Data driven audit

Case studies and field experiences

Participants' contributions

Case studies and field experiences

Day 1

Understanding the data model
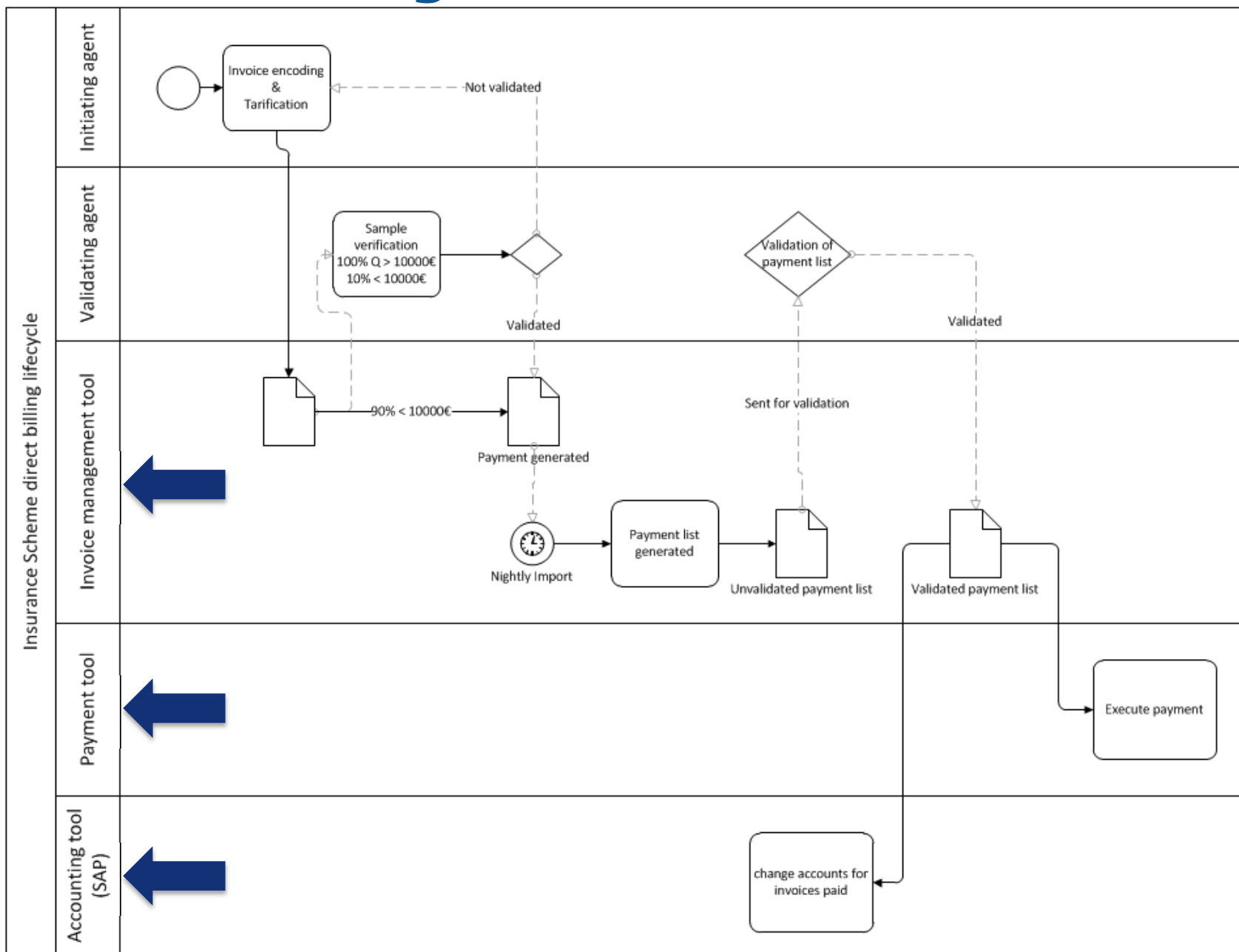
Use case integrated audit IAS

Benedict Van Rompuy
July 2018

Consultancy engagement data analytics

- *Topic: Accounting for invoices technical interventions*

- *Integrated team : accountancy and IT expertise required*

Understanding the data flow

Understanding the data model

INVOICES	
PK	**INVOICE_ID**
FK1	INVOICE_SCANNR
	INVOICE_TYPE
	PERSON_ID
	PERSON_ID_OBTAINED_FROM
	INVOICE_DATE
	DATE_START
	DATE_END
	LAST_EXAM_DATE
	TOTAL_AMOUNT_CURRENCY
	TOTAL_AMOUNT_IN_EURO
	AMOUNT_TO_BPAID
	INVOICE_COMMENT
	INVOICE_REFERENCE
	SYSADMIN_DATE_MODIFY
	SYSADMIN_REASON_MODIFY
	TARIFICATION_BYPASSED_FLAG
	PAYMENT_DATE

INVOICE INJECTIONS	
PK	**INVOICE_INJECTION**
FK2	PAYMENT_ORDER_ID
	PAYMENT_AFFILIATE
FK1	INVOICE_SCANNR
	INVOICE_TYPE
	DATE_CREATE
	USER_CREATE
	DATE_PAYMENT

PAYMENT ORDERS	
PK	**PAYMENT_ORDER_ID**
	PAYMENT_DATE
	PAYMENT_AFFILIATE
	AMOUNT_TO_BPAID
	AFFILIATE_BANK_NUMBER
	AFFILIATE_COUNTRY
	AFFILIATE_BANK
	PAYMENT_ORDER_EXECUTED_FLAG

TARIFICATION COMMENTS	
PK	**COMMENT_ID**
	OBJ_CD
	OBJ_ID
	OST_CD
	PER_ID_DEFINER
	PER_ID_CLOSER
	CMT_CONTENT
	DATE_CREATE
	USER_CREATE
	DATE_MODIFY
	USER_MODIFY
	SYSADMIN_DATE_MODIFY
	SYSADMIN_REASON_MODIFY

- *Test requested : average throughput time for invoice payments?*

Conclusion

- *Design of tests prone to error when not understanding the business process*

- *Crucial to understand IT tools and processes behind*

- *Unexpected recommendations stemming from business analytics efforts*

Risk-based audit planning

A data driven scoring method for selecting audited entities

Angelo Maria Quaglini
Corte dei conti – Abruzzo Regional Audit Chamber

CORTE DEI CONTI

Building up the annual audit plan on local municipalities and provinces: challenges, needs and data sources

- Main challenges:

  - Number of auditable entities: 309 local jurisdictions (municipalities and provinces);

  - Insufficient HR involved in the audit (3 magistrates and 5 auditors).

- Main needs:

  - Identify the most critical situations in order to prioritize our audit;

  - Reduce the time-lag between the audit and the ongoing management.

- Data sources:

  - Several dataset are available: Public administration database (including financial statements figures), cash transactions dataset (so called SIOPE), questionnaires filled by the local entities' external auditors.

Forward looking analysis based on cash flaws

- Cash transactions records are available on a daily basis;

- Opportunity to monitor the municipalities without delays since liquidity shortages show financial problems and anticipate financial distress;

- Massive volume of records (all the single transactions for all the public administration);

- Software to determine the mid-year cash flaws for all the municipalities;

- It allows us to:

  - Detect municipalities with liquidity problems;

  - Set a threshold for cash advances by the treasury bank above which there is a high risk that the entity will not be able to return it at year end (20% of the current income);

  - Order the identified entities to take the necessary measures to solve the problems by the end of the year.
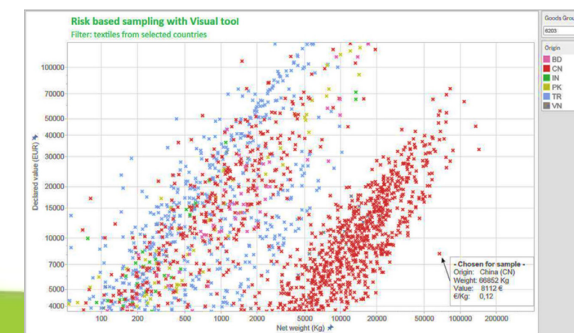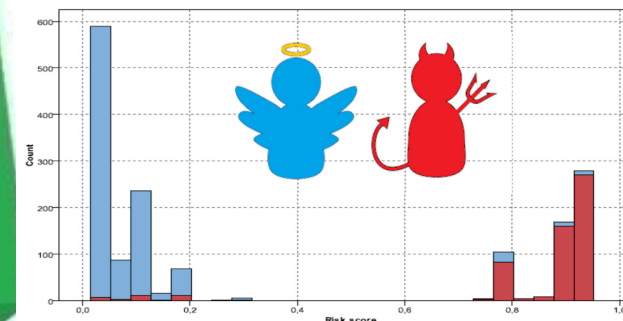
Finding risk and fraud in EU customs

with Risk-Based Sampling and Visual Data Analysis

Carlos Soler – Head of task

Jesús Nieto – Technical support

European Court of Auditors

How? With Risk models – with Data Mining Tool

Experience → Risk model → Risk scores/predictions
findings + sample

Experience

year	id	info	Audit
2016	1	lorem	✔
2016	37	ipsum	✘
2016	128	dolor	✘
2016	378	sit	✔

1- train

year	id	info	Audit
2016	526	amet	✔
2016	812	consectetur	✘

2- test

Data Mining Tool

Risk **model**

xx% accuracy

3- use

New population

year	id	info
2017	1	adipiscing
2017	2	elit
2017	3	sed
2017	4	do
2017	5	eiusmod
2017	6	tempor

Risk prediction

year	id	info	Risk
2017	1	adipiscing	✓
2017	2	elit	✓
2017	3	sed	✗
2017	4	do	✓
2017	5	eiusmod	✓
2017	6	tempor	✗

+ sample

Risk-based sample with Data Visualisation
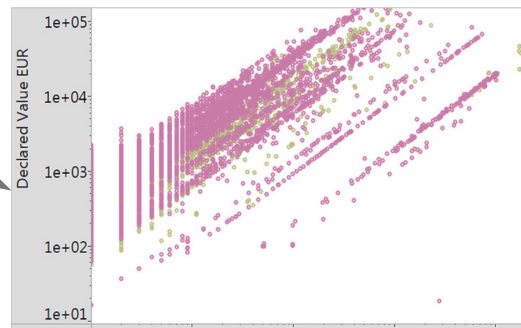
Experience &
professional judgement→
Risk model

Risk predictions
+ sample

New population

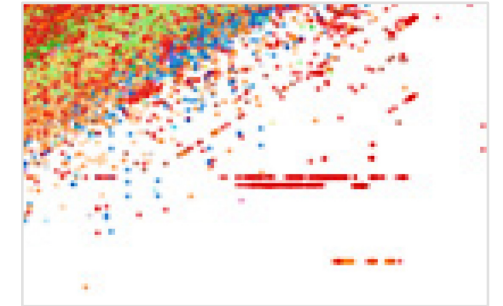| year	id	info
2017	1	adipiscing
2017	2	elit
2017	3	sed
2017	4	do
2017	5	eiusmod
2017	6	tempor

**Data
Visualisation**

Conclusions

Technical

1. **Data Visualisation** can show risks

2. **Visual Risk-Based Sampling** can be very effective

Business

3. **Hits** increase the

CREDIBILITY of

"you should improve controls"

Case studies and field experiences

Day 2

Evaluating government policy in absence of clearly defined goals: combining output indicators

Examples from the Belgian Court of Audit

Evaluating effectiveness of government policy

What if they are not (clearly) defined?

- Not evaluate
- Use own definition/indicator/…
- Combine multiple "suboptimal" solutions

Combining output indicators: examples from the Belgian Court of Audit

1. Equal education opportunities in regular primary education
2. Governing bodies of school in primary and secondary education (ongoing audit)

Open questions

- Consequences for acception of the audit?
- Is it an auditor's task to collect/compose indicators or should the audit be limited to existing ones?
- Other ways to collect indicators?

Data analytics for the audit of EU Funds' projects -
challenges and opportunities

Presentation in the context of the *Summer School in public auditing and accountability. Data mining and analytics: what implications for auditing*
Pisa, Italy, July 2018

EY
Building a better
working world

Innovating the audit to deliver
quality and value.
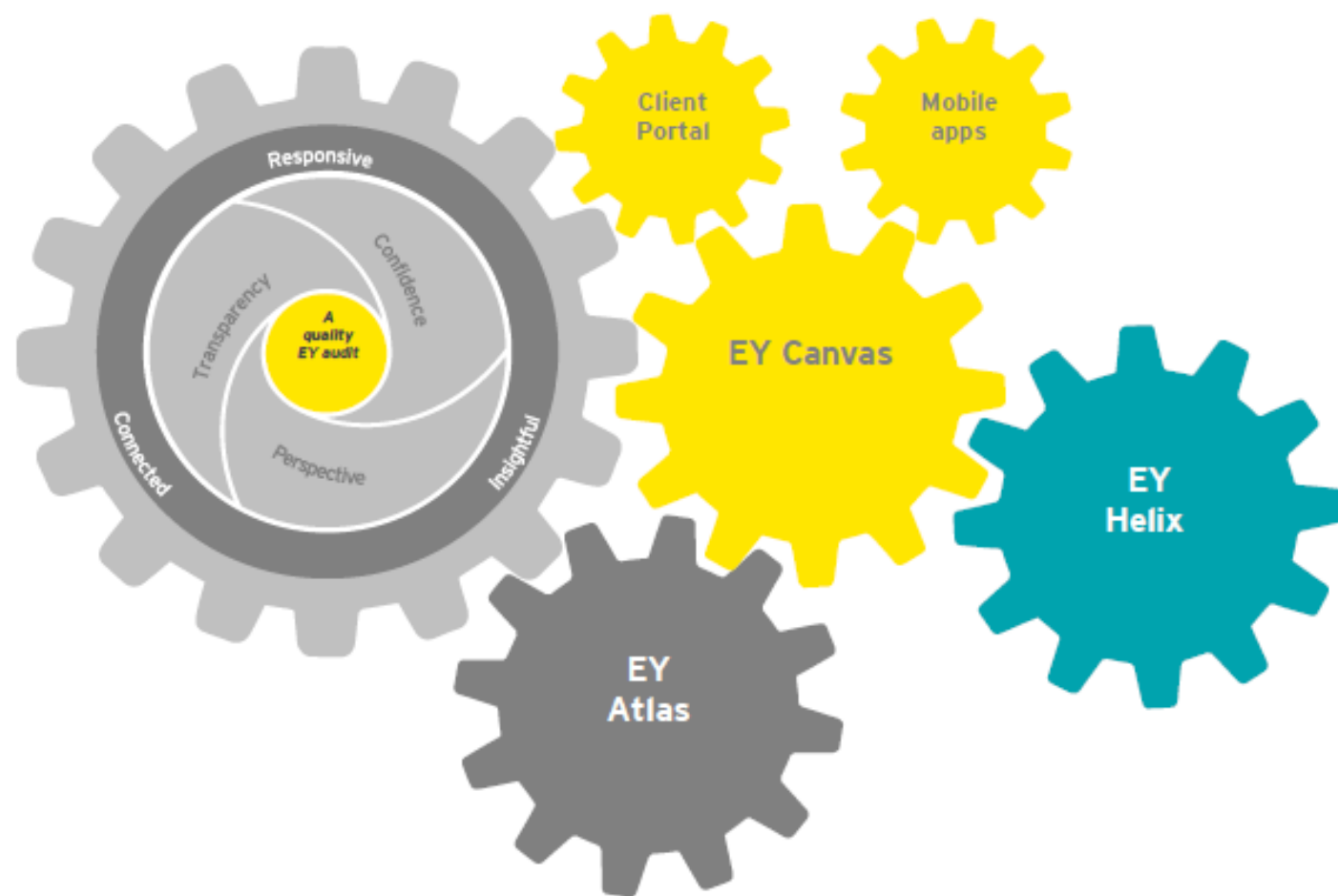
EY Canvas, our global online audit platform lies at the heart of the audit. EY Canvas **Client Portal** is integrated into EY Canvas and connects our clients directly with our auditors. EY Canvas **mobile apps** connect our people around the globe and allow them to support our clients on the go.

EY Atlas provides our auditors and our clients with the latest accounting and auditing content.

EY Helix, our suite of data analytics, are embedded in our audit process and help us provide deeper insights to our clients.

Opportunities

- **An approach for the digital age**
  The appeal of using data analytics likely to increase with beneficiaries that are themselves working with new tools and approaches.

- **Alignment with new ways of delivering financial support**
  Traditional ways of delivering funding are changing: **from actual cost and focus on grants to simplified cost options and blended instruments.** Emphasis on **performance.**

- **Full analysis of project expenditure**
  Change of approach to audit, from samples and focus on extrapolations to full testing and focus on correlations and analysis of patterns.

- **Likely easier detection and treatment of problematic eligibility aspects**
  A full analysis of the GL may reveal cases of **double funding.** Comprehensive income analysis may help detect the receipt of **operating grants.** Detailed analysis is facilitated in the case of **actual indirect costs.**

- **Decrease of administrative burden on auditee**
  With the decrease in intensity and volume of substantive testing, the time investment of the auditee is likely to decrease.
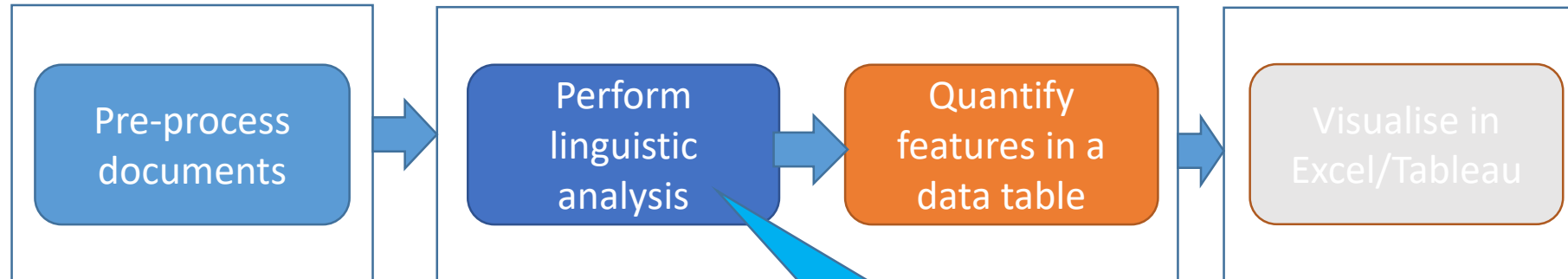
EY

Text mining pipeline

Pre-process documents → Perform linguistic analysis → Quantify features in a data table → Visualise in Excel/Tableau

- Named Entity Recognition
- Vocabulary analysis
- Sentiment analysis

Typical dilemmas:

- Off-the-shelf vs Custom developed

- Free vs Commercial

- Public/open vs Private/internal data

- Artificial vs Natural intelligence ☺

Start small, think BIG!

Find out if there's a „data lab" or somebody already experimenting with text mining within your organisation and team up with them. If not, be a champion yourself!

The dataset enrichment (1/2)

Document Set

Document

Clustering (K-means)

Topic Modeling (LDA)

Categorisation (JRC EuroVOC)

Summarisation (Gensim)

The dataset enrichment (2/2)

Paragraph

Keywords, Categories, Concepts, Entities (Watson)

Doc2Vec (Spacy - GloVe)

Sentence

Sen2Vec (Spacy - GloVe)

The evaluation

Final result is way too complex!
(ex: 15 documents, just the main category...)

Interpretation needs a better user interface!

Case studies and field experiences

Day 3

Gaining new insights by combining existing data sources in new ways
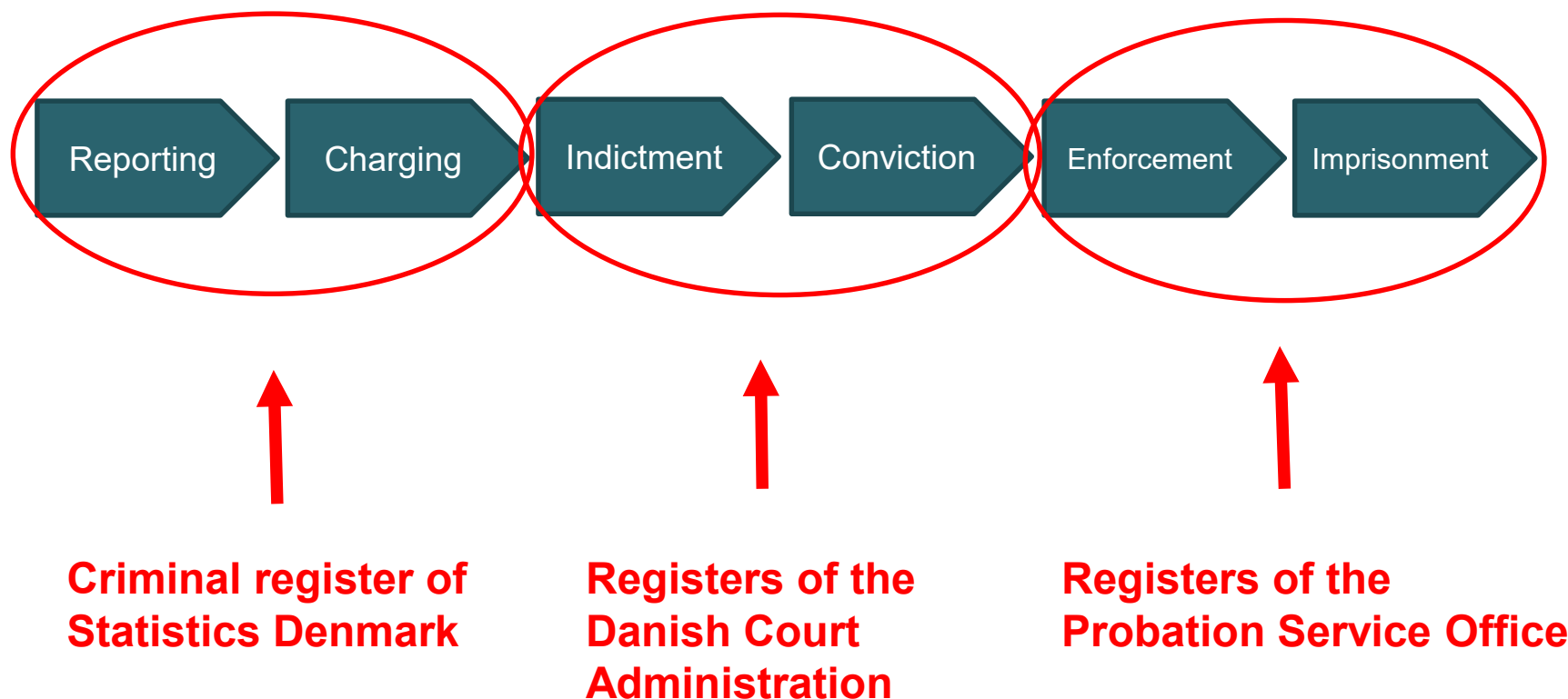National Audit Office of Denmark

Niels Dyhrberg-Nørregaard
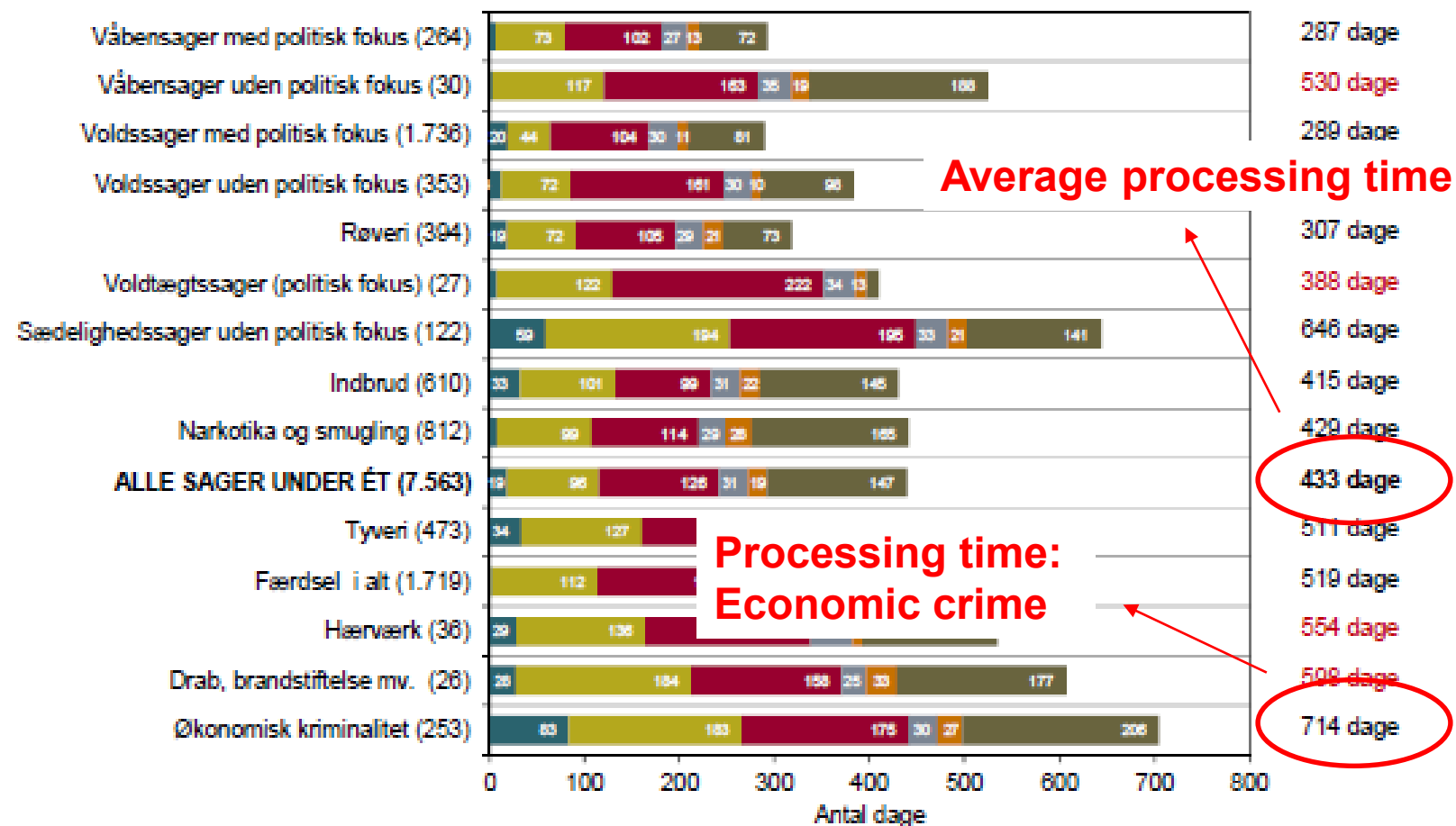*PhD, Senior Advisor*

Zuzanna Thiel
Senior Advisor

RIGSREVISIONEN

Tracing cross-institutional flows:
Processing time of criminal cases #1

- How much processing time from a criminal case is reported until the convicted person begins serving a sentence?

| Reporting | Charging | Indictment | Conviction | Enforcement | Imprisonment |

Criminal register of Statistics Denmark

Registers of the Danish Court Administration

Registers of the Probation Service Office

33

Tracing cross-institutional flows:
Processing time of criminal cases # 2

Average processing time

**Processing time:
Economic crime**

Swedish NAO experience

with data driven performance audit and model for long term predictions
 - PISA 2018

NILS LUNDH

HELENA FRÖBERG

SWEDISH NATIONAL
AUDIT OFFICE

Introduction

› Increased focus on data driven performance audit and quantitative methods in recent years

› Focus on:

– Methods for establishing causality in effect studies

– Better availability to micro data from public entities

– Using analysis of quantitative data to achieve more credible audit reports

› Big data analysis

– No current ongoing projects or strategies involving "big data"

– Probable causes:

• *Uncertain legal status (can we collect data for this purpose?)*

• *"Audit culture" (can we use big data in performance audit?)*

• *Current staff has other competences (economists, political scientists, sociologists)*

• *Data needed may not be structured for big data analysis*

SWEDISH NATIONAL
AUDIT OFFICE

Abolition of audit obligation for small limited companies

› Objective: impact evaluation of abolition of audit obligation

› Data sources: Statistics Sweden, Swedish companies registration office, Bisnode AB and Swedish inspectorat of auditors.

› Coarsened exact matching to identify comparative group

› Measured growth, errors in annual reports, risk of tax evasion

› Result: reform did not stimulate growth, errors increased, companies that opted out were mainly in industries/sectors of high risk of economic crime and tax evasion

SWEDISH NATIONAL
AUDIT OFFICE

SESIM – model for long term prediction

› SNAO manages a dynamic simulation model called SESIM

› Developed by government in cooperation with researchers (1997)

› Forecast horizon up to 100+ years

› For analysis of: pension system, ageing issues, lifecycle redistribution, generational redistribution

› **Dynamic ageing:** changes in individual, and household, characteristics over time by simulation of estimated statistical models.

› www.sesim.org (in English)

› Ongoing projects in performance audit using SESIM:
  – Government interventions for immigrant women outside of the labour force
  – Effects of fees in the premium pension on future payouts for individuals

SWEDISH NATIONAL
AUDIT OFFICE

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

- DK NAO experience
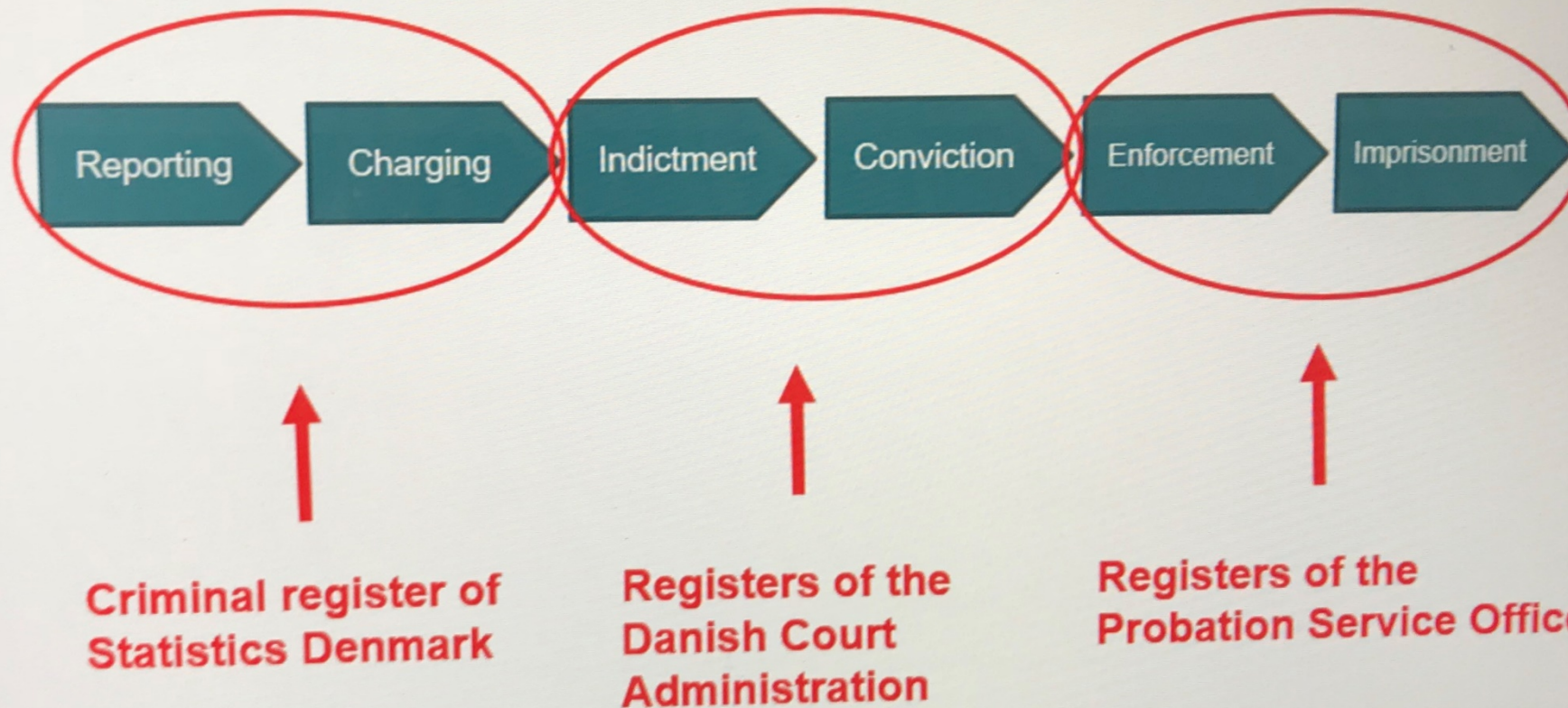
Tracing cross-institutional flows: Processing time of criminal cases # 2
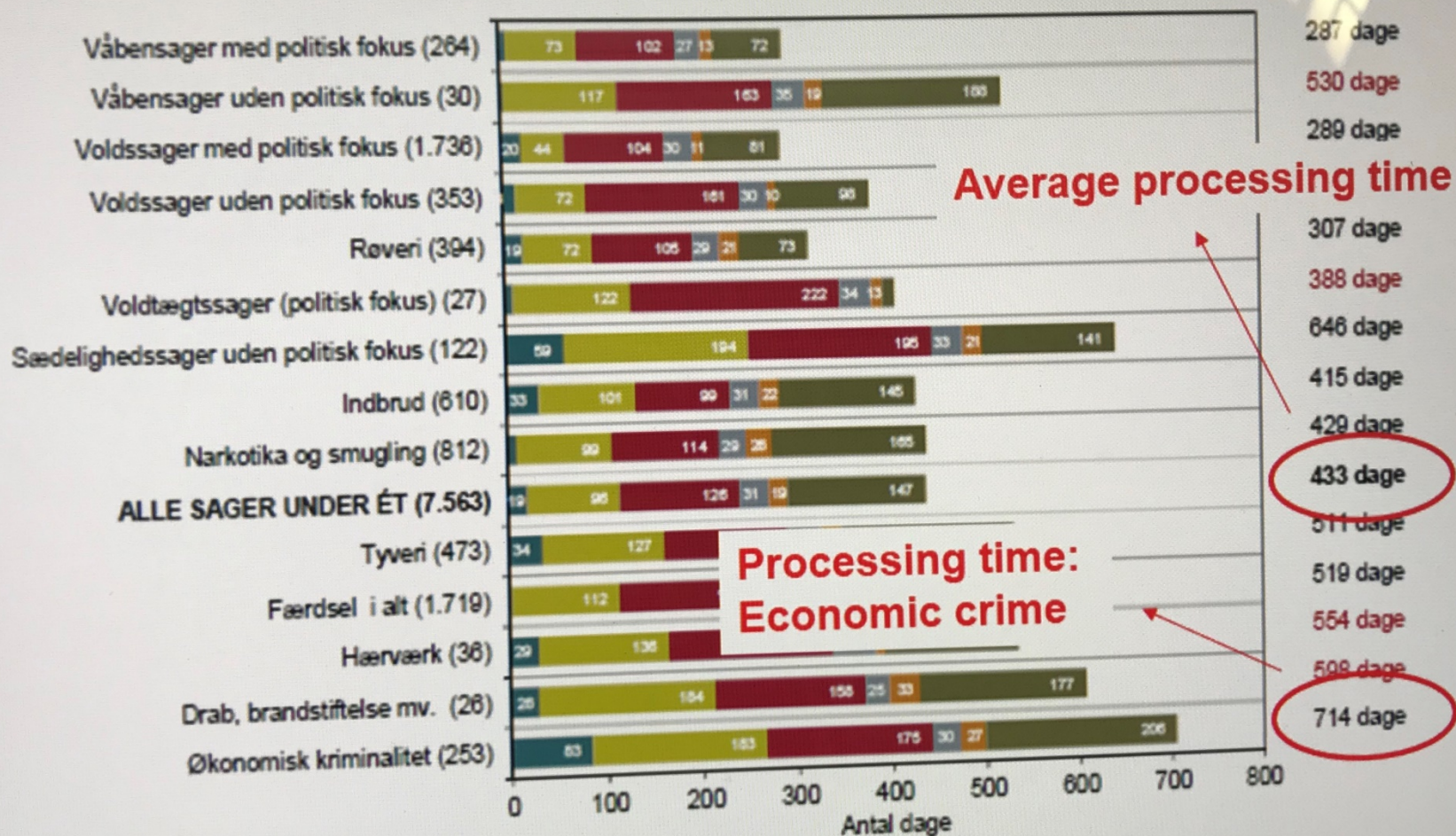
Process Mining (1)

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

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

Generate → *Extract* → *Compare* ↔
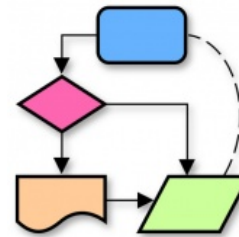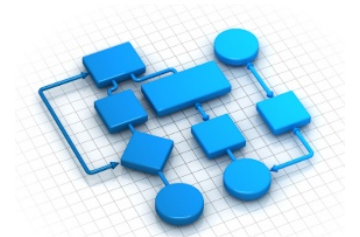
Information Systems → Event data → *De facto* model → *De jure* model

The case (1)

Audit of a logistics process

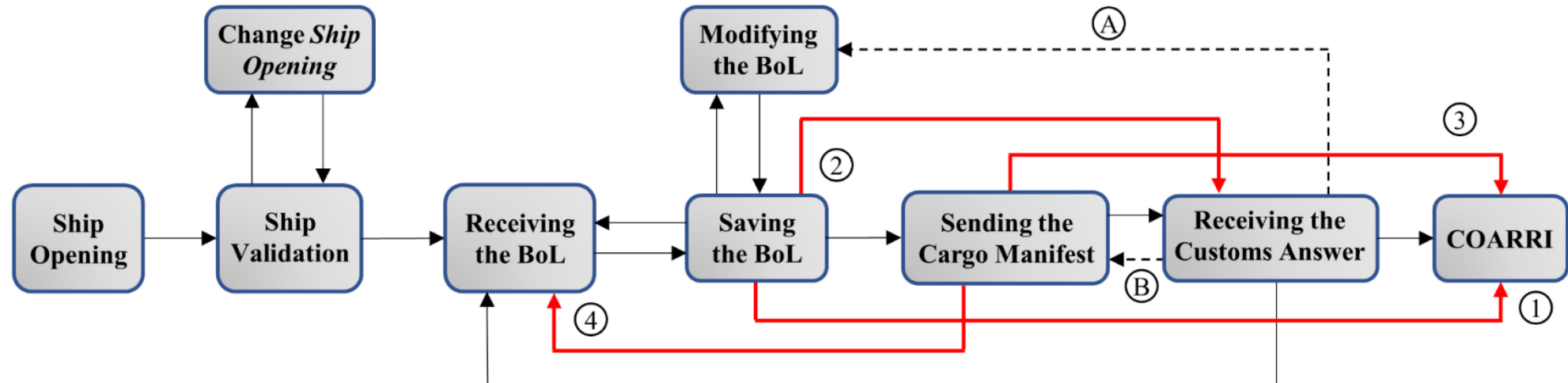*To check the compliance with the prescribed document and process flow*
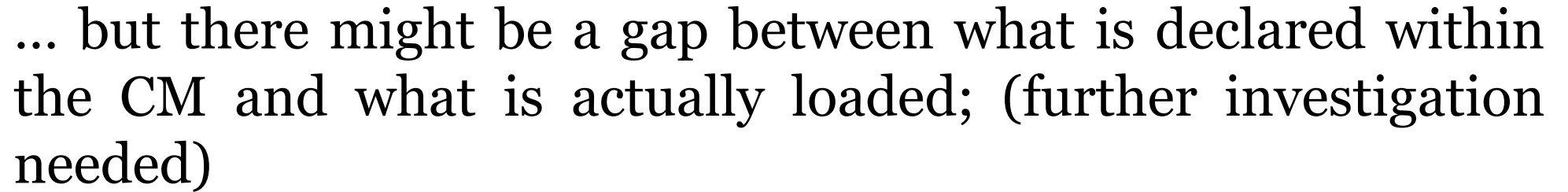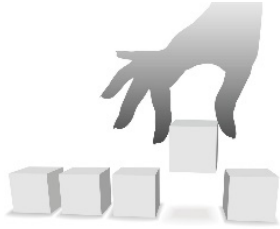
Customs   Control bodies   Regulations   Monitoring

Logistics port process of a Mid-sized Mediterranean port

❶ Data-intensive;

❷ Critical to value creation;

❸ I like ports and sea.

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)

Lessons learnt about PM (1)
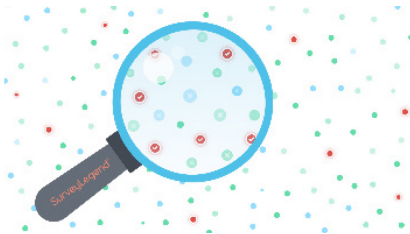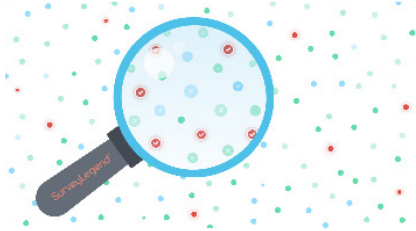
No need for sampling;

Strongly automatable;

Little invasive;

Extremely high granularity;

High granularity is a double-edged weapon!
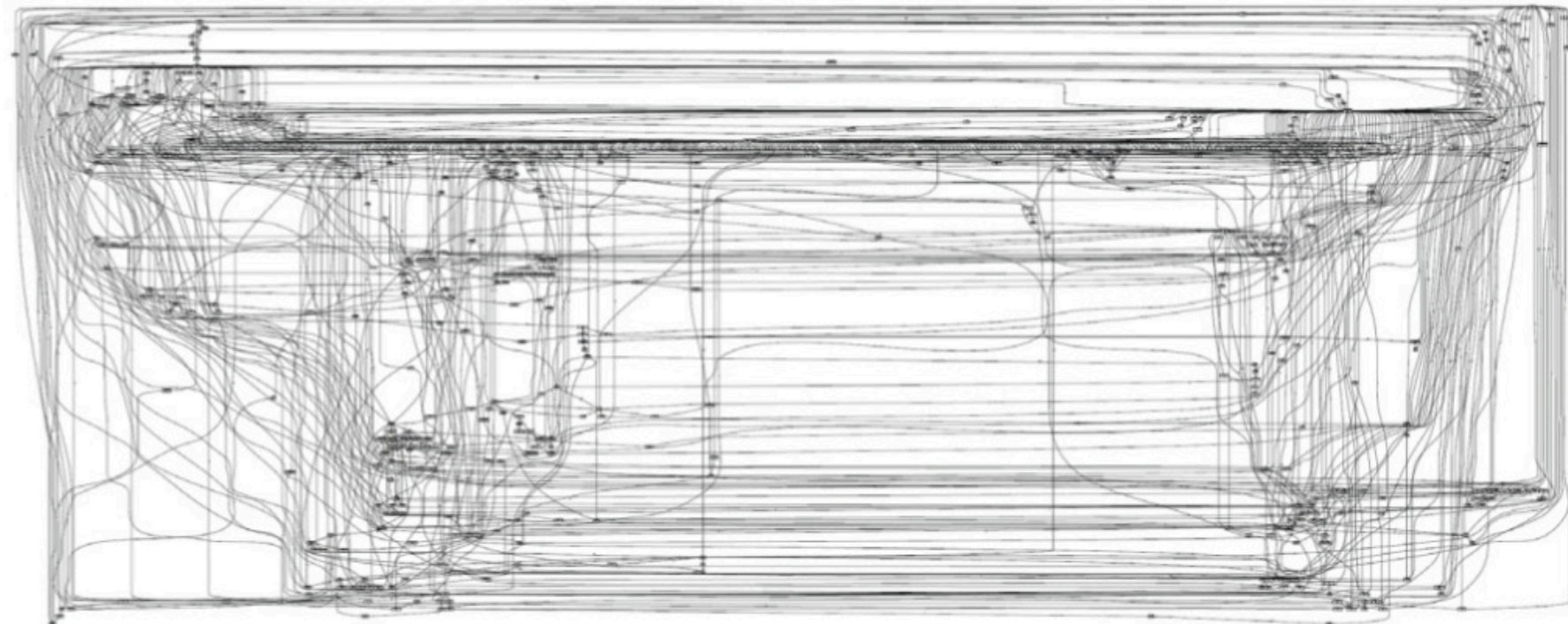
Lessons learnt about PM (3)

- 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...)

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.

Some notes

- Technology is not the problem. Choosing the right one is
- Integration of skills and mutual understanding
- Access to the right data in the right form
- The power is in the human brains