

Data Quality in Process Mining



Sareh Sadeghianasl
Queensland University of Technology (QUT), Australia

About Me

- BSc and MSc of Computer Software Engineering from FUM
- PhD candidate in Event Log Data Quality at QUT
- Experienced in process mining and data quality projects from Australian organizations
- BPM@QUT
 - Event Data Quality
 - Process Mining
 - Robotic Process Automation



Sareh Sadeghianasl

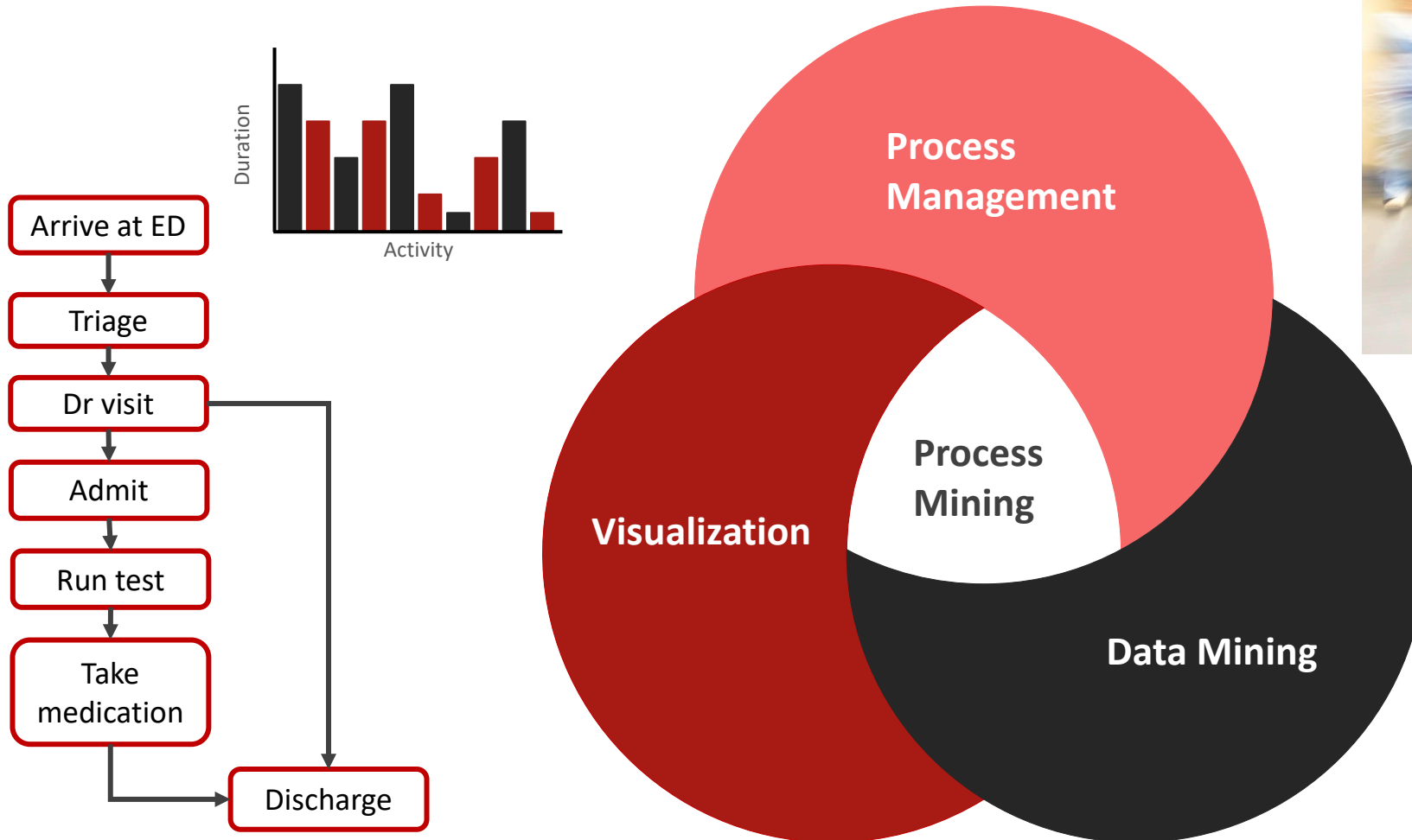
Researcher in Process Mining and
Data Quality

Agenda

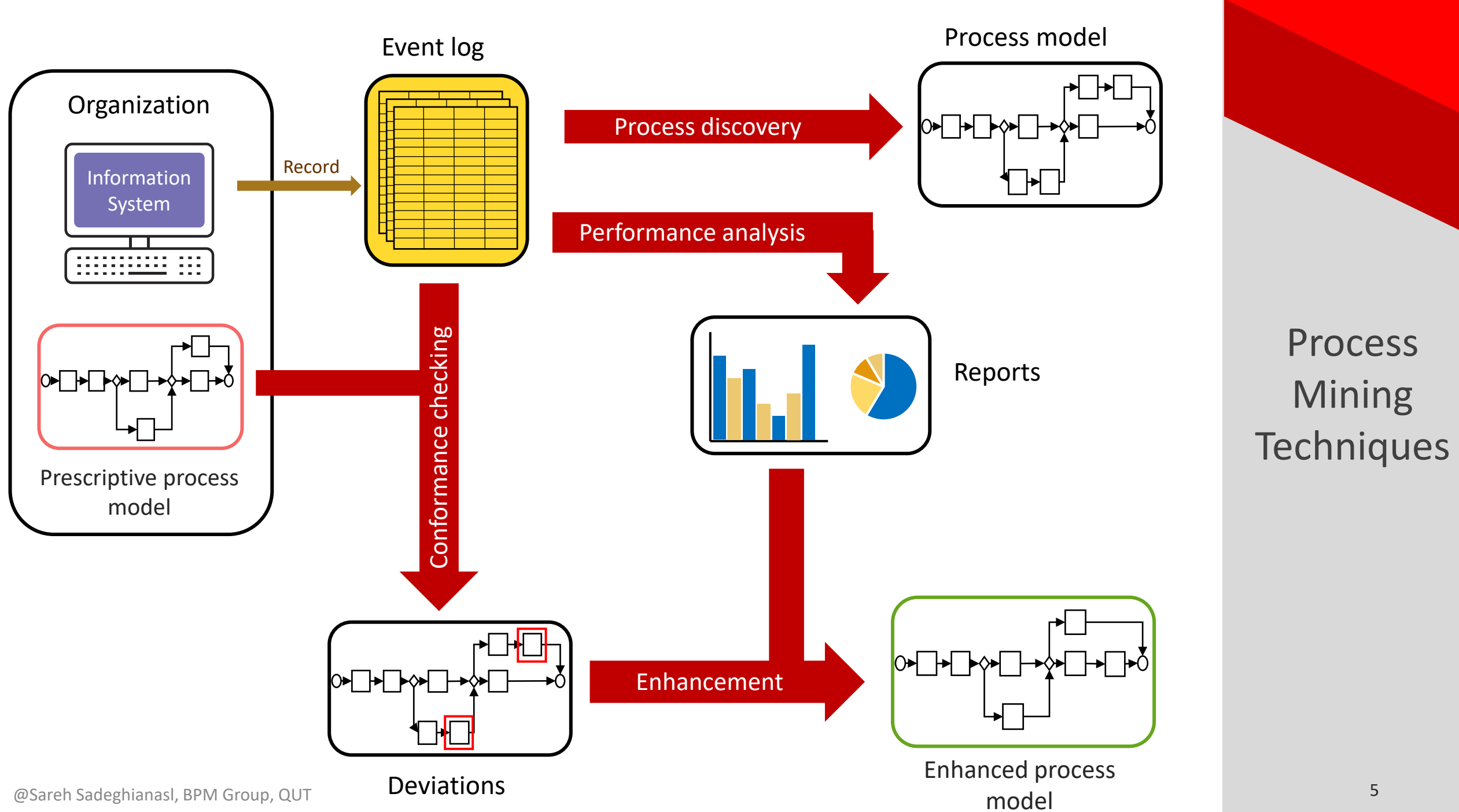
Today, we'll discuss ...

- Process mining
- Data quality: why does it matter?
- Event Log Imperfection Patterns
- Approaches to detect and repair
- A Gamified Crowdsourcing approach

Process Mining



Case ID	Activity	Timestamp
2033480	Arrive at ED	10/2/2020 13:45
2033480	Triage	10/2/2020 13:55
2033480	Dr visit	10/2/2020 14:20
2033480	Run test	10/2/2020 15:10
2033480	Admit	10/2/2020 17:01
2033480	Take medication	10/2/2020 17:50



Process mining use cases

01

Insurance



- Best practice in claim processing
- Factors influencing performance

02

Healthcare



- Different patient journey for similar symptoms?
- Road trauma process, conformance to Trauma By-pass Guidelines

03

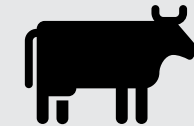
Education



- Timely completion of PhD journey (leave, extension, examination)

04

Livestock



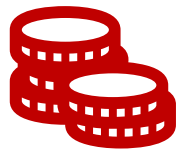
- From farm to packed meat, conformance to food safety guidelines

What about DATA?



Data Quality: A Problem in the Real World

“Data quality and people remain the two main hurdles for widespread adoption” – Prof. Wil van der Aalst, 2020 (<https://www.gartner.com/en/documents/3991229/market-guide-for-process-mining>)



COSTLY

- Bad data costs the US economy \$3.1 trillion per year (IBM)
- 84% of CEOs are concerned about the quality of the data that is basis for decisions (Forbes)

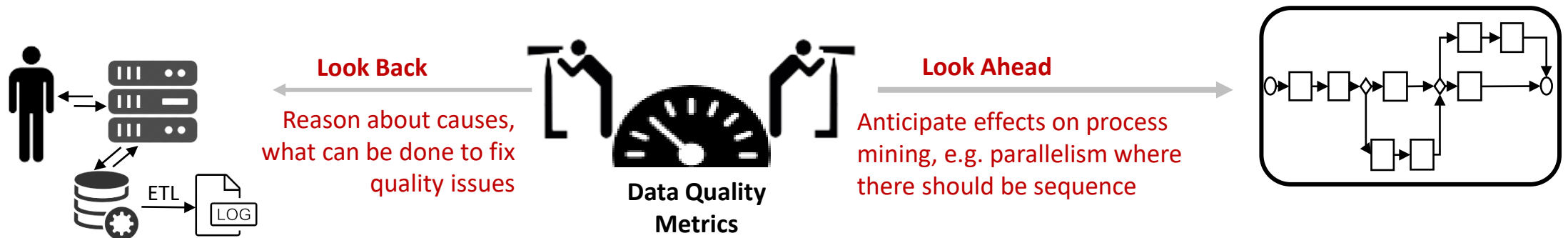


TIME-CONSUMING

- Recent XES survey: 60% of the effort is spent on data preparation for process mining (<https://www.tf-pm.org/resources/xes-standard/xes-2-0-workshop>)

Data Quality: Questions of Interest

- How do we **recognise** data quality issues in event data?
- How do we **assess** the quality?
- How can we **quantify** it's impact on process mining analyses?
- How do we **improve** the quality?
- How can we **prevent** the occurrence of data quality issues?



Event log structure

- The starting point of process mining
- Consists of a set of *events* recorded in a process of an organization
- *Event: What* happened for *who* and *when*
- Minimum (critical) elements

- Activity

- Case id

- Timestamp

Case id	Event id	Properties				
		Timestamp	Activity	Resource	Cost	...
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
	35654427	07-01-2011:14.24	reject request	Pete	200	...
2	35654483	30-12-2010:11.32	register request	Mike	50	...
	35654485	30-12-2010:12.12	check ticket	Mike	100	...

Process
instance
/ case

Event Log Imperfection Patterns

- 1) Form-based event capture
- 2) Inadvertent time travel
- 3) Unanchored event
- 4) Scattered event
- 5) Elusive case
- 6) Scattered case
- 7) Collateral events
- 8) Polluted labels
- 9) Distorted labels
- 10) Synonymous labels
- 11) Homonymous labels

Suriadi, S., Andrews, R., ter Hofstede, A. H., & Wynn, M. T. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information systems*, 64, 132-150.

A systematic view of commonly found quality issues in an event log

Form-based event capture

Episode ID	Event	Timestamp	Description	...
ID1	Primary Survey	2012-11-23 15:42:38
ID1	Airway Clear	2012-11-23 15:42:38
ID1	...	2012-11-23 15:42:38
	Primary Survey	2012-11-24 09:58:33
	Airway Clear	2012-11-24 09:58:33
	...	2012-11-24 09:58:33
	Procedure 1	2012-11-24 09:58:33	Completed on
			2012-11-24 06:58:34	

These events are recorded on a form ...

... and all have the same timestamp.

Root cause

- Multiple events are captured through an e-form.
- They all have the same timestamp (the time the user clicks the 'save' button in the e-form)

Effect

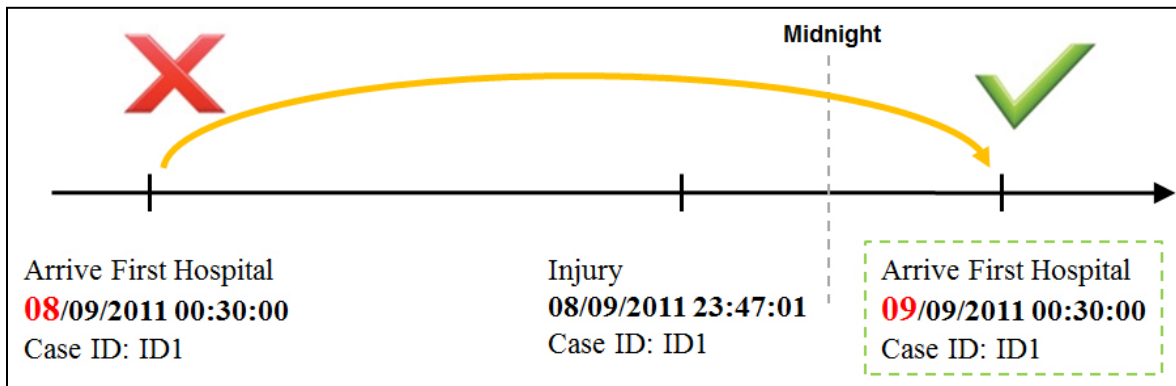
- The actual ordering of events is lost
- Same-time events are considered as parallel in the discovered process model

Andrews, R., Suriadi, S., Ouyang, C., & Poppe, E. (2018, October). Towards event log querying for data quality. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"* (pp. 116-134). Springer, Cham.

Inadvertent time travel

Episode ID	Activity	Timestamp	...
ID1	Arrival first hospital	2011-09-08 00:30:00
ID1	Injury	2011-09-08 23:47:01
...
ID1	Operation	2011-09-09 16:30:00

'Midnight' problem. Time portion correct but date part in error.



Root cause

- Wrong timestamp because of the proximity of the correct and incorrect value
- Human error
- E.g., events that happen just after midnight.

Effect

- Incorrect event order
- Incorrect process models discovered
- Inaccurate performance analysis

Andrews, R., Suriadi, S., Ouyang, C., & Poppe, E. (2018, October). Towards event log querying for data quality. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"* (pp. 116-134). Springer,.

Unanchored event

Event timestamps in dd/mm/yyyy format are imported ...

Original Data			
caseID	Activity	Timestamp	Description
1234567	Progress note	01/09/2013 21:53:25
1234567	Medical note	02/09/2013 01:11:25
1234567	Therapy	12/11/2013 16:08:23
1234567	Discharge letter	14/11/2013 16:43:29
....

... as mm/dd/yyyy format

Parsed Data			
caseID	Activity	Timestamp	Description
1234567	Progress note	09/01/2013 21:53:25	Progress notes
1234567	Medical note	09/02/2013 01:11:25
1234567	Discharge letter	11/14/2013 16:08:23
1234567	Therapy	12/11/2013 16:43:29
....

Root cause

- Timestamp format is not what the event log parsing tool expects

Effect

- Incorrect timestamp
- Incorrect process model
- Inaccurate performance analysis

Suriadi, S., Andrews, R., ter Hofstede, A. H., & Wynn, M. T. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information systems*, 64, 132-150.

Collateral events

caseID	Activity	Timestamp
1234567	Adjust recovery cost	19/06/2014 12:15:18
1234567	Adjust recovery cost	19/06/2014 12:16:53
1234567	Email	19/06/2014 12:19:25
....
1234567	Pay assessor fee	19/06/2014 12:22:...
1234567	Adjust admin cost	19/06/2014 12:22:48

All events refer to single process step 'Pay Insurance Claim Assessor'.

Root cause

- Multiple low-level activities with very close timestamps
 - Recorded by different systems
 - Fired by a system when an event occurs

Effect

- Overly complex process models
- Hinders the extraction of meaningful conclusions

Andrews, R., Suriadi, S., Ouyang, C., & Poppe, E. (2018, October). Towards event log querying for data quality. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"* (pp. 116-134). Springer, Cham.

Polluted labels

CaseID	Activity	Timestamp	...
xxxx	Notification of Loss - AAAA Incident No. aaaa	xxxx-xx-xx xx:xx:xx
xxxx	Notification of Loss – BBBB Incident No. bbbb	yyyy-yy-yy yy:yy:yy
xxxx	Notification of Loss – CCCC Incident No. cccc	zzzz-zz-zz zz:zz:zz
.....	<u>Notification of Loss</u> – <u>DDDD</u> Incident No. <u>dddd</u>	



- Root cause:**
- Free-text data entry (with a recommended label)

- Effect:**
- Overly complex process models
 - Models over-fitting event logs in conformance checking

Suriadi, S., Andrews, R., ter Hofstede, A. H., & Wynn, M. T. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information systems*, 64, 132-150.

Distorted labels

CaseID	Activity	Timestamp	...
1234567	a/w inv to cls	xxxx-xx-xx xx:xx:xx
8912345	a/w inv to cls.	yyyy-yy-yy yy:yy:yy
1234567	XX – Further Information Required	zzzz-zz-zz zz:zz:zz
8912345	XX – Further Infomation Required

Root cause

- Free-text data entry

Effect

- Similar to polluted labels
- Activities that are the same are treated differently in
 - Discovery
 - Conformance
 - Performance analysis

Suriadi, S., Andrews, R., ter Hofstede, A. H., & Wynn, M. T. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information systems*, 64, 132-150.

Synonymous labels

Hospital A Event Log			
CaselD	Activity	Timestamp	Description
1234567	Medical Assign	7/9/2013 14:50:30	Seen by physician
....
1234567	Troponin	7/9/2013 15:39:32	Blood test
....

Hospital B Event Log			
CaselD	Activity	Timestamp	Description
1234567	DrSeen	7/9/2013 00:52:25	Seen by physician
8912345	Blood test - Troponin	7/9/2013 02:04:51	Blood test
....

Syntactically different labels, but semantically similar activities.

Root cause

- Free-text data entry
- Data from different systems

Effect

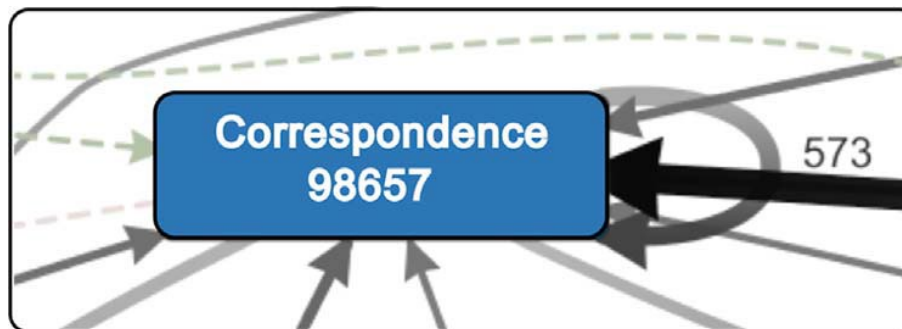
- Similar to polluted, and distorted labels
- Activities that are the same are treated differently in
 - Discovery
 - Conformance
 - Performance analysis

Andrews, R., Suriadi, S., Ouyang, C., & Poppe, E. (2018, October). Towards event log querying for data quality. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"* (pp. 116-134). Springer.

Homonymous labels

CaseID	Activity	Timestamp	Description
1234567	Triage Assessment	06/09/2013 12:33:17
1234567	Progress Note	06/09/2013 13:10:23
1234567	Discharged	06/09/2013 13:15:00
1234567	Triage Assessment	13/09/2013 07:24:36
1234567	Triage Assessment	13/09/2013 07:28:51

Syntactically the same labels, but semantically different activities.



Root cause

- They are usually captured by the system, e.g. when a triage form is viewed or edited

Effect

- Different activities are grouped into one in:
 - Discovery
 - Conformance
 - Performance analysis

Andrews, R., Suriadi, S., Ouyang, C., & Poppe, E. (2018, October). Towards event log querying for data quality. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"* (pp. 116-134). Springer, Cham.

Scattered event

Event log 1			
caseID	Activity	Timestamp	Description
1234567	Surgical Procedure	21/09/2011 08:11:25	Stent insertion
....
1234567	Procedure start-time	21/09/2011 08:11:25	0:2011092010480000:0.000000:0:0
1234567	Procedure end-time	21/09/2011 08:11:25	0:2011092010590000:0.000000:0:0

These attribute values ...

Event log 1			
caseID	Activity	Timestamp	Description
1234567	Stent insertion	20/09/2011 10:59:00	
....

... can be used to construct a new event.

Root cause

- Free-text data fields, manual data entry

Effect

- Missing information that could enrich insights from a process mining task
- Incomplete process models

Suriadi, S., Andrews, R., ter Hofstede, A. H., & Wynn, M. T. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information systems*, 64, 132-150.

Scattered case

Event Log			
caseID	activity	timestamp	description
1234567	Progress Note	7/09/2013 00:50:30
1234567	Medical Note	7/09/2013 00:52:25
1234567	Vital Signs	7/09/2013 03:52:48
....
8912345	Therapy	12/09/2013 15:59:32
8912345	Procedure 1	13/09/2013 10:20:01

Order Table			
caseID	activity	timestamp	description
1234567	Order ECG	7/09/2013 00:52:25	
8912345	Order Procedure 1	9/09/2013 14:04:51
....

Case Summary Table			
caseID	visit type	timestamp	record number
1234567	Emergency	6/09/2013 23:47:00	RN1234567
8912345	Hospital	8/09/2013 13:45:00	RN1234567
....

Apparently 2 distinct 'caseID's' ...

... scattered across multiple sources ...

... that can be linked by a common 'record number'.

Root cause:

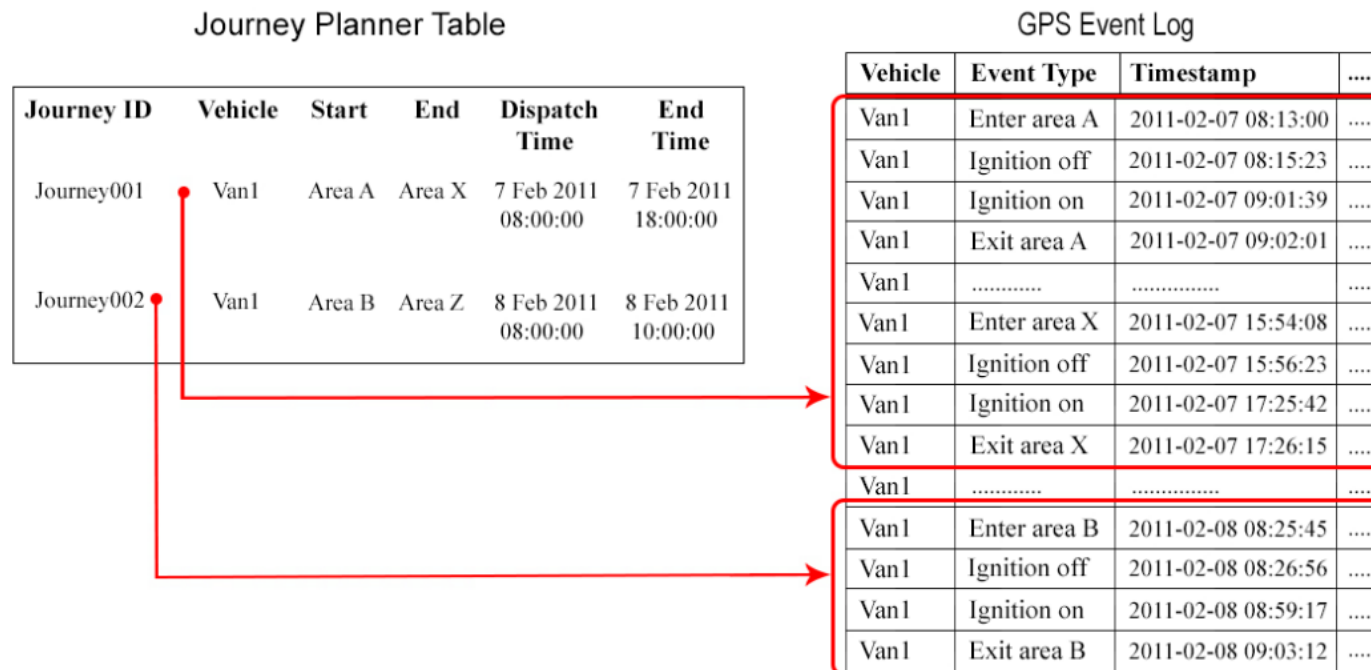
- Event data recorded in multiple systems

Effect:

- Information that could enrich insights from a process mining task
- Incomplete process models

Suriadi, S., Andrews, R., ter Hofstede, A. H., & Wynn, M. T. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information systems*, 64, 132-150.

Elusive case



Root cause

- Event data derived from a system that is not process-aware (e.g. a GPS tracking system)

Effect

- Prevents conducting process mining analysis

Suriadi, S., Andrews, R., ter Hofstede, A. H., & Wynn, M. T. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information systems*, 64, 132-150.

Detection and repair: Timestamp issues



Event automaton (Conforti et al., 2020)

Common ordering in the log used for incorrect ordered ones



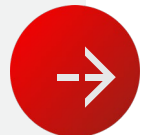
Order anomaly, statistical anomaly (Dixit et al. 2018)

Identify candidates of incorrect ordered events and deploy user input to detect and repair them



Probability distribution function (Van der Aa et al., 2020)

Over all possible total orders of known partial orders of events



Timestamp quality metrics (Fischer et al., 2020)

Accuracy, completeness, consistency, uniqueness

Log, trace, activity, and event level

Detection and repair: Label issues

Same syntax, different semantic (Homonymous)

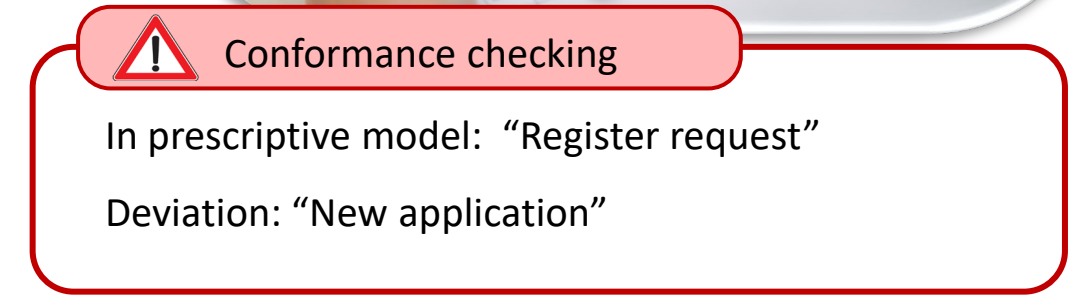
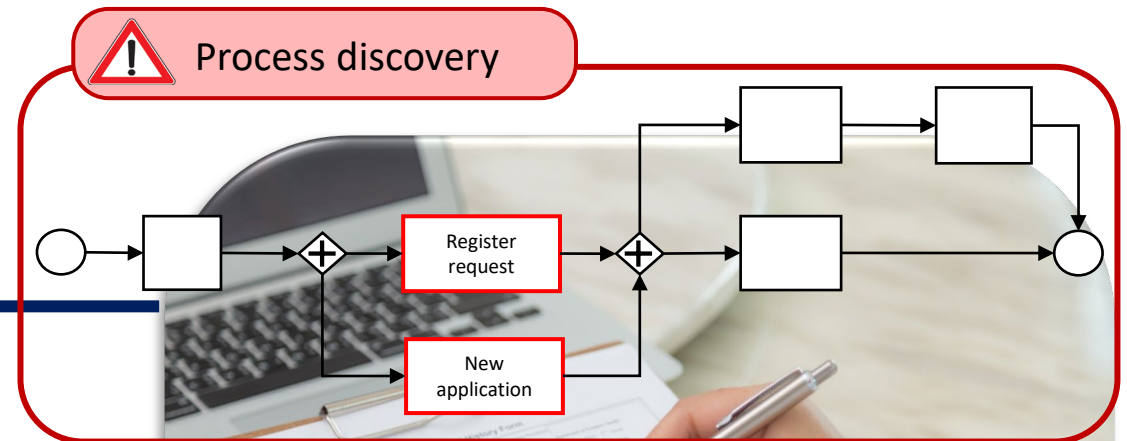
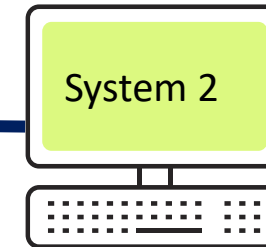
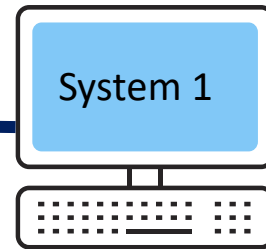
- Process model level
 - Labels with different neighbours in the model (Vazquez-Barreiros et al., 2015; Sanchez-Charle et al., 2016)
- Event log level
 - Using timestamps of event (Tax et al., 2017)
 - Using control flow of events (Lu et al., 2016)

Different syntax, same semantics (Synonymous, Polluted, Distorted)

- Process model/event log level
 - Using a domain ontology or dictionary (Cairns et al., 2014; Koschmider et al., 2015; Pittke et al., 2015)

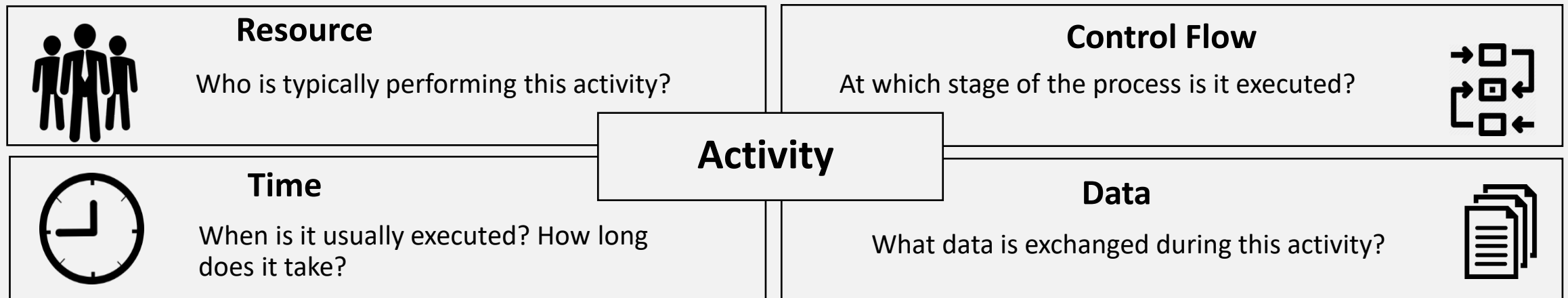
Imperfect Activity Labels

Case ID	Activity
1	Register request
1	examine thoroughly
1	Check ticket
1	...
2	New application
2	Check pass
2	decide
2	...



Identifying Imperfect label candidates

Activity Context



$$\mathcal{R} = (r_1, r_2, r_3, r_4, r_5)$$

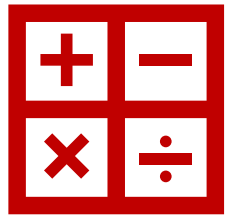
Resource distance of a and b:
The difference between their resource PDF

<i>a</i>	0	0	0	0.5	0.5
<i>b</i>	0.25	0.5	0.25	0	0

$$D_{re}(a, b) = \frac{2}{2} = 1$$

Sadeghianasl, S. Hofstede, A.H.M, Wynn, M. Suriadi, S.: A Contextual Approach to Detecting and Repairing Synonymous and Polluted Activity Labels in Process Event Logs, *International Conference on Cooperative Information Systems (CoopIS)*, Rhodes, Greece, 2019, pp. 76-94

Problem of computational approaches

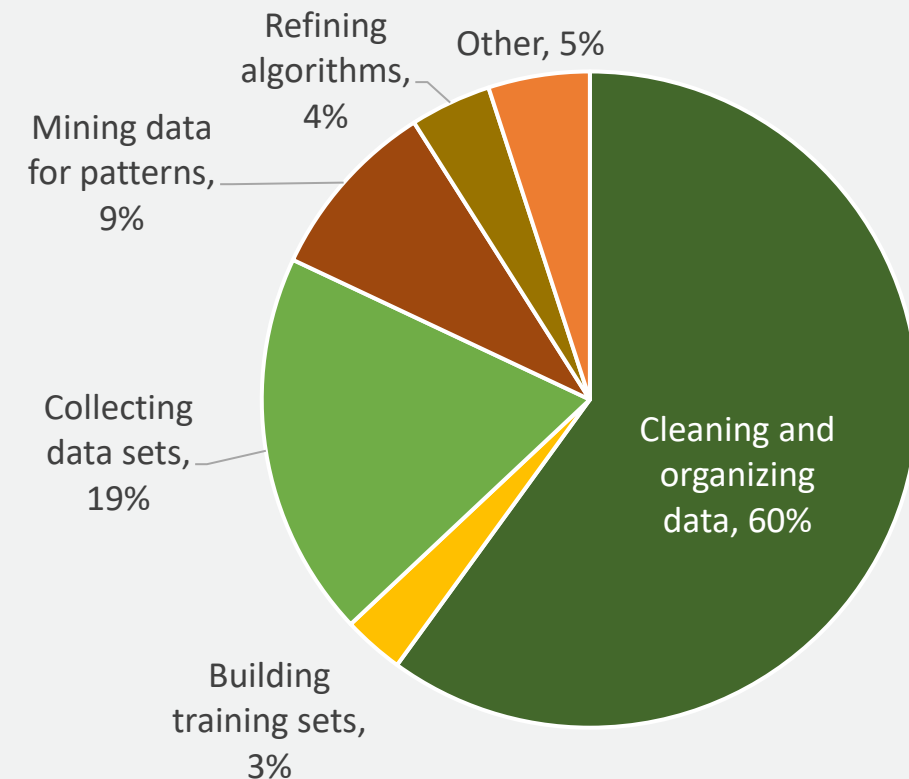


Computational approaches:
low effectiveness in real-life
scenarios (Klinkmüller & Webber,
2021; Rodríguez et al, 2016)

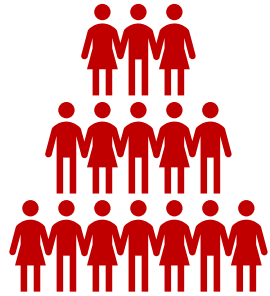


Domain knowledge: required
Domain expert: in-demand,
expensive, and time-poor
(Scibona, 2018; Wohlgenannt et al,
2016)

Data cleaning: “the most time consuming
and the least enjoyable data science task”
(Gil, 2016)



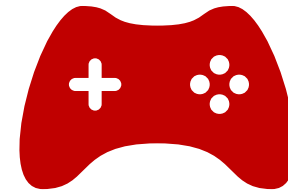
Our solution



Crowdsourcing

A diversity of knowledge of a large group of people (Howe, 2006)

No restriction to the physical location of domain experts



Gamification

Using game elements in non-gaming system (Deterding et al., 2011)

Increasing user engagement → providing better advice

Motivational Drives

Based on the Octalysis framework for gamification design (Chou, 2019)

Development & Accomplishment



Overcoming challenges, achieving goals and mastery

Points, badges, progress bars

Epic Meaning & Calling



Doing something bigger than themselves, e.g. saving the world

Narration

Social Influence & Relatedness

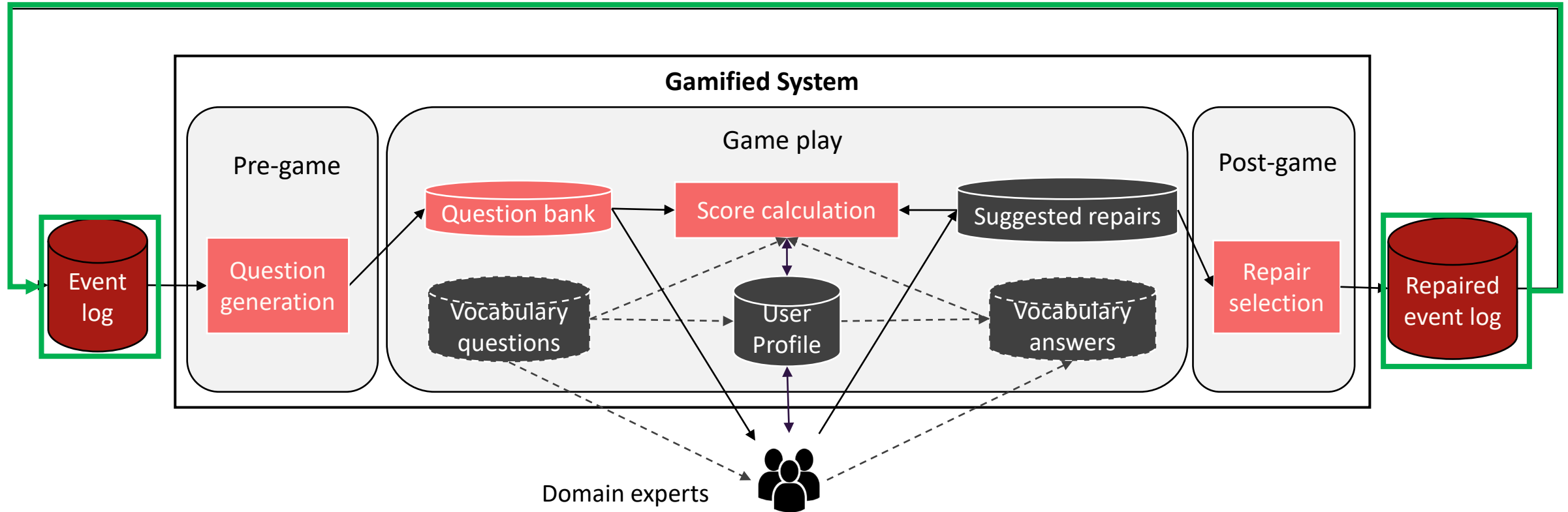


Follow a social norm, i.e. what every one else is doing

Social reports, leaderboard

Sadeghianasl, S. Hofstede, A.H.M, Suriadi, S. Turkay, S.: Collaborative and Interactive Detection and Repair of Activity Labels in Process Event Logs, *International Conference on Process Mining (ICPM)*, Padua, Italy, 2020, pp. 41-48

The Quality Guardian Framework



Sadeghianasl, S. Hofstede, A.H.M, Suriadi, S. Turkey, S.: Collaborative and Interactive Detection and Repair of Activity Labels in Process Event Logs, *International Conference on Process Mining (ICPM)*, Padua, Italy, 2020, pp. 41-48

The Quality Guardian Games

The Quality Guardian

Select all options that have the same meaning as "Clear Inv" from the following options.

Option	Control Flow	Resource	Temporal	Data	Absolute frequency	Relative frequency	Popularity among players
Change Final Invoice Indicator	2%	0%	0%	0%	11	0%	▼ 1 out of 21
Clear Invoice	64%	99%	99%	99%	97197	6.09%	▼ 21 out of 21
SRM: Transaction Completed	0%	0%	0%	0%	8	0%	▼ 1 out of 21
Remove Invoice	68%	99%	99%	99%	38879	2.44%	▼ 16 out of 21
Eliminate Inv	65%	99%	99%	99%	19439	1.22%	▼ 14 out of 21
Clr Invoice	78%	99%	99%	99%	19439	1.22%	▼ 17 out of 21
SRM: Incomplete	0%	0%	0%	0%	6	0%	▼ 0 out of 21
Change Rejection Indicator	0%	0%	0%	0%	2	0%	▼ 0 out of 21
SRM: Held	0%	0%	0%	0%	6	0%	▼ 0 out of 21

Control Flow Resource Temporal Data Absolute frequency Relative frequency Popularity among players

@Sareh Sadeghianasl, BPM Group, QUT 32

Topic: Claim Review Activities

Level 1/3

Drag and drop cards into groups. You can add more groups. All cards in a group should have the same meaning. Then click on group names to rename them. The group names should represent the meaning of their cards.

Insurance charge Claims reserve Premium Balance sheet reserve

Terms of settlement Deed of release

Group 1

Group 1 Feedback

How many players have grouped these terms together so far?

6 out of 8

The absolute and relative frequencies of these terms in data:

URGENT REVIEW REQUEST: 1, 0%, please review asap: 1, 0%.

Average context Similarities:

URGENT REVIEW REQUEST, please review asap: 37%

@Sareh Sadeghianasl, BPM Group, QUT 33

Sadeghianasl, S. Hofstede, A.H.M, Suriadi, S. Turkay, S.: Collaborative and Interactive Detection and Repair of Activity Labels in Process Event Logs, *International Conference on Process Mining (ICPM)*, Padua, Italy, 2020, pp. 41-48

Sadeghianasl, S. Hofstede, A.H.M, Wynn, M. Turkay, S.: Gamifying Activity Label Repair in Process Event Logs, Submitted to an International Journal.



Select all options that have the same meaning as "Clear Inv" from the following options.

Change Final Invoice Indicator

2%
0%
0%
0%

11 0% ♥ 1 out of 21 0%

Clear Invoice

64%
99%
99%
95%

97197 6.09% ♥ 21 out of 21 95%

SRM: Transaction Completed

0%
0%
0%
3%

8 0% ♥ 1 out of 21 3%

Remove Invoice

68%
99%
99%
95%

38879 2.44% ♥ 16 out of 21 95%

Eliminate Inv

65%
99%
99%
94%

19439 1.22% ♥ 14 out of 21 94%

Clr Invoice

78%
99%
99%
95%

19439 1.22% ♥ 17 out of 21 95%

SRM: Incomplete

0%
0%
0%
0%

6 0% ♥ 0 out of 21 0%

Change Rejection Indicator

0%
0%
0%
0%

2 0% ♥ 0 out of 21 0%

SRM: Held

0%
0%
0%
0%

6 0% ♥ 0 out of 21 0%

Control Flow Resource Temporal Data Absolute frequency Relative frequency ♥ Popularity among players

Topic: Claim Review Activities



Drag and drop cards into groups. You can add more groups. All cards in a group should have the same meaning. Then click on group names to rename them. The group names should represent the meaning of their cards.

Insurance charge Claims reserve Premium Balance sheet reserve

Terms of settlement Deed of release

Group 1 +

Group 1 Feedback

How many players have grouped these terms together so far?

6 out of 8

The **absolute** and **relative** frequencies of these terms in data:

URGENT REVIEW REQUEST: **1**, **0%**.
please review asap: **1**, **0%**.

Average context Similarities:

URGENT REVIEW REQUEST, please review asap: 37% ?

Gamifying Activity Ontology Creation



Formalize domain knowledge about activities



Semantic relations

Synonymy

Antonymy

Hyponymy (kind-of)

Meronymy (Part-of)



Can be used for activity label repair

Spin the wheels to reveal a question. Then answer that by selecting the semantic relation between the two activity labels. Overall, you have 6 attempts to spin, answer, and earn points.

SPIN Attempts left: 1

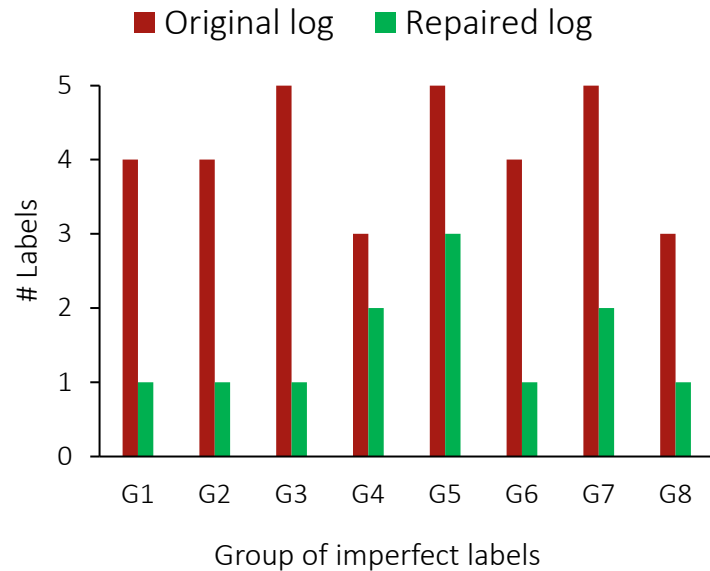
8 phase decision irrevocable Choose... 99 phase archived case Save Hint

Graph

Filter relations: Synonym, Antonym, Hypernym/hyponym, H

Sadeghianasl, S. Hofstede, A.H.M, Wynn, M. Turkay, S. Myers, T.:Process activity ontology learning from event logs through gamification, IEEE Access, 9, 165865-165880.

Results of Activity Quality Improvement

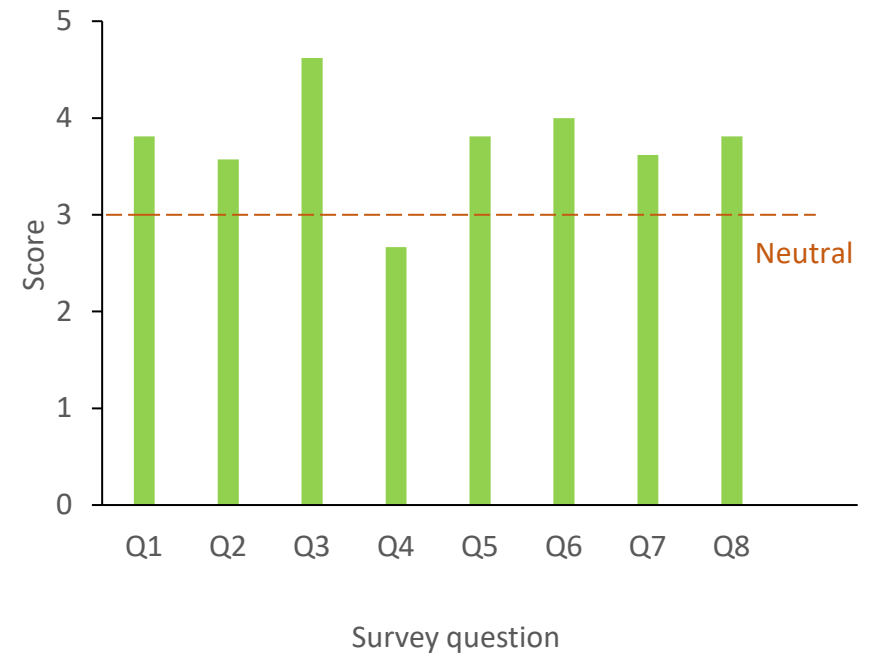


Sadeghianasl, S. Hofstede, A.H.M, Suriadi, S. Turkay, S.: Collaborative and Interactive Detection and Repair of Activity Labels in Process Event Logs, *International Conference on Process Mining (ICPM)*, Padua, Italy, 2020, pp. 41-48

Sadeghianasl, S. Hofstede, A.H.M, Wynn, M. Turkay, S.: Gamifying Activity Label Repair in Process Event Logs, Submitted to an International Journal.

Results of User Engagement

Survey Question	Assessed Feature (Sweetser and Wyeth, 2005)
Q1	The clarity of the game goals
Q2	The usefulness of the feedback
Q3	The ease of use
Q4	The usefulness of knowing crowd views
Q5	The ability to control game actions and interface
Q6	The overall engagement
Q7	The knowledge development
Q8	The required concentration



Sadeghianasl, S. Hofstede, A.H.M, Suriadi, S. Turkay, S.: Collaborative and Interactive Detection and Repair of Activity Labels in Process Event Logs, *International Conference on Process Mining (ICPM)*, Padua, Italy, 2020, pp. 41-48

Summary



Event log imperfection patterns

- What are they? Root causes/ effects



Detect and repair quality issues in event logs

- Timestamp, label issues



Gamification solution to detecting and repairing label quality issues



Goals

- Convey the importance of data quality
- To prevent them, and to understand the effect they can have on process mining analysis
- Make reliable decision for our organizations

Thank you

- Acknowledgement
 - QUT BPM group

Sareh Sadeghianasl,
Queensland University of Technology
(QUT), Australia
s.sadeghianasl@qut.edu.au

References (1)

1. Suriadi, S., Andrews, R., ter Hofstede, A. H., & Wynn, M. T. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information systems*, 64, 132-150.
2. Sadeghianasl, S. Hofstede, A.H.M, Wynn, M. Suriadi, S.: A Contextual Approach to Detecting and Repairing Synonymous and Polluted Activity Labels in Process Event Logs, *International Conference on Cooperative Information Systems (CoopIS)*, Rhodes, Greece, 2019, pp. 76-94
3. Sadeghianasl, S. Hofstede, A.H.M, Suriadi, S. Turkay, S.: Collaborative and Interactive Detection and Repair of Activity Labels in Process Event Logs, *International Conference on Process Mining (ICPM)*, Padua, Italy, 2020, pp. 41-48
4. Sadeghianasl, S. Hofstede, A.H.M, Wynn, M. Turkay, S. Myers, T.: Process activity ontology learning from event logs through gamification, *IEEE Access*, 9, 165865-165880.
5. Andrews, R., Suriadi, S., Ouyang, C., & Poppe, E. (2018, October). Towards event log querying for data quality. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"* (pp. 116-134). Springer, Cham.
6. Andrews, Robert, Fahame Emamjome, Arthur HM ter Hofstede, and Hajo A. Reijers. "Root-cause analysis of process-data quality problems." *Journal of Business Analytics* (2021): 1-25.
7. P. M. Dixit, S. Suriadi, R. Andrews, M. T. Wynn, A. H. M. ter Hofstede, J. C. A. M. Buijs, and W. M. P. van der Aalst, "Detection and Interactive Repair of Event Ordering Imperfection in Process Logs", in *International Conference on Advanced Information Systems Engineering (CAiSE)*, pp. 274–290, Tallinn, Estonia, Jun. 2018
8. Conforti, Raffaele, Marcello La Rosa, Arthur HM Ter Hofstede, and Adriano Augusto. "Automatic repair of same-timestamp errors in business process event logs." In *International Conference on Business Process Management*, pp. 327-345. Springer, Cham, 2020.
9. Fischer, Dominik Andreas, Kanika Goel, Robert Andrews, Christopher Gerhard Johannes van Dun, Moe Thandar Wynn, and Maximilian Röglinger. "Enhancing event log quality: detecting and quantifying timestamp imperfections." In *International Conference on Business Process Management*, pp. 309-326. Springer, Cham, 2020.

References (2)

9. Van der Aa, Han, Henrik Leopold, and Matthias Weidlich. "Partial order resolution of event logs for process conformance checking." *Decision Support Systems* 136 (2020): 113347.
10. J. de San Pedro and J. Cortadella, "Discovering Duplicate Tasks in Transition Systems for the Simplification of Process Models", in *International Conference on Business Process Management (BPM)*, ser. Lecture Notes in Computer Science, M. La Rosa, P. Loos, and O. Pastor, Eds., vol. 9850, pp. 108–1124. Rio de Janeiro, Brazil: Springer, Sep. 2016.
11. B. Vázquez-Barreiros, M. Mucientes, and M. Lama, "Mining Duplicate Tasks from Discovered Processes", in *International Workshop on Algorithms & Theories for the Analysis of Event Data (ATAED)*, ser. CEUR Workshop Proceedings, W. M. P. van der Aalst, R. Bergenthum, and J. Carmona, Eds., vol. 1371, pp. 78–82. Brussels, Belgium: CEUR-WS.org, Jun. 2015.
12. X. Lu, D. Fahland, F. J. van den Biggelaar, and W. M. van der Aalst, "Handling Duplicated Tasks in Process Discovery by Refining Event Labels", in *International Conference on Business Process Management (BPM)*, ser. Lecture Notes in Computer Science, M. La Rosa, P. Loos, and O. Pastor, Eds., vol. 9850, pp. 90–107. Rio de Janeiro, Brazil: Springer, Sep. 2016.
13. N. Tax, E. Alasgarov, N. Sidorova, W. M. P. van der Aalst, and R. Haakma, "Generating Time-Based Label Refinements to Discover More Precise Process Models", Eindhoven University of Technology, Philips Research, Tech. Rep., Oct. 2017, <https://arxiv.org/pdf/1705.09359.pdf>, Accessed at 2018.8.2.
14. Cairns, Awatef Hicheur, Joseph Assu Ondo, Billel Gueni, Mehdi Fhima, Marcel Schwarcfeld, Christian Joubert, and Nasser Khelifa. "Using semantic lifting for improving educational process models discovery and analysis." In *SIMPDA*, pp. 150-161. 2014.
15. Koschmider, Agnes, Meike Ullrich, Antje Heine, and Andreas Oberweis. "Revising the vocabulary of business process element labels." In *International Conference on Advanced Information Systems Engineering*, pp. 69-83. Springer, Cham, 2015.
16. Pittke, Fabian, Henrik Leopold, and Jan Mendling. "Automatic detection and resolution of lexical ambiguity in process models." *IEEE Transactions on Software Engineering* 41, no. 6 (2015): 526-544.