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

Process-Mining-enabled audit of Information Systems

The case of a Mediterranean port

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

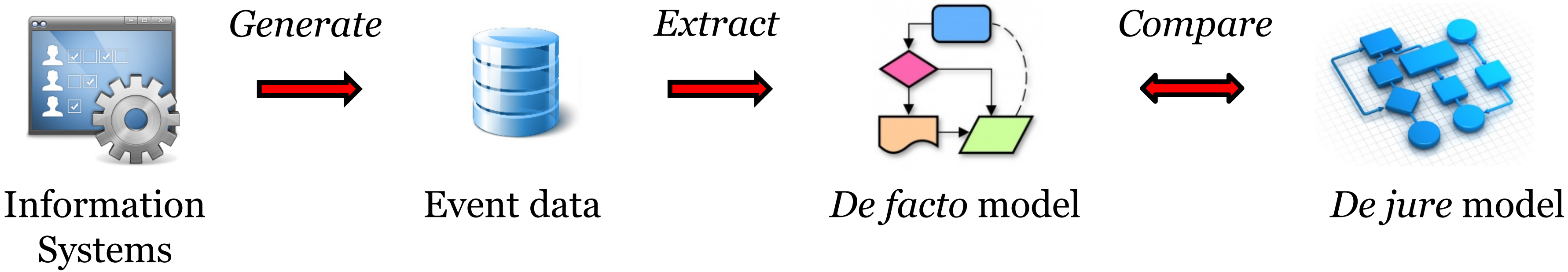
Agenda

- Overview about Process Mining (PM)
- Context of the case
- Results and discussion
- Lessons learnt
- Further PM capabilities in auditing

Process Mining (1)

Process Mining is a set of data-driven techniques for analysing and enhancing Information-System-supported business processes

Information Systems (ISs) do not behave as they are supposed to do



Process Mining (2)

Scope of PM applications

Every business process supported by an IS (workflow system)

Healthcare

Accounting

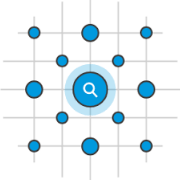
Logistics

Operations

... and others

Process Mining (3)

PM types

 Discovery

 Conformance

 Enhancement

PM types can be combined!

Discovery

+

Conformance

=

 **Audit**

The case (1)

Audit of a logistics process

To check the compliance with the prescribed document and process flow

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

Logistics port process of a Mid-sized Mediterranean port

1. Data-intensive;
2. Critical to value creation;
3. I like ports and sea.

The case (2)

Mid-sized Mediterranean port

> > 800000 TEUs and
> $> 30 \cdot 10^6$ tons **in 2016**

1st semester **2017**: increase in number of vessels, and decrease in TEUs

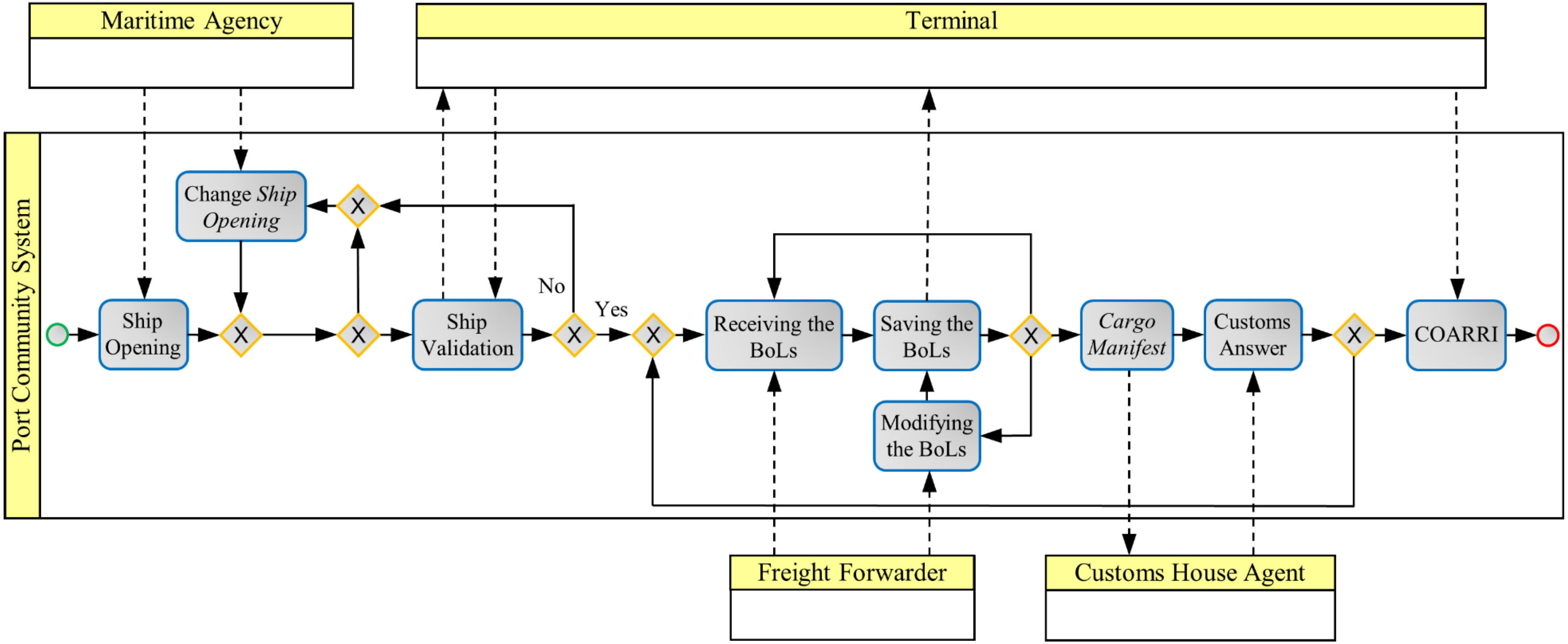
Port Community System

Open electronic platform enabling information exchange among port stakeholders through a single point of data entry

- Export;
- Import;
- National cabotage;
- Port taxes;
- Certified weighing.

The case (3)

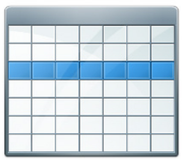
De jure model



The case (4)

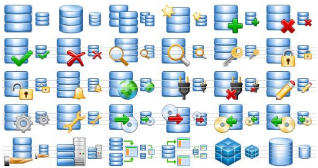
Dump of data

> 15 GB



52 tables

15 views



10^2 attributes

Heading

Ciphered

The case (5)

Excerpt of the export data

The case (6)

01 / Jan / 2016
30 / Jun / 2016

Data cleansing and filtering

Final dataset

686 cases (vessels)
171630 events

Complexity Index
(Zerbino et al. 2018)

$$\mu = \frac{number\ of\ process\ instances}{total\ process\ instances} = \frac{624}{686}\left[\frac{instances}{instances}\right] = 90.9\%$$

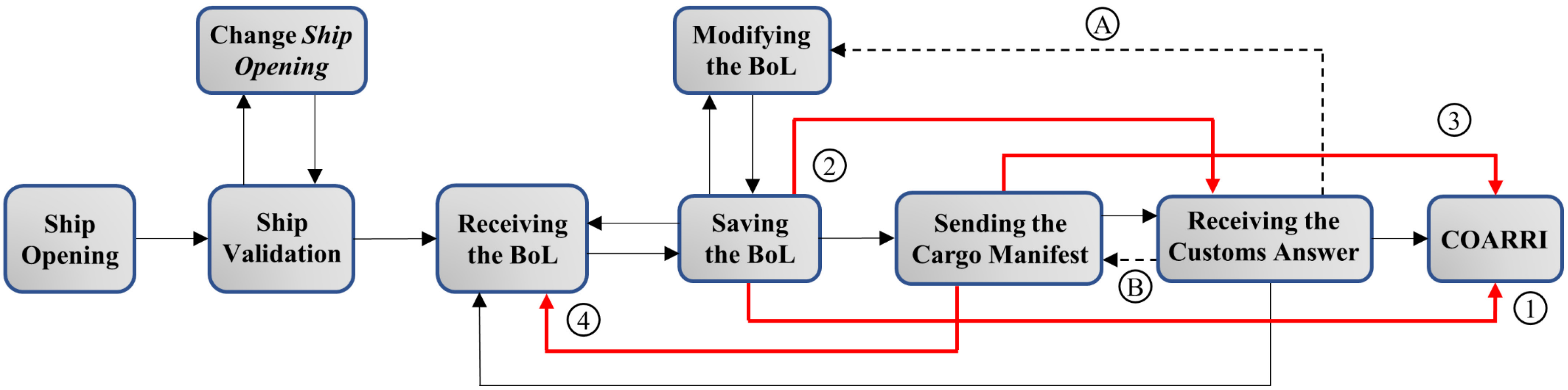
The case (7)

Screenshot of the event log

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

Results (1)

De facto model



Correct flow

Minor nonconformance

Major nonconformance

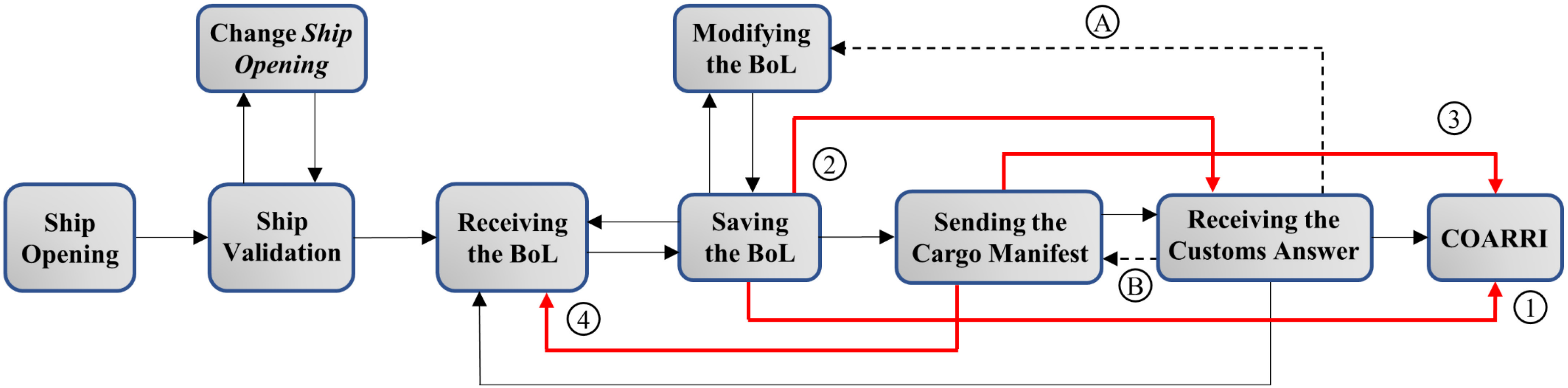
Results (2)

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

Discussion (1)

Arches A and B

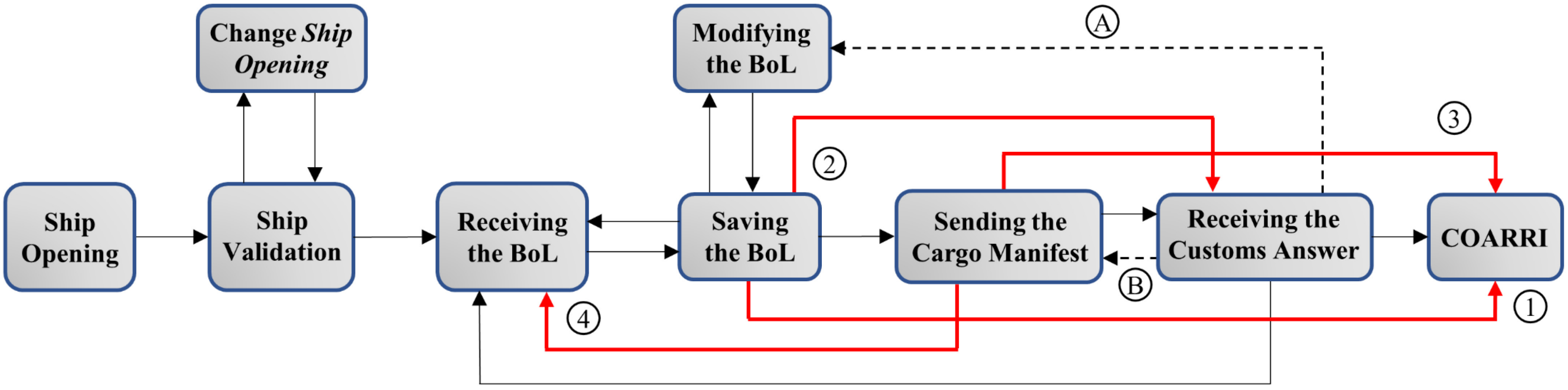


The *de jure* requires a feedback to *Receiving the BoL*

Arches A and B were consciously introduced by the programmer

Discussion (2)

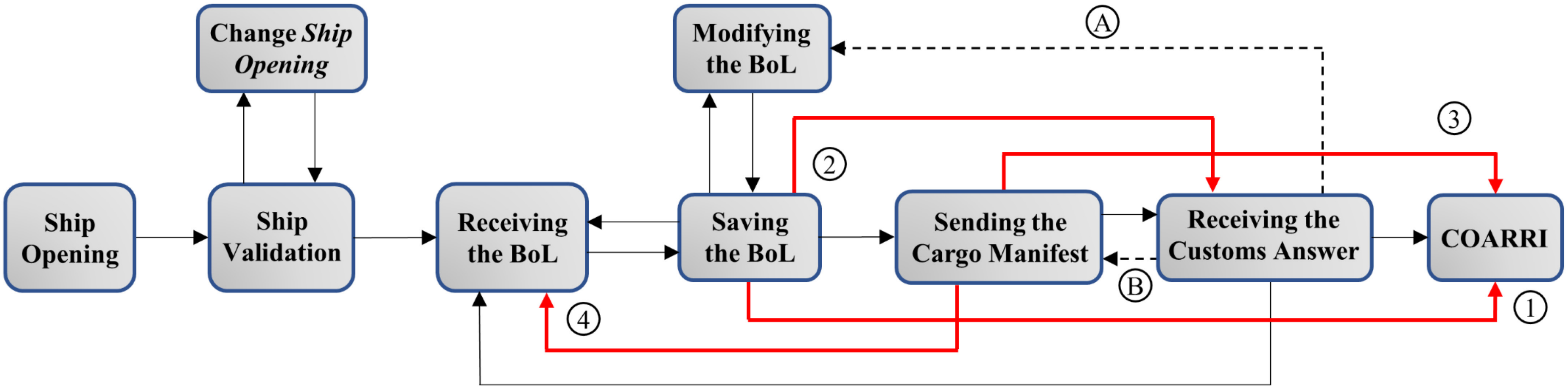
Arch 1 *(case frequency = 54)*



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

Discussion (3)

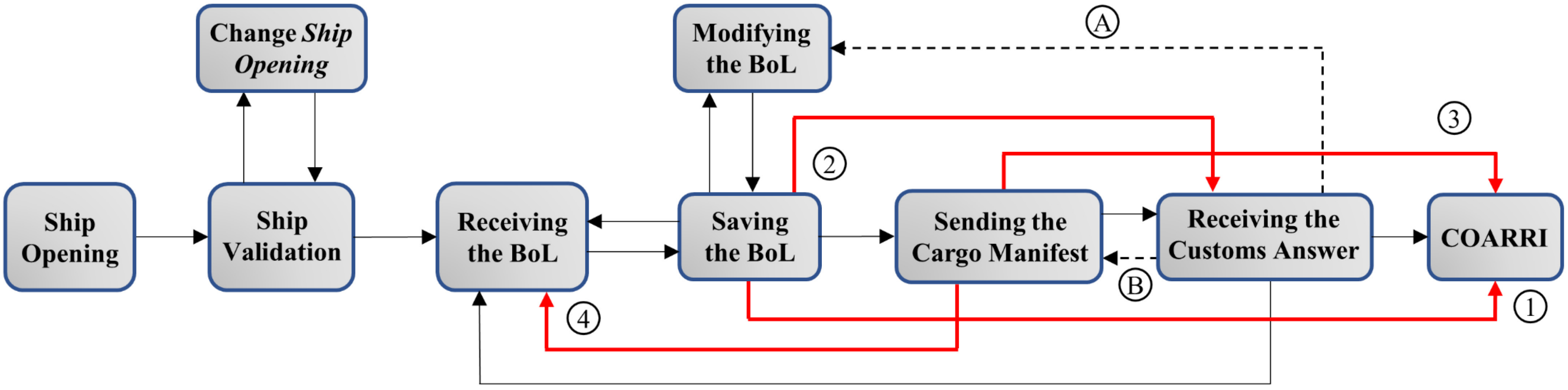
Arch 2 *(case frequency = 244)*



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

Discussion (4)

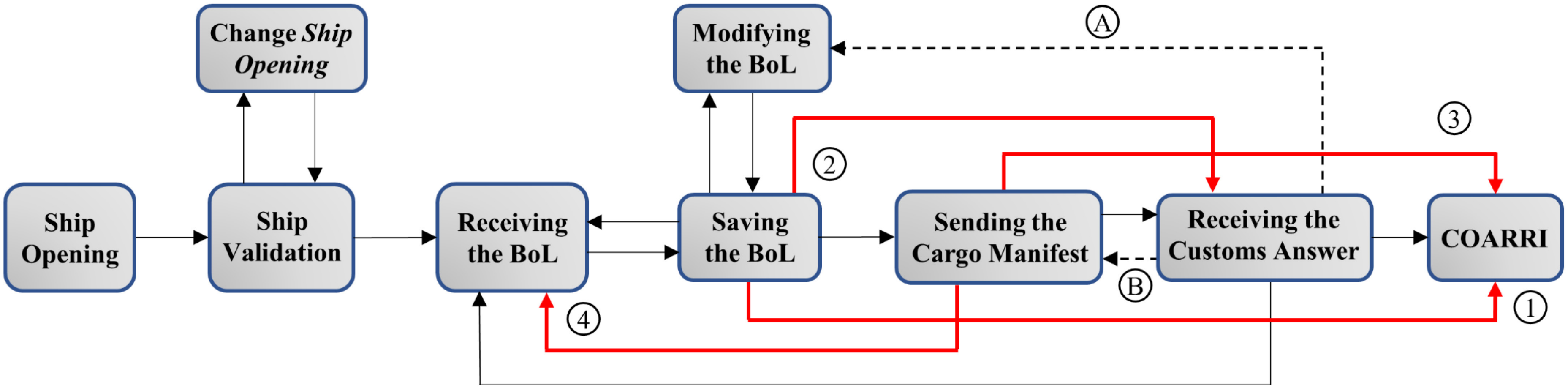
Arch 3 *(case frequency = 29)*



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

Discussion (5)

Arch 4 *(case frequency = 214)*

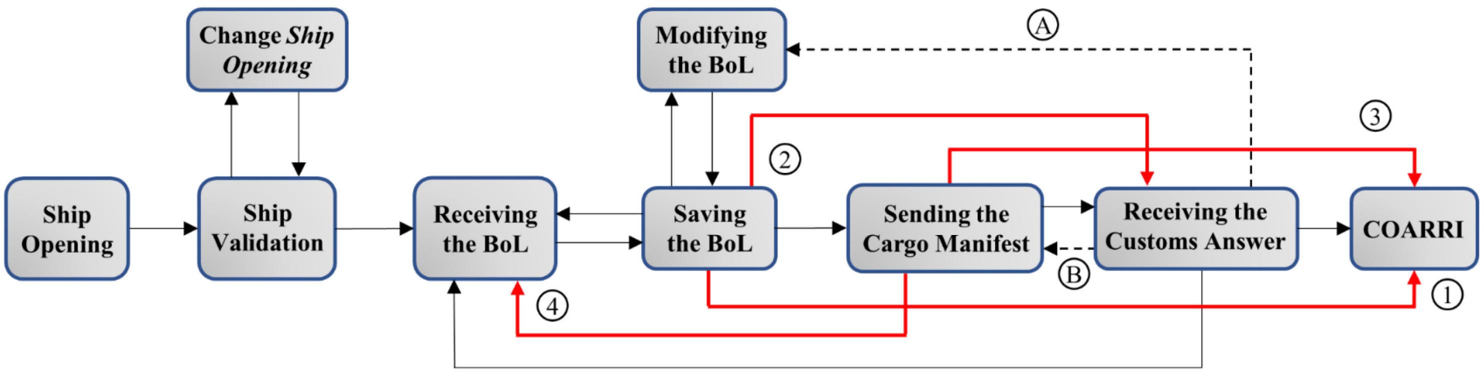


- All the cases present a CM;
- In 22 cases, the BoL were modified without updating the CM or checking for its correctness.

Implications (1)

No issues with Customs Answer were detected …

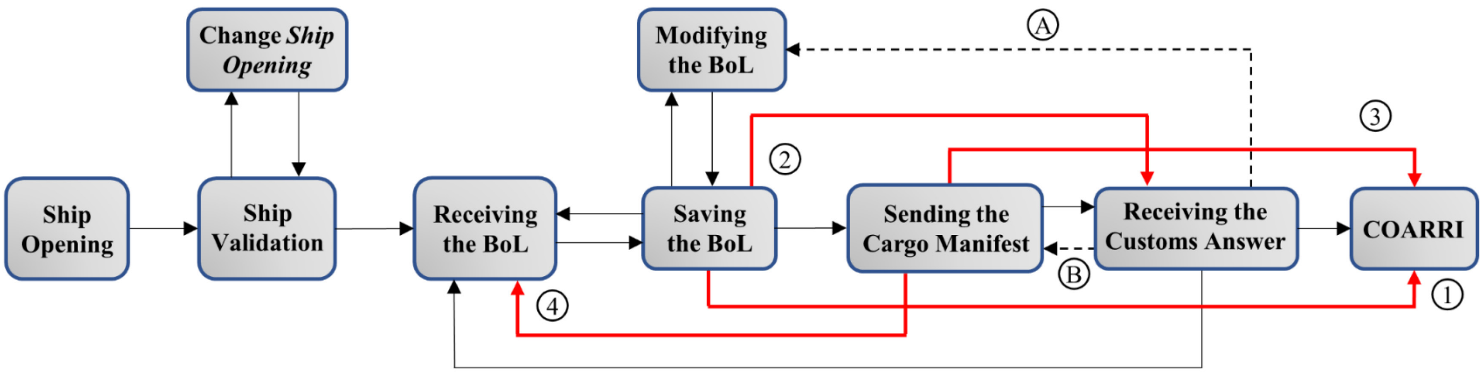
… but there might be a gap between what is declared within the CM and what is actually loaded; (further investigation needed)



Implications (2)

Possible interventions

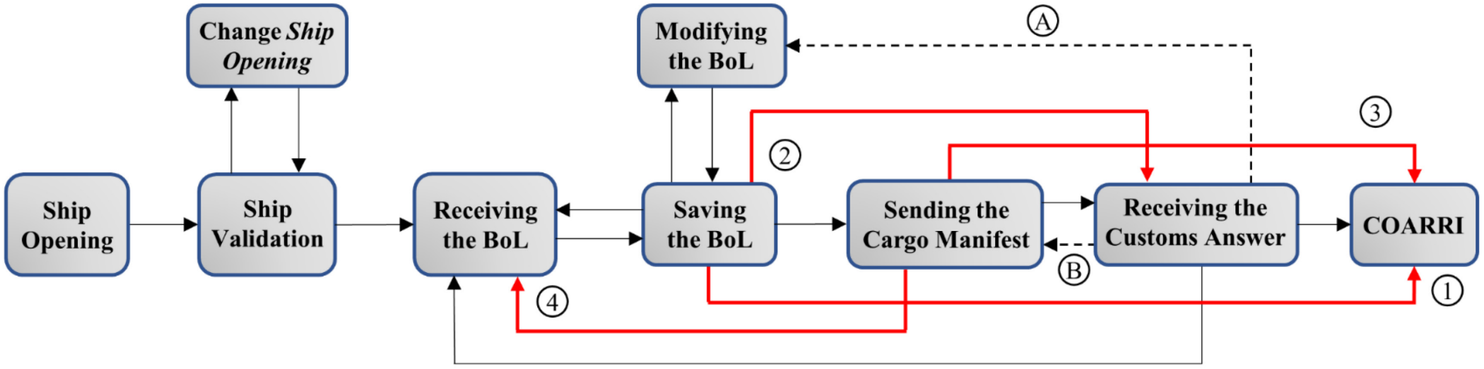
- To impede modifications in the BoL if the CC is positive;
- No feedback to the BoL process after the submission of the CM but before receiving the Customs Answer;



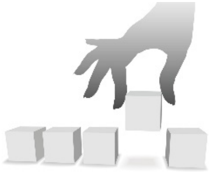
Implications (3)

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);

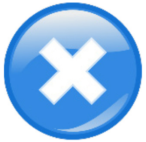


Lessons learnt about PM (1)

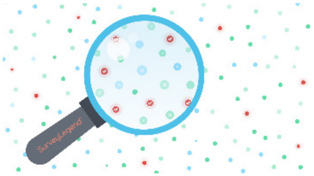


No need for sampling;

Strongly automatable;

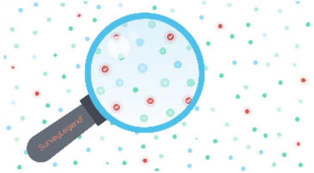


Little invasive;

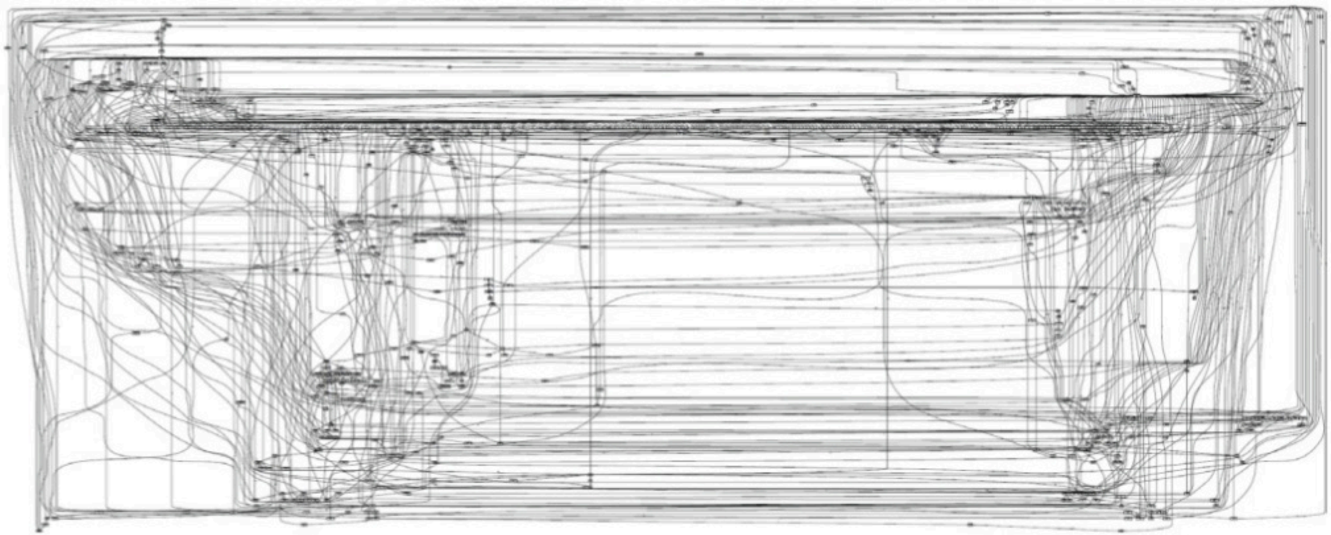


Extremely high granularity;

Lessons learnt about PM (2)



High granularity is a double-edged weapon!



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...)

Lessons learnt about PM (4)

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.

Thank you

Pierluigi Zerbino, PhD

pierluigi.zerbino@ing.unipi.it

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.

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