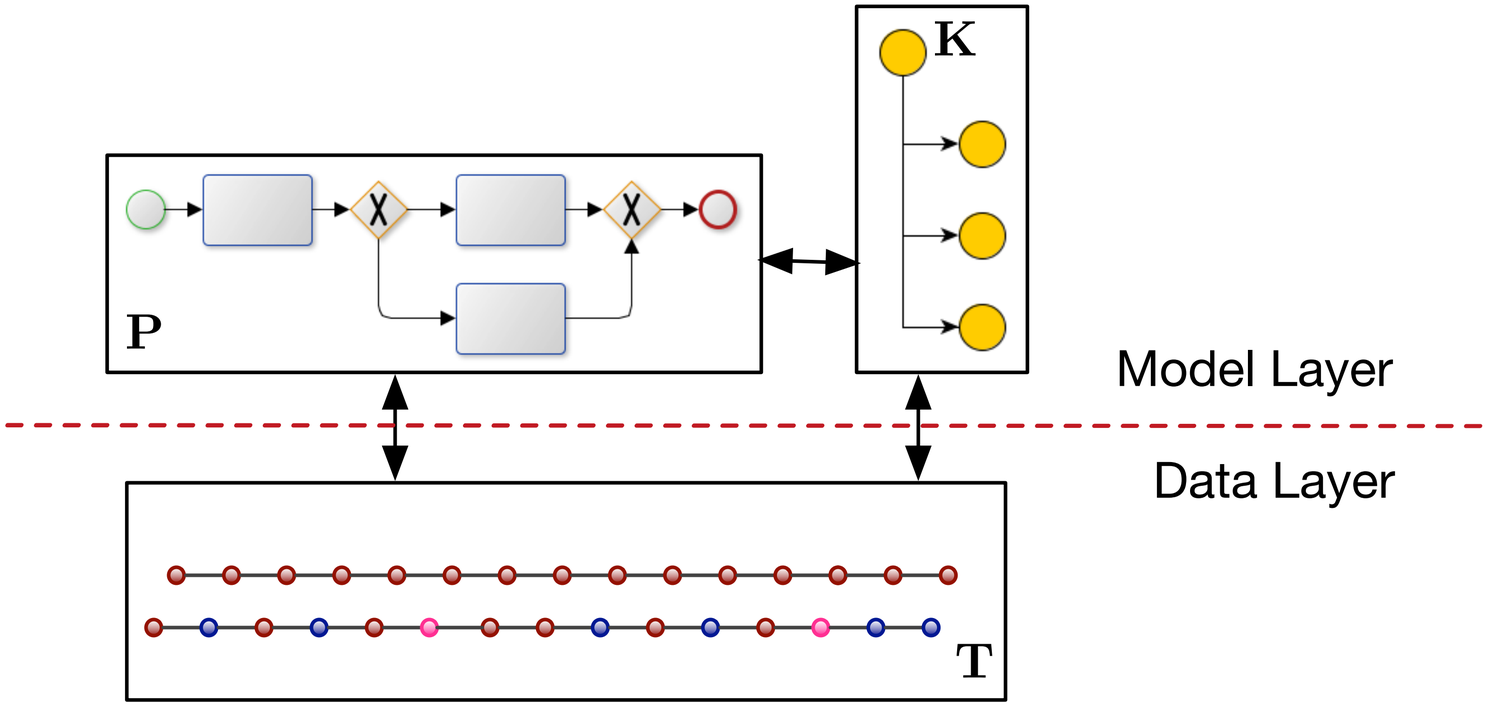
Our approach / contributions:

- Challenge 1: **Integrated OWL 2 / RDF model of P + K + T queried with SPARQL**
- Challenge 2: **OWL 2 reasoning** for making explicit inferrable knowledge
- Challenge 3: Implementation based on **SW triplestores**

The Proposed Model: an Integrated View

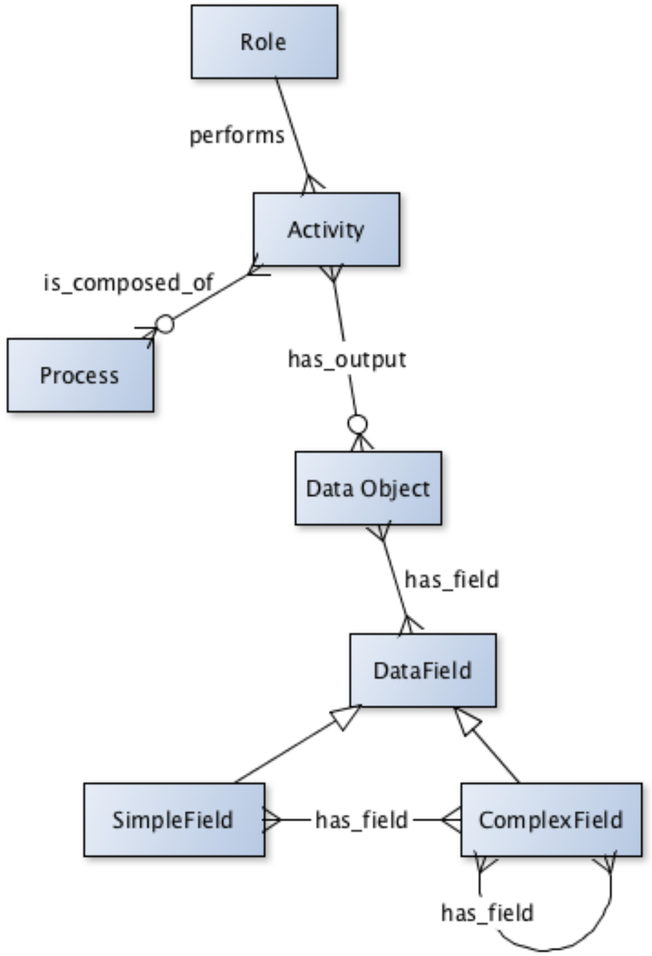
Reconciliation of knowledge and information related to different dimensions:

- BPMN Ontology
- Domain Ontology
- Trace Ontology

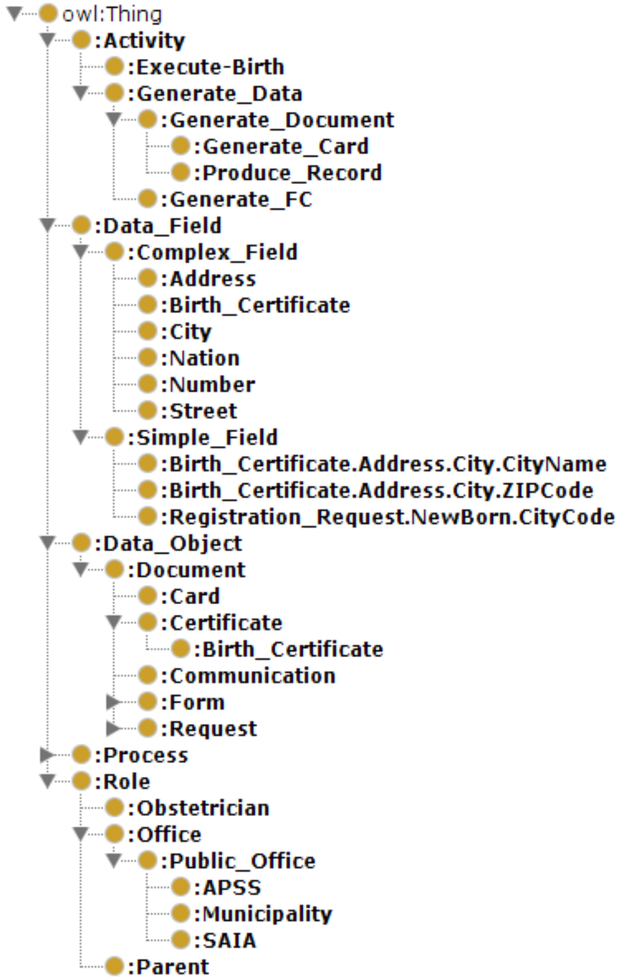


Piergiorgio Bertoli, Francesco Corcoglioniti, Chiara Di Francescomarino, Mauro Dragoni, Chiara Ghidini, Marco Pistore:
Semantic modeling and analysis of complex data-aware processes and their executions. Expert Syst. Appl. 198: 116702 (2022)

The Domain Ontology

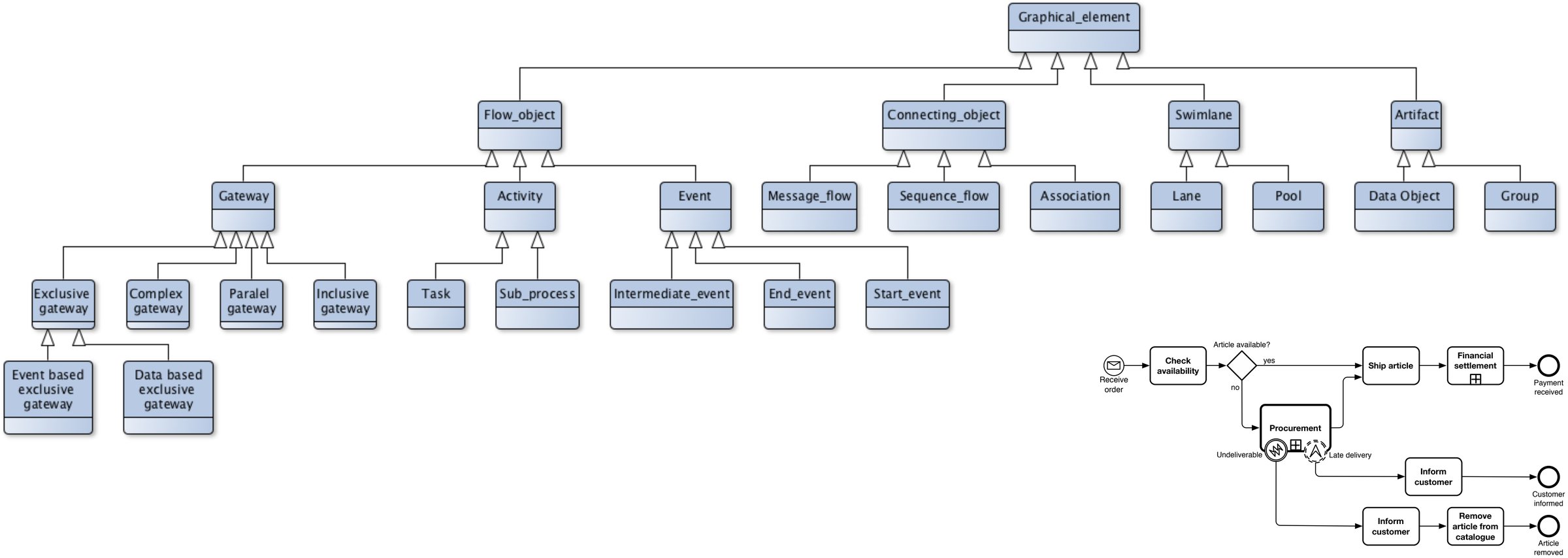


The core



The specific

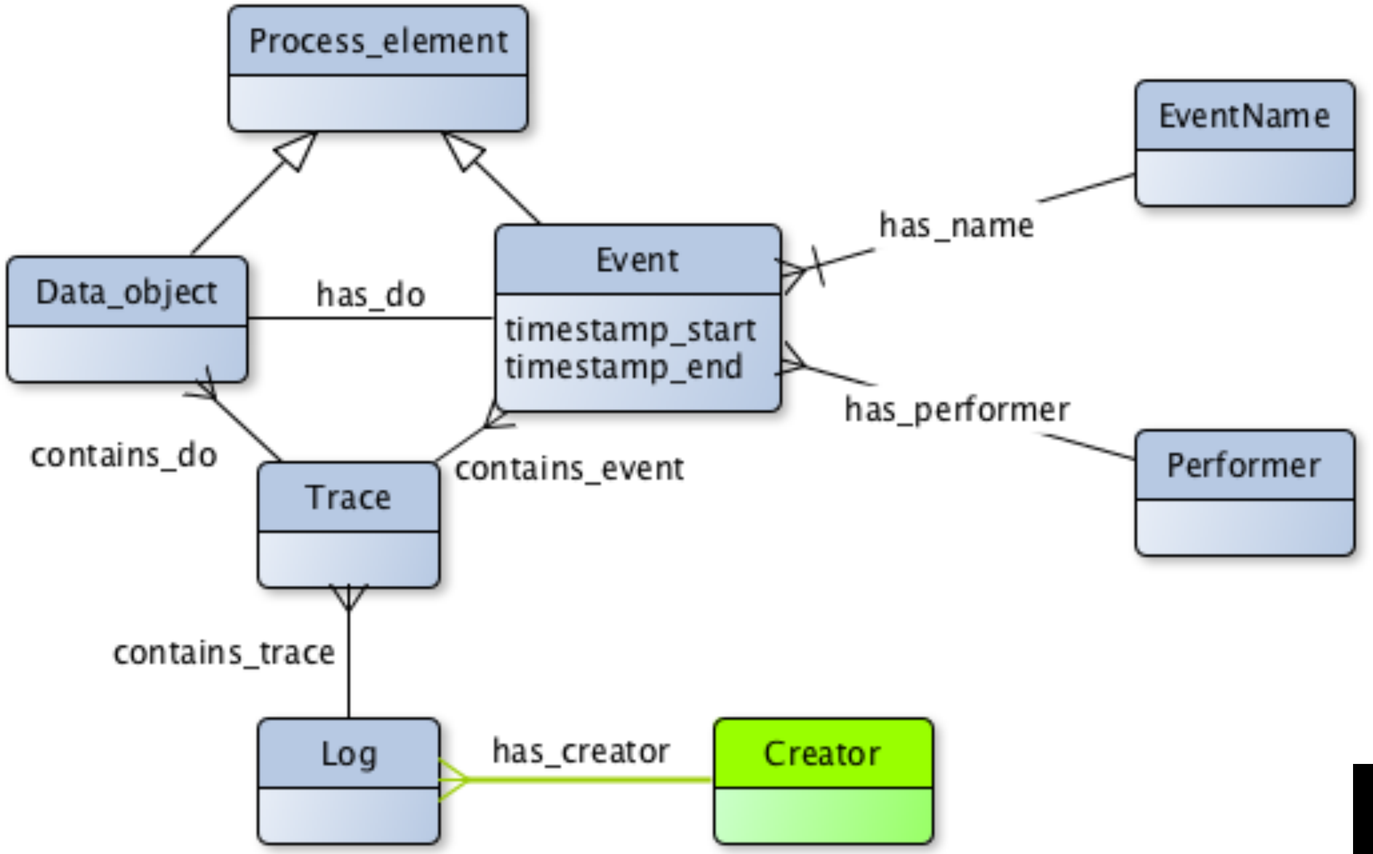
The BPMN Ontology



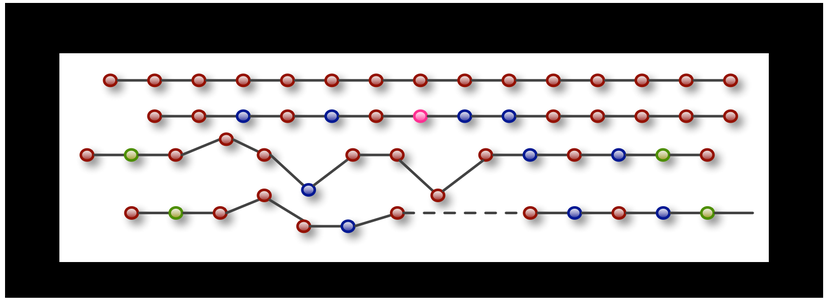
Rospocher, M., Ghidini, C., Serafini, L.: An ontology for the business process modelling notation.
In: 8th International Conference on Formal Ontology in Information Systems (FOIS 2014), 22-25 September 2014, Rio de Janeiro, Brazil. (2014)

Chiara Di Francescomarino, Chiara Ghidini, Marco Rospocher, Luciano Serafini, Paolo Tonella:
Semantically-Aided Business Process Modeling. ISWC 2009: 114-129

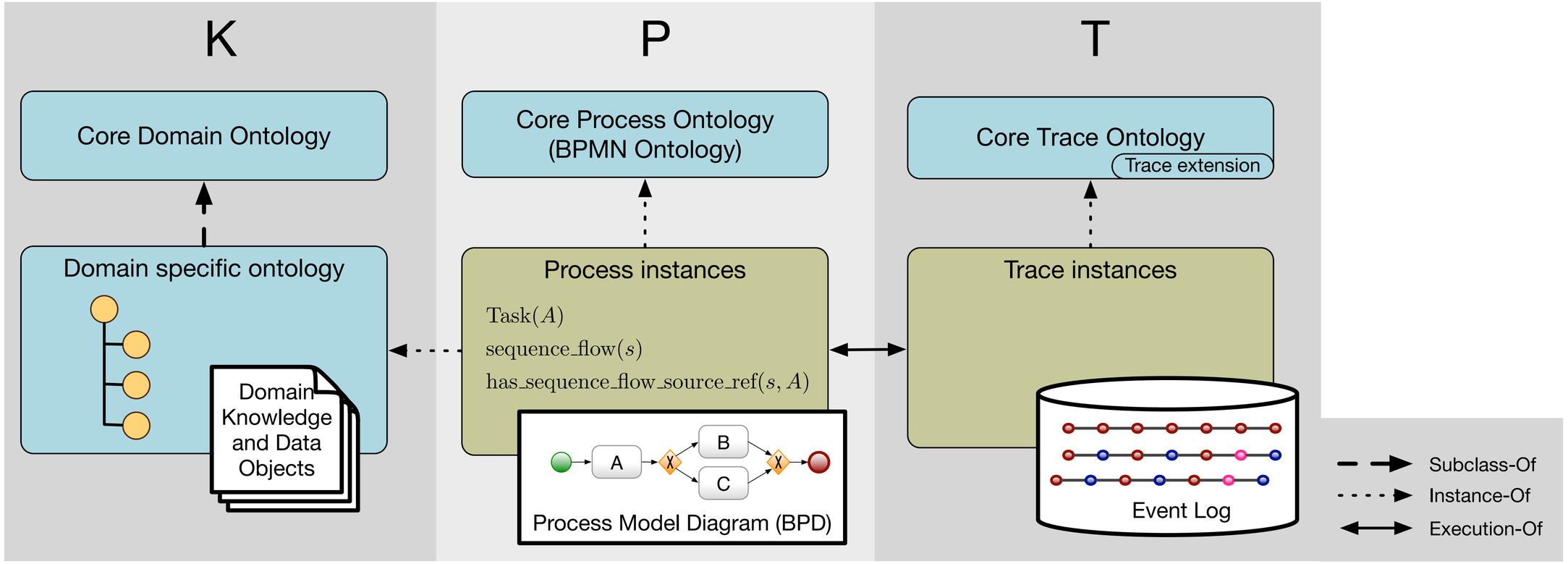
The Trace Ontology



Execution Traces



The Integrated Ontological Model

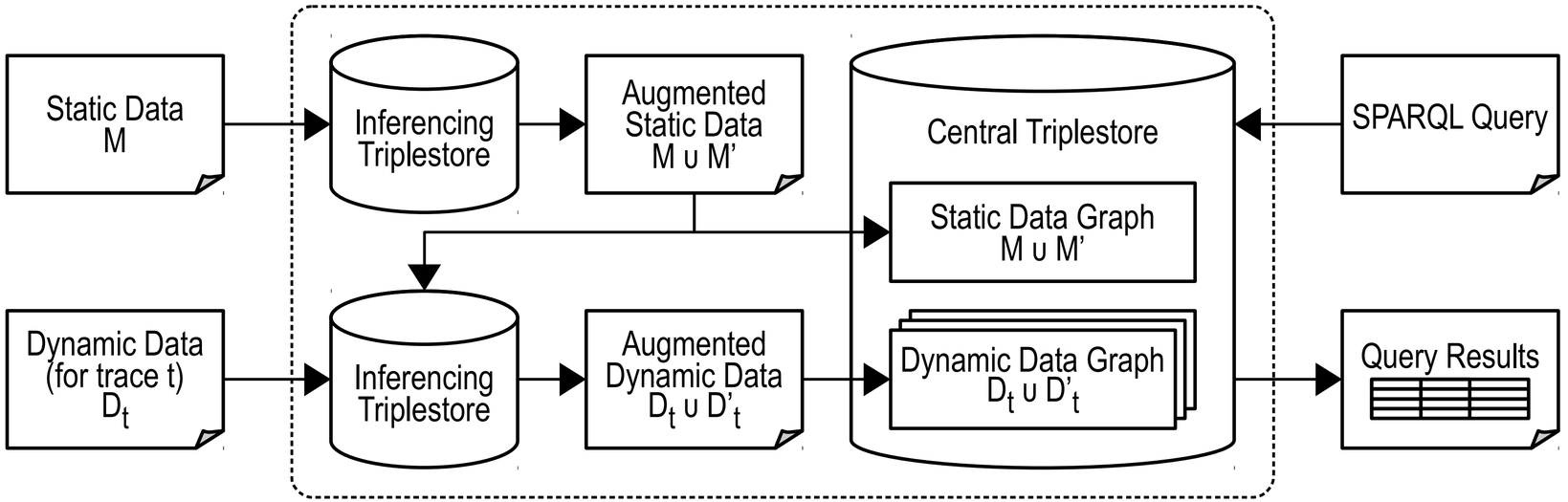


The Architectural Solution

Challenges to cope:

- collect trace data at fast rate
- answer to complex queries

Investigated **solution**: architecture based on **triplestores**



Semantic Web technologies work!

Traces	Stored triples			Storing		Querying	
	Asserted	Inferred	Total	Throughput	Total time	Avg. Time Q.4	Avg. Time Q.8
1500	3062349	1895471	4957820	37.89 trace/min	2426.88 s	324 ms	41.4 ms
10500	21910269	13057464	34967773	37.41 trace/min	16851.21 s	881.4 ms	26.2 ms
42000	87503538	52045200	139548738	37.34 trace/min	67537.95 s	4510.0 ms	105.0 ms

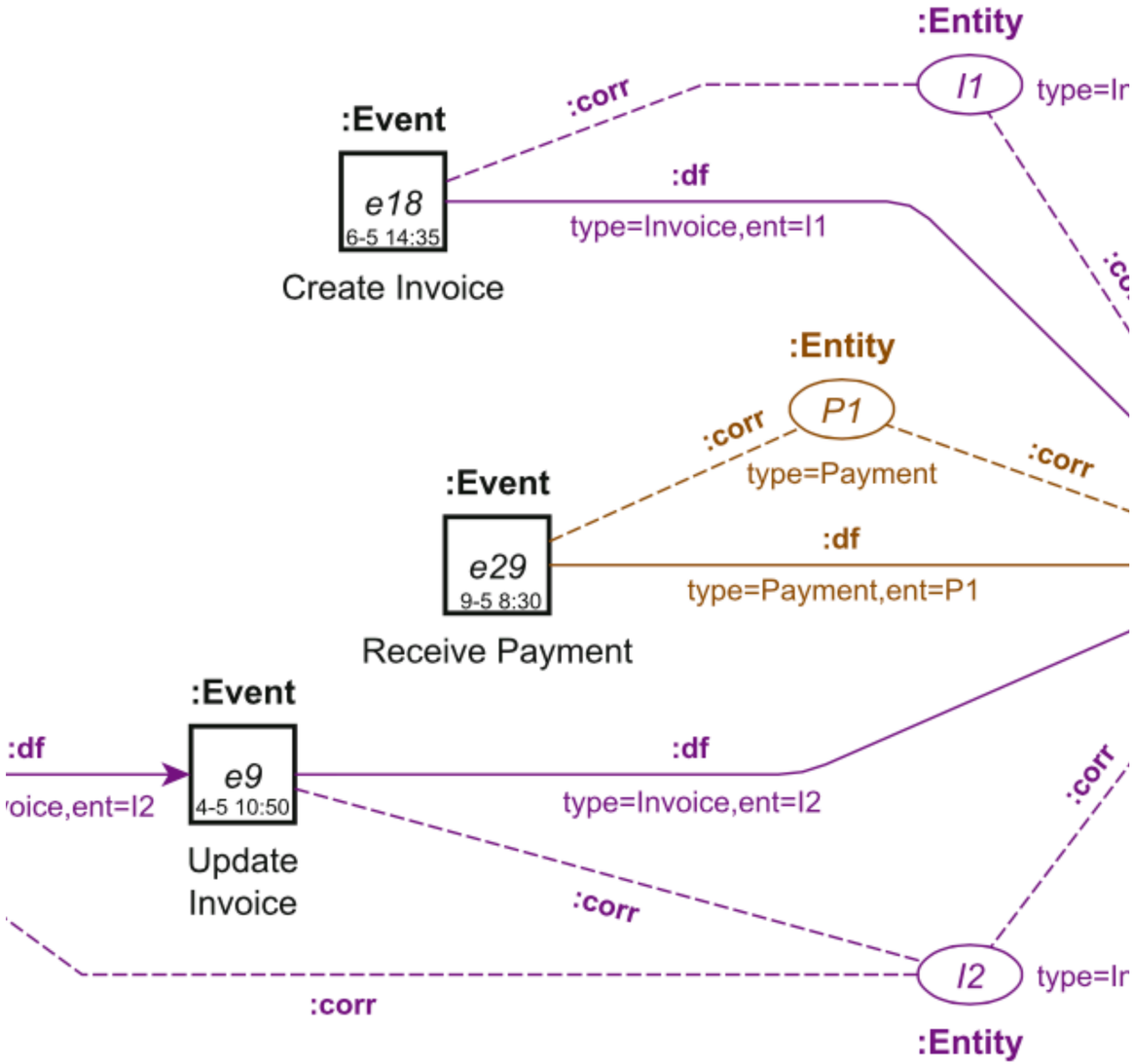
Daily, weekly, and monthly load.

Throughput independent of the load

Time required for queries is acceptable for real-time usage

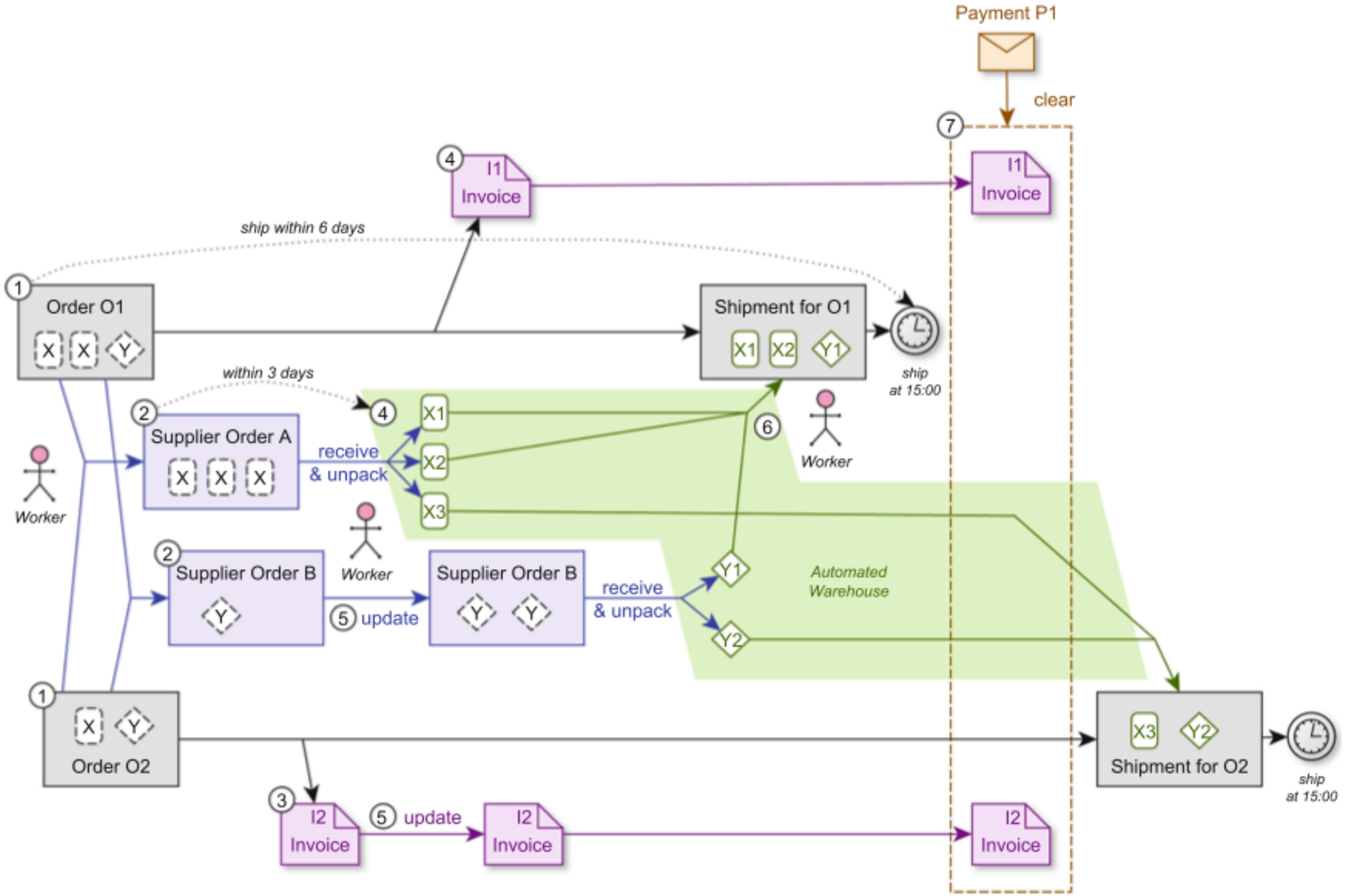
Piergiorgio Bertoli, Francesco Corcoglioniti, Chiara Di Francescomarino, Mauro Dragoni, Chiara Ghidini, Marco Pistore: Semantic modeling and analysis of complex data-aware processes and their executions. Expert Syst. Appl. 198: 116702 (2022)

Event Knowledge Graphs



Dirk Fahland: Process Mining over Multiple Behavioral Dimensions with **Event Knowledge Graphs**. Process Mining Handbook 2022: 274-319

Multi-entity Processes



Consider a retailer who took two Orders for multiple Items from the same customer: the customer first places Order O1 for 2 items X and 1 item Y , and shortly afterwards Order O2 for 1 item X and 1 item Y ….

Credits to Dirk Fahland

Multi-entity Processes

Table 1. Event table of events underlying the event log of Table 2.

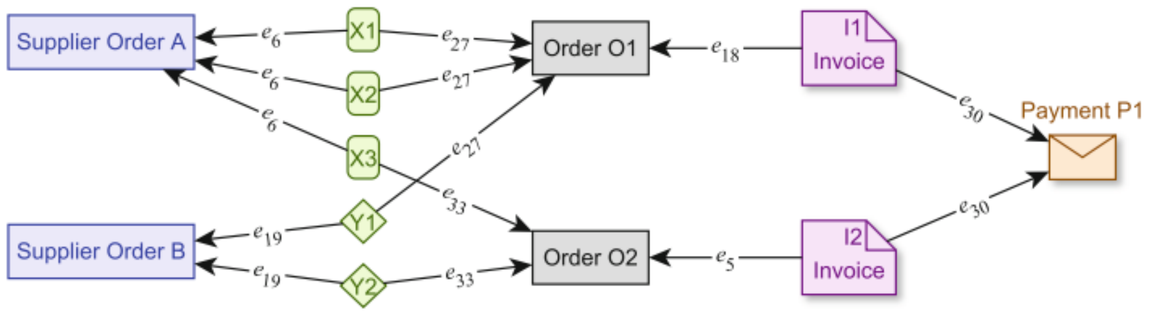
EventID	Activity	Time	Actor	Order	Supplier Order	Order Details	Item	Invoice	Payment
e1	Create Order	01-05 09:05	R1	O1		2·X, 1·Y			
e2	Create Order	01-05 09:30	R1	O2		1·X, 1·Y			
e3	Place SO	01-05 11:25	R1		A	3·X			
e4	Place SO	02-05 11:55	R3		B	1·Y			
e5	Create Invoice	03-05 16:15	R3	O2				I2	
e6	Receive SO	00-01 10:00	R2		A		X1,X2,X3		
e7	Update SO	04-05 10:25	R1	O2	B	2·Y			
e8	Unpack	00-01 10:30	R2		A		X3		
e9	Update Invoice	04-05 10:50	R2					I2	
e10	Unpack	04-05 11:00	R2		A		X1		
e11	Unpack	04-05 11:15	R2		A		X2		
e18	Create Invoice	06-05 14:35	R3	O1				I1	
e19	Receive SO	07-05 10:10	R2		B		Y1,Y2		
e20	Unpack	07-05 10:45	R2		B		Y1		
e21	Unpack	07-05 11:00	R2		B		Y2		
e27	Pack Shipment	07-05 17:00	R4	O1			X1,X2,Y1		
e28	Ship	08-05 15:00	R4	O1					
e29	Receive Payment	09-05 08:30	R5						P1
e30	Clear Invoice	09-05 08:45	R5					I1,I2	P1
e33	Pack Shipment	09-05 11:45	R4	O2			X3,Y2		
e34	Ship	09-05 15:00	R4	O2					

How do I translate these data into linear traces?

What do I lose?

Credits to Dirk Fahland

Multi-entity Processes



Events relate to objects!

Table 1. Event table of events underlying the event log of Table 2.

EventID	Activity	Time	Actor	Order	Supplier Order	Order Details	Item	Invoice	Payment
e1	Create Order	01-05 09:05	R1	O1		2·X, 1·Y			
e2	Create Order	01-05 09:30	R1	O2		1·X, 1·Y			
e3	Place SO	01-05 11:25	R1		A	3·X			
e4	Place SO	02-05 11:55	R3		B	1·Y			
e5	Create Invoice	03-05 16:15	R3	O2				I2	
e6	Receive SO	00-01 10:00	R2		A		X1,X2,X3		
e7	Update SO	04-05 10:25	R1	O2	B	2·Y			
e8	Unpack	00-01 10:30	R2		A		X3		
e9	Update Invoice	04-05 10:50	R2					I2	
e10	Unpack	04-05 11:00	R2		A		X1		
e11	Unpack	04-05 11:15	R2		A		X2		
e18	Create Invoice	06-05 14:35	R3	O1				I1	
e19	Receive SO	07-05 10:10	R2		B		Y1,Y2		
e20	Unpack	07-05 10:45	R2		B		Y1		
e21	Unpack	07-05 11:00	R2		B		Y2		
e27	Pack Shipment	07-05 17:00	R4	O1			X1,X2,Y1		
e28	Ship	08-05 15:00	R4	O1					
e29	Receive Payment	09-05 08:30	R5						P1
e30	Clear Invoice	09-05 08:45	R5					I1,I2	P1
e33	Pack Shipment	09-05 11:45	R4	O2			X3,Y2		
e34	Ship	09-05 15:00	R4	O2					

Credits to Dirk Fahland

Event Knowledge Graphs

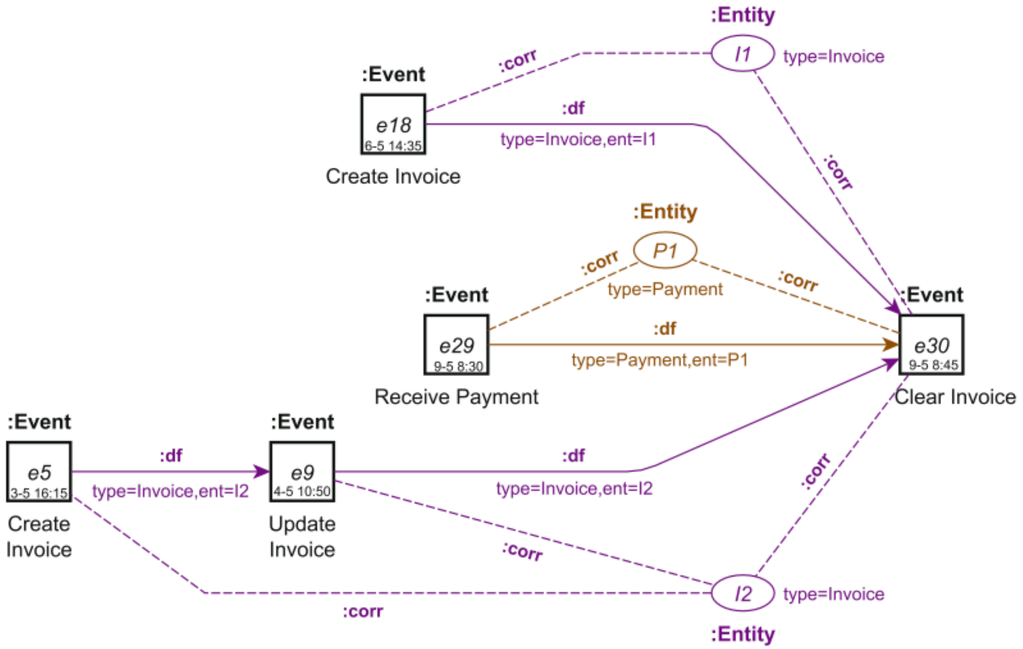


Table 1. Event table of events underlying the event log of Table 2.

EventID	Activity	Time	Actor	Order	Supplier Order	Order Details	Item	Invoice	Payment
e1	Create Order	01-05 09:05	R1	O1		2·X, 1·Y			
e2	Create Order	01-05 09:30	R1	O2		1·X, 1·Y			
e3	Place SO	01-05 11:25	R1		A	3·X			
e4	Place SO	02-05 11:55	R3		B	1·Y			
e5	Create Invoice	03-05 16:15	R3	O2				I2	
e6	Receive SO	00-01 10:00	R2		A		X1,X2,X3		
e7	Update SO	04-05 10:25	R1	O2	B	2·Y			
e8	Unpack	00-01 10:30	R2		A		X3		
e9	Update Invoice	04-05 10:50	R2					I2	
e10	Unpack	04-05 11:00	R2		A		X1		
e11	Unpack	04-05 11:15	R2		A		X2		
e18	Create Invoice	06-05 14:35	R3	O1				I1	
e19	Receive SO	07-05 10:10	R2		B		Y1,Y2		
e20	Unpack	07-05 10:45	R2		B		Y1		
e21	Unpack	07-05 11:00	R2		B		Y2		
e27	Pack Shipment	07-05 17:00	R4	O1			X1,X2,Y1		
e28	Ship	08-05 15:00	R4	O1					
e29	Receive Payment	09-05 08:30	R5						P1
e30	Clear Invoice	09-05 08:45	R5					I1,I2	P1
e33	Pack Shipment	09-05 11:45	R4	O2			X3,Y2		
e34	Ship	09-05 15:00	R4	O2					

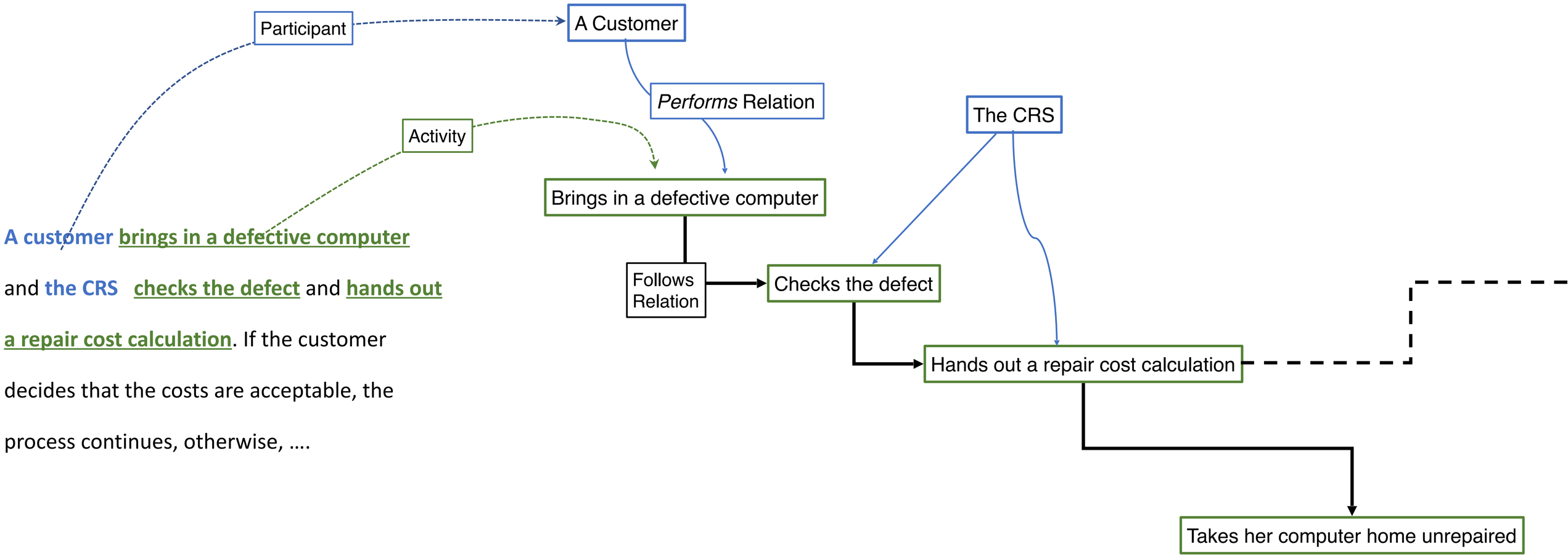
Objects “justify” follows relations and correlate to events

Credits to Dirk Fahland

Assisted Process Knowledge Graph Building Using Pre-Trained Language Models

Patrizio Bellan, Mauro Dragoni, Chiara Ghidini: Extracting Business Process Entities and Relations from Text Using Pre-trained Language Models and In-Context Learning. EDOC 2022: 182-199

Process information extraction from text



Process information extraction from text – The Challenges

1. **NO Data Available**
2. **Challenging Entities**

Factual event

The concert of Pink Floyd was in Venice

Extraction of instances

VS

The candidate sends the application to the HR office by email

The candidate **sends** the application to the HR office by email

The candidate **sends the application** to the HR office by email

The candidate **sends the application** to the HR office **by email**

The candidate sends the application to the HR office **by email**

…..

The candidate sends the application to the HR office by email

N.B. *"talk to potential investors"* IS a goal and NOT an activity

Why pre-trained language models?

1. **NO Data Available**: pre-trained language models solve data problem
2. **Challenging Entities**: They contain the conceptual representation we need

Problem Solved! ????

Why pre-trained language models?

1. **NO Data Available**: pre-trained language models solve data problem
2. **Challenging Entities**: They contain the conceptual representation we need

Problem Solved! ???? **NO**

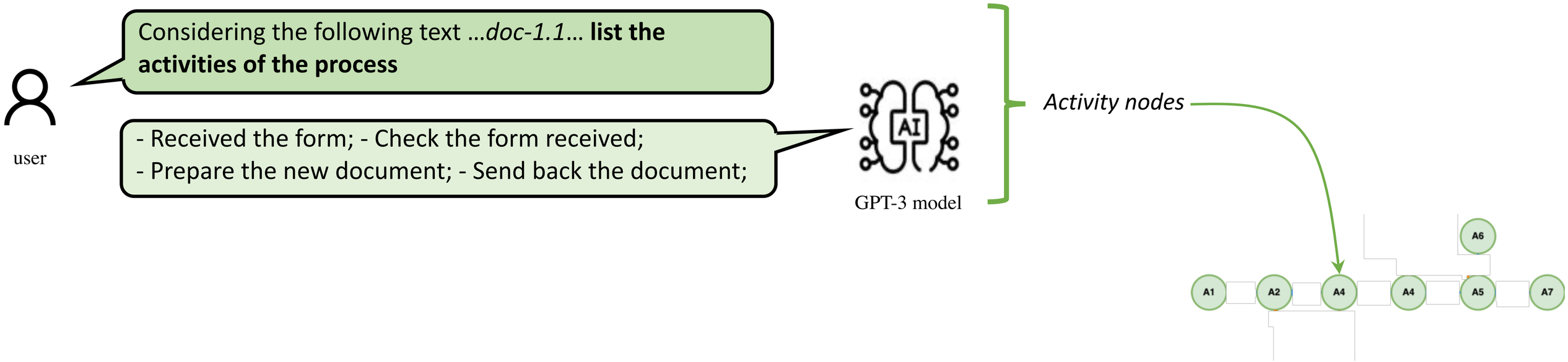
We tried to use Zero-shot, but it does not work!!!

	Precision	Recall	F1 score
Activity	0.43	0.25	0.32
Follows	0.00	0.00	0.00
Actor	0.38	0.37	0.36
Performs	0.27	0.43	0.32

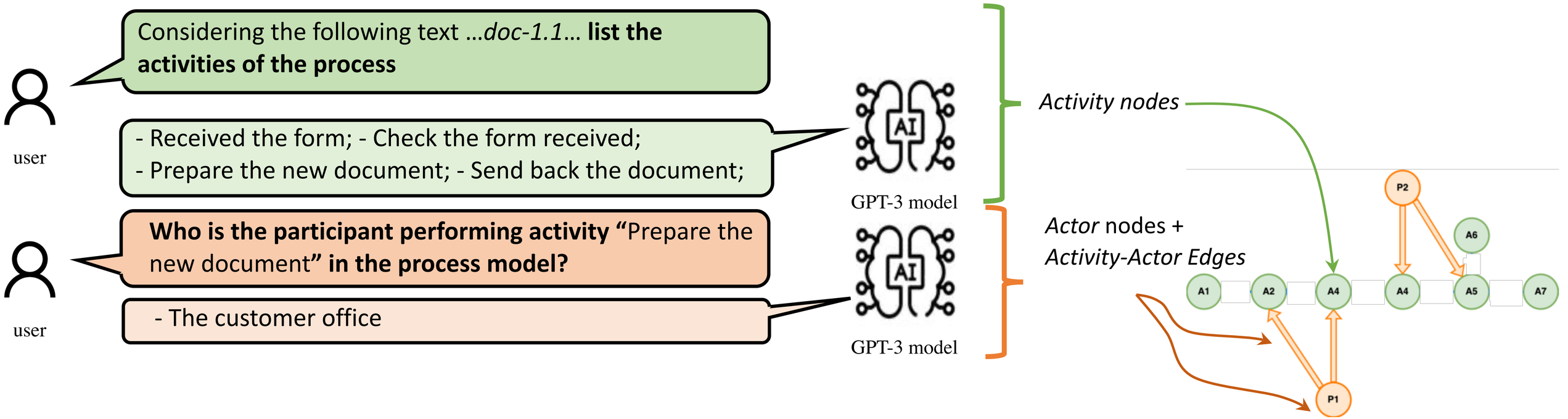
Challenges for using generative pre-trained language models

1. How to pose **questions**
2. How to build the input (**prompt** construction)
3. How to **evaluate** results

1- How to pose questions – Step 1 – Activity Node



1- How to pose questions – Step 2 – Actor Node and Performs Relation



1- How to pose questions – Step 3 – Follows Relation

user

Considering the following text …*doc-1.1*… **list the activities of the process**

GPT-3 model

- Received the form; - Check the form received;
- Prepare the new document; - Send back the document;

Activity nodes

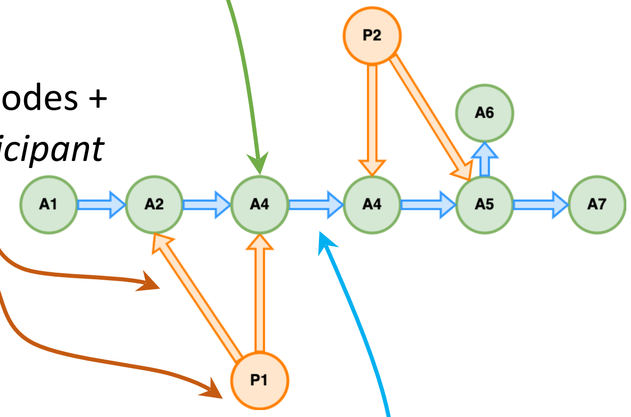
user

Who is the participant performing activity “Prepare the new document” **in the process model?**

GPT-3 model

- The customer office

Participant nodes + *Activity-Participant Edges*



user

Considering the list of process activity described in the text, does activity “Received the form” **immediately follow activity** “Prepare the new document” **in the process model?**

GPT-3 model

no

Activity edg

Good results for activity, actors, and performing relation. Still work to do with temporal relations and gateways.

Summing up!

- Work-Processes
- Process Mining
- …… and I
- **Graphs, Semantics, Knowledge Extraction & Processes**
  - Semantic modeling and analysis of process executions
  - Event Knowledge Graphs
  - Extraction of Process Knowledge Graphs from text