UNIVERSITY OF PISA
Department of Energy, Systems, Territory and Construction Engineering

Process-Mining-enabled audit of Information Systems:
a case in a Mediterranean port

Pierluigi Zerbino, PhD
pierluigi.zerbino@ing.unipi.it

Further details in: Zerbino et al. (2018), "*Process-mining-enabled audit of information systems: Methodology and an application*", Expert Systems with Applications, Vol. 110, pp. 80-92

1

Agenda


The case


Results and discussion


Lessons learnt


Further PM capabilities in auditing

2

Why to audit port operations (1)

30/01/2019 650 kg of cocaine Origin: Honduras Destination: Spain

3

Why to audit port operations (2)

31/01/2019 Origin: Colombia Destination: Spain

4

Why to audit port operations (3)

23/06/2019 | 55 tons of hazardous waste | Origin: Venezuela | Destination: Italy

5

The case (1)

Objective of the case

To audit an IS-supported port logistics process

To check the compliance with the prescribed document and process flow

Customs | *Control bodies* | *Regulations* | *Monitoring*

Port Community System

Export

National cabotage

...

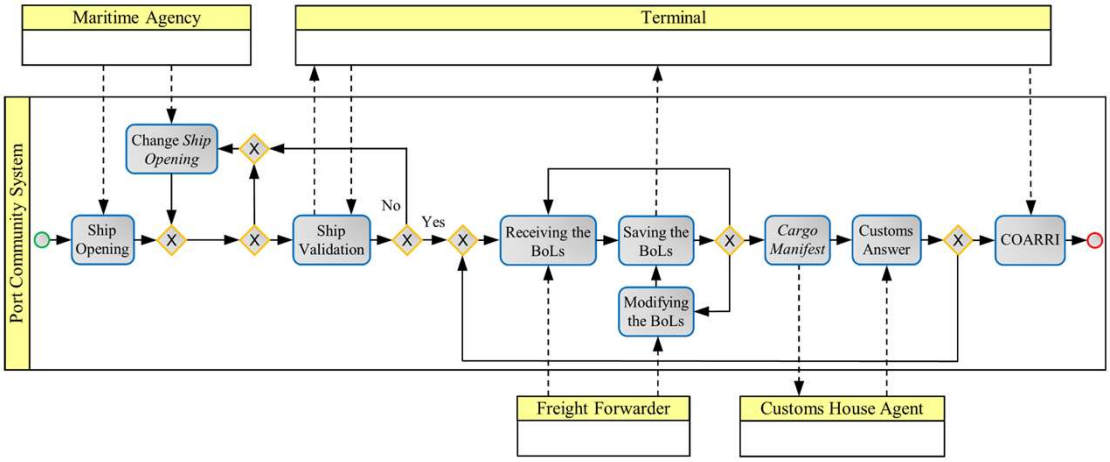
Port taxes

Import

6

The case (2)

De jure model



7

The case (3)

Dump of data

> 15 GB

52 tables

15 views

10^2 attributes

Encrypted

8

The case (4)

Content removed for privacy concerns

9

The case (5)

Excerpt of the export data

Content removed for privacy concerns

10

The case (6)

01 / Jan / 2016
30 / Jun / 2016

Data cleansing and filtering

Final dataset

686 cases (vessels)
171630 events

Same data attributes of the previous analyses

Timestamp	Activity	Lifecycle	IMO	Route
29/04/2016 10:18:49:775	Customs Answer	Start	925xxxx	04xxxx
29/04/2016 10:18:56:887	Customs Answer	Complete	925xxxx	04xxxx
29/04/2016 10:20:31:098	Ship validation	Start	936xxxx	06xxxx
29/04/2016 10:20:31:169	Ship validation	Complete	936xxxx	06xxxx
29/04/2016 10:24:37:131	COARRI	Start	920xxxx	CXxxxx
29/04/2016 11:01:39:073	Customs Answer	Start	947xxxx	41xxxx

11

The case (7)

Algorithm selection

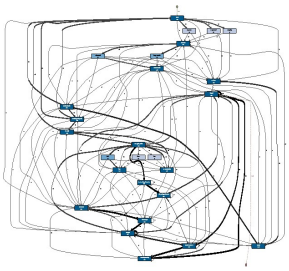
Several Process Mining algorithms, *e.g.* Heuristic Miner (Weijters, van der Aalst, & De Medeiros, 2006) or Genetic Miner (Lang et al., 2008), yield hard-to-interpret process maps in complex contexts

Content removed for privacy concerns

12

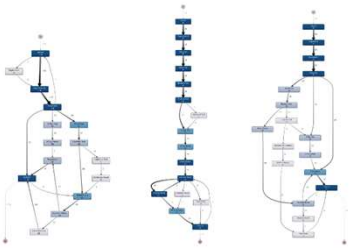
The case (8)

Spaghetti processes



VS

Lasagna processes



13

The case (9)

Algorithm selection

Scientific research concerning complexity metrics in business processes is still ongoing!

Complexity Index (Zerbino et al. 2018) $\mu = \frac{number\ of\ process\ deviations}{total\ process\ instances} = \frac{624}{686}\left[\frac{instances}{instances}\right] = 90.9\%$

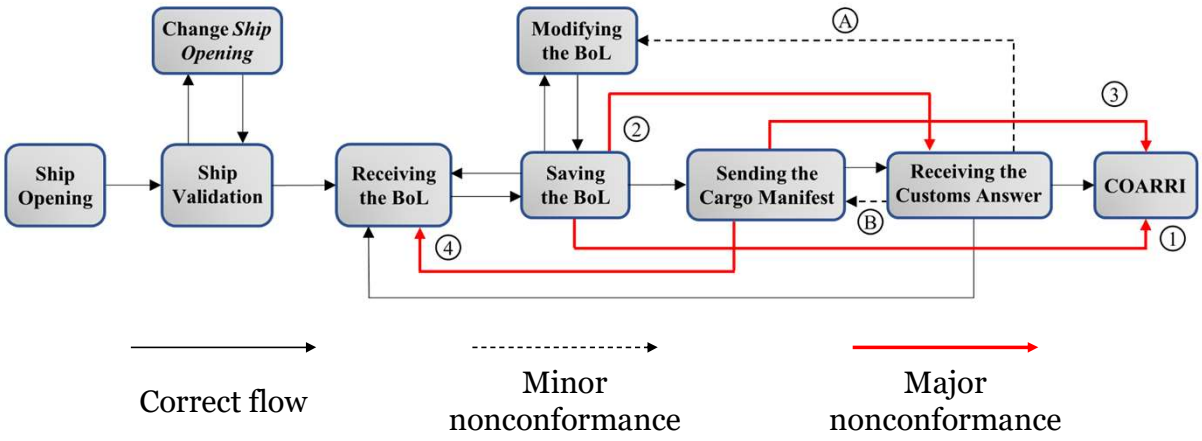
Fuzzy Miner algorithm suitable for complex contexts (Günther and van der Aalst, 2007; De Weerdt et al., 2012)

Disco

14

Results (1)

De facto model



15

Results (2)

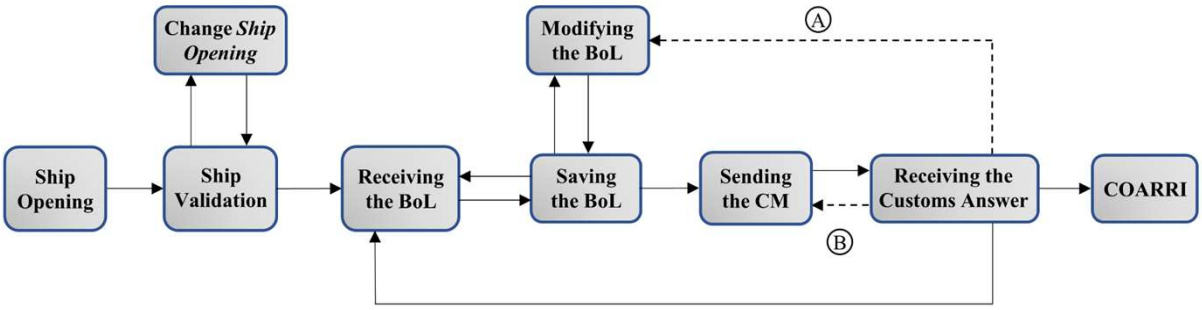
Statistics

Code	From	To	Absolute Frequency	Case Frequency	Total Duration	Median	Mean	Max	Min
A	Receiving the Customs Answer	Modifying the BoL	119	104	17.8 days	15 min	3.6 hrs	48.4 hrs	3.4 s
B	Receiving the Customs Answer	Sending the CM	322	211	29.9 days	14.3 min	2.2 hrs	4 days	14.4 s
1	Saving the BoL	COARRI	54	54	17.8 wks	23.7 hrs	55.5 hrs	25.9 days	66 min
2	Saving the BoL	Receiving the Customs Answer	459	244	9 days	3.3 min	28.3 min	49.4 hrs	218 ms
3	Sending the CM	COARRI	29	29	36.1 days	25.5 hrs	29.9 hrs	4.8 days	94.7 min
4	Sending the CM	Receiving the BoL	352	214	6.2 days	2.7 min	25.3 min	45.4 hrs	796 ms

16

Discussion (1)

Arches A and B



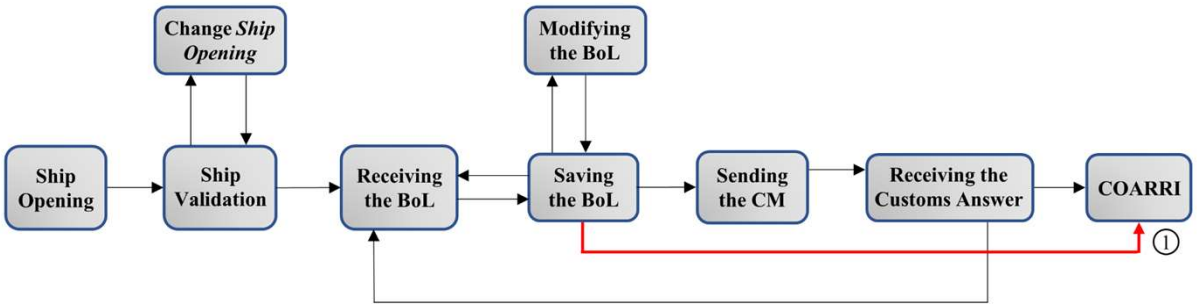
The *de jure* requires a feedback to *Receiving the BoL*

Arches A and B were consciously introduced by the programmer

17

Discussion (2)

Arch 1 *(case frequency = 54)*

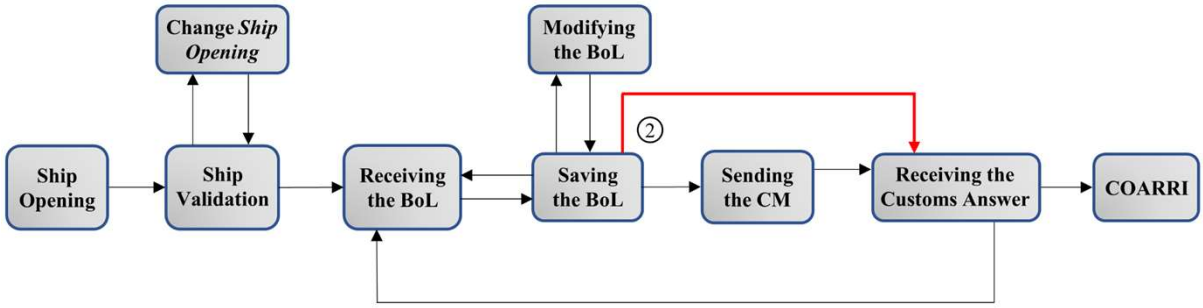


- No Cargo Manifest (CM) in 9 cases;
- No updates in the CM after the negative answer from the Customs, in the other cases.

18

Discussion (3)

Arch 2 *(case frequency = 244)*

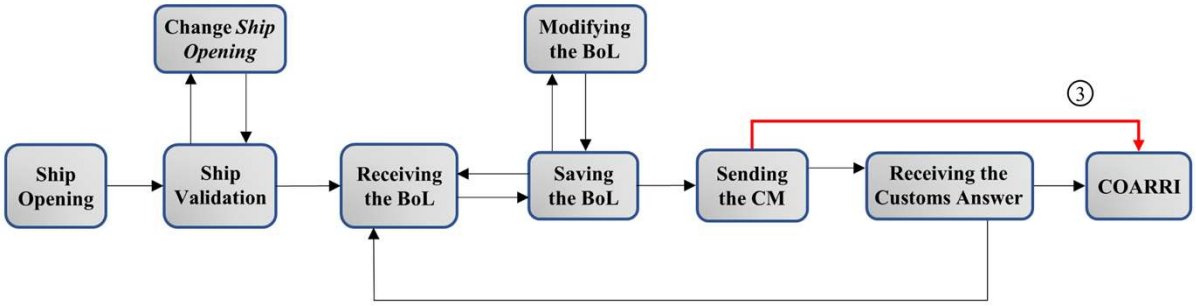


- The 244 process instances did not actually skip the CM;
- Possible dummy deviations.

19

Discussion (4)

Arch 3 *(case frequency = 29)*

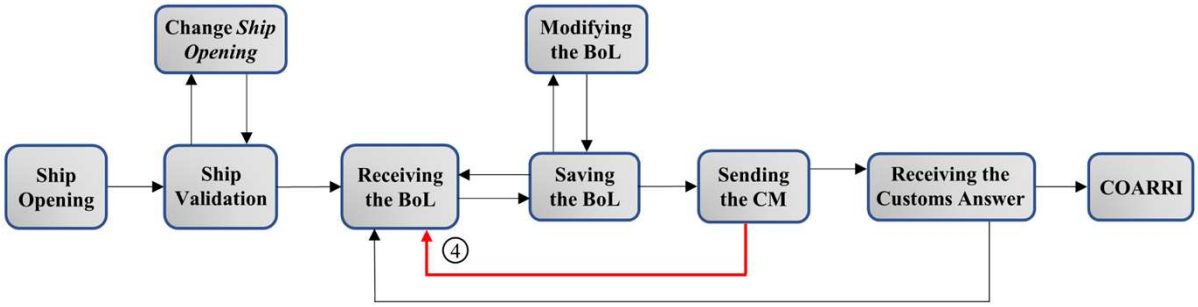


- 22 cases did skip the Customs check in the last iteration (akin to Arch 2);
- 7 cases did not perform a Customs Check.

20

Discussion (5)

Arch 4 *(case frequency = 214)*



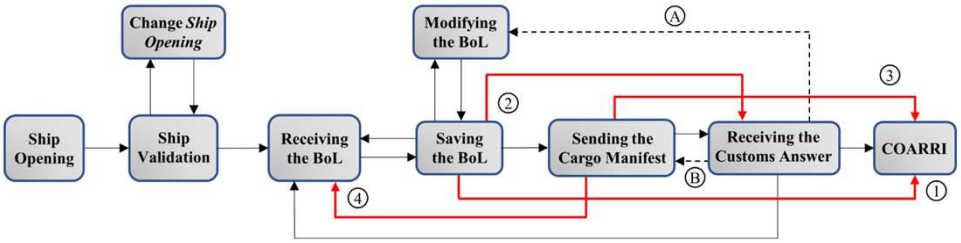
- All the cases present a CM;
- In 22 cases, the BoL were modified without updating the CM or checking for its correctness.

21

Discussion (6)

No issues with Customs Answer were detected …

… but there might be a gap between what is declared within the CM and what is actually loaded (further investigation needed);



22

Discussion (7)

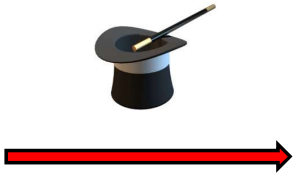
23

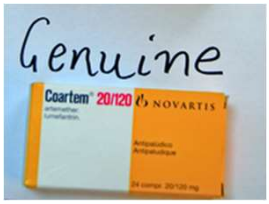
Discussion (7)

https://www.ilmattino.it/napoli/cronaca/napoli_business_camorra_racket_bare_cinesi_morti-4532732.html
https://www.dailynews24.it/napoli-pentito-svela-che-fine-fanno-i-cinesi-morti-in-citta/

24

Discussion (7)





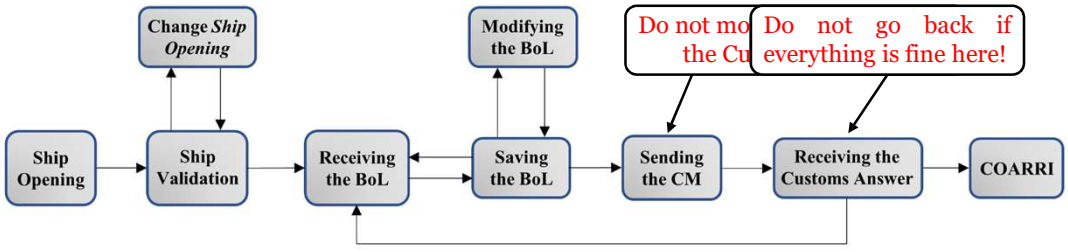
https://www.lastampa.it/cronaca/2019/06/10/news/farmaci-contraffatti-per-bambini-scoperti-in-un-container-nel-porto-di-genova-1.36540042

25

Implications (1)

Possible interventions

- To impede modifications in the BoL if the CC is positive;
- No feedback to the BoL process should be allowed after the submission of the CM but before receiving the Customs Answer;

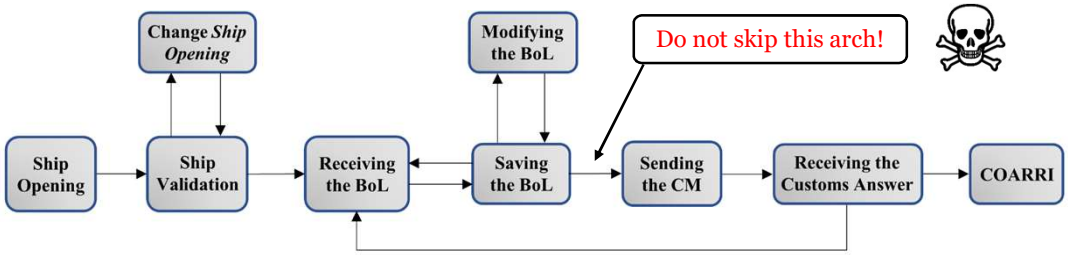


26

Implications (2)

Possible interventions

- To force the users to update the CM before its submission when the BoLs are modified;
- Reduce data entry errors with a double data entry (Barchard and Pace, 2011);



27

Lessons learnt about PM (1)

No need for sampling;

Strongly automatable;

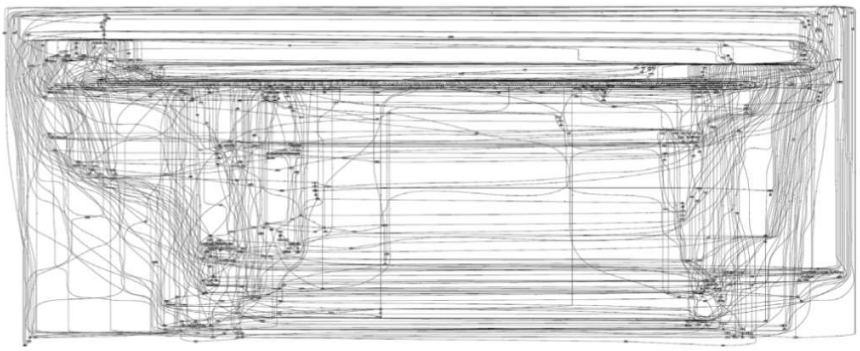
Little invasive;

Extremely high granularity;

28

Lessons learnt about PM (2)

High granularity is a double-edged weapon!



29

Lessons learnt about PM (3)

Additional issues

- Gap between the IS process flow and the actual process flow
- Issues in multi-stakeholder contexts (data ownership and privacy, authorizations, liability)
- Need for specific skills (data analysis, algorithm selection, PM-related know-how, attribute selection, data cleansing…)

30

Future developments for PM-enabled auditing


Multi-sourcing from Internet of Things


Scalability to big Data?


Integration with management dashboard


On-line, real-time auditing


Predictive auditing

31

UNIVERSITY OF PISA
Department of Energy, Systems, Territory and Construction Engineering

Thank you

Pierluigi Zerbino, PhD


pierluigi.zerbino@ing.unipi.it

32

References

Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. In *Computers in Human Behavior*, 27, pp. 1834–1839.

De Weerdt, J., De Backer, M., Vanthienen, J., & Baesens, B. (2012). A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Information Systems*, Vol. 37 No. 7, pp. 654-676.

Günther, C. W., & van der Aalst (2007). Fuzzy mining – adaptive process simplification based on multi-perspective metrics. In *Business process management*. In Lecture notes in computer science: 4714, pp. 328-343.

Lang, M., Bürkle, T., Laumann, S., & Prokosch, H.-U. (2008). Process mining for clinical workflows: Challenges and current limitations. Studies in Health Technology and Informatics, 136, pp. 229–234.

Weijters, A. J. M. M., van der Aalst, W. M. P., & De Medeiros, A. K. A. (2006). Process mining with the heuristics miner algorithm. In Process mining with the heuristics miner algorithm: 166 (pp. 1-34). Technische Universiteit Eindhoven. Tech.Rep.WP.

Zerbino, P., Aloini, D., Dulmin, R., Mininno, V. (2018). Process-Mining-enabled audit of Information Systems: methodology and an application. *Expert Systems with Applications*, 110, pp. 80-92.