



# AUDIT DATA ANALYTICS AND PROCESS MINING

37 WCARS

GOLD COAST – AUSTRALIA

SEPTEMBER 14 2016

SOME RESEARCH EFFORTS







# OUTLINE

- Audit data analytics – ADA -> EDA- CGA
- Process Mining

# AUDIT DATA ANALYTICS

TREVOR STEWART AND QI LIU



# AUDIT DATA ANALYTICS (ADA)

## ONE WAY TO DEFINE...

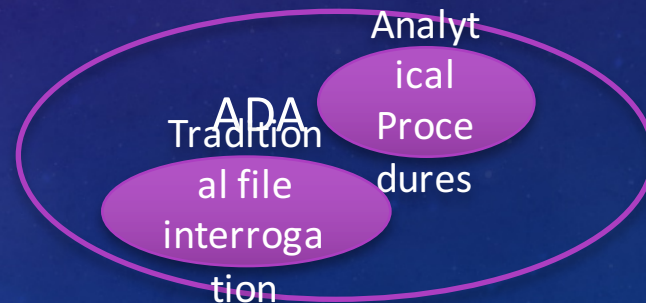
Audit Data Analytics (ADA) is the analysis of data underlying financial statements, together with related financial or non-financial information, for the purpose of identifying potential misstatements or risks of material misstatement.

ADA includes methodologies for:

- Identifying and analyzing anomalies in the data
- Identifying and analyzing patterns in the data including outliers
- Building statistical (e.g., regression) or other models that explain the data in relation to other factors and identify significant fluctuations from the model
- Synthesizing pieces of information from disparate analyses and data sources into wholes that are greater than the sum of their parts for purposes of overall evaluation

ADA defined in this way  
*includes:*

- Analytical Procedures (AU-C 520)—preliminary, substantive, and FS review—including reasonableness testing
- Traditional file



# ADA MODE CAN BE *EXPLORATORY* OR *CONFIRMATORY*

	Exploratory mode	Confirmatory mode
When	Planning	Performance
Question	What is going on here? Do the data suggest something might have gone wrong? Where do the risks appear to be? What assertions should we focus on?	Do the data conform with and thus confirm my model for what ought to be?
Approach style	Bottom-up, inductive, few starting assumptions, assertion-free	Top-down, deductive, model-driven, starts with development of model based on assertions to be tested
Methods	Graphical visualizations used to discover patterns in and understand the data—possibly several to get different viewpoints	Comparison of actual data to model taking into account materiality, desired assurance and assertions being tested; more mathematical than graphical
Results	Identified risks, areas of focus, potential models for confirmatory stage	Identified anomalies, unexpected patterns, outliers and other significant deviations



# ADA EXAMPLES

## Exploratory

- Cluster analysis
- Text and data mining
- Data visualization
  - Scatterplots
  - Scatterplot matrices
  - Line charts
  - Spread charts
  - Needle graphs
  - Small multiples of graphics
  - Heat maps
  - Treemaps
  - Relationship maps

## Confirmatory

- Analytical procedures
  - Regression analysis, ratio analysis
  - Reasonableness tests
- Recalculations
- Traditional file interrogation
  - Footing, extending
  - Duplicate detection
  - Out-of-range detection
  - Other 100% tests
- Journal entry testing (SAS 99)

Exploratory and confirmatory ADA is a spectrum of analytics and the processes are iterative, starting with exploratory

Technology	Immediate Application	TPR effect
Scanning and OCR	Contract evaluation	Income Verification Income forecasting All electronic records
Electronic records and plentiful computing	Full population feature testing	Process mining Preventive audit
Cloud	Ubiquitously available data for corporate applications	Shared corporate applications among companies Need for application standards across industries
Blockchain	Cybercoins	Public recordkeeping in multiple locations Real time audit in many locations and organizations simultaneously
Smart Contracts, large data stores, plentiful computing		Contract execution monitoring Automatic monitoring of process variances

## TECHNOLOGICAL PROCESS RETROFITTING (TPR)



The background is a dark blue gradient with a subtle pattern of white dots. Overlaid on the left side are several concentric circles and arcs in a lighter blue color. Some of these arcs have degree markings, such as 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, and 260. There are also small white arrows pointing in various directions, suggesting a circular or rotational theme.

# PROCESS MINING

MIEKE JENS (HASSELT UNIVERSITY)

MICHAEL ALLES (RUTGERS UNIV.)

# AN EXAMPLE OF AN EVENT LOG OF AN INVOICE

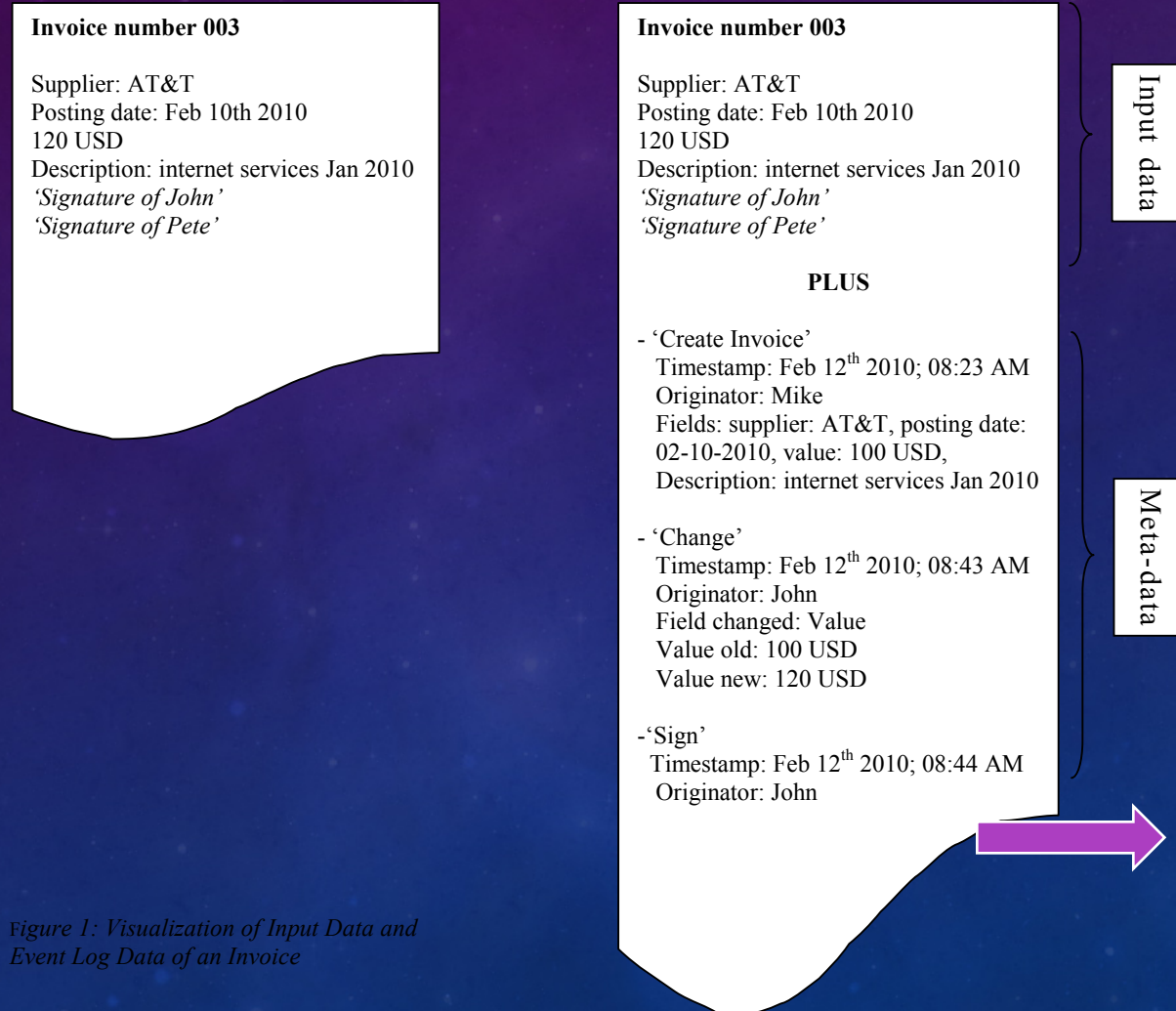
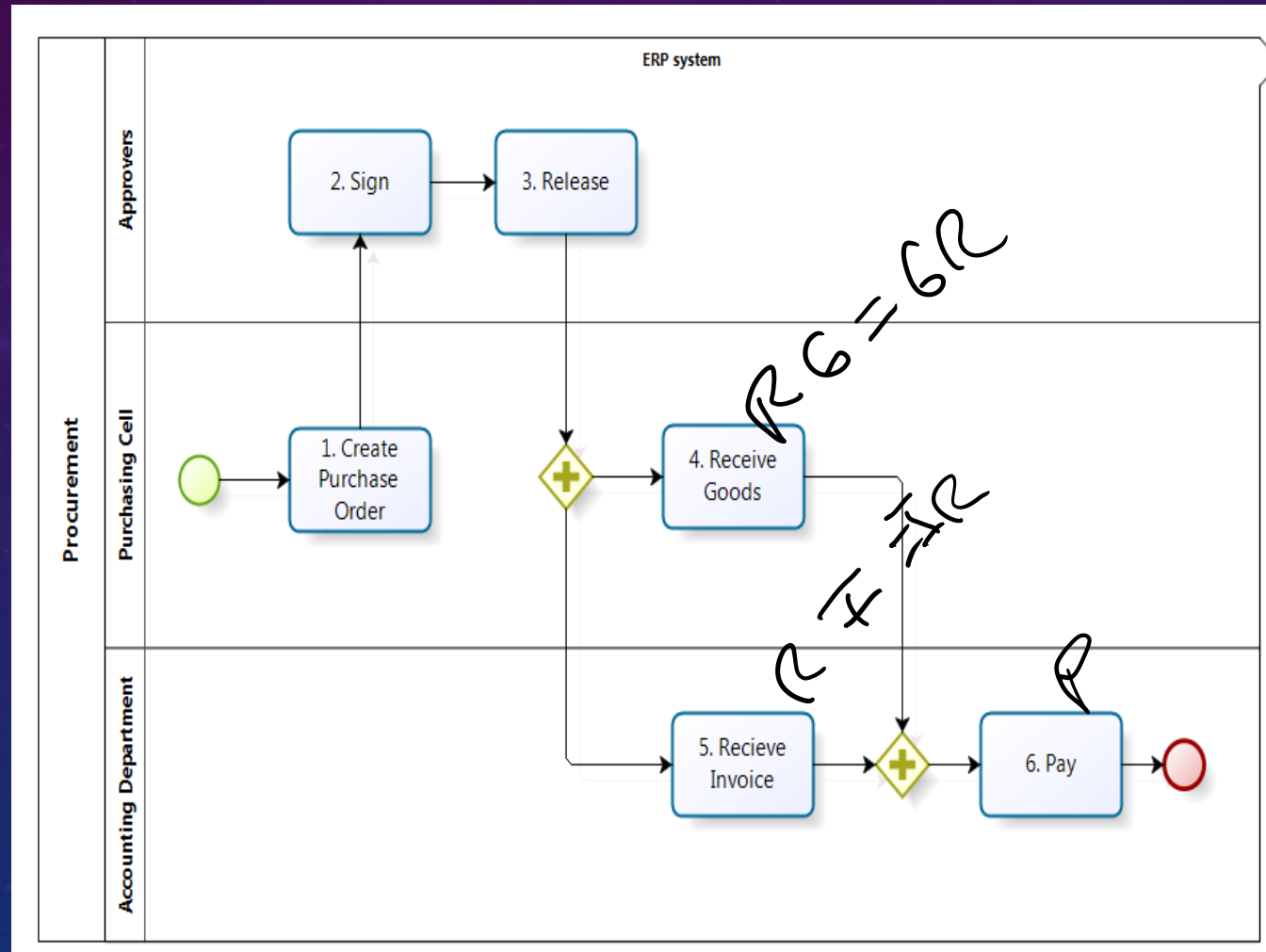


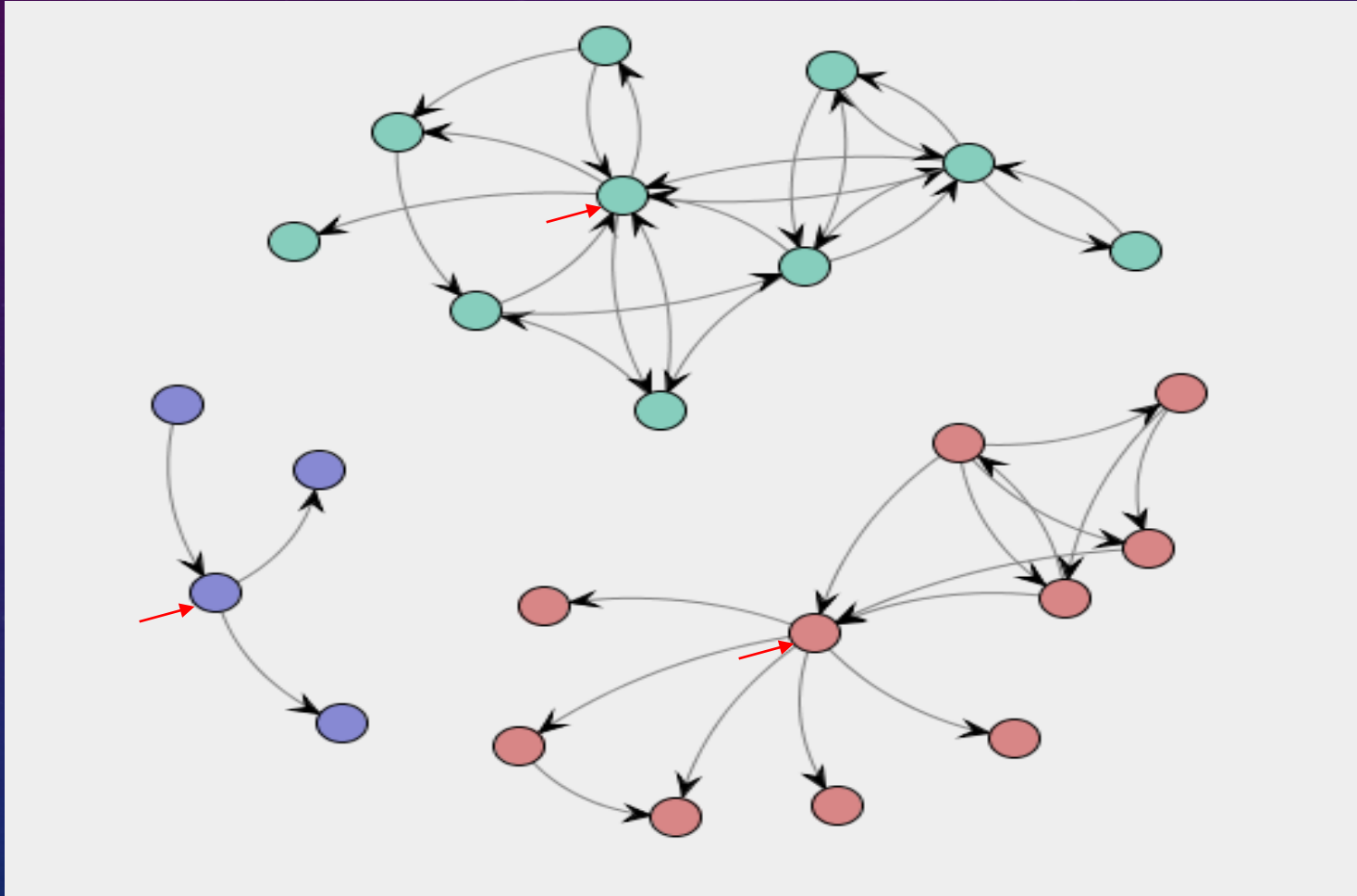
Figure 1: Visualization of Input Data and Event Log Data of an Invoice



# DESIGNED (“IDEAL”) PROCESS MODEL

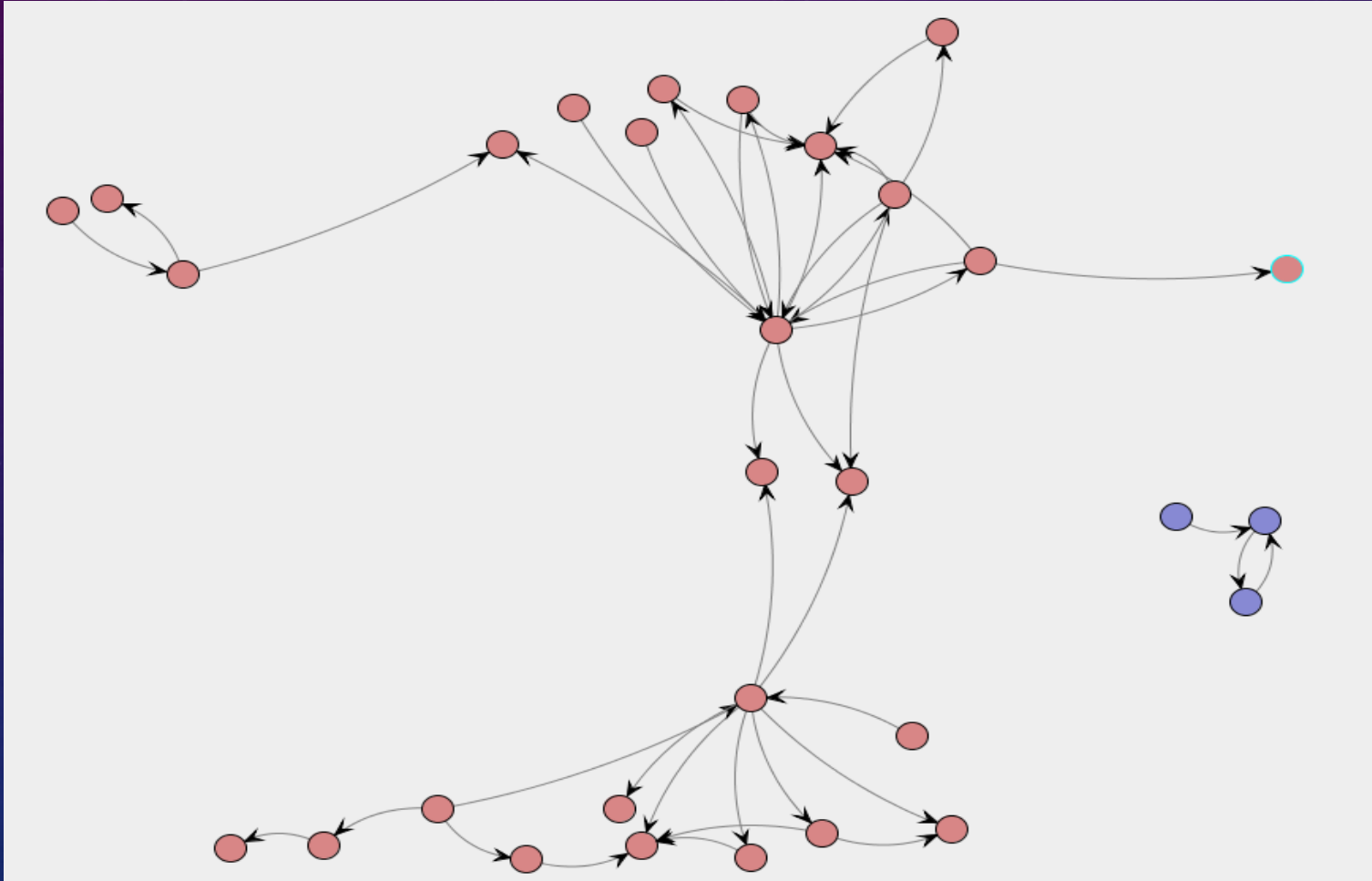


# SOCIAL NETWORK OF 175 CASES BY THREE INDIVIDUALS VIOLATING SOD





# SOCIAL NETWORK OF THE 742 CASES WITHOUT *SIGN* AND IN VIOLATION OF SOD CONTROLS







Rutgers  
Business  
School

Rutgers  
Business  
School  
Best Leadership

PLEASE DO NOT AUDITION  
ORATED BY  
THE SAGE UNIVERSITY  
NEW JERSEY CAMPUS





# **APPLY PROCESS MINING TO EVALUATE INTERNAL CONTROL EFFECTIVENESS AUTOMATICALLY**

**MIKLOS A. VASARHELYI AND TIFFANY CHIU**

# INTRODUCTION

- Process mining can add value and improve the performance of auditing; specifically, it could improve the effectiveness of internal control (Yang and Hwang 2006; Kopp and Donnell 2005; Jans et al. 2009, 2011, 2013, 2014).
- Apply process mining to evaluate internal control effectiveness:
  - (1) Determine the controls required for the business process including the rules for acceptable and unacceptable variants (e.g., the variant is unacceptable if the purchase order has been released without sign).
  - (2) Highlight the weakness of internal control by automatically extracting the unacceptable variants.
  - (3) Examine the log information from unacceptable variants to discover possible causes of violation (e.g., conduct social network analysis or conformance check).



# EXAMPLE OF EVENT LOGS

Characteristics of Event	
(1) Activity	The activity taking place during the event, e.g. sign
(2) Process Instance	The process instance of the event, e.g. invoice
(3) Originator	The originator, or party responsible for the event, e.g. name of the action owner
(4) Timestamp	The timestamp of the event, e.g. date/time of the event (2006-11-07T10:00:36)

Process Instance	Activity	Event Type	Timestamp	Originator
450000000190	Create PO	Complete	02 Feb 2006	John
450000000190	Change Line	Complete	30 Nov 2006	John
450000000190	Sign	Complete	05 Dec 2006	Paul
450000000190	Release	Complete	06 Dec 2006	Anne
450000000190	GR	Complete	05 Jan 2007	John
450000000190	IR	Complete	15 Jan 2007	Matt
450000000190	Pay	Complete	16 Feb 2007	Marianne
450000000210	Create PO	Complete	23 Jan 2007	Doug
...				

# EXAMPLE OF VARIANTS (1/2)

The screenshot displays a software interface with three main sections. The left sidebar lists variants, with 'Variant 1' highlighted. The middle section shows a list of cases, with the first case selected. The right section displays details for the selected case, including a table of activities.

**Left Sidebar: Variants (980)**

- Complete log  
All cases (26185)
- Variant 1**  
10943 cases (41.79%)
- Variant 2  
6668 cases (25.46%)
- Variant 3  
2488 cases (9.5%)
- Variant 4  
646 cases (2.47%)
- Variant 5  
640 cases (2.44%)
- Variant 6  
490 cases (1.87%)
- Variant 7  
393 cases (1.5%)

**Middle Section: Cases (10943)**

- 450039741940**  
6 events
- 4500397495780  
6 events
- 45003965696410  
6 events
- 45004018321840  
6 events
- 45004018413910  
6 events
- 450039741930  
6 events
- 45003965696420  
6 events
- 4500397495770  
6 events

**Right Section: Case Details (450039741940)**

Case with 6 events

Graph Table

	Activity	Resource	eventtype	Value_PO	Quantity_PO	Quantity_GR	Value_Pay
1	Create PO	U35824	complete	8138	1085022	1085022	8100
2	Sign	G19091	complete	8138	1085022	1085022	8100
3	Release	U42242	complete	8138	1085022	1085022	8100
4	GR	G35730	complete	8138	1085022	1085022	8100
5	IR	G10849	complete	8138	1085022	1085022	8100
6	Pay	G10849	complete	8138	1085022	1085022	8100



# EXAMPLE OF VARIANTS (2/2)

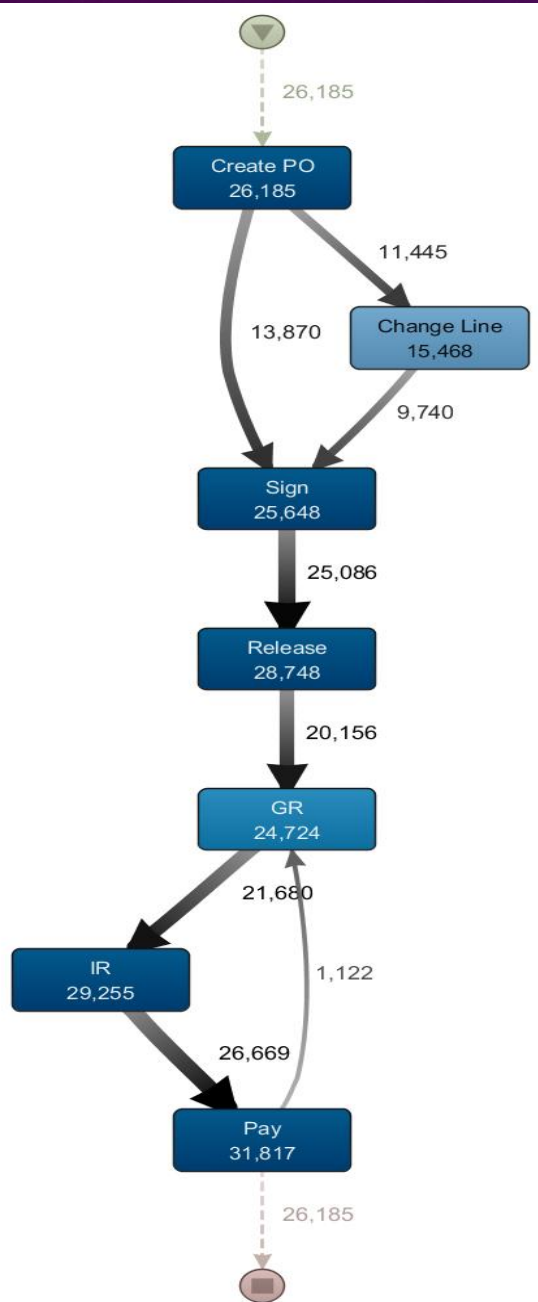
E1		fx Variant						
	A	B	C	D	E	F	G	H
1	Case ID	sequenceNO	Activity	Resource	Variant	eventtype	timestamp	Value_PO
2	450039741940	1	Create PO	U35824	Variant 1	complete	10/1/2007	8138
3	450039741940	2	Sign	G19091	Variant 1	complete	12/1/2007	8138
4	450039741940	3	Release	U42242	Variant 1	complete	1/15/2007	8138
5	450039741940	4	GR	G35730	Variant 1	complete	1/16/2007	8138
6	450039741940	5	IR	G10849	Variant 1	complete	1/17/2007	8138
7	450039741940	6	Pay	G10849	Variant 1	complete	1/18/2007	8138
8	4500397495780	1	Create PO	U21356	Variant 1	complete	10/1/2007	1170
9	4500397495780	2	Sign	U29598	Variant 1	complete	11/1/2007	1170
10	4500397495780	3	Release	G13307	Variant 1	complete	12/1/2007	1170
11	4500397495780	4	GR	U21356	Variant 1	complete	1/29/2007	1170
12	4500397495780	5	IR	G55584	Variant 1	complete	8/2/2007	1170
13	4500397495780	6	Pay	G55584	Variant 1	complete	2/14/2007	1170
14	45003965696410	1	Create PO	U45859	Variant 1	complete	8/1/2007	606
15	45003965696410	2	Sign	G16977	Variant 1	complete	9/1/2007	606
16	45003965696410	3	Release	U29598	Variant 1	complete	9/1/2007	606
17	45003965696410	4	GR	U45859	Variant 1	complete	12/1/2007	606
18	45003965696410	5	IR	G15330	Variant 1	complete	1/18/2007	606
19	45003965696410	6	Pay	G15330	Variant 1	complete	1/24/2007	606

# Project Overview





# DATASET OVERVIEW



<b>Events</b>	<b>181,845</b>
<b>Cases</b>	<b>26,185</b>
<b>Activities</b>	<b>7</b>
<b>Activities Detail</b>	(1) Create PO (2) Sign (3) Release (4) GR (5) IR (6) Pay (7) Change Line
<b>Variants</b>	<b>980</b>
<b>Mean Case Duration</b>	<b>46.2 Day</b>
<b>Start</b>	<b>01/02/2007</b>
<b>End</b>	<b>01/25/2008</b>

# CLASSIFICATION RULES

Classification Rules - Unacceptable Variants		
Categories	Sub-categories	Description
Skipping Activities	Skipping Purchase Order (PO)	Missing activity "PO" in the sequence. For example: Sign-Release-GR-IR-Pay
	Skipping Sign	Missing activity "Sign" in the sequence. For example: PO-Release-GR-IR-Pay
	Skipping Release	Missing activity "Release" in the sequence. For example: PO- Sign -GR-IR-Pay
	Skipping Goods Receipt (GR)	Missing activity "GR" in the sequence. For example: PO-Sign-Release-IR-Pay
	Skipping Invoice Receipt (IR)	Missing activity "IR" in the sequence. For example: PO-Sign-Release-GR-Pay
	Skipping Payment (Pay)	Missing activity "Pay" in the sequence. For example: PO-Sign-Release-GR-IR



# CLASSIFICATION RULES

Classification Rules - Unacceptable Variants		
Categories	Sub-categories	Description
Repetitive Activities	Repeat Purchase Order (PO)	More than one "PO" occurs in the sequence. For example: PO-PO-Sign-Release-GR-IR-Pay
	Repeat Sign	More than one "Sign" occurs in the sequence. For example: PO-Sign-Sign-Release-GR-IR-Pay
	Repeat Release	More than one "Release" occurs in the sequence. For example: PO-Sign-Release-Release-GR-IR-Pay
	Repeat Goods Receipt (GR)	More than one "GR" occurs in the sequence. For example: PO-Sign-Release-GR-GR-Pay
	Repeat Invoice Receipt (IR)	More than one "IR" occurs in the sequence. For example: PO-Sign-Release-GR-IR-IR-Pay
	Repeat Payment (Pay)	More than one "Pay" occurs in the sequence. For example: PO-Sign-Release-GR-IR-Pay-Pay

# CLASSIFICATION RULES

Classification Rules - Unacceptable Variants		
Categories	Sub-categories	Description
Switching Activities	Change Line not followed by Sign	Sign does not occur after change line in the sequence. For example: PO-Sign-Release-Change Line-GR-IR-Pay
	Change Line not followed by Release	Release does not occur after change line in the sequence. For example: PO-Sign-Release-Change Line-GR-IR-Pay
	GR occurs before Sign	GR occurs before Sign in the sequence. For example: PO-GR-Sign-Release-IR-Pay
	GR occurs before Release	GR occurs before Release in the sequence. For example: PO-Sign-GR-Release-IR-Pay
	IR occurs before Sign	IR occurs before Sign in the sequence. For example: PO-IR-Sign-Release-GR-Pay
	IR occurs before Release	IR occurs before Release in the sequence. For example: PO-IR-Release-GR-Pay



# CLASSIFICATION RULES

Classification Rules - Unacceptable Variants		
Categories	Sub-categories	Description
Switching Activities	Pay occurs before Sign	Pay occurs before Sign in the sequence. For example: PO-Pay-Sign-Release-IR-Pay
	Pay occurs before Release	Pay occurs before Release in the sequence. For example: PO-Pay-Release-GR-IR
	Pay occurs before GR	Pay occurs before GR in the sequence. For example: PO-Sign-Release-IR-Pay-GR
	Pay occurs before IR	Pay occurs before IR in the sequence. For example: PO-Sign-Release-GR-Pay-IR
Additional Activity	Any additional activity that is not in the standard procedure	The sequence include unexpected activity which should not be in the sequence based on standard process. For example: PO-Sign-Release-GR-IR-Pay-(Other additional activity)

# RESULTS

Skipping Activities		
Activities	Variants	CaseID
Create Purchase Order	0	0
Sign	40	3443
Release	3	3
Goods Receipt	157	4706
Invoice Receipt	0	0
Pay	0	0

Repetitive Activities		
Activities	Variants	CaseID
Create Purchase Order	0	0
Sign	379	1094
Release	209	680
Goods Receipt	450	548
Invoice Receipt	455	527
Pay	650	1,830



# RESULTS - SWITCHING ACTIVITIES

Switching Activities		
Sub-categories	Variants	CaseID
Change Line not followed by Sign	544	4293
Change Line not followed by Release	494	1547
GR occurs before Sign	0	0
GR occurs before Release	0	0
IR occurs before Sign	5	9
IR occurs before Release	15	131
Pay occurs before Sign	1	1
Pay occurs before Release	2	3
Pay occurs before GR	8	8
Pay occurs before IR	0	0

# RESULTS – SOCIAL NETWORK ANALYSIS

Social Network Analysis - SOD		
Description	CaseID	Variants
Same person Sign and Release	473	138
Same person Release and GR	175	12
Same person GR and IR	0	0





# SUMMARY OF PROJECT UPDATE AND DATA REQUEST

- Summary of project update
  - We are in the process of classifying our current data into acceptable and unacceptable variants
  - There will be several further analyzes being conducted under each category.
- Data request
  - Event log data



**Thank you!**



