

*The Future of Audit: A Perspective from  
the Rutgers AICPA Data Analytics  
Research Initiative*

Miklos Vasarhelyi, Rutgers University  
Summer School in Public Auditing and  
Accountability

July 26<sup>th</sup>, 2018

Pisa, Italy

## Outline

- The CarLab
- Work in the public area
- What the students think of accounting
- Big Data
- Analytic Methods
- Robotic Process Automation
- Exogenous data
- Imagineering the new audit
- Artificial<sup>2</sup> Intelligence and cognitive computing

**RUTGERS BUSINESS SCHOOL  
THE CARLAB (CONTINUOUS  
AUDIT AND REPORTING  
LABORATORY)**

BRIGHAM YOUNG  
UNIVERSITY

## The Ranking of Rutgers in the Accounting Areas

Areas	Ranking 2008-2013	Ranking 2002-2013	Ranking 1990- 2013
AIS	#1 out of 179	#1 out of 207	#1 out of 241
Audit	#6 out of 320	#7 out of 370	#11 out of 438
Financial	#70 out of 356	#89 out of 406	#83 out of 470
Managerial	#120 out of 286	#80 out of 346	#66 out of 413
Tax	#53 out of 129	#76 out of 178	#79 out of 246
Other	#35 out of 171	#18 out of 248	#25 out of 341

## CarLab Analytic Research

<u>Choosing apps</u>	<u>Predictive Analytics with Weather data</u>	<u>Audit data analytics and EDA</u>	<u>Envisaging the future of audit and Big Data</u>	<u>Text Mining</u>	<u>Monitoring Unibanco's branches</u>
<u>Visualization</u>	<u>Process Mining at Gamma Bank</u>	<u>Expert System for P-Card</u>	<u>Logit regression for control risk assessment</u>	<u>Exceptional Exceptions</u>	<u>Client Retention Project</u>
<u>Litigation prediction</u>	<u>Fraud Risk Assessment using EDA</u>	<u>Detecting duplicate records</u>	<u>Continuity equations</u>	<u>Predictive Audit</u>	Credit card Default prediction
Insurance Analytics	<u>Multidimensional clustering for fraud detection</u>	<u>Rule-based selection for transitory accounts</u>	<u>Continuity Equations at HCA</u>	<u>XBRL</u>	<u>Insurance Analytics</u>
Cognitive Decision Aids	<b>AI: Deep Learning</b>	<u>Robotic Process Automation (RPA)</u>	<u>Intelligent Process Automation (RPA)</u>	<u>Blockchain and Smart contracts</u>	Cluster Analysis of US States

## Recent research of PhD students

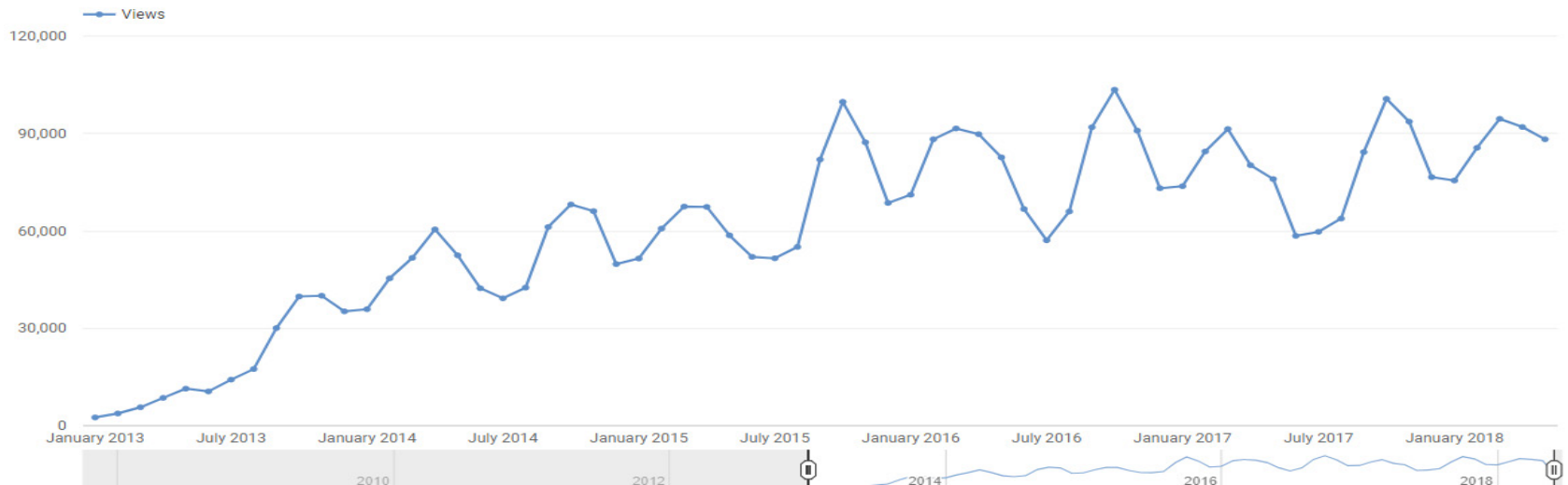
<b>Name</b>	<b>Title</b>
<b>Abdulrahman Alrefai</b>	<b>Formalization of Internal Control Assessment: A Process Mining Application</b>
<b>Ahmad AlQassar</b>	<b>Resisting Change in the Audit Profession: Two Case Studies from Multi-National Firms</b>
<b>Andrea Rozario</b>	<b>Examination of Audit Planning Risk Assessments Using Verbal Protocol Analysis: An Exploratory Study</b>
<b>Cheng Yin</b>	<b>Privacy-Preserving Information Sharing within an Audit Firm</b>
<b>Deniz Appelbaum</b>	<b>Using Drones in Internal and External Audits: An Exploratory Framework</b>
<b>Feiqi Huang</b>	<b>Audit Evidence Index Project</b>
<b>He Li</b>	<b>Are External Auditors Concerned about Cyber Incidents? Evidence from Audit Fees</b>
<b>Jiahua Zhou</b>	<b>The Survived Companies with Going Concern Are Really Different from Those Bankrupted</b>
<b>Jun Dai</b>	<b>Towards Blockchain-based Accounting and Assurance</b>
<b>Jun Dai</b>	<b>Imagineering Audit 4.0</b>
<b>Zhaokai Yan</b>	<b>Impact of Data Analytics on Managerial Accounting Using Balanced Scorecard Framework</b>
<b>Yunsen Wang</b>	<b>An Application of Blockchain Technology to Fraud Detection</b>
<b>Yue Liu</b>	<b>Risk Analysis Based on 10-K Item 1a</b>
<b>Ting Sun</b>	<b>The Performance of Sentiment Features of 10-K MD&amp;As for Financial Misstatement Prediction</b>
<b>Tiffany Chiu</b>	<b>Apply Process Mining to Evaluate Internal Control Effectiveness Automatically</b>
<b>Qiao Li</b>	<b>Rule-Based Decision Support System for Audit Planning and Audit Risk Assessment</b>
<b>Lu Zhang</b>	<b>Interactive Data Visualization for Error and Fraud Detection: Case Studies and Practice Implications</b>

<https://www.youtube.com/playlist?list=PLauepKFT6DK9vKn7-eKxzmxBegpe8v8xw>

# **THE RUTGERS DIGITAL LIBRARY**

## Usage

<http://raw.rutgers.edu/RADL.html>







# **SOME WORK IN THE PUBLIC AUDIT AREA**

## Some work in the public area

- XBRL work with the GASB
- Big data analytics cooperation with the TCU (Brazilian Public Audit Organization)
- FSS (Korean SEC)
- Research workshops with FGV – Brasil
- Dialogues with the SEC

# CVM – Brazilian Stock Exchange

- Disclosure Initiative
  - not enough relevant information;
  - too much irrelevant information;
  - and ineffective communication of information provided
  - Compare how FRs are similar with the IFRS
- Impairment Evaluation:
  - Text mining to extract information from footnotes
  - How was evaluated
  - Interests rate
  - Revenue Growth

# SICONFI – Brazilian National Treasure Dept

- Standard XBRL taxonomy based on IPSAS
- Second Layers taxonomies for local rules
- Balance Sheet Reconciliation for municipalities
  - +5000 municipalities
  - Standard XBRL taxonomy for all municipalities, states and federal government

# Santa Catarina State

- Continuous audit on payroll
- 150 thousands employees
- 46% of all expenses
- Continuous Process Mining

# **WHAT THE STUDENTS THINK OF ACCOUNTING**

## Student view of the profession

- Tedious
- Very long crazy hours
- Low (starting) pay



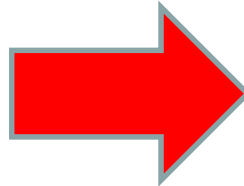
# ~~Tedious~~

- Robotic Process Automation (RPA)
- Reframing of the audit towards assurance in many aspects
- Re-Skilling of the auditor

# ~~Very long crazy hours~~

- Continuous audit
  - Audit by exception
- Technologization of assurance
- Re-Skilling of the auditor
- The assurance eco-system

~~Low (starting) pay~~

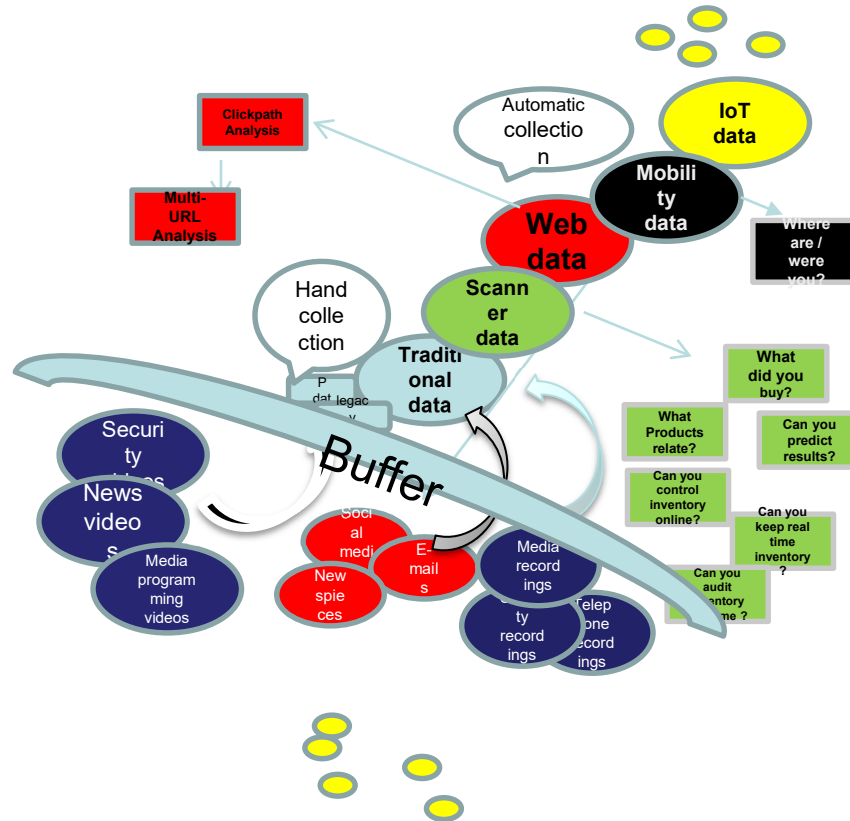




# BIG DATA

## BIG DATA

An evolving framework towards automatic data collection and exogenous data usage



**5 Vs: Volume, Variety, Veracity, Value, & Velocity**

# **ANALYTIC METHODS THE RADAR PROJECT**

# Rutgers AICPA Data Analytics Research Initiative



The mission of RADAR is to facilitate the further integration of data analytics into the audit process, and to demonstrate through research how this can effectively lead to advancements in the public accounting profession.

Additional information can be found at: <http://raw.rutgers.edu/radar>

# Participants

- Big Four
  - KPMG
  - Deloitte
  - EY
  - PWC
- Other Firms
  - GT
  - Crowe Horwath
  - BDO
  - RSM/US
- AICPA
- CPA Canada



# Rutgers AICPA Data Analytics Research Initiative

- **Research Projects**
  - *Multidimensional Audit Data Selection*
  - *Process Mining*
  - *Visualization*

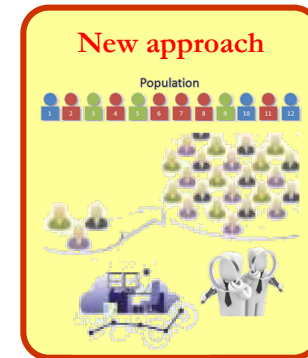
# Multidimensional Audit Data Selection (MADS)



## BACKGROUND



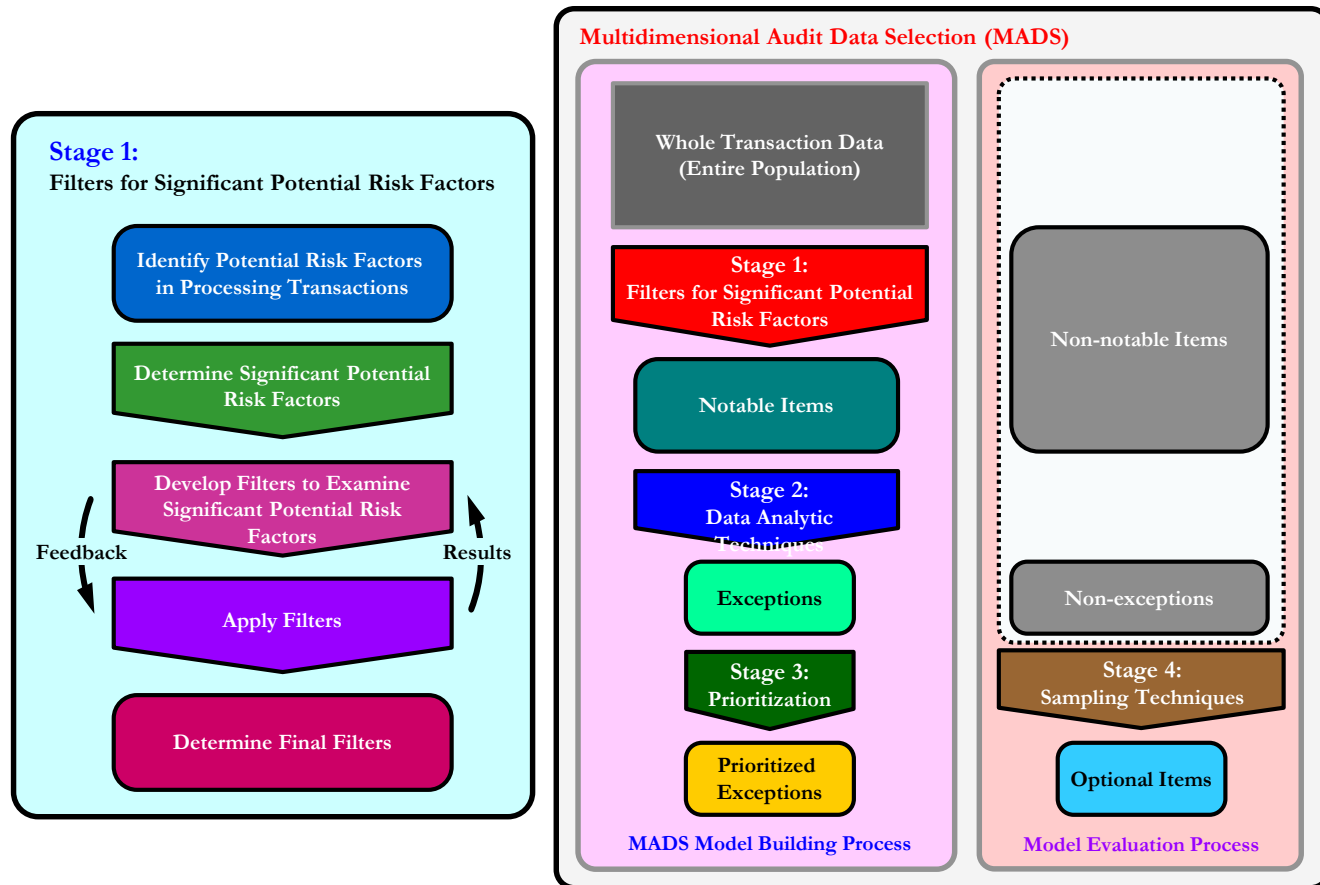
Advance in data processing ability & data analytic techniques allows auditors to evaluate the entire population instead of examining just a chosen sample.



- BUT, often generate large numbers of outliers.
- Impractical for auditors to investigate entire outliers

- Crucial to develop a method that can help auditors effectively deal with large amounts of data, but also assist them to efficiently handle a massive number of outliers.

# MADS ANALYTIC FRAMEWORK



# **Evaluating the Effectiveness of Internal Control using Process Mining**

**Tiffany Chiu and Miklos A. Vasarhelyi**

October, 2017

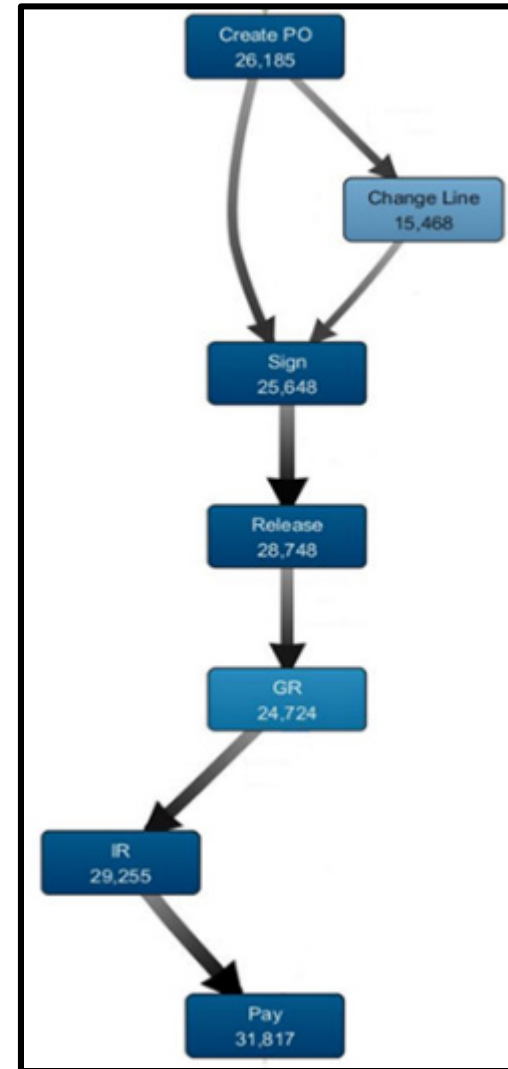
## Introduction

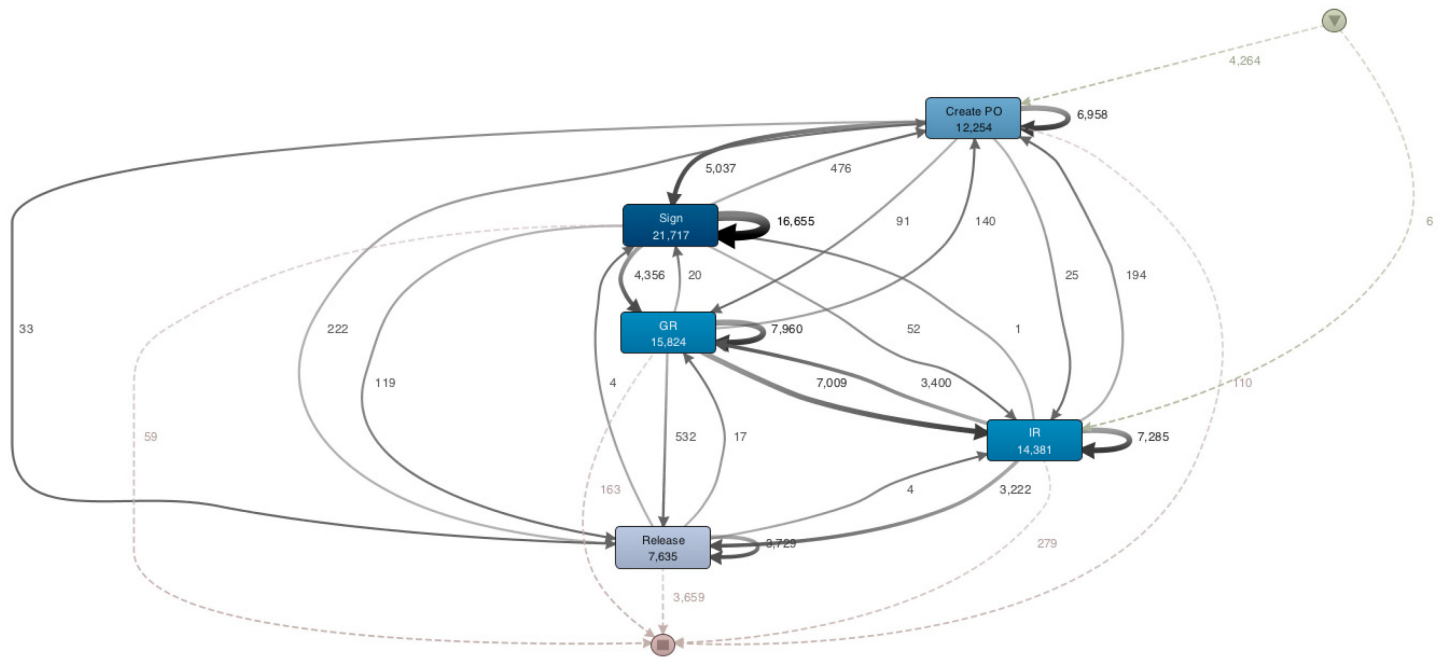
- This project aims at adopting process mining to evaluate the effectiveness of internal control using a real-life event log.
- The evaluation is based on the full population of event logs and contains four analyses:
  - (1) *Variant analysis* that identifies acceptable and notable variants.
  - (2) *Segregation of duty analysis* that examines process instances and employees that violate segregation of duty controls.
  - (3) *Personnel analysis* that investigates employees who are involved in multiple potential control violations.
  - (4) *Timestamp analysis* that detects time related issues such as the ones performed during the weekends and process instances that have lengthy process duration.

## Dataset Overview

<b>Event</b>	<b>181,845</b>
<b>Process Instance</b>	<b>26,185</b>
<b>Activity</b>	<b>7</b>
<b>Activity Detail</b>	(1) Create PO (2) Sign (3) Release (4) GR (5) IR (6) Pay (7) Change Line
<b>Variant</b>	<b>980</b>
<b>Mean Process Instance Duration</b>	<b>46.2 Days</b>
<b>Start</b>	<b>01/02/2007</b>
<b>End</b>	<b>01/25/2008</b>

31







## What is Process Mining of Event Logs?

- Process mining technique refers to using event log to analyze business process.
  - **Event Log** is defined as “a chronological record of computer systems activities which are save to a file on the system. The file can be reviewed by the system administrator to identify users’ actions on the system or processes which occurred on the system” (FAS: Federation of American Scientists)

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
...				

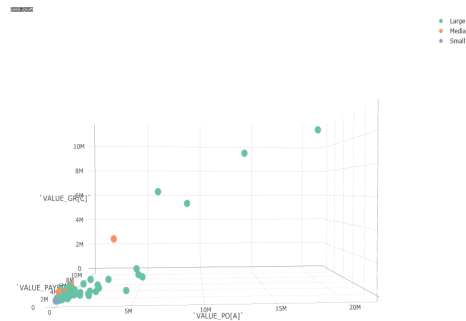


THE STATE UNIVERSITY  
OF NEW JERSEY

# Visualization as Audit Evidence

Qi Liu, Heejae Lee,  
Zamil S. Alzamil  
RADAR  
December 6th, 2017

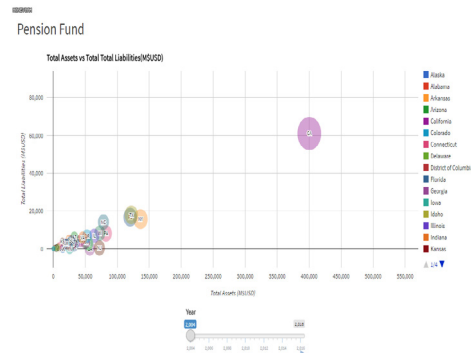
## Dynamic Visualization as Audit Evidence 3D scatter Interactive Visualization



Interactive Analysis\_R.html

- Use 3D scatter plot to investigate relationship between more than three values and identify potential risks
- Provides more information than using 2D plot
- Process Mining Data Log
  - Value of Purchase Order
  - Value of Payment
  - Value of Goods Received

## Dynamic Visualization as Audit Evidence (cont'd) Time Series Interactive Visualization

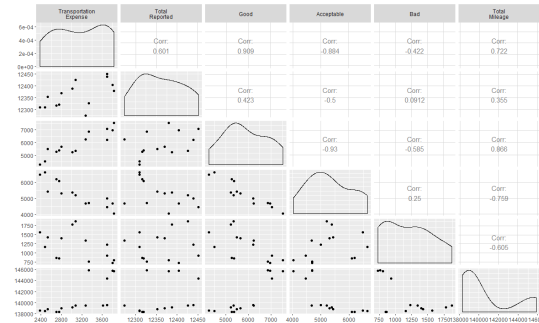


- Investigate the time change of more than 2 values
- Select only cases you want to examine with Interactive Visualization Technique
- See how the target changes compared with other cases
- 50 States Comprehensive Annual Financial Report(CAFR) Pension Fund Balance Sheet (2004 – 2016)
  - Total Liabilities
  - Total Assets



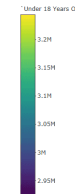
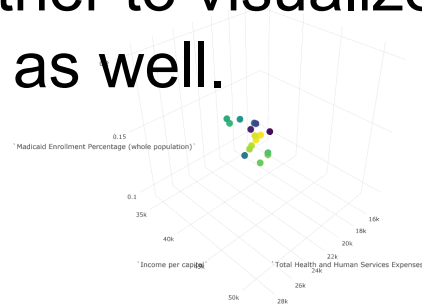
- Positive
  - Total Reported(IRI)
  - Good(IRI)
  - Total Mileage
- Negative
  - Acceptable(IRI)
  - Bad(IRI)

## Correlation with Transportation Expense



# Expectation Model for healthcare expense

- A set of 3D models can be used together to visualize the model as well.



3dplot\_HealthExpense\_Income\_Medicaid.html

# RPA and IPA in Auditing

Abigail Zhang & Andrea Rosario  
Rutgers CarLab

# What is RPA (Robotic Process Automation)?

- McKinsey (2017) describes RPA as “A software automation tool that automates routine, definable, repeatable, high-volume, rule-based and standardized tasks such as data extraction and cleaning through existing user interfaces.”
- RPA robots conduct work the same way that humans do through the software presentation layer



# RPA pilot project – confirmation process

Request - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW 福昕阅读器 feiqi huang

Clipboard Font Alignment Number Styles Cells Editing

Calibri 11 A A Wrap Text Text Normal Bad Good Neutral Calculation Check Cell

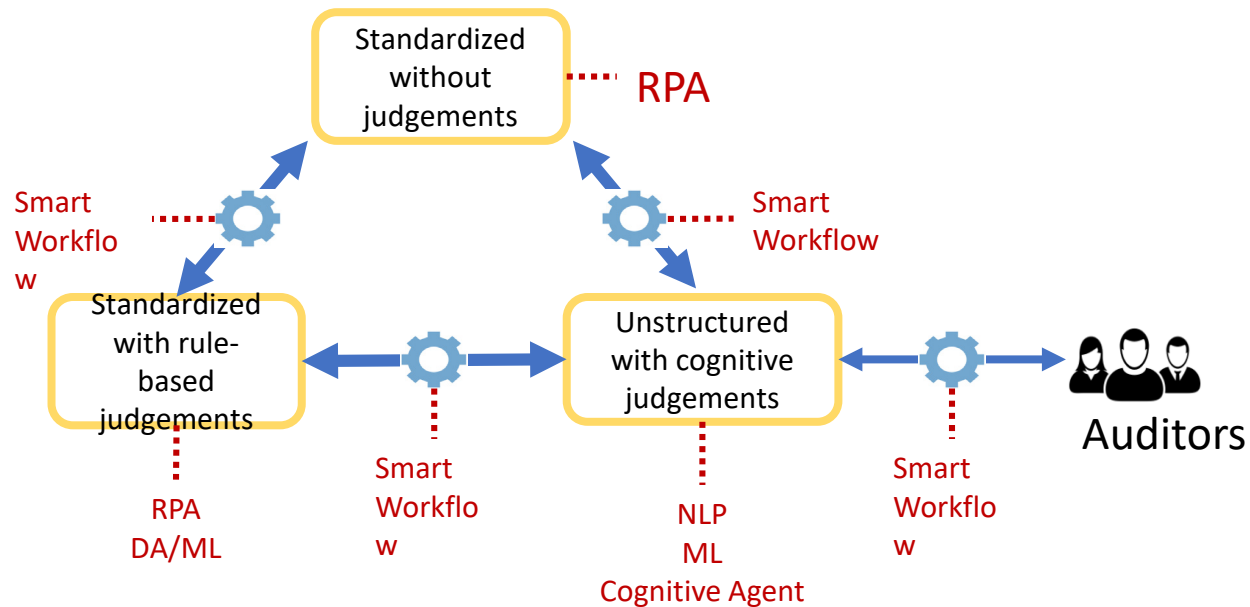
AutoSum Fill Sort & Find & Filter Select

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
	Client	Signer Job Title	First Name	Last Name	Address	City	State	Zip code	Email	Phone	Bank Name	Form	Account ID	Date	Status	Confirm	Balance
1																	
2	Simply Soups Inc.	Manager	Jennings	Lou	177 Washington Lane	Cherry Hill	NJ	08034	lou.jennings@ssoups.com	609-555-5555	Fifth Federal	Asset	675-42223	12/31/2016			
3	Simply Soups Inc.	Manager	Jennings	Lou	177 Washington Lane	Cherry Hill	NJ	08034	lou.jennings@ssoups.com	609-555-5555	Sparkasse-Frankfurt	Asset	44-322711	12/31/2016			
4	Simply Soups Inc.	Manager	Jennings	Lou	177 Washington Lane	Cherry Hill	NJ	08034	lou.jennings@ssoups.com	609-555-5555	American NorthWest Bank	Asset	05-198305	12/31/2016			
5	Simply Soups Inc.	Manager	Jennings	Lou	177 Washington Lane	Cherry Hill	NJ	08034	lou.jennings@ssoups.com	609-555-5555	BNY Federal	Asset	061-22031	12/31/2016			
6																	
7																	
8																	
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	
22																	
23																	
24																	
25																	
26																	
27																	
28																	
29																	
30																	
31																	
32																	

Sheet1

READY 85%

# IPA Ecosystem in audit



# What is IPA?

- “IPA mimics activities carried out by humans and, over time, learns to do them even better.”

## 5 Core Technologies Embedded in IPA (McKinsey, 2017)

### Robotic Process Automation (RPA)

A software automation tool that automates routine, definable, repeatable, high-volume, rule-based and standardized tasks such as data extraction and cleaning through existing user interfaces.

### Smart Workflow

A process-management software tool that integrates tasks performed by groups of humans and machines (for instance, by sitting on top of RPA to help manage the process)

### Machine Learning/Advanced Analytics

Algorithms that identify patterns in structured data, such as daily performance data, through “supervised” and “unsupervised” learning.

### Natural Language Generation

Software engines that create seamless interactions between humans and technology by following rules to translate observations from data into prose.

### Cognitive Agents

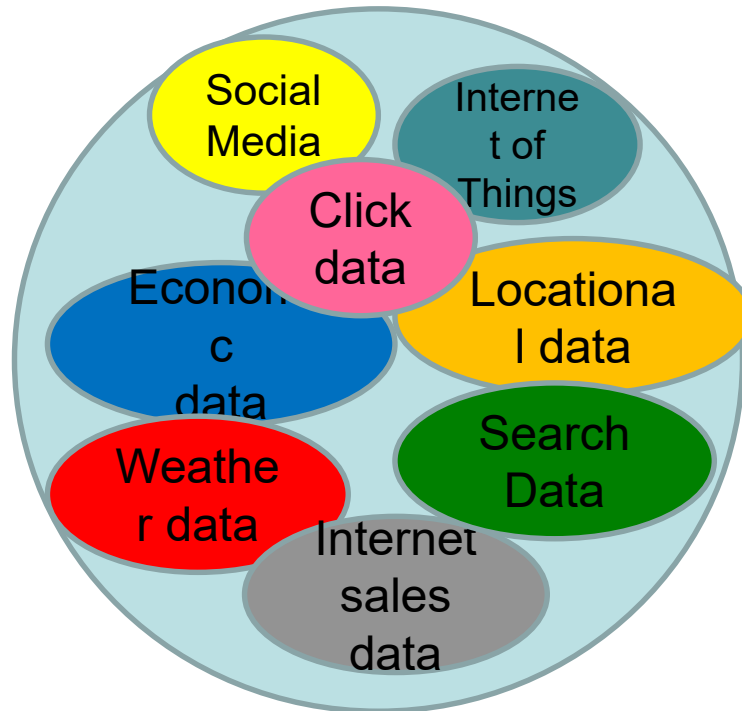
Technologies that combine machine learning and natural-language generation to build a completely virtual workforce (or “agent”) that is capable of executing tasks, communicating, learning from data sets, and even making decisions based on “emotion detection.”

## Exogenous data analytics for Auditing

Miklos A. Vasarhelyi  
Helen Brown Liburd

Rutgers Business School

# Exogenous Data



ED may be of easier access

ED is likely less tamperable

ED relationships will be stochastic

ED is a form of confirmation

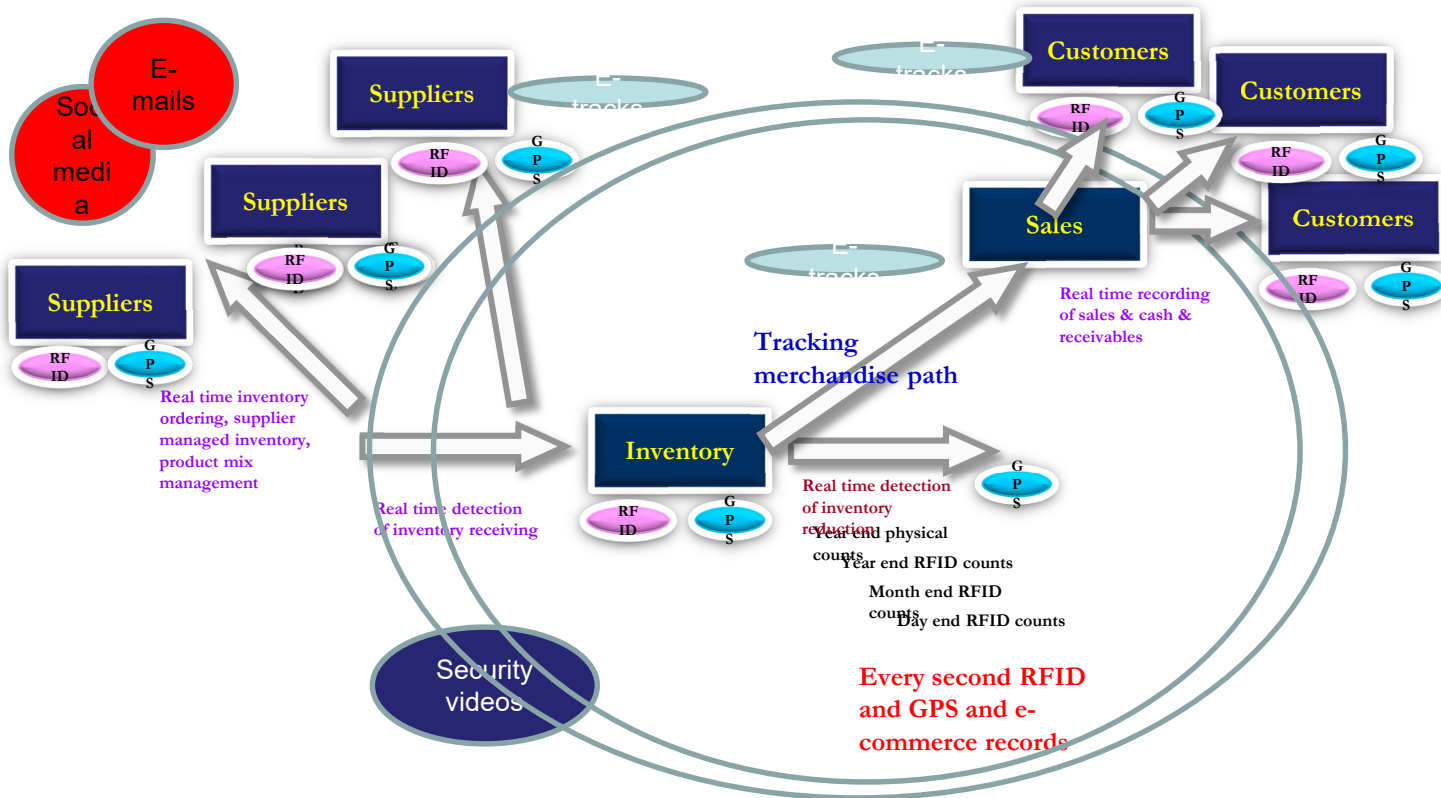
ED may complement many current procedures

ED may create many new procedures

- Forget about privacy.... Its gone....
- Fortunately you are not very interesting
- Technology giveth ....
- Technology taketh

Jun Dai and Miklos A. Vasarhelyi  
SWUFE and RBS

# **IMAGINEERING THE NEW AUDIT**



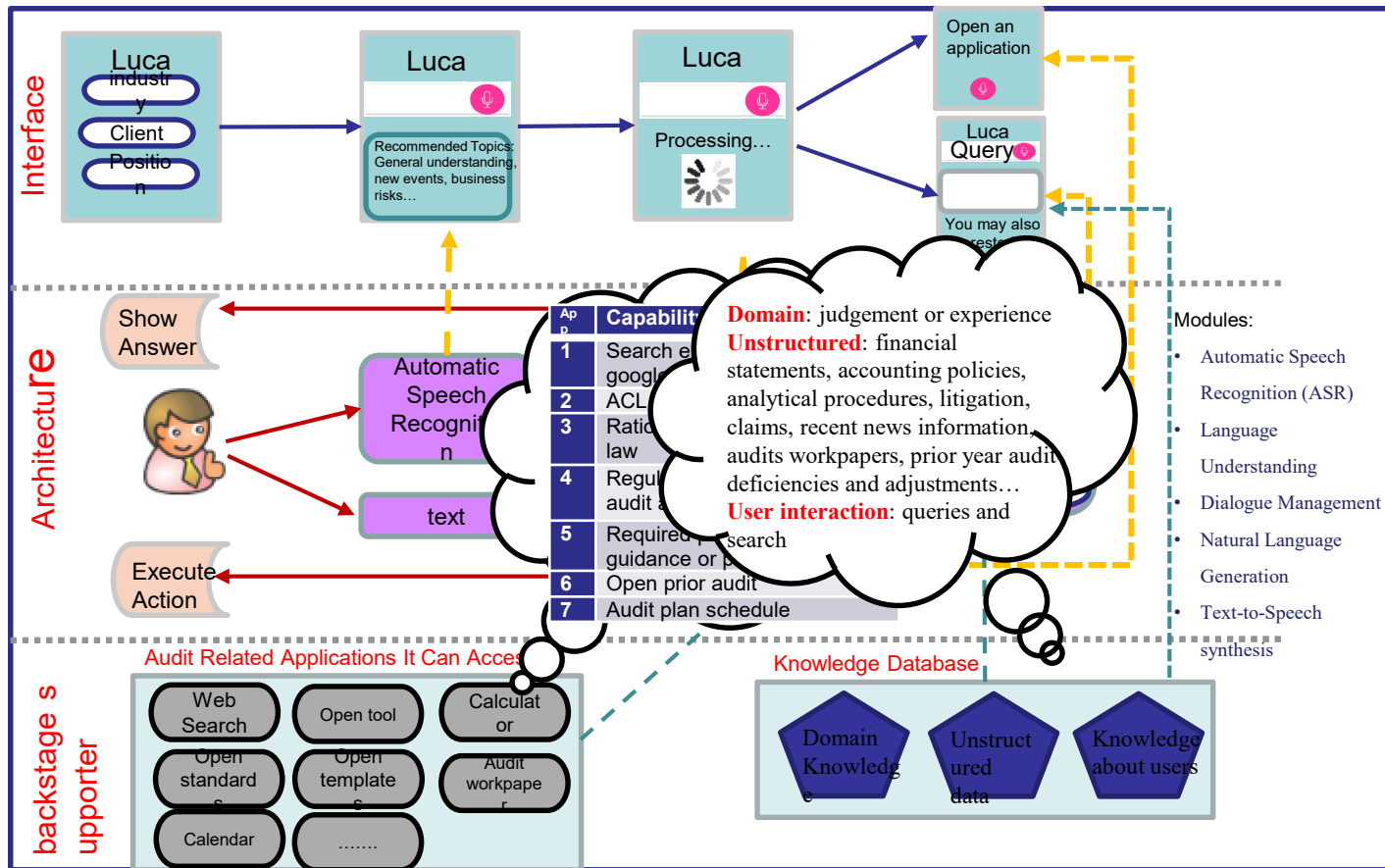


# **ARTIFICIAL INTELLIGENCE / COGNITIVE ASSISTANTS**

# AI in auditing

- A wide range of applications
- Large firms are doing deep learning (e.g. Watson)
- Major benefits will come from
  - Cognitive computing decision aids a la LUCA
  - IPA
- The highest returns come from production line and straightforward decision automation
- Technological process reframing
  - E.g. contracts...

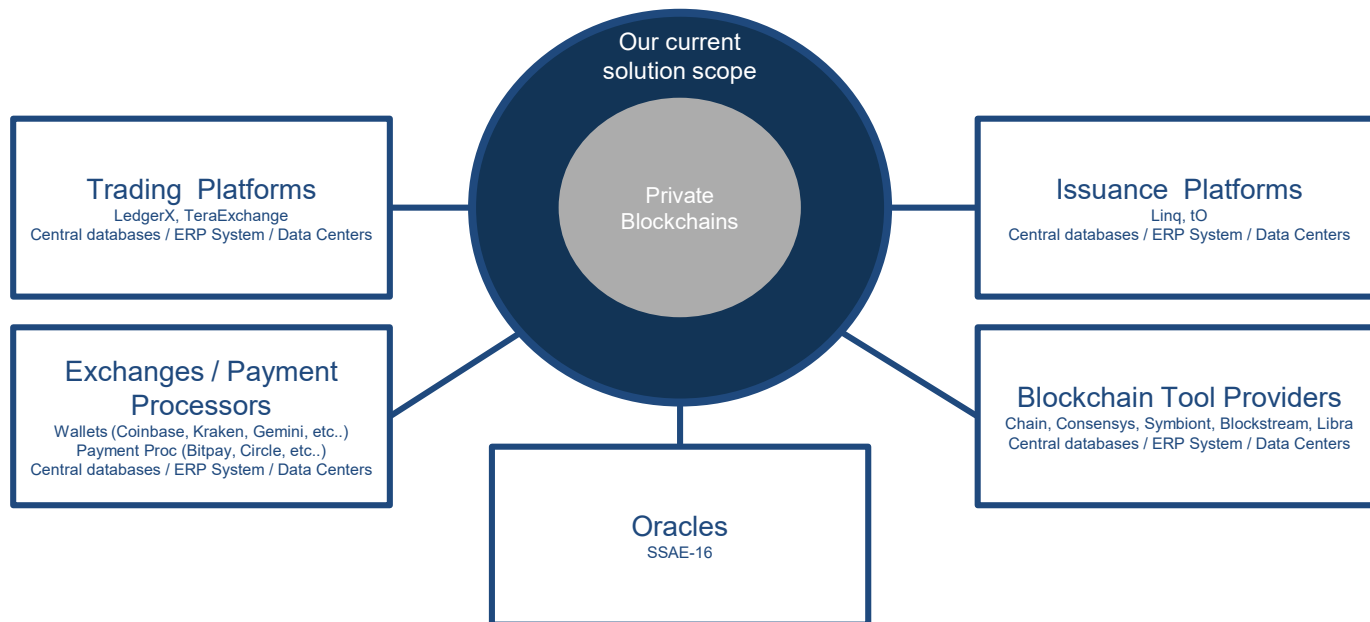
## Architecture of the Proposed Audit Cognitive Assistant (SIRI-like LUCA)



From LIBRA's ROD BRENNAN

# **AUDIT BLOCKCHAIN ECOSYSTEM**

## Initial Scope - Libra Blockchain Audit Tools



# Smart Contracts

- Smart contracts are programs that are designed to be self executing and automatically monitoring of contractual conditions.
- Smart contracts existed prior to blockchain but became feasible when incorporated into a blockchain environment.
- Once a smart contract is programmed it is able to draw from both data within the blockchain, or through oracles to examine data external to the blockchain.
- Contract provisions will instantly be activated based on these relevant conditions.

# Smart Contracts: An Example

- One example of a smart contract application could be in bank loans.
- In this example the interest rate may be adjustable based on a companies credit rating.
- A smart contract will be continuously monitoring rating agencies.
- Once the rating changes the contract will automatically adjust the interest rate or call the loan.

# CONCLUSIONS



# Computerization of Occupations

Occupations Susceptible to Computerization	
Occupation	Probability
Recreational Therapists	0.28%
Physicians and Surgeons	0.42%
Career/Technical Education Teachers, Secondary School	0.88%
Sales Managers	1.30%
Chief Executives	1.50%
Writers and Authors	3.80%
Nuclear Engineers	7.00%
Police and Sheriff's Patrol Officers	9.80%
Flight Attendants	35.00%
Economists	43.00%
Police, Fire, and Ambulance Dispatchers	49.00%
Commercial Pilots	55.00%
Librarians	65.00%
Medical Secretaries	81.00%
Executive Secretaries and Executive Administrative Assistants	86.00%
Taxi Drivers and Chauffeurs	89.00%
Accountants and Auditors	94.00%
Cashiers	97.00%
Bookkeeping, Accounting, and Auditing Clerks	98.00%
Tax Preparers	99.00%

Adapted from: "The Future of Employment: How Susceptible are Jobs to Computerisation?" (Frey and Osborne, 2013)

**Ability to automate, % of time spent on activities<sup>1</sup> that can be automated**  
by adapting currently demonstrated technology



<sup>1</sup>Our analysis used "detailed work activities," as defined by O\*NET, a program sponsored by the US Department of Labor, Employment and Training Administration.

# Conclusions

- Disruption will come first from exogenous data and then from cognitive assistants
- AI will take longer to really affect
- External validation is a new way of thinking
- Labor replacement will come mainly from RPA
- Technological process retrofitting will be necessary on the audit process and in standards
- The purchase/ acquisition of data will require substantive resources but will not depend on the client IT personnel
- Auditors will also create collection needs for IoT



THE STATE UNIVERSITY  
OF NEW JERSEY

# Design of Audit Apps for Government Procurement Contract Fraud Detection

**Jun Dai**

**Rutgers University**

**Qiao Li**

**Rutgers University**

**Miklos A. Vasarhelyi**

**Rutgers University**

# Contents

- ❖ Introduction
- ❖ Background information
- ❖ Objective
- ❖ Proposed apps
- ❖ Illustrations
- ❖ Future study

# Introduction

## *Government procurement:*

- 10%-15% of GDP; 7 trillion dollars annually in U.S.
- Not always Open and Transparent
- Fraud schemes:  
bid rigging, bribery, kickbacks, cost mischarging, defective pricing,  
product substitution ...

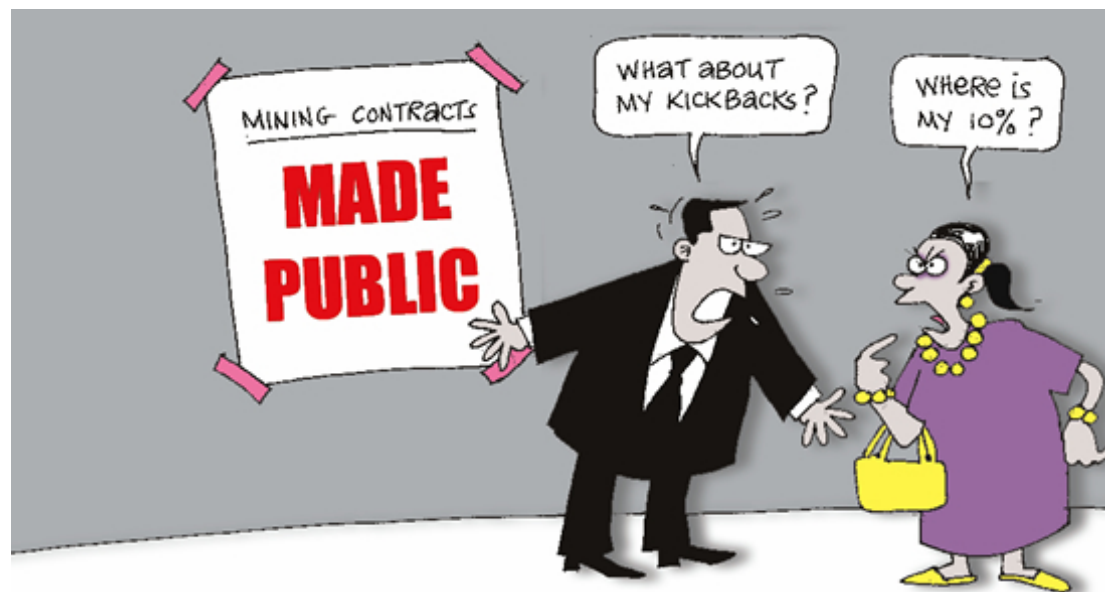


# Introduction

What data to use?

Who has interest?

How to detect fraud?



# Background

## *Open Data Initiatives*

- Make info available and transparent
- 45 countries and 163 international regions
- U.S.
  - Data.gov
  - 39 states and 46 cities and counties
  - formats: Excel, CSV, XML, API, HTML, op
- Government procurement data:
  - China: ccgp.gov.cn
  - Australian: tenders.gov.au
  - Canada: buyandsell.gc.ca
  - Brazil : dados.gov.br
  - UK: gov.uk

### Data Standards :

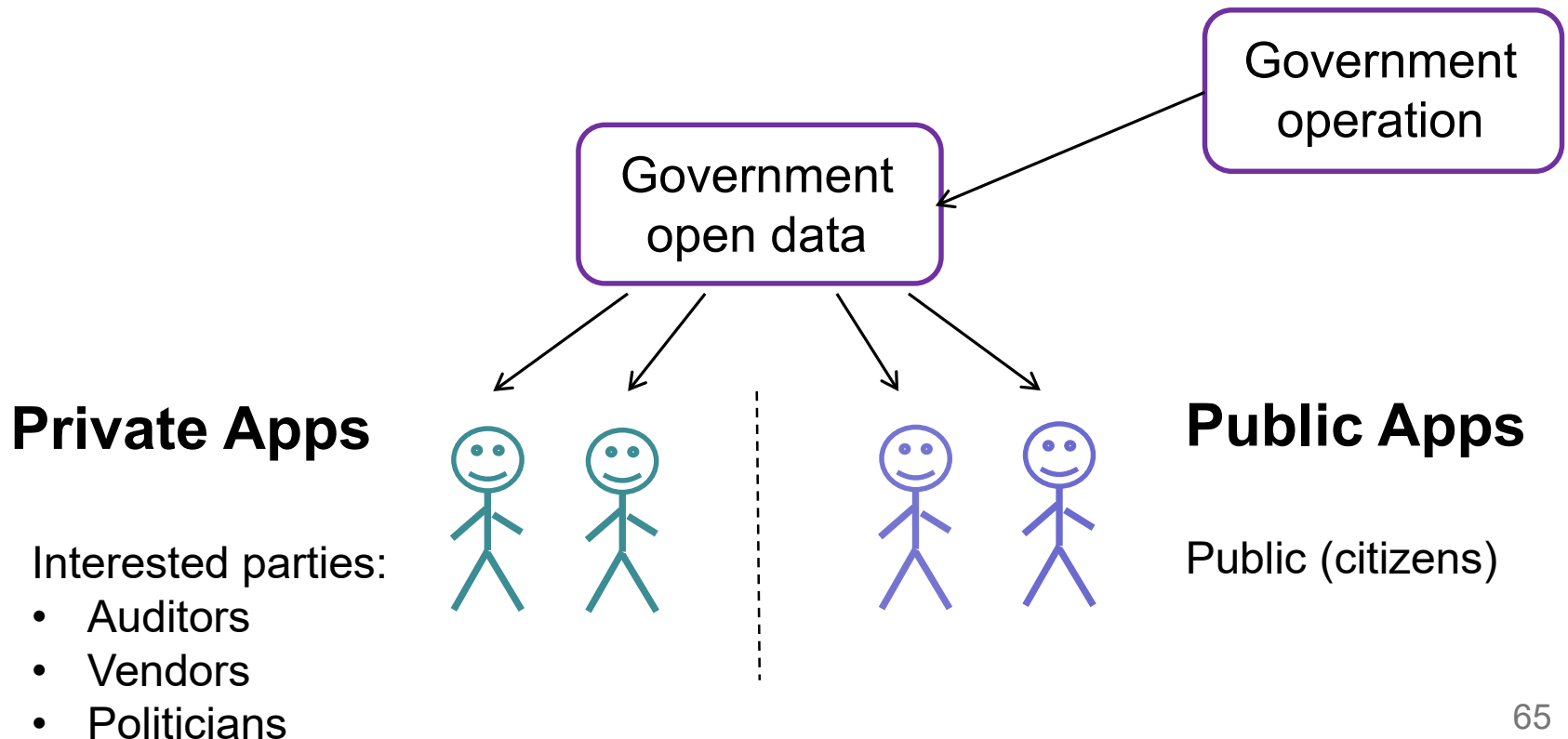
- Different data structure, format, and field names
- AICPA: Audit Data Standards
- Help build data model for ERP and audit apps



# Background

## *“Armchair Auditor”*

-- Crowdsourcing analysis of government data (DE O’Leary, 2015)



# Background

## *“Armchair Auditor”*

- Pilot projects:
  - 2 English councils: Isle of Wight council and Hull City
  - calculated government payments information
- Achievements
  - in 2011, a group of activists uncovered a £1.3m audit scandal at their local council (Patrick, 2011; Patrick, 2011)
- Barrier: quality and comparability of information



## Objective

Although we have open government data,  
few studies discuss:

- how to use
- what tools

### This paper :

Propose a list of audit apps that help professional auditors and citizens to

- **analyze** open government procurement data
- **find out** suspicious contracts which have higher probability of fraud, and detect exception and anomaly

## Why Audit Apps

What is it

- Formalized audit procedures that are performed through computer scripts (Dai et al. 2014)

Example

- Caseware and ACL: test journal entries, account payable, assets, etc

Advantages

- simplify data analytics procedures, require few user interactions, improve audit quality

No apps for open government data analysis or for non-professional auditors such as “armchair auditors”

# Proposed Apps for Government Expenditure Audit

## Guidance:

- **Procurement Fraud Handbook**
  - prepared by the General Services Administration (GSA) Office of Inspector General (OIG) (GSA, 2012)
- **The Brazilian General Accounting Office (GAO) -TCU**

## Fraud Type : Bid Rigging

No.	Purpose of App	Data Needed	Fraud Indicator
1	Monopoly check	Market data	only very fewer suppliers
2	Contract prices comparison (gov. VS other clients)	Prices to different clients	Contractor submit higher price bids to government for exactly same product /service
3	Distribution and gap analysis on all bidding prices (winner & losing bidders)	Pricing info of all bidders	A large gap exists
4	Bidders withdraw detection (in a short time period)	Bidding process information	Qualified bidders inexplicably withdraw valid bids, especially if multiple bids are withdrawn in the same time frame
5	Bids wining history check	Statistic contract data	a certain contractor always or never wins a bid, or all contractors win an equal volume of contracts over time

## Fraud Type: Bribery, Kickbacks, and Conflicts of Interest

No.	Purpose of App	Data needed	Fraud Indicator
1	Relationship check (gov. personnel VS contractor)	Background information of both parties	employment of contractor or sub-, or their family member in government personnel
2	Orders changes check (times, costs)	order change data	Many change orders; Change orders with a high percent of original costs

## Fraud Type:

### Changing for products not used or services not rendered

No.	Purpose of App	Data needed	Fraud Indicator
1	<b>Duplicate billings check</b> (for same products or services)	billings	Duplicate billings for the same products or services
2	<b>Address check</b> (company's & delivery)	addresses	Delivery location is not the office, plant, or job site
3	<b>Geographic information check</b> (identify invoices that indicate work at multiple distant job sites on same day)	invoices	Employees bill at multiple distant job sites on same day
4	<b>Weird working hours check</b>	invoices	Employees bill for more hours than typically worked in a day <sup>72</sup>



## Other apps:

No.	Purpose of the app	Data needed	Fraud indicator
1	Contract reliable values check (unusual “0” and tiny)	initial values of contracts	Unusual number in the values, such as 0, 0.01,0.05
2	Contractor qualification check (“blacklist” companies)	Contractor information, “blacklist”	Contractor once occurred in the “blacklist”
3	“Waived bidding” contracts check	Bidding type information	firm has very high percentage of “waived bidding” contracts in all contracts with gov
4	Data Completeness and Integrity Check	contracts data	Contract records lost important contracting information

# Illustrations

## Data:

Contracts of Brazil federal government from 1989 to 2014 from SIASG  
(Brazilian public federal procurement information system)

## 1. Descriptive Analysis

Software : Qlik Sense Enterprise

-- dashboard for visualization

## Qlik Sense

### Descriptive Analysis App

#### Modalidade da Licita...

- 01: CONVITE
- 02: TOMADA DE PREÇOS
- 03: CONCORRÊNCIA
- 04: CONCORRÊNCIA INTER...
- 05: PREGÃO
- 06: DISPENSA DE LICITAÇÃO
- 07: INEXIGIBILIDADE DE LI...

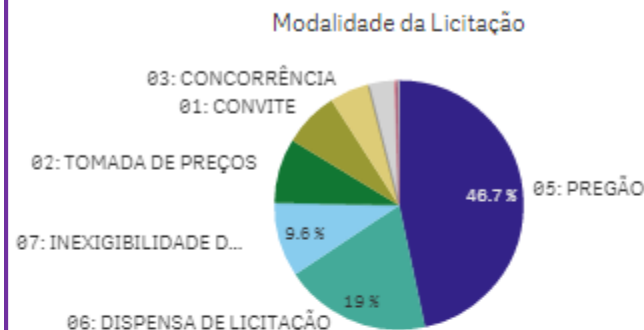
#### Contratada

- Fornecedor 00.000.000/00...
- Fornecedor 00.000.000/00...
- Fornecedor 00.000.000/00...

#### UASG

- 020001: SENADO FEDERAL
- 060001: STM\_ SUPERIOR T...
- 060020: STM-3A.AUDITORI...

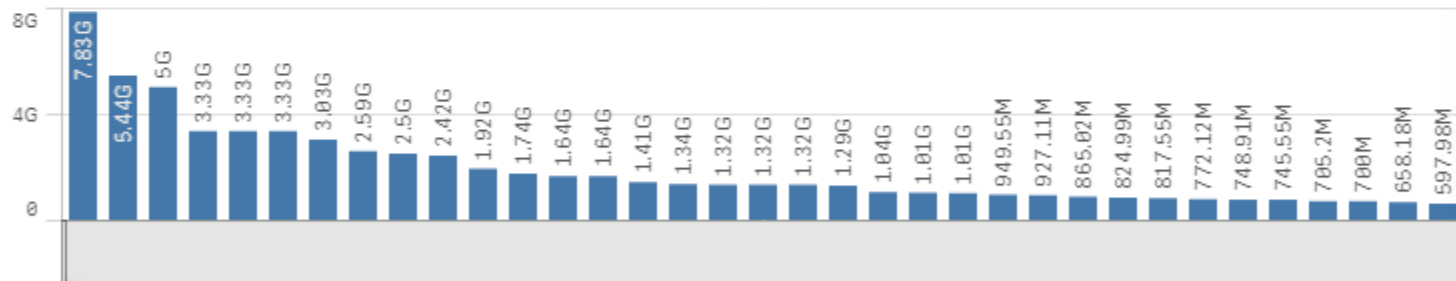
#### Analyze by Bidding Mode



#### Contract Value by bidding Mode

Modalidade da Li...	avg([Valor inicial])
-	3496645
99: null	67013472
44: CONCORRÊNCIA INTERNACIONAL POR TÉCNICA E PREÇO	5866328.8
04: CONCORRÊNCIA INTERNACIONAL	5737919.4

#### Top contract value



#### Analysis by UASG

153261: HOSPITAL CLINICAS/UF MG	153163: MEC - UNIV. FED DE SANTA CATARINA...	160069: COMAND. LOGISTICO	154421: FUND. UNIVERSIDADE FEDERAL VA...	153010: MEC-CEF ET-CENT. FED.ED.T EC.CELS...	153173: FND- MEC-FUND ONAC.DE DESENVOLVIMENTO DA EDUCACAO/DF	05: PREGÃO	06: DISPENSA DE LICITAÇÃO
---------------------------------	--	---------------------------	--	--	--	------------	---------------------------

## 2. Data Completeness and Integrity Check

Software : Caseware IDEA

-- Integrity Check for Missing Contractors

### App script

```
Sub Main
  Call ExcelImport() 'C:\Users\Administrator\Downloads\integrity check.xlsx
  Call DirectExtraction() 'integrity check3-Sheet2.IMD
End Sub

'File - Import Assistant: Excel
Function ExcelImport
  Set task = Client.GetImportTask("ImportExcel")
  dbName = "C:\Users\Administrator\Downloads\integrity check.xlsx"
  task.FileToImport = dbName
  task.SheetToImport = "Sheet2"
  task.OutputFilePrefix = "integrity check3"
  task.FirstRowIsFieldName = "TRUE"
  task.EmptyNumericFieldAsZero = "FALSE"
  task.PerformTask
  dbName = task.OutputFilePath("Sheet2")
  Set task = Nothing
  Client.OpenDatabase(dbName)
End Function

'Data: Direct Extraction
Function DirectExtraction
  Set db = Client.OpenDatabase("integrity check3-Sheet2.IMD")
  Set task = db.Extraction
  task.IncludeAllFields
  dbName = "integrity check - Contractors.IMD"
  task.AddExtraction dbName, "", "@IsBlank( CONTRATADA )"
  task.CreateVirtualDatabase = False
  task.PerformTask 1, db.Count
  Set task = Nothing
  Set db = Nothing
  Client.OpenDatabase (dbName)
End Function
```

### Sample results

	IDENTIFICADOR_DO_CONTRATO
1	11460650000011984
2	15404753000011984
3	15404753000011986
4	17011650000011988
5	51208450000011992
6	51211150000011992
7	17011950000011990

Integrated results:

For contracts that lost contractor records, 90% belong to waived bidding

In 470,683 contracts,

- 35,516 contracts lose contractor information
- 6,167 contracts lose bidding mode
- 1,000 contracts lost valid dates

## 3. Anomaly Detection

Software: Caseware IDEA

-- unusual initial values

### App script

```
Sub Main
    Call ExcellImport() 'C:\Users\Administrator\Downloads\initial value.xlsx
    Call DirectExtraction() 'initial value2-Sheet1.IMD
End Sub

'File - Import Assistant: Excel
Function ExcellImport
    Set task = Client.GetImportTask("ImportExcel")
    dbName = "C:\Users\Administrator\Downloads\initial value.xlsx"
    task.FileToImport = dbName
    task.SheetToImport = "Sheet1"
    task.OutputFilePrefix = "initial value2"
    task.FirstRowIsFieldName = "TRUE"
    task.EmptyNumericFieldsAsZero = "FALSE"
    task.PerformTask
    dbName = task.OutputFilePath("Sheet1")
    Set task = Nothing
    Client.OpenDatabase(dbName)
End Function

'Data: Direct Extraction
Function DirectExtraction
    Set db = Client.OpenDatabase("initial value2-Sheet1.IMD")
    Set task = db.Extraction
    task.IncludeAllFields
    dbName = "small value.IMD"
    task.AddExtraction dbName, "", "VALOR_INICIAL < 0.1"
    task.CreateVirtualDatabase = False
    task.PerformTask 1, db.Count
    Set task = Nothing
    Set db = Nothing
    Client.OpenDatabase(dbName)
End Function
```

### Sample results

IDENTIFICADOR_DO_CONTRATO	VALOR_INICIAL
17007854000011994	0.00
17005854000011996	0.00
15326654000011996	0.00
15325454000011996	0.00
15326654000011996	0.00
15325454000011996	0.00
15301752000011996	0.00
20100454000012000	0.00
15303252000011996	0.00
25502652000012000	0.00

Integrated results:

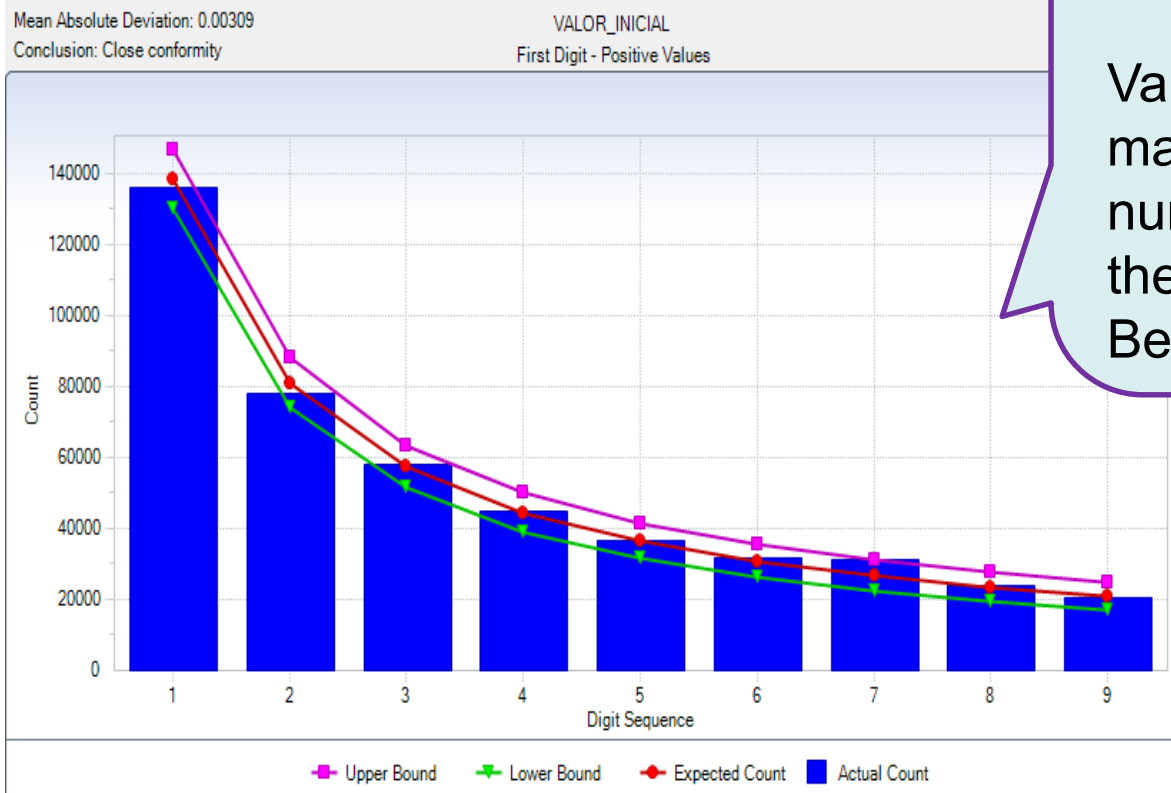
501 contracts that have "0" value after removing contracts pertaining to government departments

527 contracts have values that <1; the values are 0.01, 0.05, 0.1, and 0.53 Brazilian real

## 3. Anomaly Detection

Software used: Caseware IDEA

-- Benford's Law



Widely used for accounting fraud detection

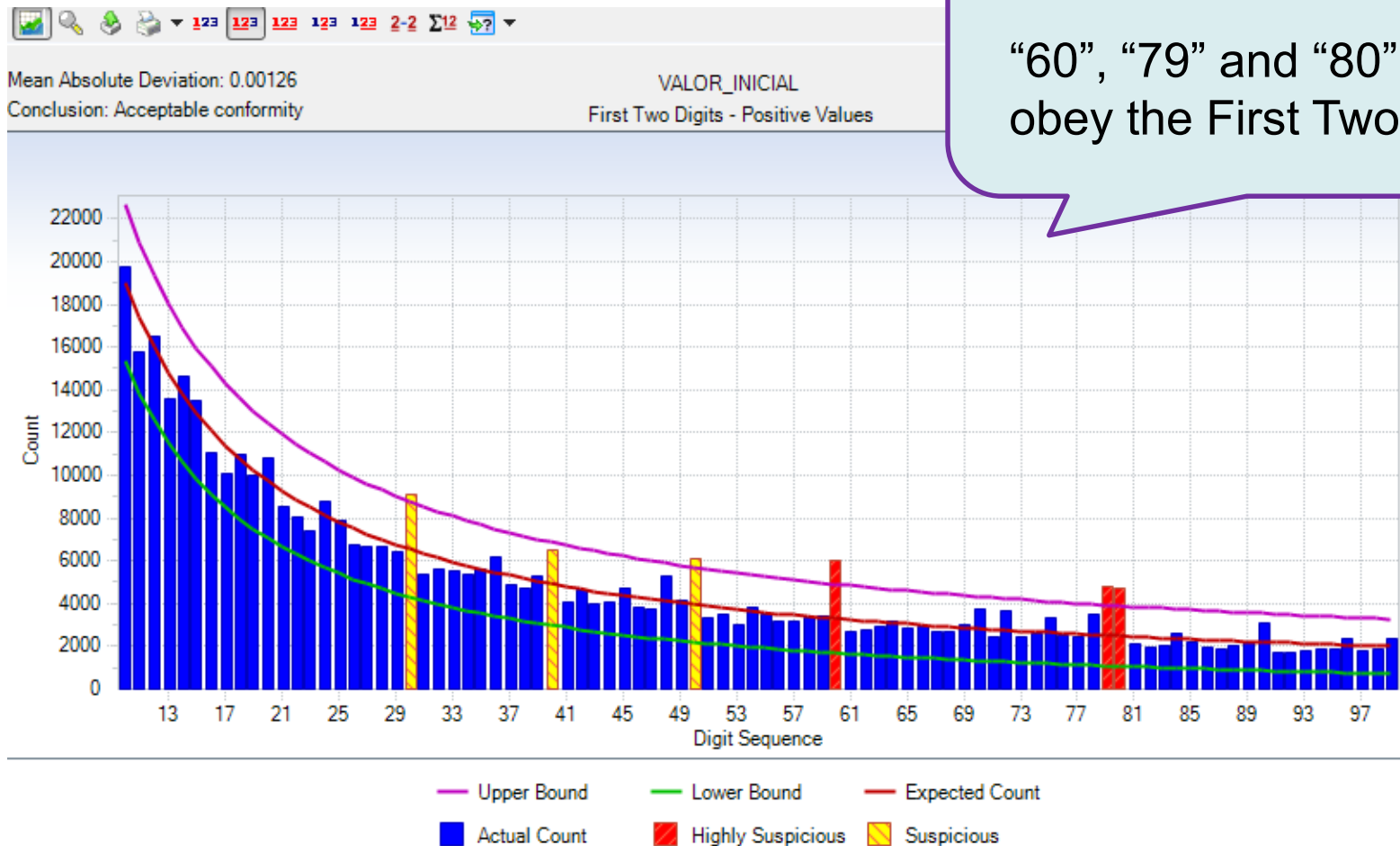
Values should come from mathematical combination of numbers (quantity  $\times$  price), they are expected to obey Benford's Law

First digit

## 3. Anomaly Detection

Software: Caseware IDEA

-- Benford's Law Check



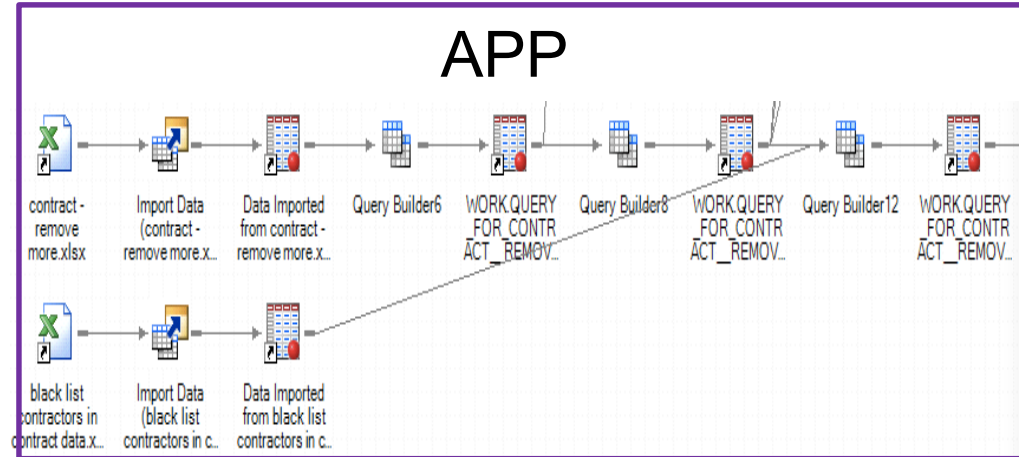
First Two Digit:

“60”, “79” and “80” do not obey the First Two Digit Law

## 3. Anomaly Detection

Software: SAS

--“black list” Contractor Detection



## Sample results

Contractor	Frequency
<b>33.000.118</b>	1717
<b>00.212.655</b>	405
<b>29.739.737</b>	404
<b>10.788.628</b>	375
<b>00.329.379</b>	345

## Integrated results:

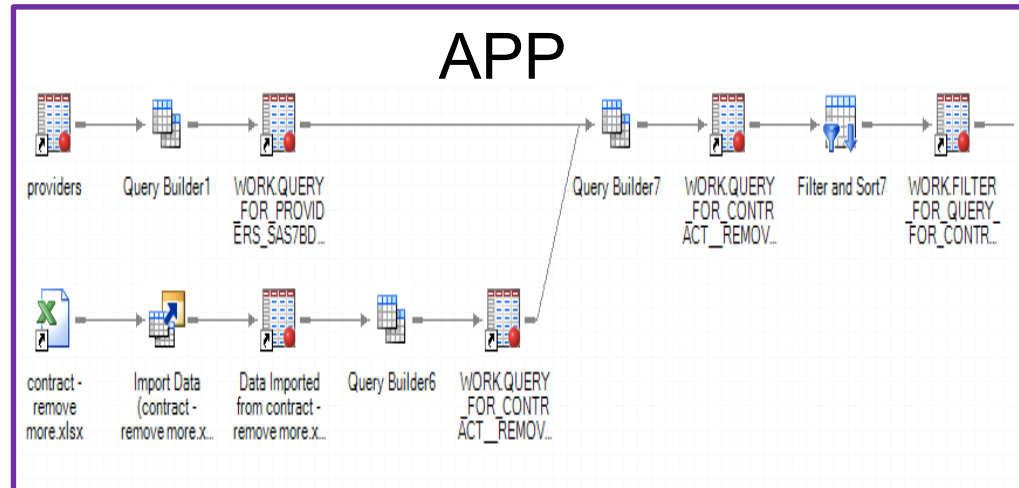
- 25,100 contracts are made with contractors listed in the blacklist
- 1,936 unique suspicious contractors (firms)



## 3. Anomaly Detection

Software: SAS

-- matching with other data sources (contractor info)



Sample results

Integrated results:

- 40,942 contracts are signed with companies that do not have any information in the contractor file
- 18,778 unique suspicious contractors

Contractor	Frequency
<b>29.213.386/0001-00</b>	164
<b>64.198.393/0001-72</b>	141
<b>59.363.937/0001-74</b>	113
<b>00.394.494/0016-12</b>	106
<b>00.656.865/0001-00</b>	103

### 3. Anomaly Detection

Software: Excel

-- Big Data Collection

legal foundation explaining why the contract can waive bidding processes

Contract ID	Bidding Mode	Objective	Link to legislation
<b>11460650000011984</b>	06: DISPENSA DE LICITAÇÃO	Contratação de imóveis para instalação da Agência do IBGE no município de Conceição do Araguaia/PA.	<a href="#">De acordo com artigo 24, Inciso X, da Lei 8.666/93 C/CR.PR. 06/96, ARTIGO 3.</a>
<b>38004450000011992</b>	06: DISPENSA DE LICITAÇÃO	Contrato de locação do imóvel da Av. Dr. Vicente Machado n.º 362 - Curitiba/PR.	<a href="#">Art. 24, Inciso X, da lei 8666/93.</a>
<b>17011650000011988</b>	07: INEXIGIBILIDADE DE LICITAÇÃO	Contrato nº 01/88 tem por objeto a locação dos imóveis nºs 26, 38 e 44 da Praça Oliveira Figueiredo, Barra do Piraí, Estado do Rio de Janeiro.	<a href="#">Decretos-Leis nos. 2300/86 e 2348/87 e Lei 6649/79</a>
<b>17011950000011990</b>	06: DISPENSA DE LICITAÇÃO	Locação dos imóveis de nos. 26, 38 e 44 da Praça Oliveira Figueiredo para abrigar a Agência da Receita Federal em Barra do Piraí	<a href="#">Decretos-Leis nos. 2300/86 e 2348/87 e Lei 6649/79</a>

## Limitations and Future Research

- Design, improve and test the apps
- Apply other data analytical tools (such as ACL and R)
- developing rule-based algorithm for improved government procurement fraud detection, applying the idea of exceptional exception (Issa, 2013) to rank suspicious contracts based on predefined rules