Summer School in Public Auditing and Accountability

Data, analytics and auditing: concepts and definitions

NICOLA CASTELLANO – FEDERICA DE SANTIS
University of Pisa

Initiative organized in co-operation with:

EUROPEAN COURT OF AUDITORS
Guardians of the EU finances

ACCA Think Ahead

Agenda

1. Big Data & Analytics: definitions
2. Big Data & Analytics and the audit process: what, when, why?
3. Pros and Cons
4. Big Data applications in auditing. Evidences from academics and practitioners:
   - Audit planning
   - Financial and compliance audit
   - Performance audit

Big Data. A definition

Douglas Laney (Gartner group) 2001. "3D Data Management: Controlling Data Volume, Velocity, and Variety" The hallmark of attempting to characterize and visualize the changes that are likely to emerge in the future. After a decade Laney updated its work (2012) "The importance of 'Big Data': A Definition". It has become Gartner's updated definition:

*Big Data is high **V**olume, high **V**elocity, and/or high **V**ariety information assets that require new forms of processing to enable enhanced decision-making, insight discovery and process optimization.*

Big Data characteristics are often summarized by **V**s:
*huge **Volume**, high **Velocity**, huge **Variety**, and uncertain **Veracity*** (Laney 2001; IBM 2012)

"**Big**" is relative according to the investigated domain. *Big Data implies that the amount of data treated is beyond the limit of what the relevant information systems can store and/or process*

Big Data and Analytics

Big Data *and* ***Analytics*** *have been used to describe the* ***data sets*** *and* ***analytical techniques*** *in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies.*

In this article we use business intelligence and analytics (BI&A) as a unified term and treat big data analytics as a related field that offers new directions for BI&A research.

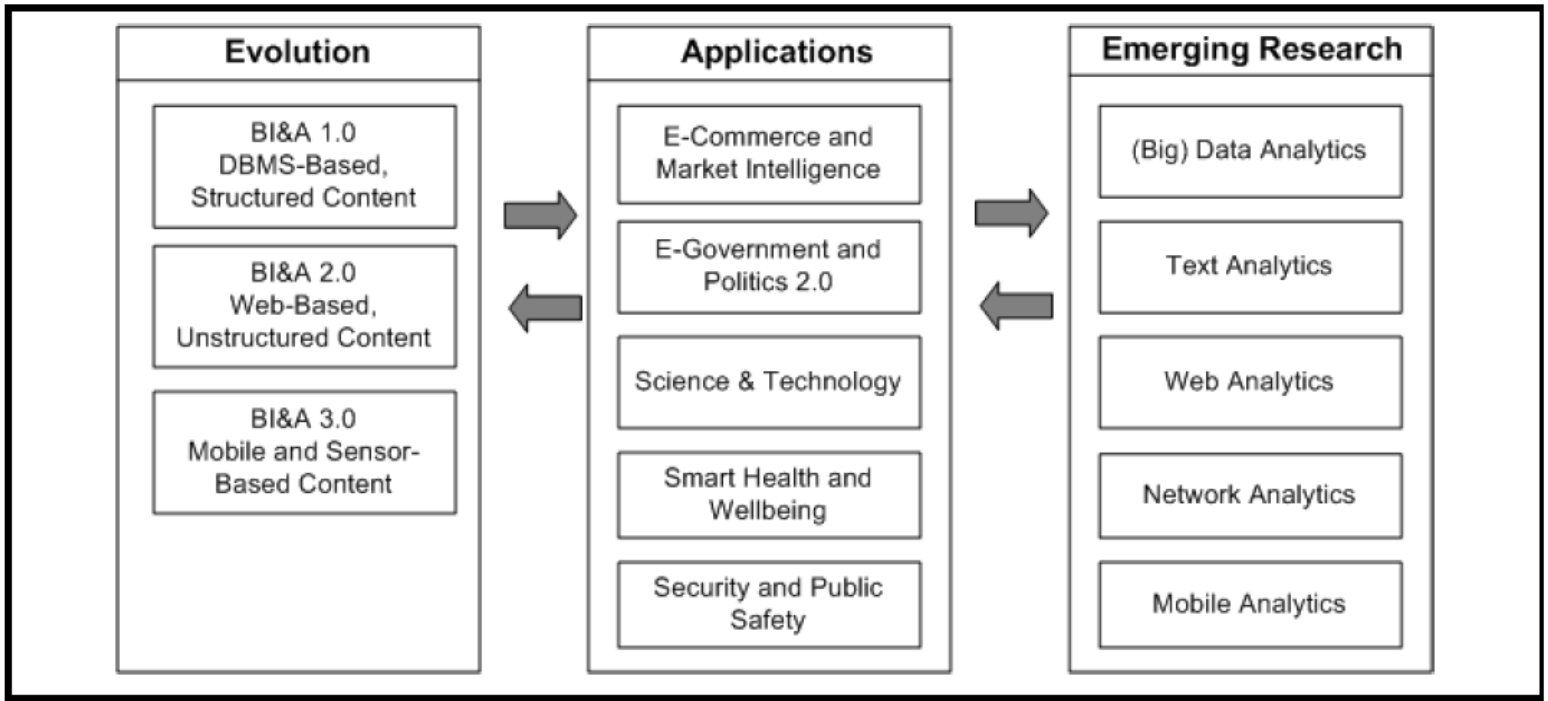


Figure 1. BI&A Overview: Evolution, Applications, and Emerging Research

(source Chen et al. 2012)

BI&A 1.0

data are mostly structured, collected by companies through various legacy systems, and often stored in commercial relational database management systems (RDBMS). The analytical techniques commonly used in these systems, popularized in the 1990s, are grounded mainly in statistical methods developed in the 1970s and data mining techniques developed in the 1980s.

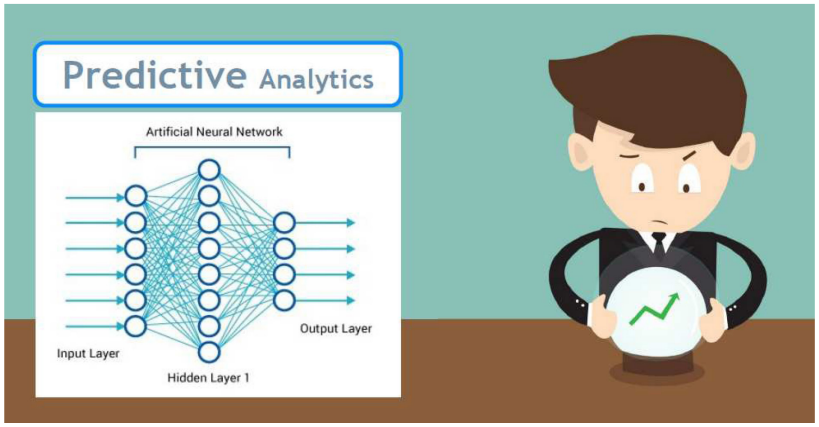
BI&A 2.0

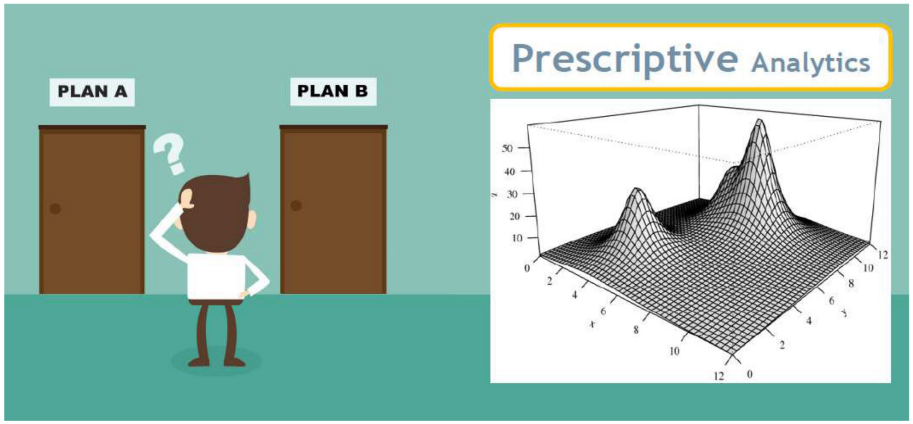
Centered on text and web analytics for unstructured web contents. The new gold mine for understanding customers' needs and identifying new business opportunities. Web 2.0 applications can efficiently gather a large volume of timely feedback and opinions from a diverse customer population for different types of businesses.

BI&A 3.0

«Mobile-BI». Mobile and sensor-based content; Location-aware analysis; Person-centered analysis; Context-relevant analysis; Mobile visualization & HCI (Human-Computer Interaction).

A taxonomy for analytics

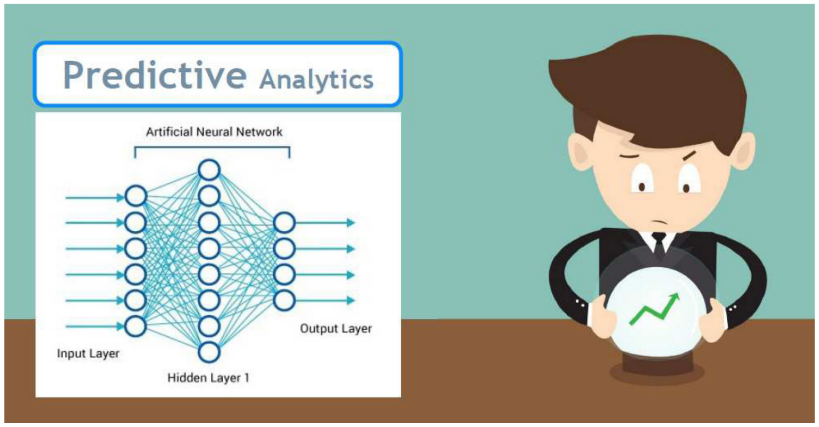




A taxonomy for analytics

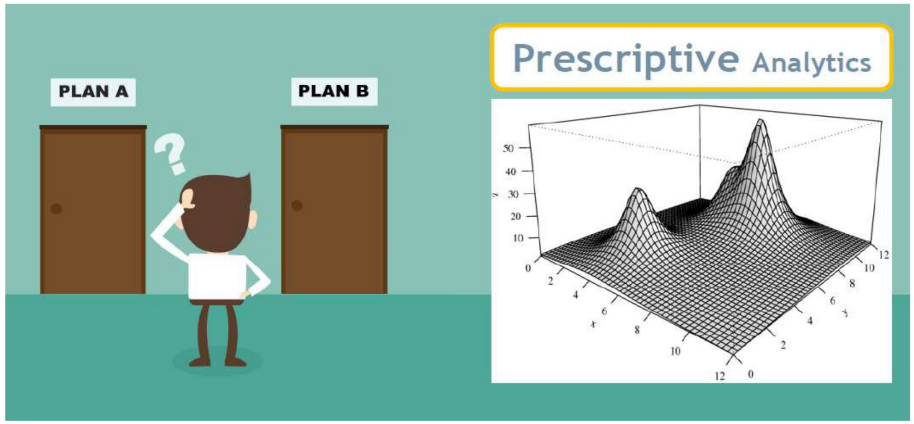
Descriptive (and diagnostic) analytics – What is happening? Why it is happening?
Traditional business intelligence (BI) and visualizations (pie-charts, bar-charts, line-graphs, tables, or generated narratives).

A taxonomy for analytics

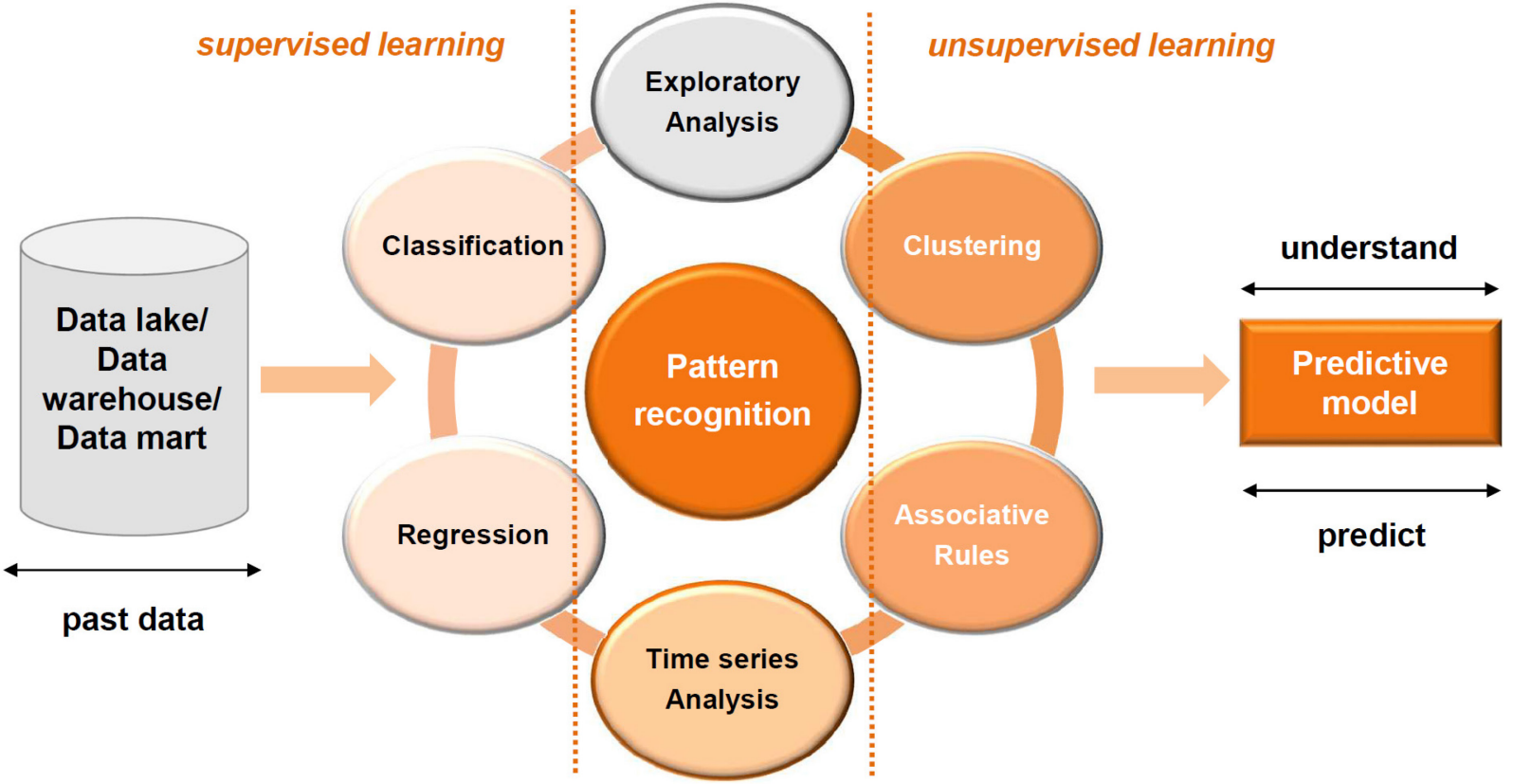


Predictive analytics – "What is going to happen?" (What is likely to happen?) Regression analysis, forecasting, multivariate statistics, pattern matching, predictive modeling, and forecasting (among others).

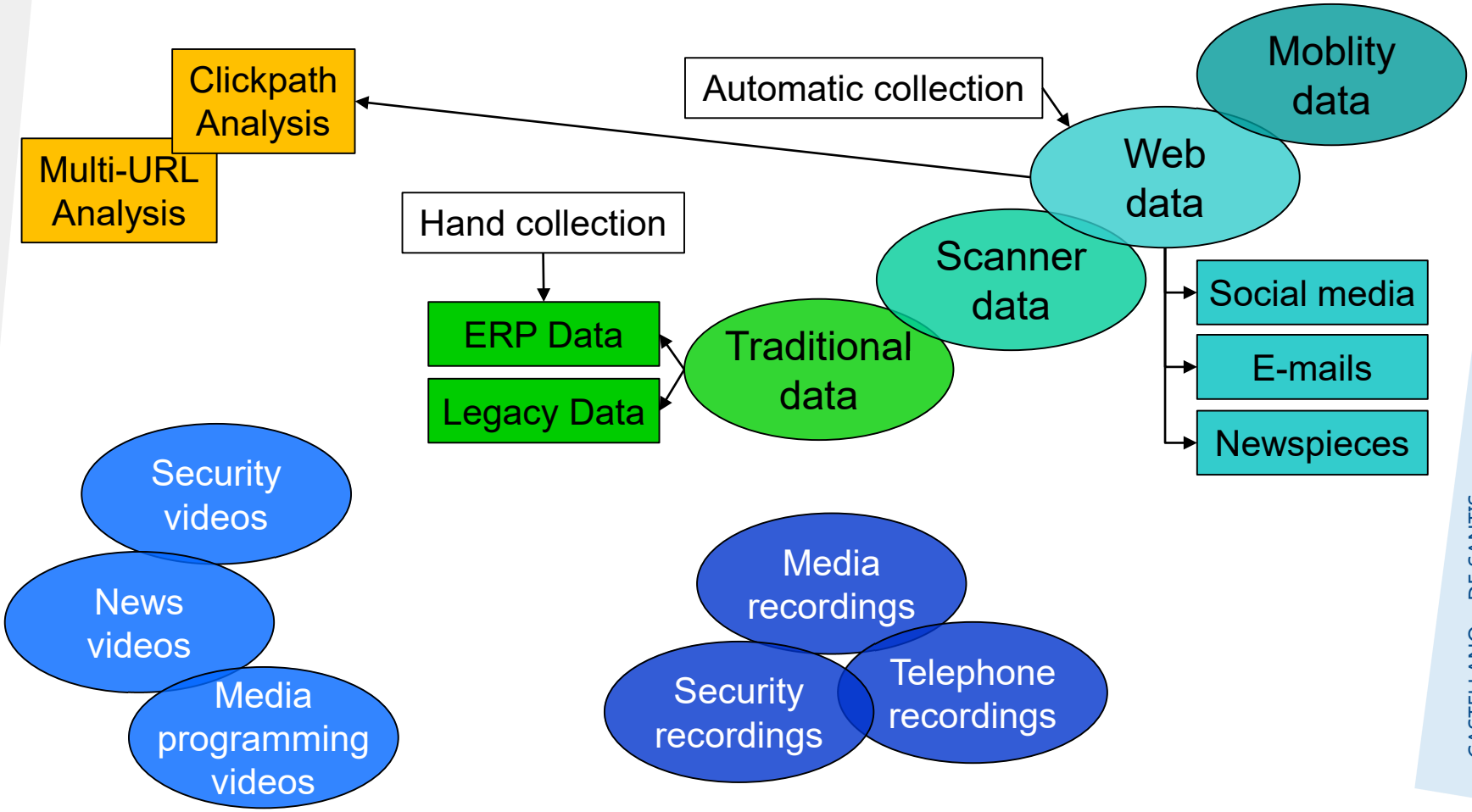
Prescriptive analytics – "What should be done?" (or What can we do to make something happen?)
Graph analysis, simulation, complex event processing, neural networks, recommendation engines, heuristics, and machine learning (among others).



Advanced analytics



Big Data & Analytics in accounting



The expanding eco-system of corporate data (source Vasarhelyi et al. 2015)

CASTELLANO – DE SANTIS

Potential impacts (Littley, 2012):

- Better forecasts of estimates, going concern, fraud, and other variables that are of concern to internal and external auditors
- Big Data as a way of increasing the effectiveness and credibility of audit
- Big Data as a way to reduce the costs/effectiveness of audits

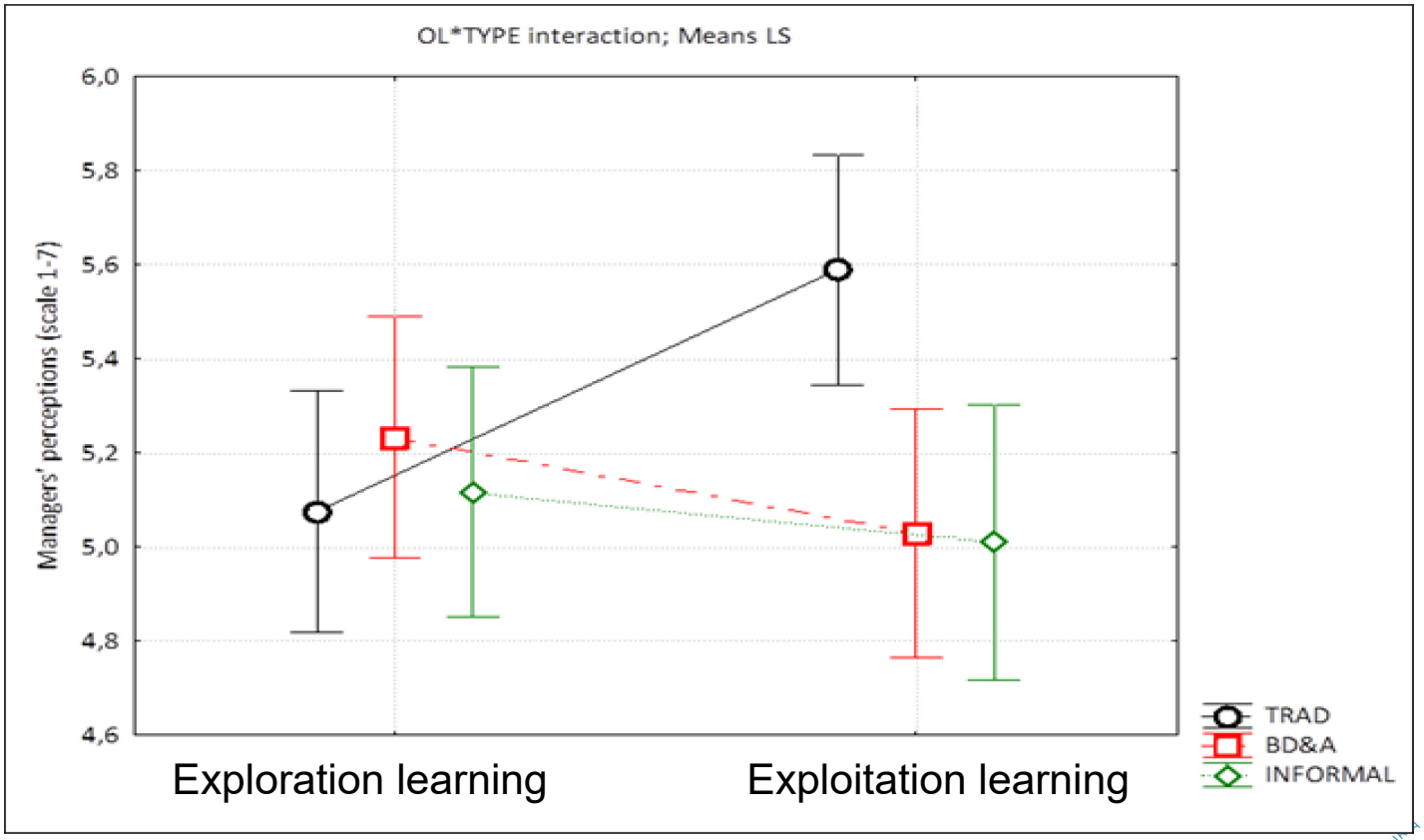
Behavioral and practical implications (Brown-Liburd et al. 2015). Decision-making based on information derived from Big Data involve interpretation and judgment, which can be influenced by the following:

- Information overload (overwhelming amount of information hinder appropriate conclusions)
- Information relevance (dilution effect - auditors must choose from a vast array of available information which items are most relevant for their audit judgments)
- Pattern recognition (providing financial auditors with contextual experience and training will improve their ability to accurately recognize patterns in data, and to correctly interpret them
- Ambiguity (variations in the amount and type of information available; source reliability and lack of causal knowledge of observed events). Ambiguity-intolerant auditors will be uncomfortable with Big Data and may avoid or downplay ambiguous information that could result in less than optimal judgments.

Managers' (Auditors') perceptions about Big Data

Perceived usefulness of **Big Data** is one of the most powerful predictors for intention to use them in management and auditing.

The managers' inclination to rely on Big Data is influenced by the kind of decision-making process (exploration vs exploitation learning) as well as by managers'individual characteristics (Castellano, Presti, Del Gobbo, 2017).



High experienced managers consider Big Data **less relevant** compared to Traditional (structured) information

Managers high confident in taking decisions supported by numbers tend to feel more "confident" in relying on "their" information (skeptical in using new forms of information).

Challenges of Big Data in Auditing

- Focus of data analysis toward recognizing patterns within large amounts of data
- Consequent to continuous auditing systems the numbers of identified exceptions and anomalies are expected to increase dramatically. Prioritization methodologies which incorporate them in decision-support systems can greatly help alleviate the burden of processing information.
- Lack of the adequate training and required **skills** to analyze Big Data.

Management accountant vs Data analyst skills: differences*

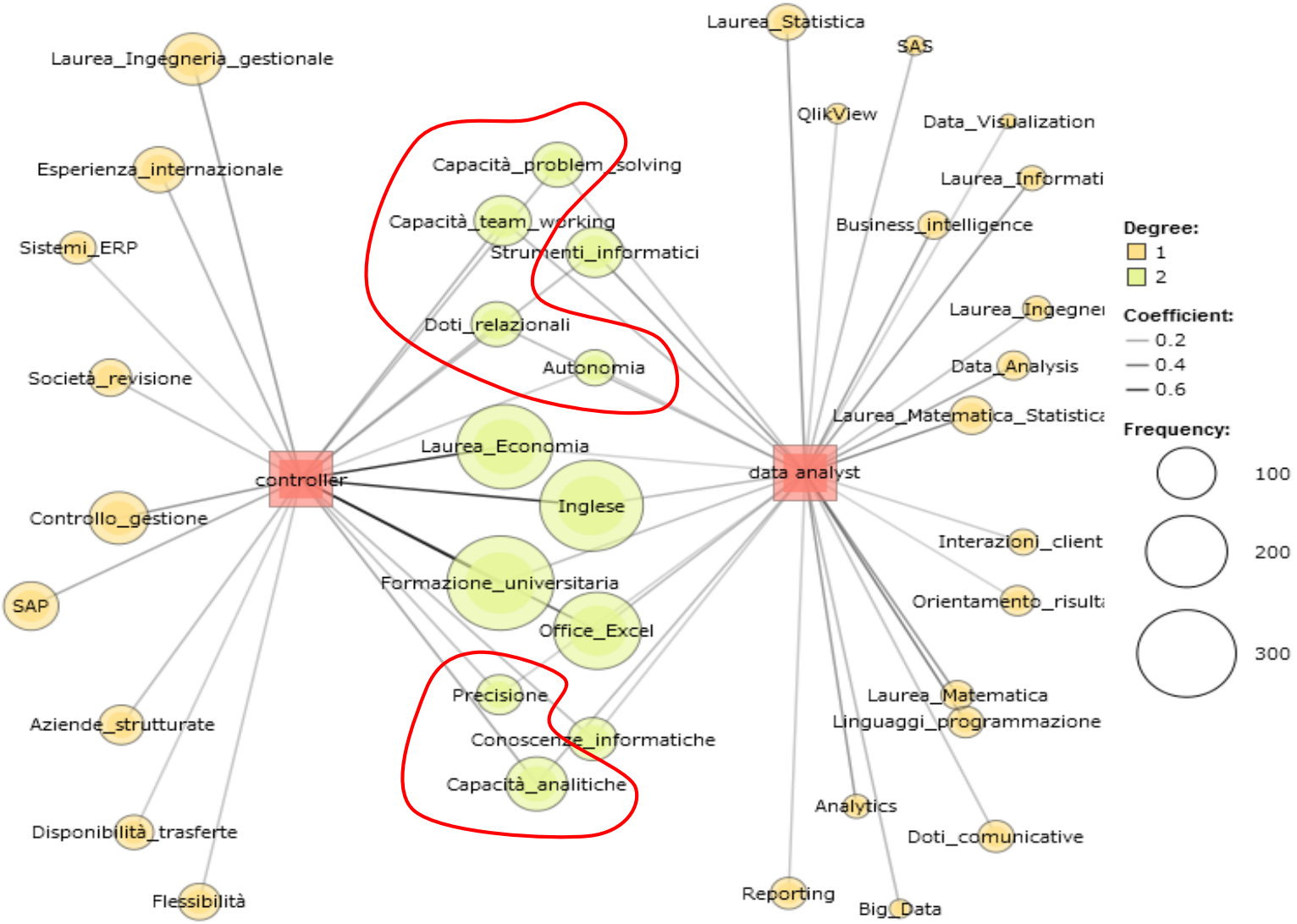
Management accountant**
Degree in Economics
Degree in Management Engineering
Experiences in Structured companies
Experiences in Auditing companies
Competences in management control
Competences in ERP

Data Analyst**
Degree in engineering
Degree in mathematics/statistics
Degree in informatics
Team working ability
Relational & comunicational skills
Interaction with customers
Informatics tools
Business Intelligence
Big Data and Analytics
Programming languages
*QlikView
*SAS
*Data_Visualization
*Data_Analysis

* Survey on job calls posted on Linkedin for positions of management accountants (nr. 392) and data analysts (97) in Italy at April 2018. – Data processed through text mining KH Coder software

** differences emerging by chi-squared test.

Management accountant vs Data analyst skills: similarities*

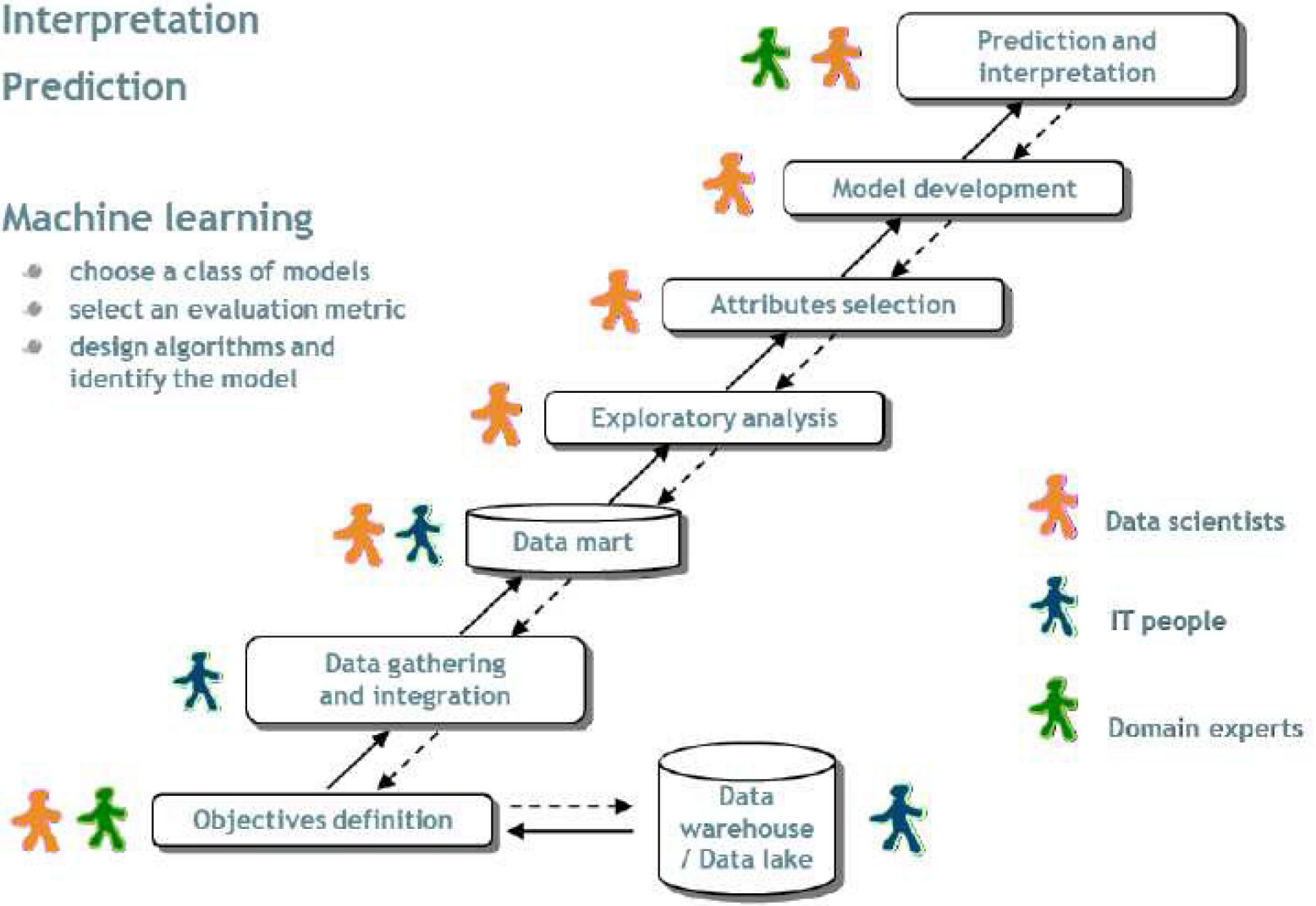


* co-occurrences network adopting Jaccard Index

Analytics development process

- Interpretation
- Prediction

- Machine learning
  - choose a class of models
  - select an evaluation metric
  - design algorithms and identify the model



CASTELLANO – DE SANTIS

- The pace of afoption of BD&A in statutory audit has been lower than in other fields (e.g. internal audit, marketing, strategic decision-making), mainly due to liability concerns and the highly regulated environment
- Using BD&A in auditing is about enhancing audit quality
- BD&A is being approached in the auditing practice with the aim of improving the efficiency and effectiveness of audits
- BD&A has the potential to represent the most significant shift in how audits are performed since the adoption of paperless audit tools and technologies

Some of the distinctive aspects of big data analytics are:

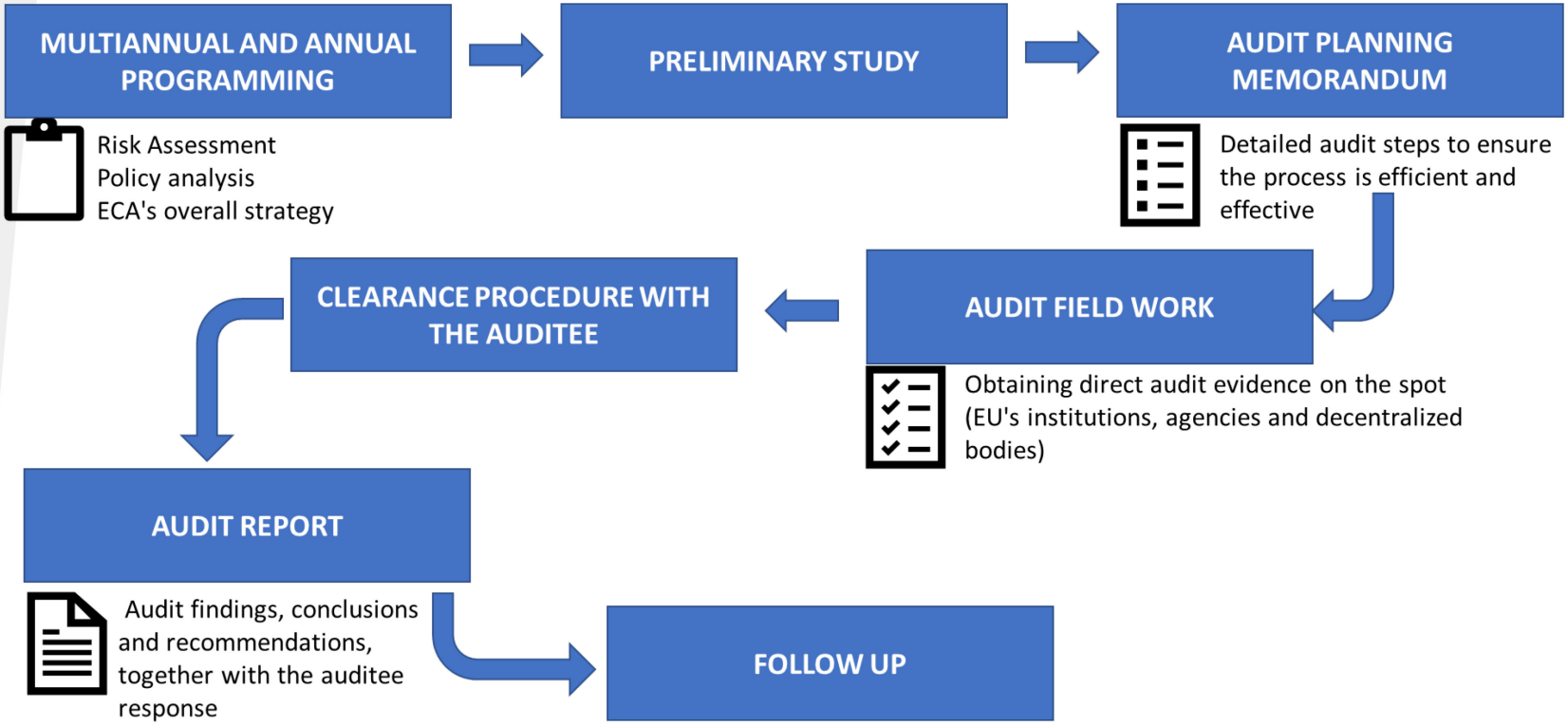
- The use of algorithms
- The opacity of data processing
- The tendency to collect "all data"
- The repurposing of data
- The use of new types of data

Different audit approaches

		APPROACH FOR ANALYSIS	
		Problem-driven	Explicative
DATA SOURCES	Traditional (financial data)	Problem-driven analysis on structured data	Data-driven analysis on structured data
	"Extended" Big Data	Problem-driven analysis on structured and unstructured data	Data-driven analysis on structured and unstructured data

Source: Richins et al. (2017)

The audit process



Financial and compliance audit: an overview

Performance audit: an overview

Big Data in the auditing field: what are the benefits?

1. **Auditors can test a (far) greater number of transactions, overcoming sample limits**

DA can increase the sufficiency of audit evidence, as it allows auditors to automate testing for transactions and, at least in theory, 100% of transactions can be testes.

2. **Auditors' focus shifts from errors in the sample to anomalies in patterns of data about the population**

Anomalies are instances where the data does not mach auditor's expectations based on his/her knowledge of the client's business. For example, the presence of sales to clients that has been identified as non-creditworthy or bankrupt, for which one could expect no sales to be recorded, constitutes an anomaly in the population of transaction.

3. **Audit quality can be increased by providing grater insights on auditee's processes**

Auditors can build a database of knowledge for each client that can be transferred from year to year (e.g. information about how the identified anomalies has been resolved would inform auditors in the following year).

4. Frauds will be easier to detect

The increased availability of auditees' data, as well as the development of more and more advanced DA tools can significantly improve auditors' ability in detecting frauds, e.g. by using tools such as Benford's Law, a mathematical principle that have proven effective in identifying fraudolent transactions.

5. Auditors can better plan the audit engagements

DA allows using NFD and external data to better inform audit planning (particularly in assessing risks), and to more effectively audit those areas that requires judgment (e.g. valuation and going concern). With specific reference to predictive ones, BD&DA allows auditors to develop predictive models that can predict future events

Big Data in the auditing field: what are the benefits?

Data analytics use examples

Compliance	• Evaluate expense reports and purchase card usage for all transactions. • Perform supplier audits by utilising line-item billing data to identify anomalies and trends to investigate. • Assess regulatory requirements. • Identify poor data quality and integrity around various data systems that are key drivers to non-compliance risks.
Fraud, risk assessment, detection and investigation	• Identify areas at high risk of fraud and assess controls. • Identify ghost employees, potential false suppliers, and related parties or employee-supplier relationships. • Highlight data anomalies that pose the greatest financial and/or reputational risk to the organisation. • Investigate the symptoms of an asset misappropriation scheme to answer the "who, what, when, where" questions.
Operational performance	• Isolate key metrics around spend analysis e.g. payment timing, forgone early-payment discounts and payment efficiency. • Perform duplicate payment analysis and recovery. • Perform revenue-assurance analysis. • Perform slow-moving inventory analysis. • Identify key performance and key risk indicators across industries and business lines.
Internal controls	• Anticipatory e.g. business continuity plan. • Detective and corrective e.g. control account reconciliations. • Directive e.g. code of conduct. • Preventative e.g. passwords, access controls. • Perform segregation of duties analysis. • Perform user access analysis. • Assess control performance. • Exception reporting e.g. identify potential outliers that would indicate control failures or weaknesses.

1. Training and expertise of auditors

The proliferation of vast amounts of (mostly non-financial) data could overwhelm the information processing capabilities of auditors, which thus require new and advanced skills in DA (e.g. pattern recognition and understanding how to treat anomalies). The DA environment will result in auditor judgment playing a much more significant role than in sample-based auditing due to the potential of anomalies evaluation, which requires a greater understanding what constitutes proper accounting in the context of the auditee's business.

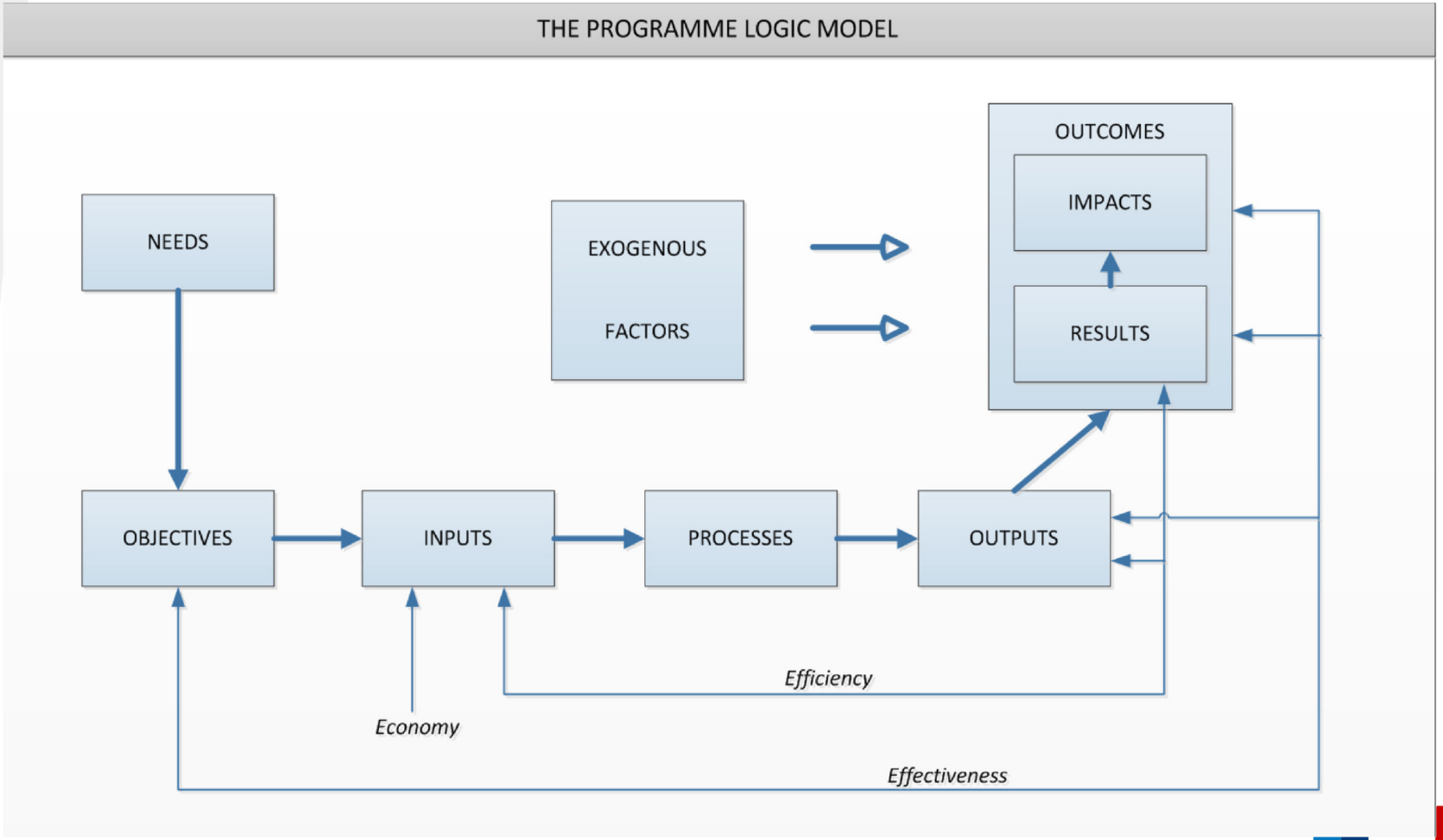
2. Data availability, relevance, and integrity

Many auditees can lack the ability to collect data in a way that is useful to the auditor, or the data might contain a lot of noise. Even when data is collected by the client, it might be unclear what level of access the auditor will be granted and what data the client will share with the auditor. Even if data can be readily provided and auditors are granted full access, they will have to consider data integrity (source's security, tampering of data). It raises concerns whether auditors can accept some inconsistencies as it scarcely reconcile with the focus in auditing to data integrity.

3. Expectations of the regulators and financial statement users

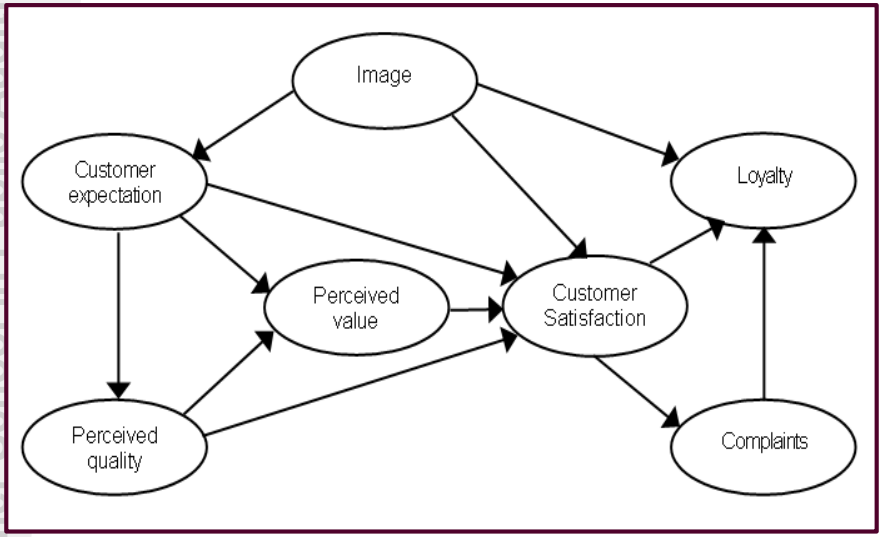
Given the opportunity to test 100% of transactions, the use of DA in auditing can exacerbate the expectation gap. This opportunity could also deprive auditors of a defense against legal liabilities (e.g. fraud detection)

Performance Auditing *is designed to determine whether the Commission and other audited entities have designed and implemented management and monitoring systems so as to optimize* ***economy***, ***efficiency*** *and* ***effectiveness*** *within the given constraints.*

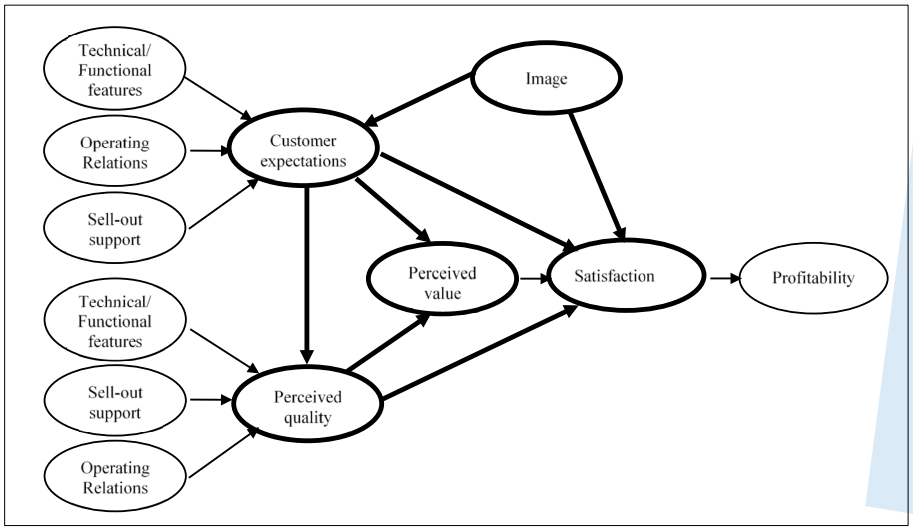


When managers discuss about the design of a causal model representing the needs of a specific communities as well as the initiatives expected to produce an impact, an evidence-based approach might provide a fact-based support useful to let the managers revise their individual experiences and beliefs (Castellano, Del Gobbo, 2018. Strategic mapping. Relationships that count. Management Decision, Vol. 56 Issue: 4, pp.908-921)

Two competing causal models…



Causal model 1



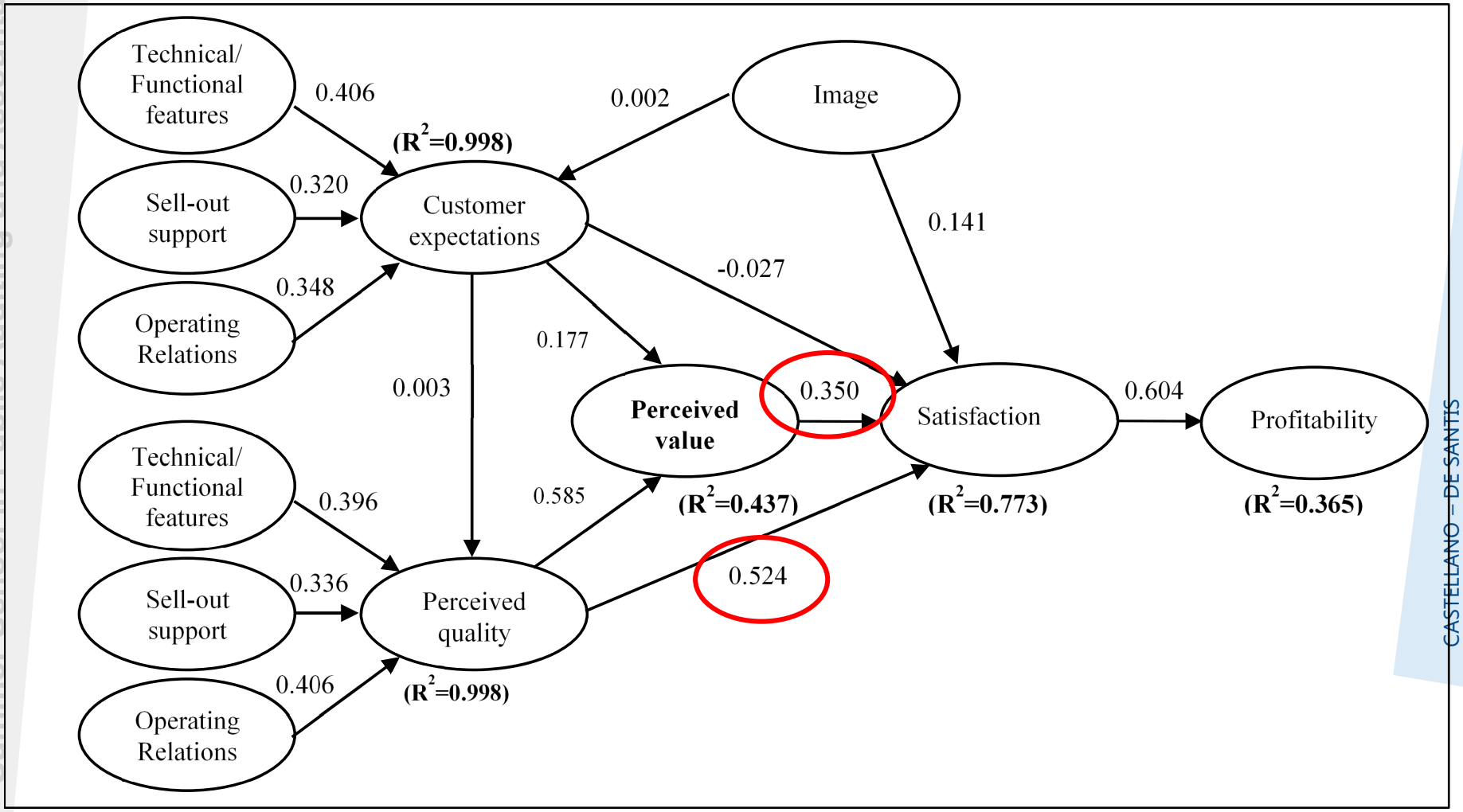
Causal model 2

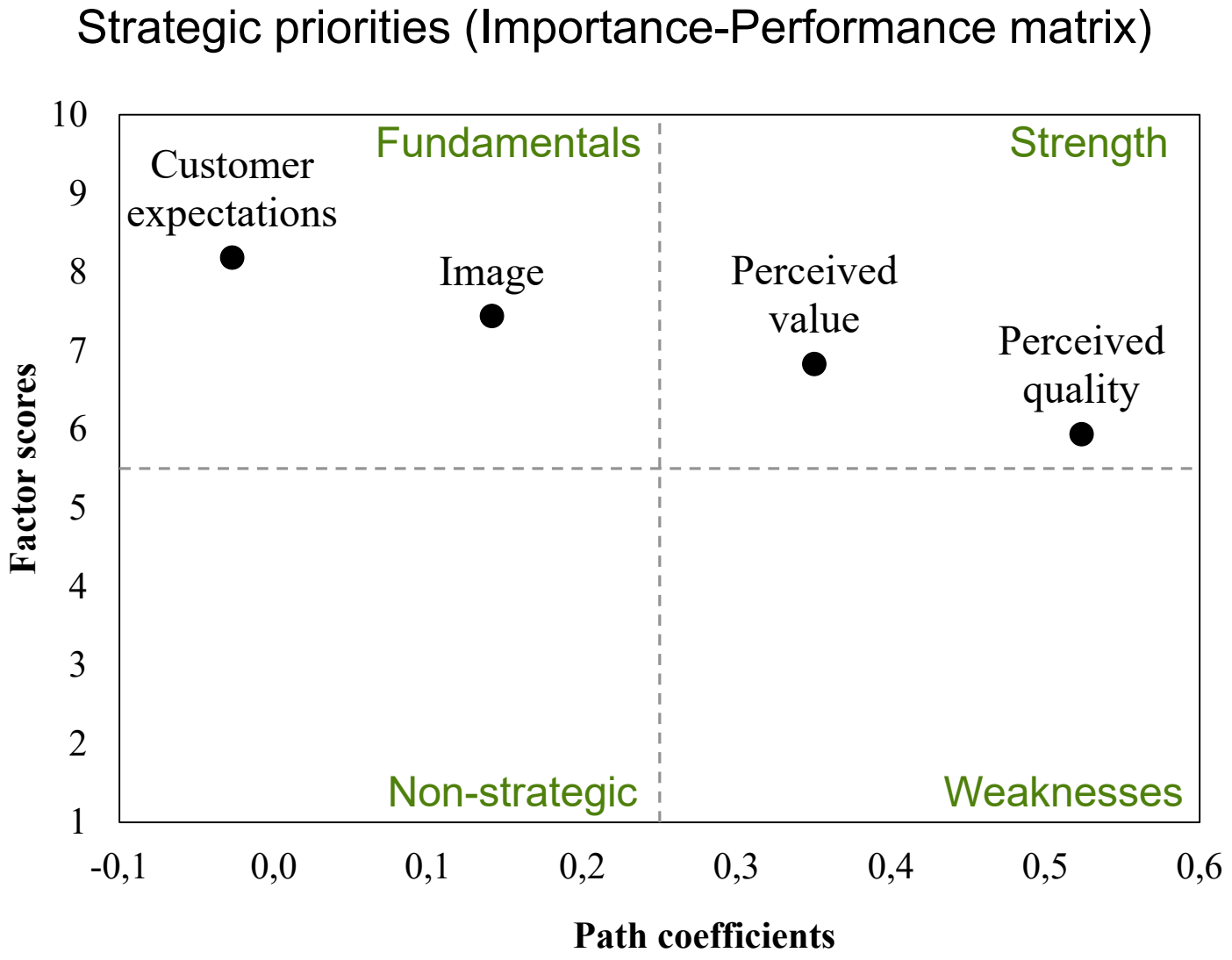
…PLS-PM is used: 1) to provide evidence about the most accurate causal model; 2) to identify the «strategic priorities»

	GoF	GoF (bootstrap)	Standard error	Critical Ratio	Upper bound (95%)	Lower bound (95%)
Causal model 1	0.483	0.489	0.036	13.543	0.409	0.561
Causal model 2	0.681	0.680	0.033	20.525	0.602	0.750

…PLS-PM – Goodness of Fit, provide a fact-based evidence that causal model 2 provides a more accurate description of the perceptions within the investigated community

Estimated magnitude of impacts





Thank you for your attention!

nicola.castellano@unipi.it

federica.desantis@unipi.it