Moving Towards CA and Big Data with Audit Analytics

Big data and the Audit Data Standard: forensic accounting implications

12th Fraud Seminar

December 1st, 2015

Miklos A. Vasarhelyi

KPMG Distinguished Professor of AIS

Director of the CarLab

Rutgers Business School

Outline

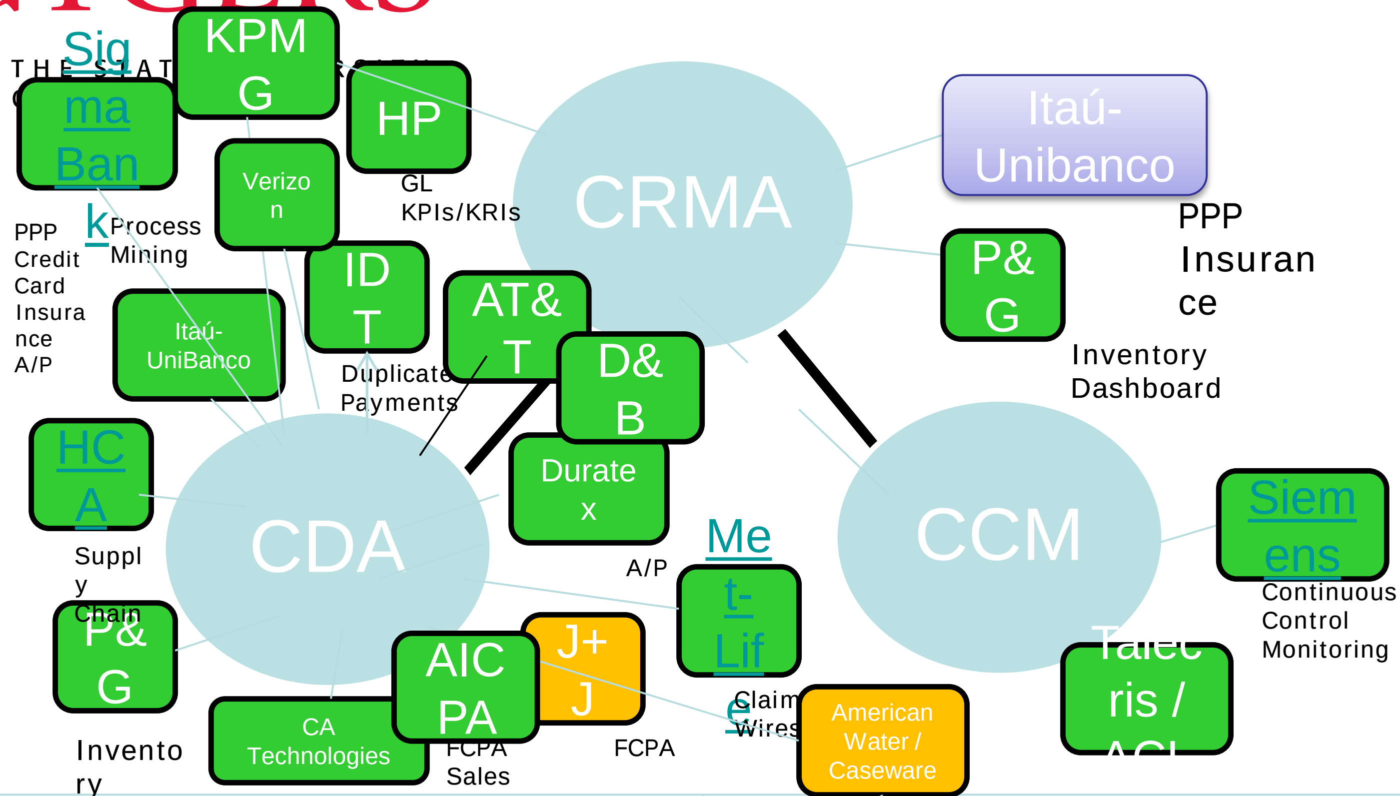
- The CarLab
- Big data and evidence
- Audit data standard and “apps”
- Imagineering
- From the IAASB presentation
- Moving towards continuous audit and data analytics
- Some additional analytics
- Conclusions

Rutgers Business School

THE CARLAB (CONTINUOUS AUDIT AND REPORTING LAB)

Visit

- http://raw.rutgers.edu
- miklosv@rutgers.edu



Audit Automation

P&G: Order to Cash

Auditor Judgment

Siemens- AAS Automation

AICPA – ADS / APS

Audit Methodologies

- **Multidimensional Clustering**
- **Process Mining**
- **Continuity Equations**
- **Predictive Auditing**
- **Visualization**
- **Analytic Playpen**

Moving Towards CA and Big Data with Audit Analytics

The CarLab

	Predictive Analytics with Weather data	Audit data analytics and EDA	Envisaging the future of audit and Big Data	Text Mining	Monitoring Unibanco's branches
Visualization	Process Mining at Gamma Bank	Expert System for P-Card	Logit regression for control risk	Exceptional Exceptions	Client Retention Project
Litigation prediction	Fraud Risk Assessment using EDA	Detecting duplicate records	Continuity equations	Predictive Audit	Credit card Default prediction
Insurance Analytics	Multidimensional clustering for fraud detection	Rule-based selection for transitory accounts	Continuity Equations at HCA	XBRL	Insurance Analytics

AND SEAPS INTO EDUCATION

Usage

semester, and additional courses being added to the library.

- New Camera Equipment

Moving Towards CA and Big Data with Audit Analytics

Continuous audit, audit analytics, and forensics

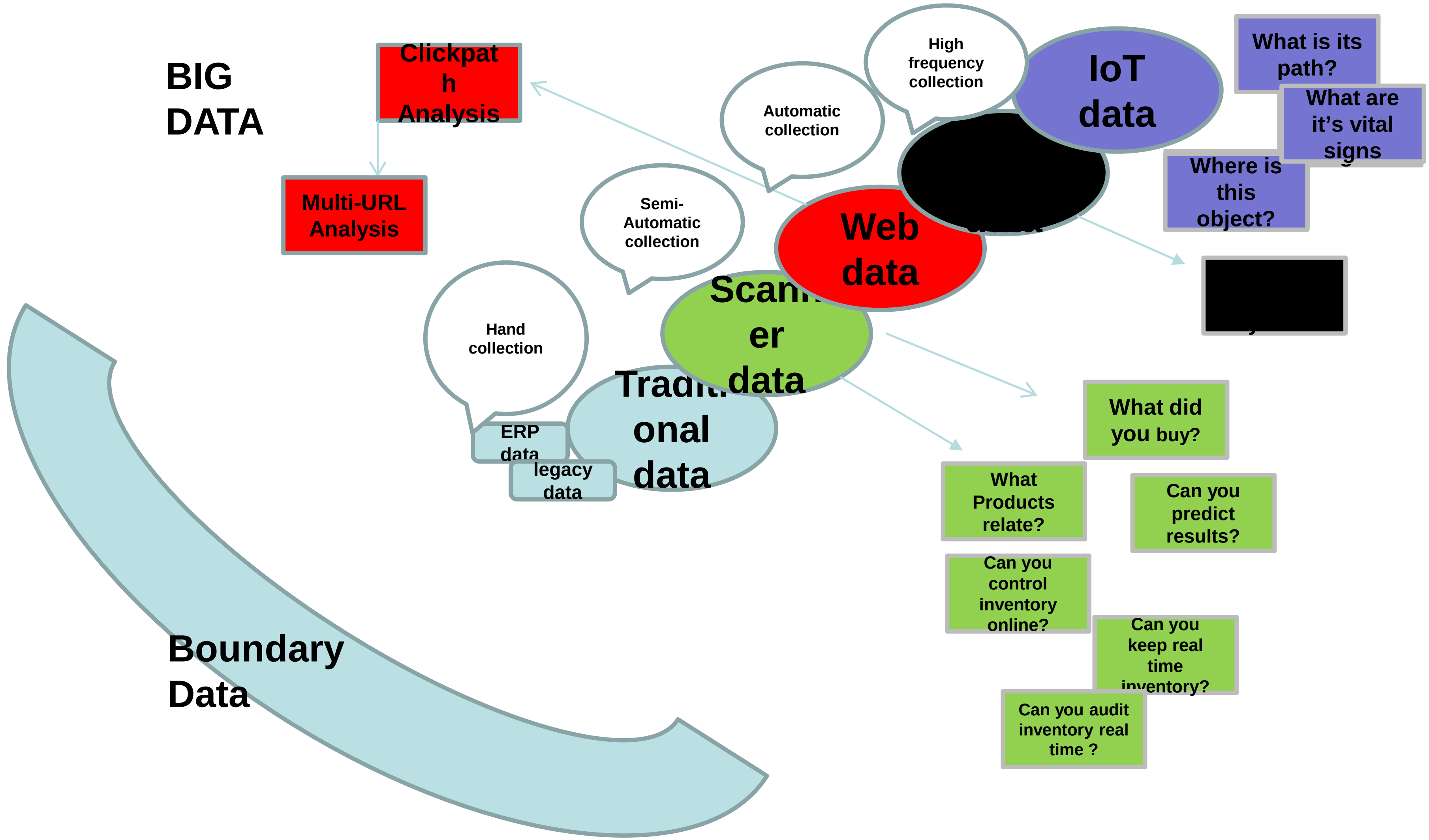
- Continuous audit is by its essence based in automation and analytic methods, although follow-ups tend to currently be automated
- The “lines of defense” are progressively being confused by automation. An outlier finding is as interesting to auditors and forensic accountants as for internal auditors and managers
- Big data generates substantive follow-up that cannot be done without analytics (exceptional exceptions), some automation, and most of times (still) human judgment
- Audit can be divided in retroactive and predictive
  - Predictive can be broken down as preventive and predictive
- Fraud examples allow “trained learning methods”

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

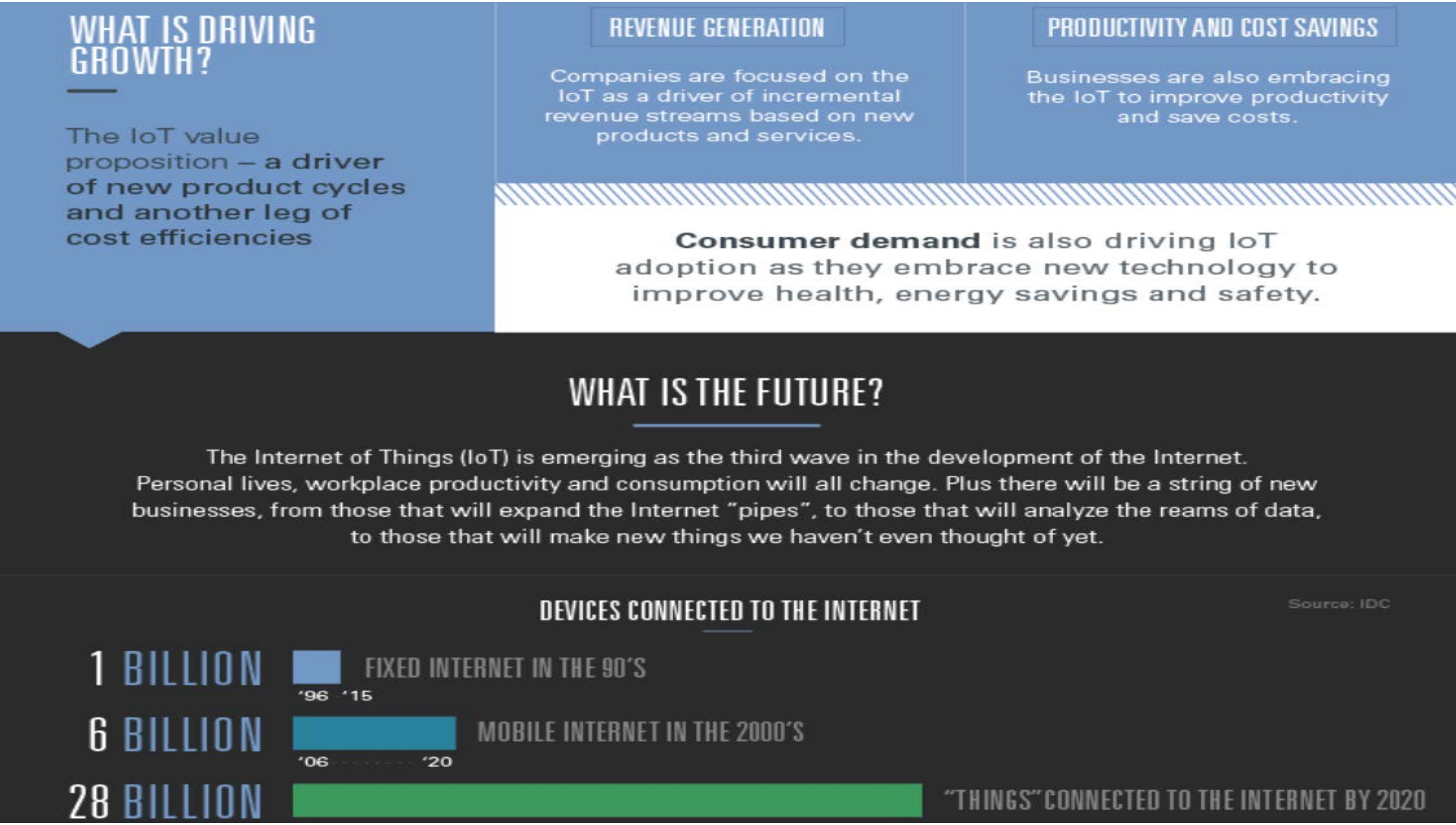
Big data and audit evidence

Helen Brown-Liburd and Miklos A. Vasarhelyi

Moving Towards CA and Big Data with Audit Analytics

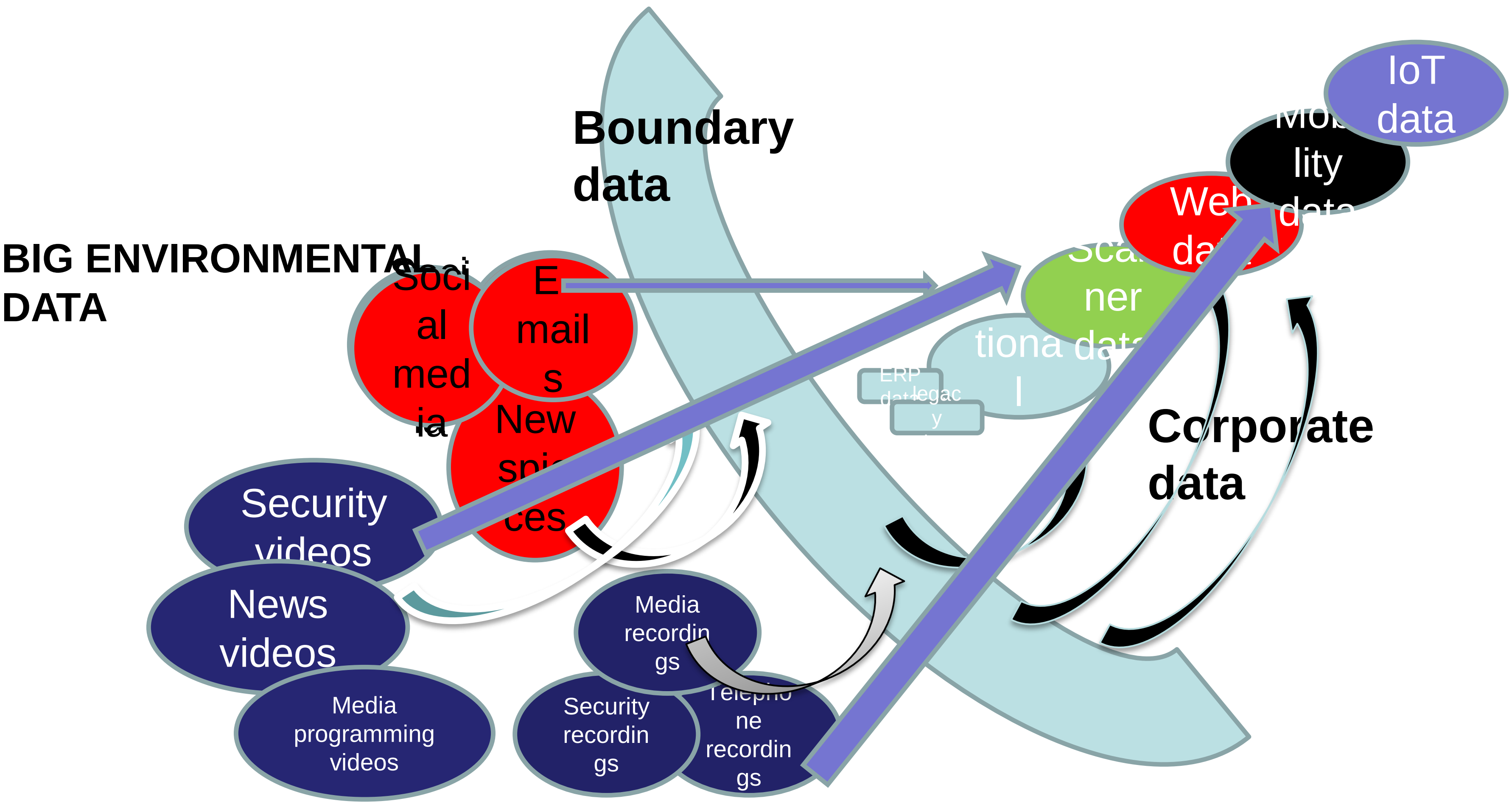


Moving Towards CA and Big Data with Audit Analytics



RUTGERS

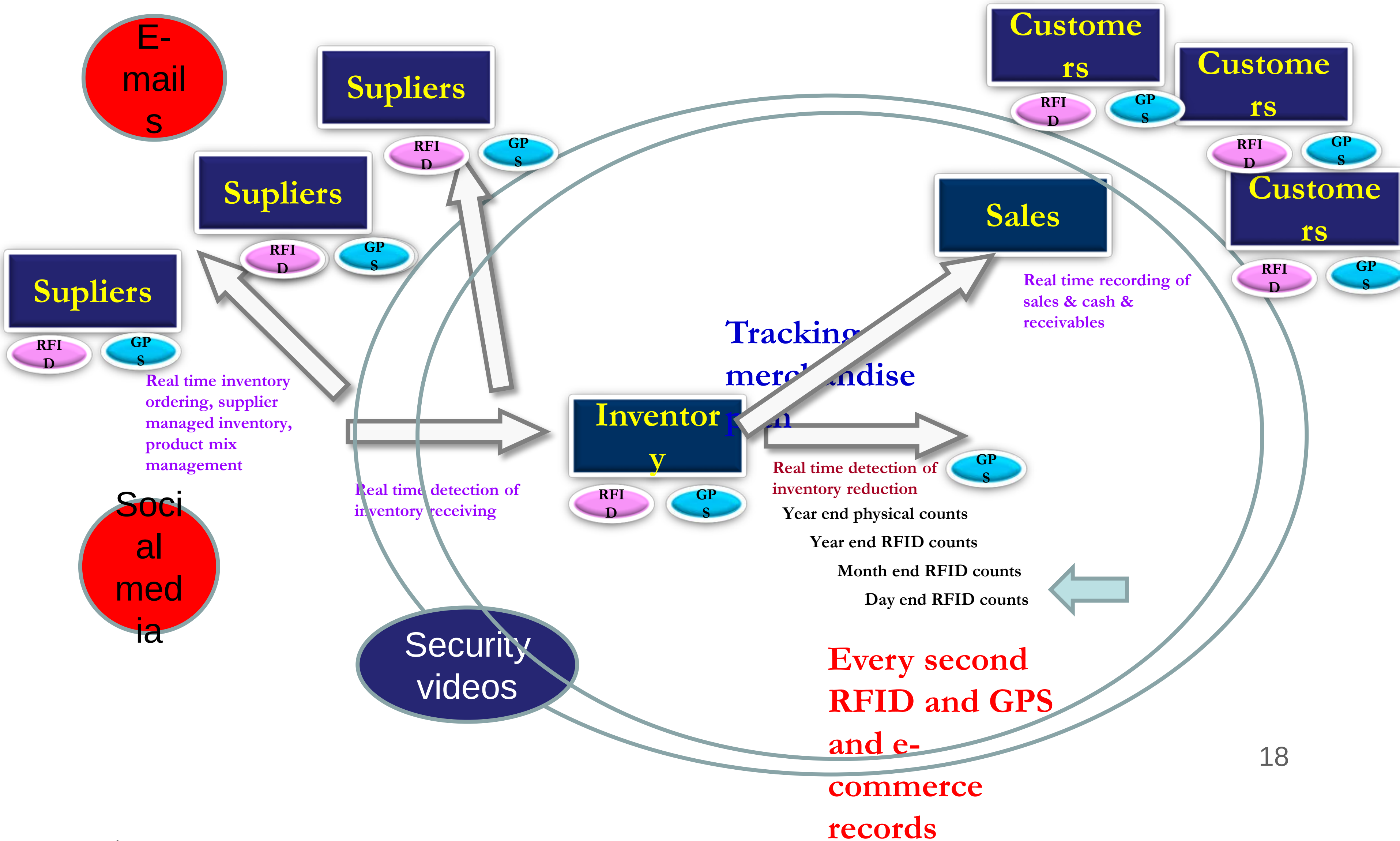
Moving Towards CA and Big Data with Audit Analytics



LOOKING AHEAD (IMAGINEERING)

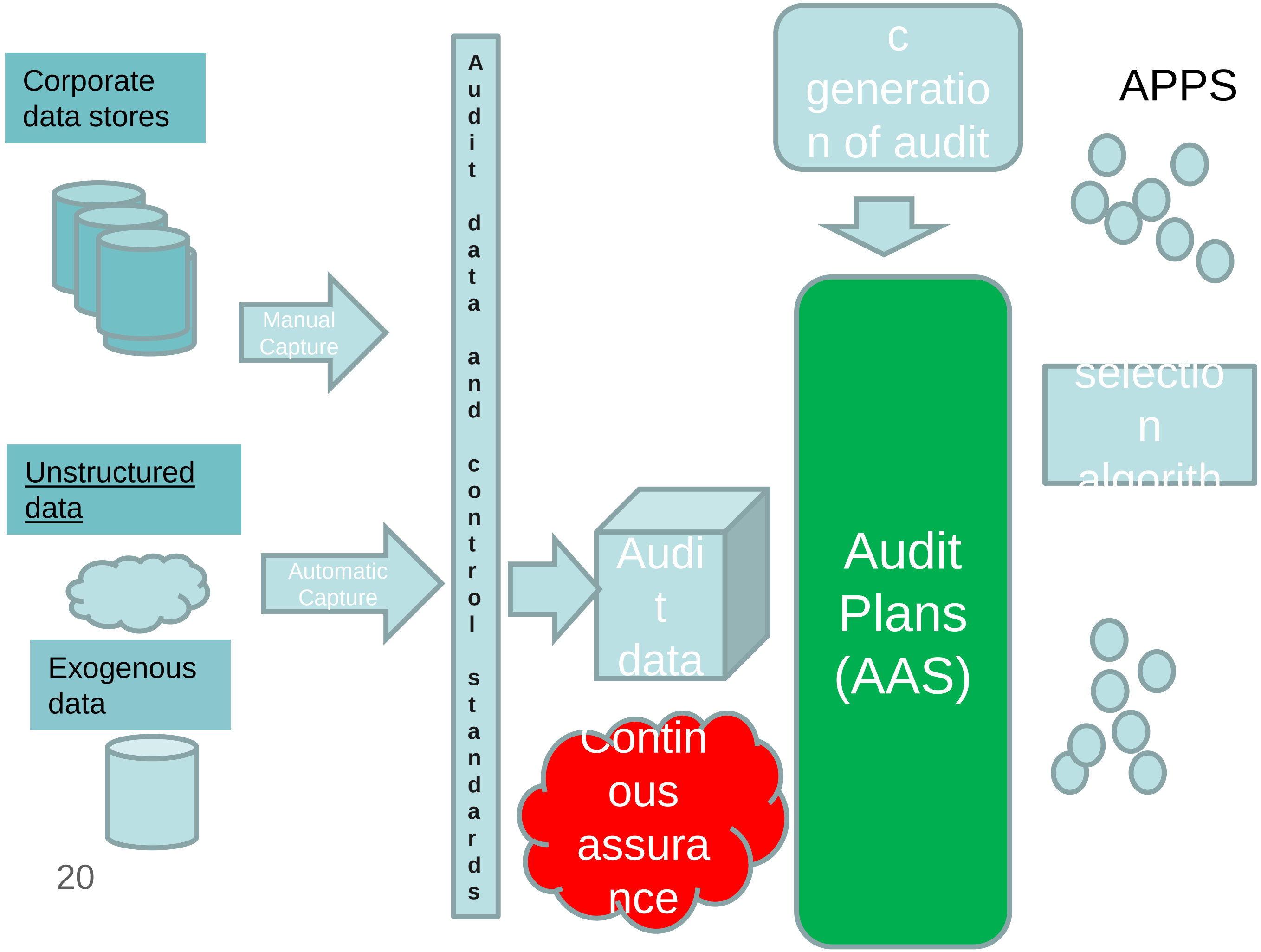
THE THINKING THAT MUST GO INTO CHANGE

Moving Towards CA and Big Data with Audit Analytics



Audit Data Standard and "Apps"

Moving Towards CA and Big Data with Audit Analytics



AUDIT EVIDENCE FROM BIG DATA

Digital environment characteristics

- Multiple use of single data strings
- Sliceable at many intervals of time
- Can be accumulated by multiple parties
- May be subject to many encryption protocols
- Fixed development costs minimum variable costs mainly composed of storage
- The effects of piggybacking
  - Code sharing
  - Multiple layers
  - Several layers of entry points

Audit Data Analytics

Bob Dohrer, IAASB Member and Working Group Chair

Miklos Vasarhelyi

Phillip McCollough

IAASB Meeting
September 2015
Agenda Item 6-A

- Procedure: For every invoice, shipping document and sales order received from customers, compare the invoiced customer, quantity, and unit price to the quantity shipped per the shipping documents and the quantity and unit price reflected in the sales order received from the customer.
- Objective: Obtaining audit evidence over the existence and accuracy of revenue. (ISA 500 paragraphs 6 and 9).
- Prior year approach: Tests of internal controls over the revenue process, substantive analytical procedures and tests of details (sampling).

Entity ABC has revenue of €125 million generated by 725,000 transactions. The three way match procedure is executed with the following results:

	Amount (€'000)	%	Number of Transactions	%
No ~~exceptions~~ differences	119,750	95.8	691,000	95.3
~~Exceptions~~ Outliers				
Quantity differences	3,125	2.5	16,700	2.3
Pricing differences	2,125	1.7	17,300	2.4

Note:

- Procedure: For every sales transaction, evaluation of any segregation of duties conflicts relative to customer master file maintenance, sales order processing, sales invoicing, sales returns / credit notes, and applying cash collections.
- Objective: Obtaining audit evidence over the effectiveness of internal control over sales processing.[1] (ISA 330 paragraphs 8–9 and ISA 315 paragraph 12)

[1] This is one of the tests of control over sales processing.

For entity ABC, an analysis of segregation of duties was executed with the following findings:

	Number of users	Amount (€'000)	Number of transactions
Population of sales	542	125,000	725,000
Instances in which same individual created sales invoice, sales return or credit note and applied cash	7	620	3,934
Instances in which same individual executed sales order processing, dispatched goods (delivery document) and applied cash	96	7,692	46,903

Note: Materiality for the audit of the financial statements as a whole is €1,000,000.

Illustration 3 – Predictive Analytic (cont.)

Clustering Using Store Sales by Peer Group

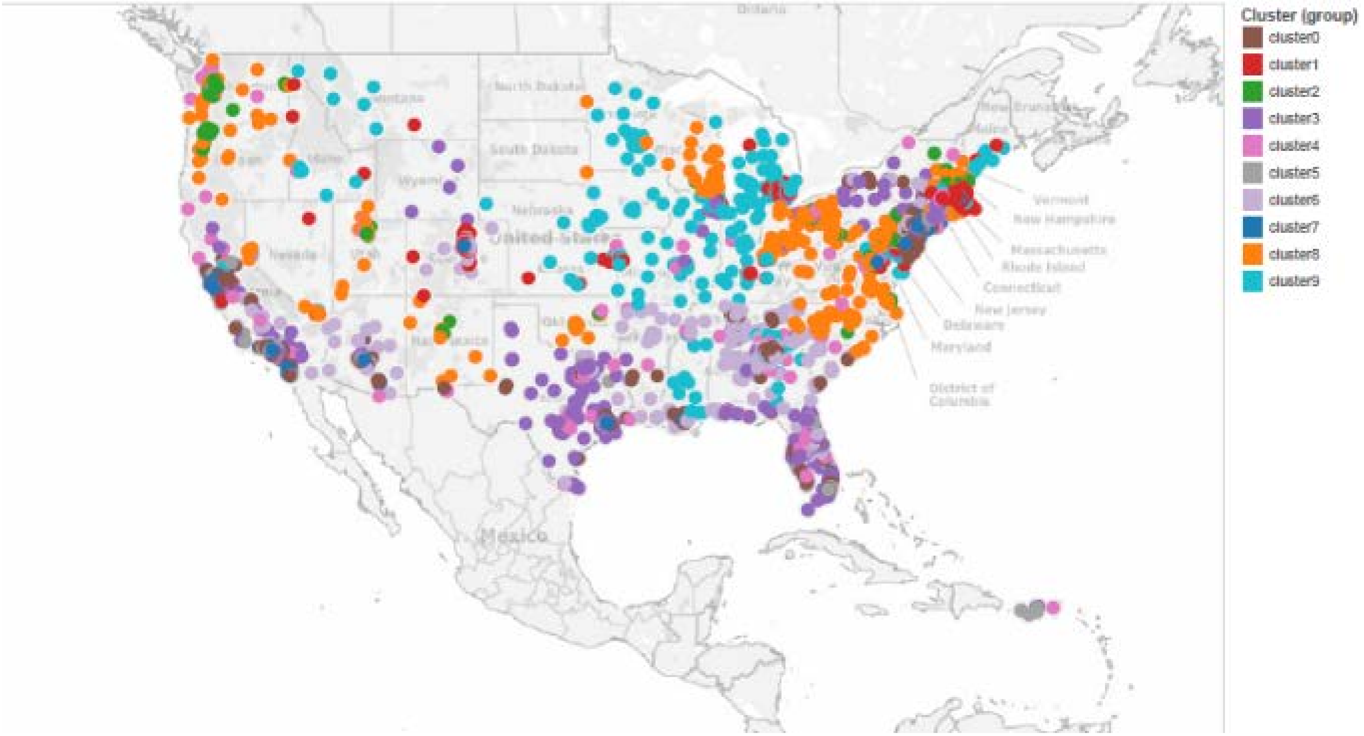
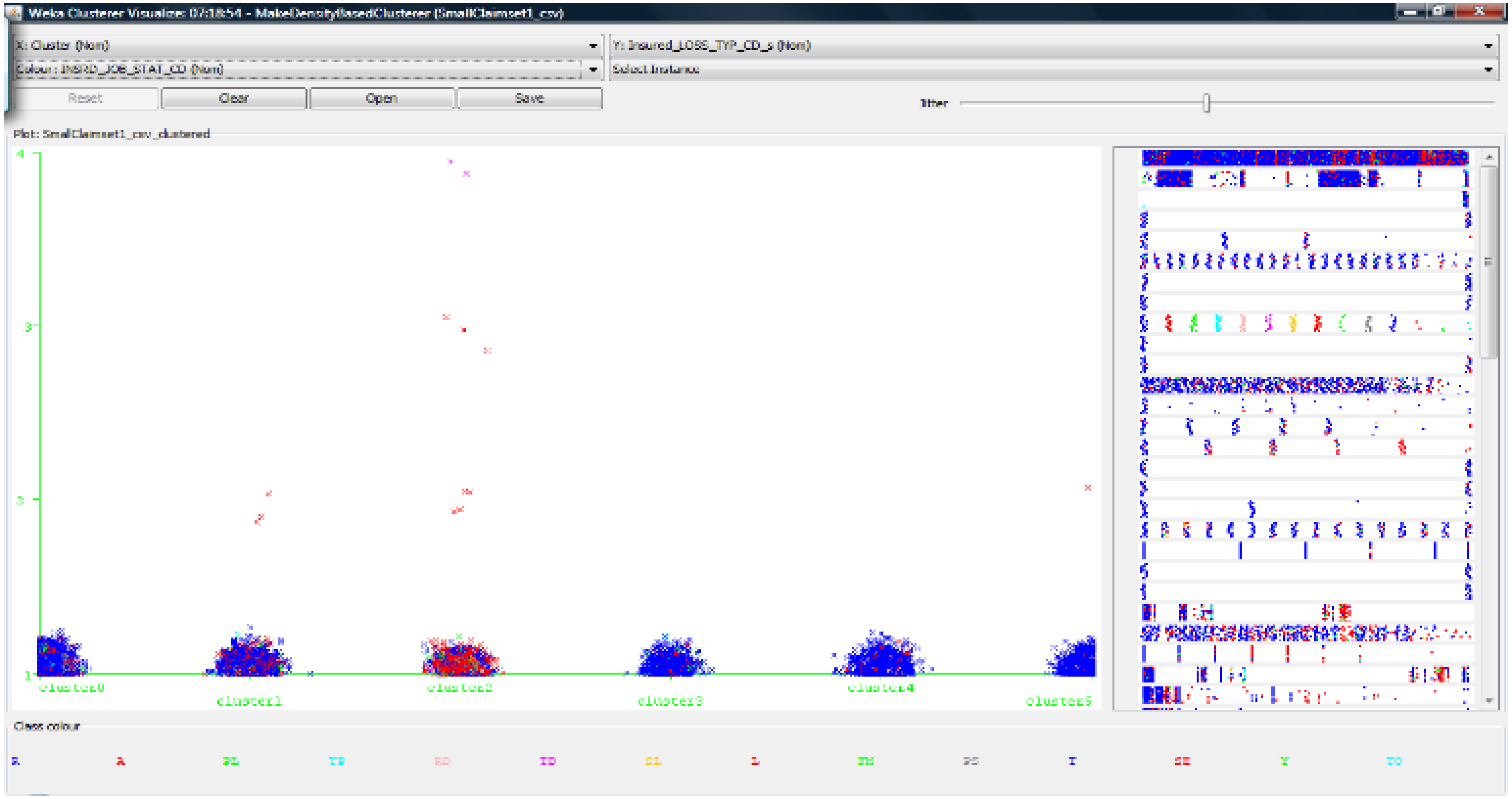
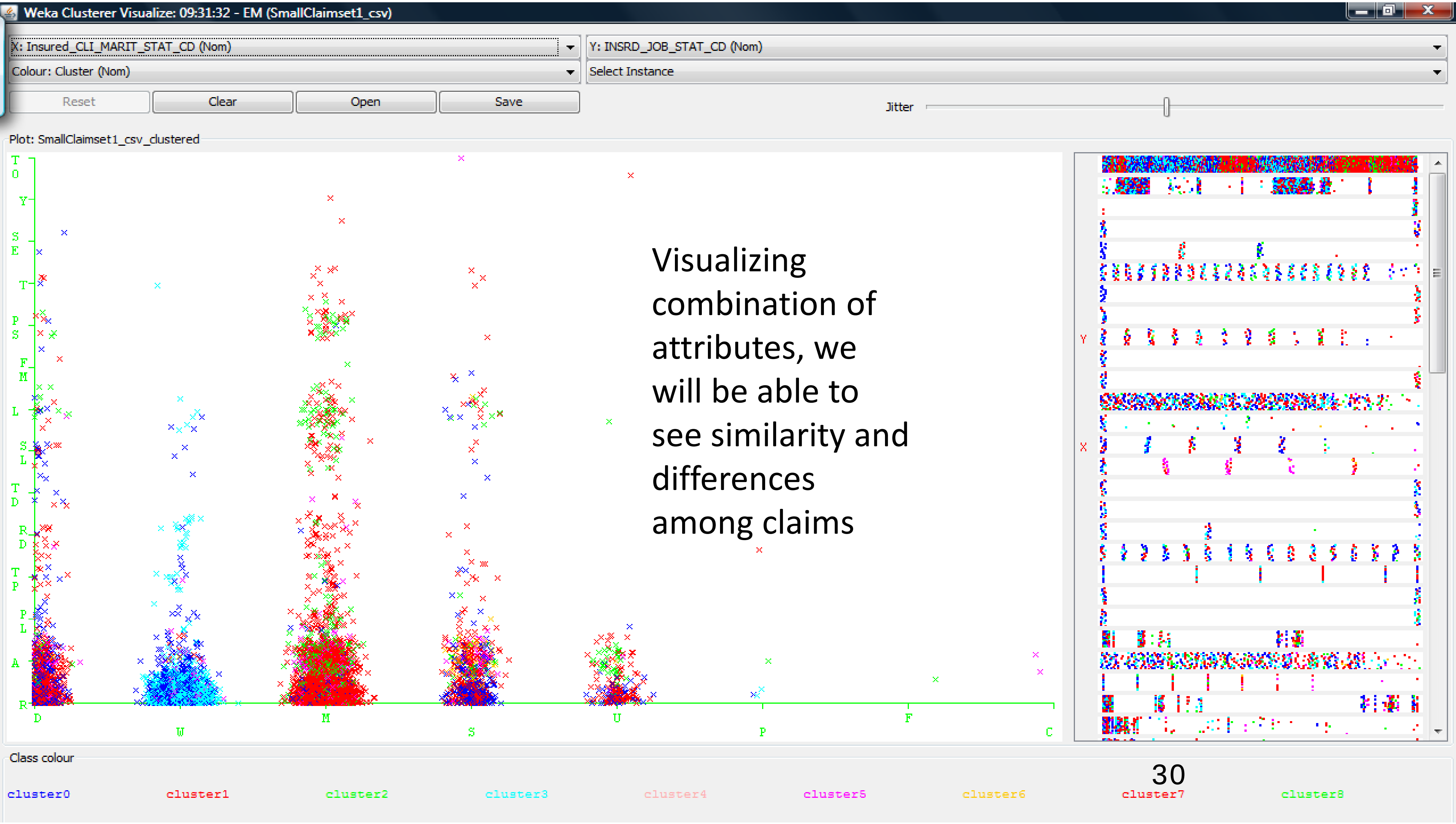


Illustration 4 – Clustering (cont.)



Moving Towards CA and Big Data with Audit Analytics



Visualizing combination of attributes, we will be able to see similarity and differences among claims

RUTGERS

THE STATE UNIVERSITY
OF NEW JERSEY

Moving towards continuous audit and big data with audit analytics: Implications for research and practice

By Deniz Appelbaum, Alexandr Kogan, and Miklos Vasarhelyi

of Rutgers, the State University of New Jersey

for the

35th World Continuous Auditing & Reporting Symposium

November 6&7, 2015

Hosted by: Rutgers Accounting Research Center (RARC) and Continuous Auditing & Reporting Laboratory (CAR-Lab)

Audit Data Analytics (ADA): One way to define…..

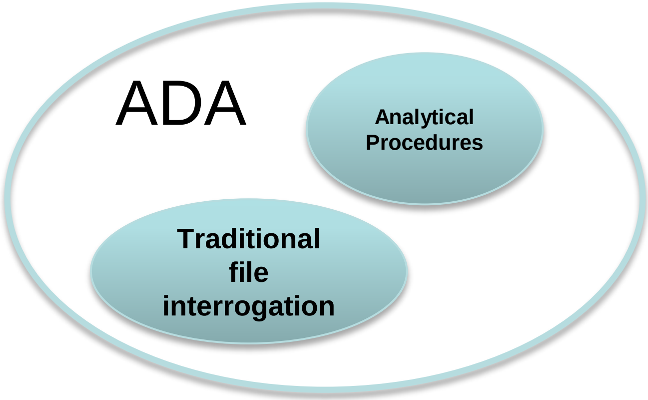
Audit Data Analytics (ADA) is the analysis of data underlying financial statements, together with related financial or non-financial information, for the purpose of identifying potential misstatements or risks of material misstatement.

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 interrogation



(Liu, 2014)

ADA *mode* can be *exploratory* or *confirmatory*

	Exploratory mode	Confirmatory mode
When	Planning	Performance
Question	What is going on here? Does the data suggest something might have gone wrong? Where do the risks appear to be? What assertions should we focus on?	Does 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

(Stewart, 2015)

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)

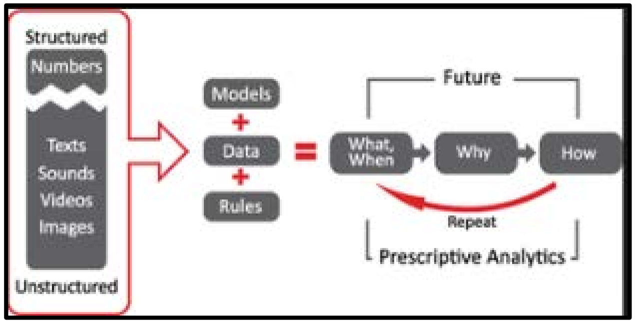
Exploratory and confirmatory ADA is a spectrum of analytics and the processes are iterative, starting with exploratory

4

(Stewart and Liu, 2014)

Audit Analytics Framework

Orientation: descriptive, predictive or prescriptive



Techniques:

- qualitative or quantitative
- deterministic or statistical
- based on unstructured, semi-structured, or structured data

The most commonly used AA techniques are those that are quantitative, statistical, and based on structured data

The dominance of quantitative techniques in AA is due to the fact that the main objective of external audit is to provide assurance on the accounting numbers. Therefore, the accounting numbers are the quantities that are the focus of AA.

(Wikipedia, 2015)

Literature Review of Analytical Procedures in the External Audit : Techniques

Machine Learning/Data Mining:	Process Mining	Simulation, Process Optimization
	Advanced Classifiers	Support Vector Machine (SVM), Artificial Neural Networks (ANN), Multilayer Feed Forward Neural Network (MLFF), Genetic Algorithm
	Rules-Based Classifiers	Expert Systems/Decision Aids, Majority Vote, AntMiner +
	Ensemble Methods	Boosting, Bagging, Bootstrap
	Decision Trees	C4.5 statistical classifier
	Bayes Classifiers/Probability Models	Bayesian Theory/Bayesian Belief Networks (BBN), Naïve Bayes, Dempster-Shafer Theory, Probability Theory
	Other	Clustering, Text Mining, Visualization, Group Method of Data Handling (GMDH)

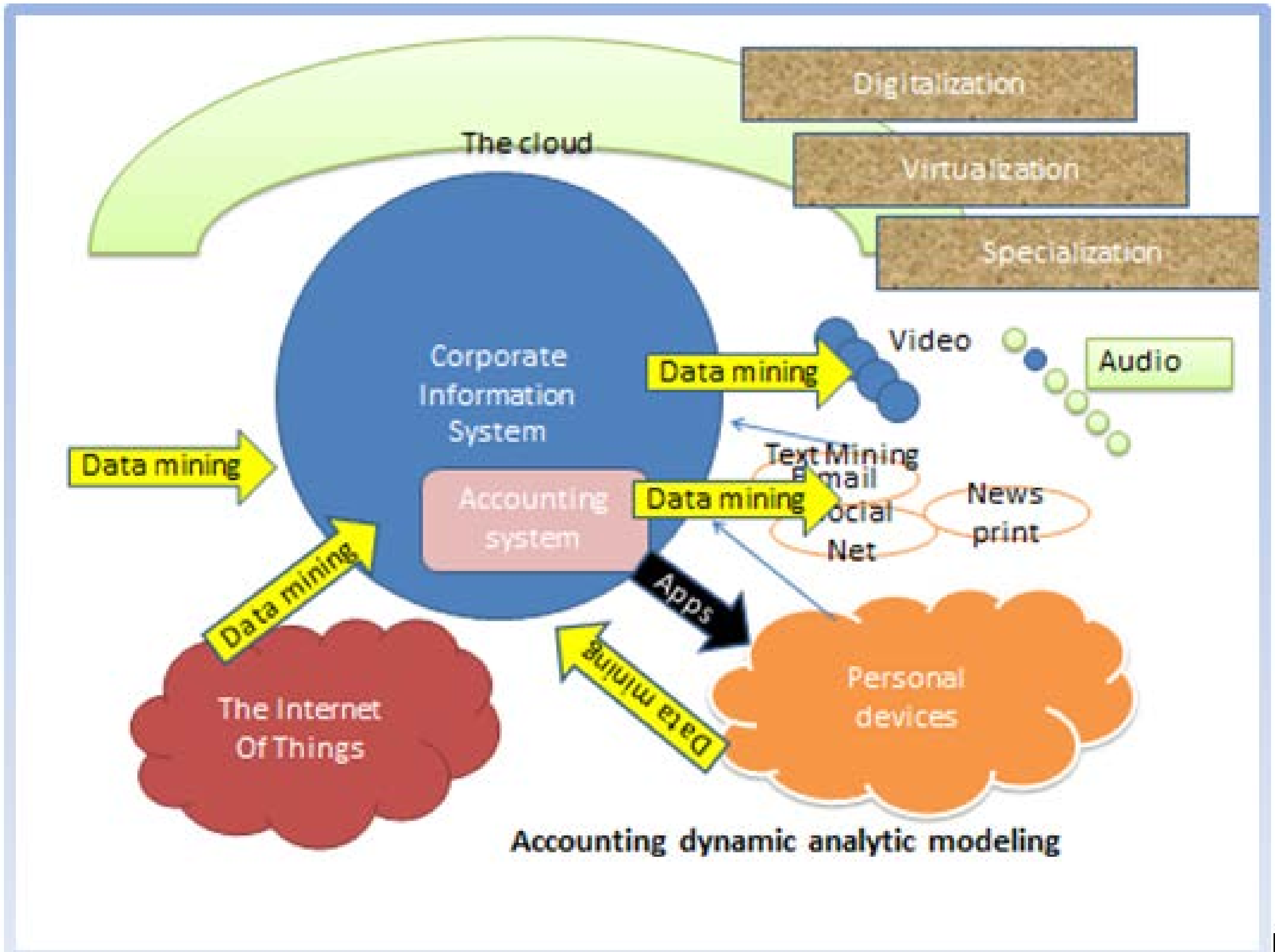
Literature Review of Analytical Procedures in the External Audit : Techniques

"Traditional" Audit Analytics:	CAATS	Transaction Tests, Data Modeling, Data Analytics
	"Traditional" Analytical Review/Analytics	Sampling, Ratio Analysis, Firm developed proprietary software
	Statistics: Log Regression	Log Regression, Step-Wise Logistic, Ordinal Regression Model
	Statistics: Linear Regression	Linear Regression
	Statistics: Time Series Regressions	Time Series Regression, Auto Regressive Integrated Moving Average (ARIMA), Box Jenkins (ARIMA), Random Walk (ARIMA), Random Walk Drift (ARIMA), Seasonal Time Series X-11, Martingale, Sub-Martingale, Single and Double Exponential Smoothing Model
	Statistics: Generalized Models	Multicriteria Decision Aid, Multivariate Distribution, Benford's Law, Descriptive Statistics, Univariate and Multivariate Regression Analysis, Structural Model, Analytical Hierarchy Process (AHP), Spearman Rank Correlation Measurements, Complementary Hypothesis Evaluation, Independent Hypothesis Evaluation, Monte Carlo Study/Simulation

Prescriptive Audit Analytics: Looking Forward

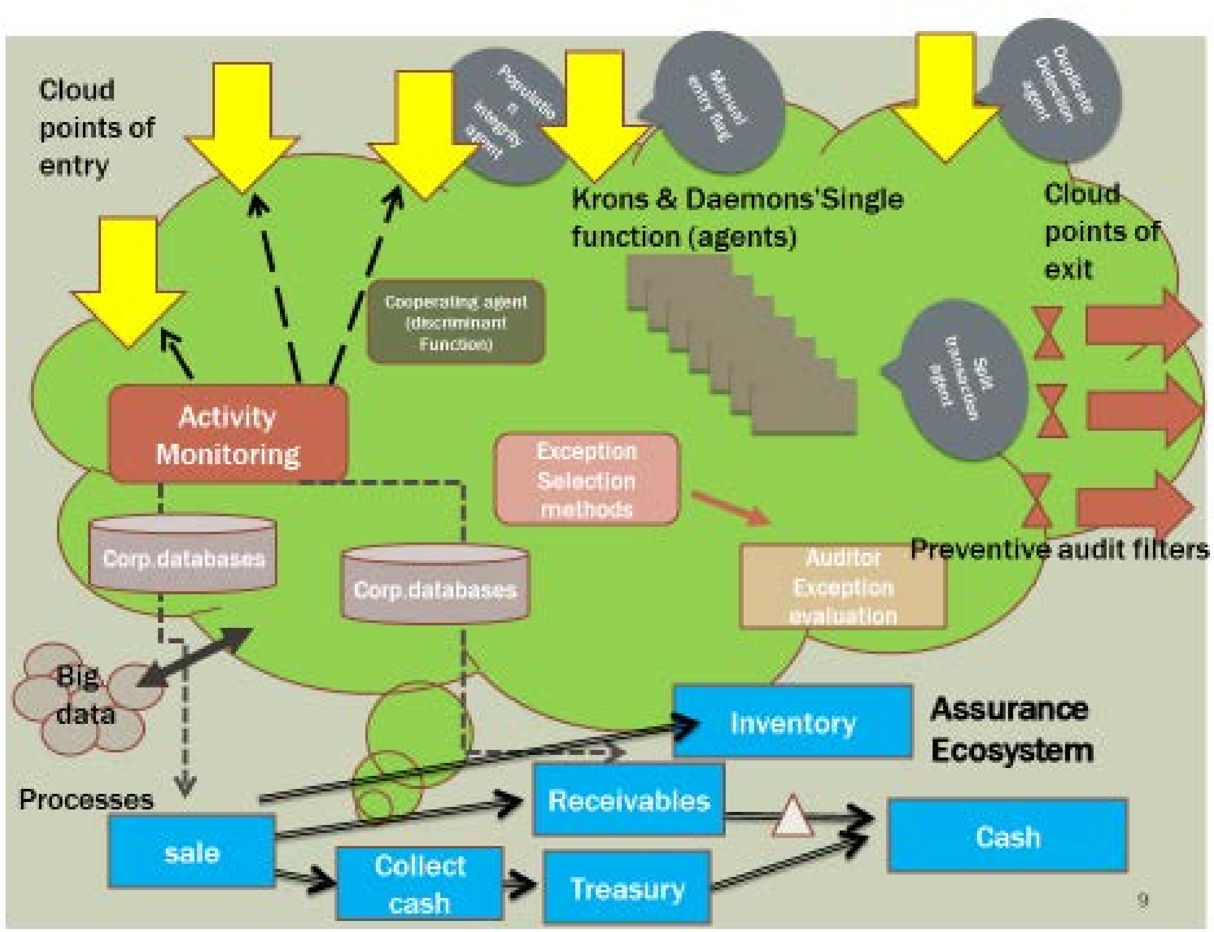
- “It has also been shown that many internal audit procedures can be automated, thus saving costs, allowing for more frequent audits and freeing up the audit staff for tasks that require human judgment (Vasarhelyi, 1983, Vasarhelyi, 1985; Alles, Kogan, and Vasarhelyi, 2002).” (AICPA, 2015)
- Audit Methods have been retroactive as most manual methods relied on some degree of manual verification of source documents or third party verification of balances thru manual confirmation
- Overall, the expected value of assurance efforts must be larger than its costs. Once manual efforts are voluminous they become very expensive.

Evolving the Environment



The evolutionary environment (adapted Liu and Vasarhelyi, 2014)

An Audit Eco-System



Audit Ecosystem (adapted from AICPA, 2015 chapter 1)

ADDITIONAL ANALYTICS

Exceptional Exceptions

Unibanco p rules

FRAUD (I.E., CC EXAMPLE)

Exceptional Exceptions
Loading factors

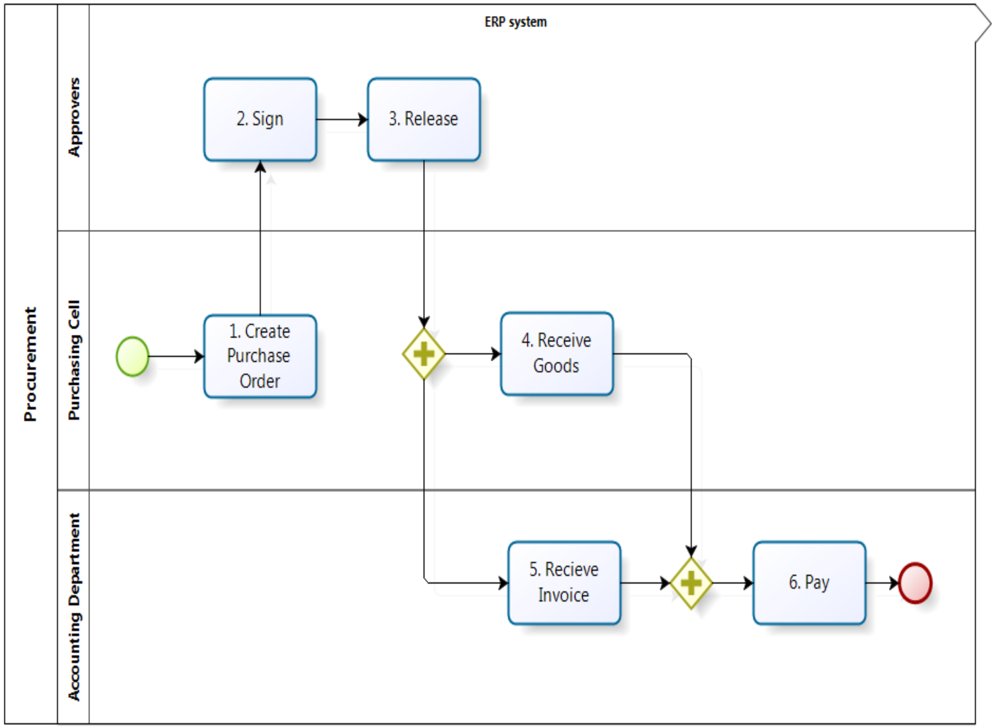
- Populations are too large for traditional sampling
- Loading factor = a. X1+ b.X2 +cX3+ dX4
- from YB Kim's work a, b, c, and d are 1
- From Hussein Issa's work on exceptional exceptions different parameters can be given different values (behavioral or other approaches)

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

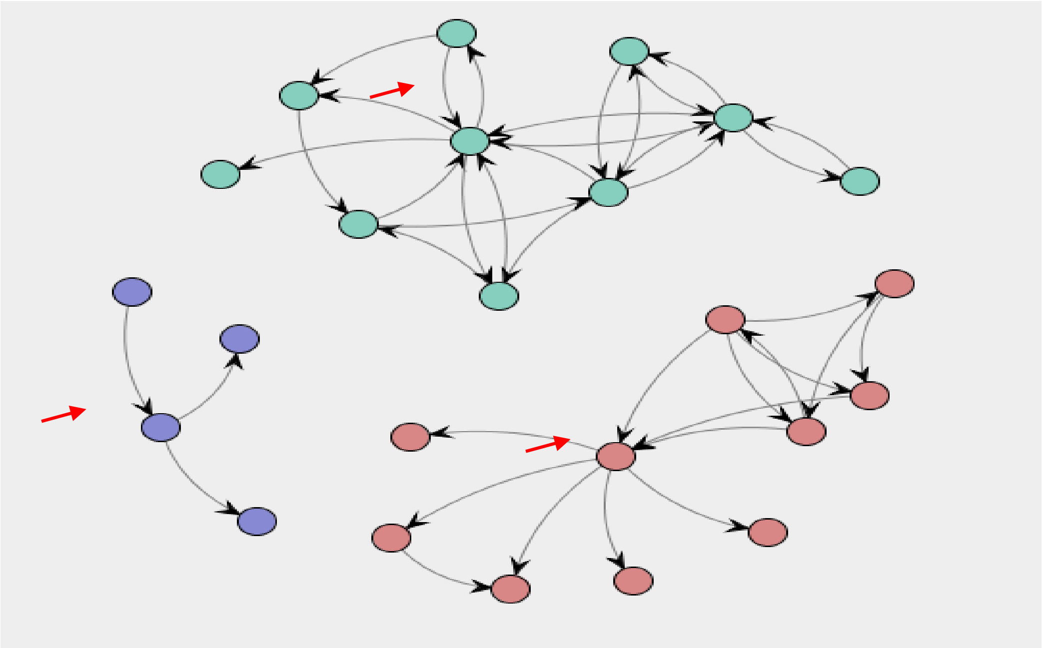
Process mining

Mieke Jens (Hasselt University)
Michael Alles (Rutgers Univ.)

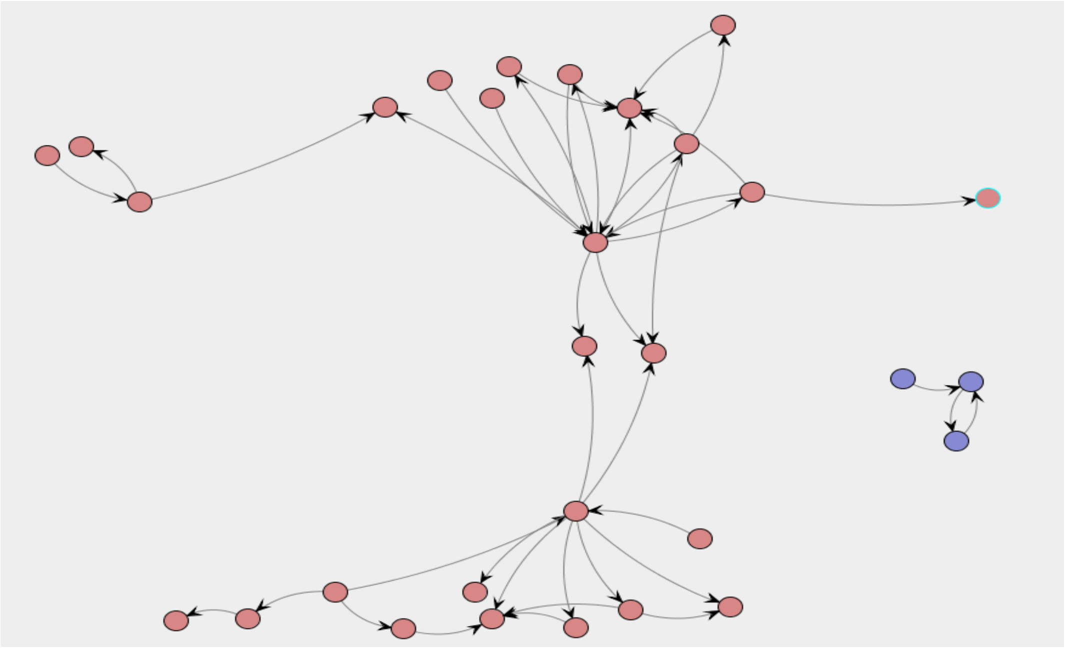
Designed ("Ideal") Process Model



Social Network of 175 cases by three individuals violating SOD



Social Network of the 742 Cases Without *Sign* and in Violation of SOD Controls



CONCLUSIONS

Some research questions

- The appropriateness of the method for a particular forensic function?
- How should the forensic function be reorganized to better use ADA?
- How can predictive technologies be used to set comparison models against which match actuals. How to set allowable variance (Vasarhelyi & Bumgartner, 2015)?
- How to set the timing of performing an assurance / forensic function?
- What types of “suspicion functions” should be set up for a preventive audit or just for transaction or account review?
- How can validation function be developed that link corporate information with big data variables to validate the dimensionality and predict variances?
- How can we migrate to a complex ecosystem and more advance assurance processes while not disrupting the current processes?