Data-Driven Oversight, Insight and Foresight

Frameworks and considerations for leveraging data analytics for audit
23 July 2018 | University of Pisa

Gavin Ugale
Portfolio Manager
Public Sector integrity Division
OECD

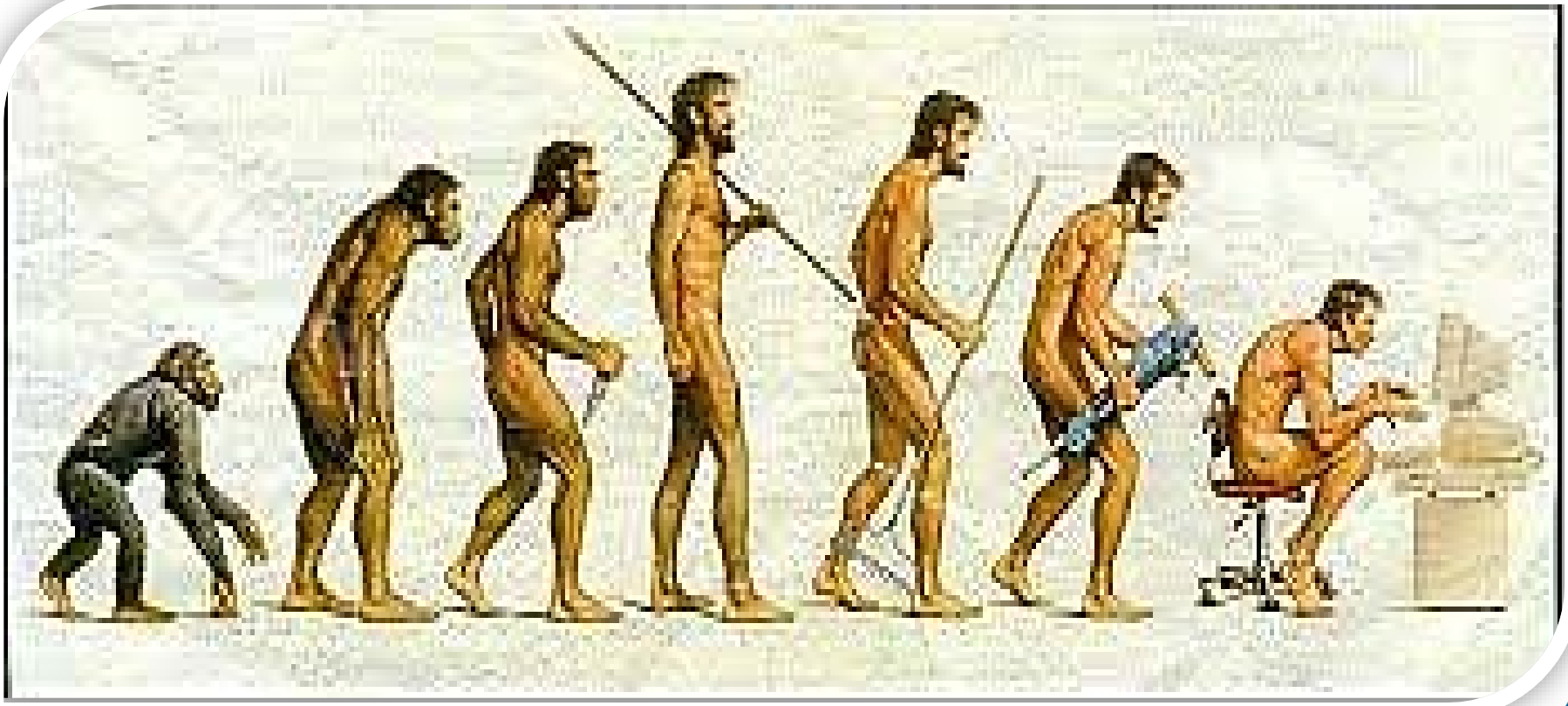
BETTER POLICIES FOR BETTER LIVES

Overview of our discussion

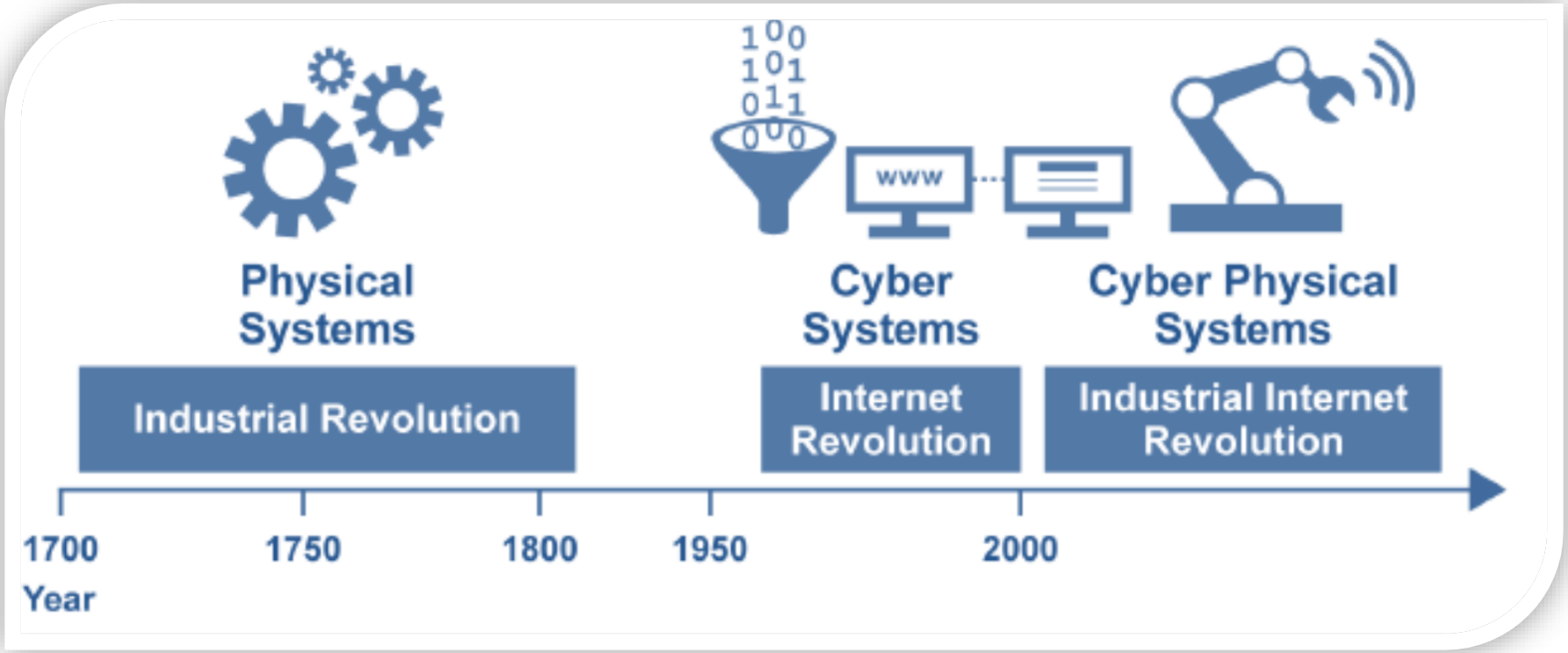
1. 10,000 Meters – The Evolving Data Landscape
2. 5, 000 Meters – Institutional Perspectives
3. 100 Meters - Data Analytics for Integrity

1. 10,000 METERS – THE EVOLVING DATA LANDSCAPE

The evolving landscape



The New Industrial Internet Revolution



Source: GAO (2016), Data and Analytics Innovation: Emerging Opportunities and Challenges, GAO-16-659SP

“Most of us will never deal with such systems”

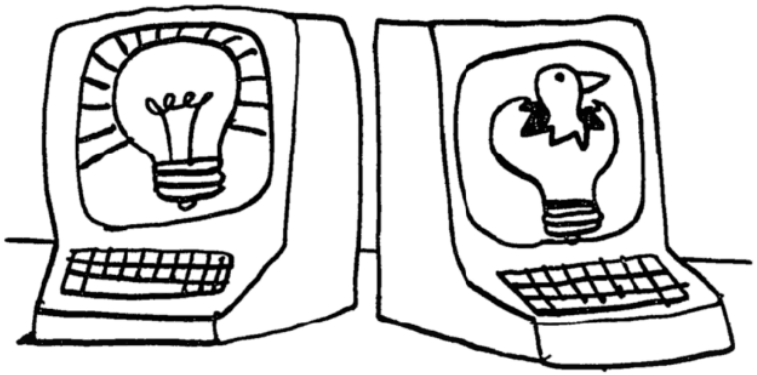
PERSONAL COMPUTERS

Artificial Intelligence: New Software Arrives

By ERIK SANDBERG-DIMENT

THE latest catchword in computing is artificial intelligence, or AI. Most researchers in the field do not promise anything as radical as self-replicating computers in the near future. Some, however, do honestly seem to feel that the reproductive facility is the only human aspect computers will be lacking a couple of decades from now. AI, then, is entering the world of hype usually attendant upon any new technology. The reality is something quite different, of course. For the most part it is yet to be seen.

Several new products employing AI techniques to make software easier to use are expected to be released this fall. Symantek has a word-processing and database package with a “natural-language front end.” That is, it is



Stuart Goldenberg

“Why are you,” I typed in. “Ask a question.” Racter replied.

ual user views its approach. I should note that the publisher

Mind Reader 3, mineral 2, minimum ;” could be made to ensue. Hitting the semicolon key again, instead of the 3 as I had intended, gave me “minimum.” Backspacing erased everything except my original “Min.” But the word window failed to reappear. Finally I erased the whole word and started over with “Min 3.” At last I had succeeded in calling forth “Mind Reader.”

Another interesting problem arose with the word “considerable.” Typing just “c” gave me a choice of five words from “can” to “computer.” Getting as far as “consi” manually, I ended up with three windowed choices, “consistent,” “consideration” and “consider.” I chose “consider,” planning to add “able.” However, after Mind Reader entered “consider,” it automatically and most obligingly skipped a space for

Source: New York Times, 17 September 1985

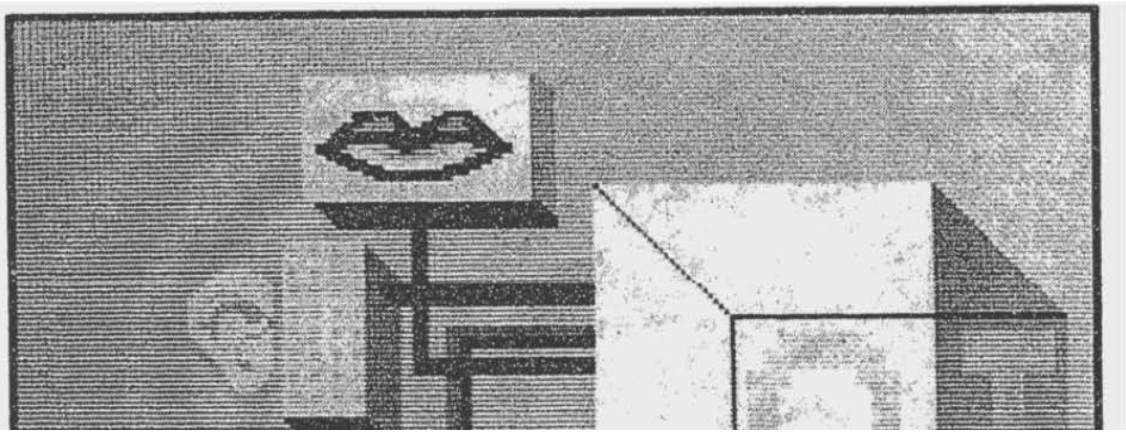
AI is “fast becoming one of the computer industry’s most overused buzzwords…”

It’s the Real Thing: Artificial Intelligence

By Bob Weinstein

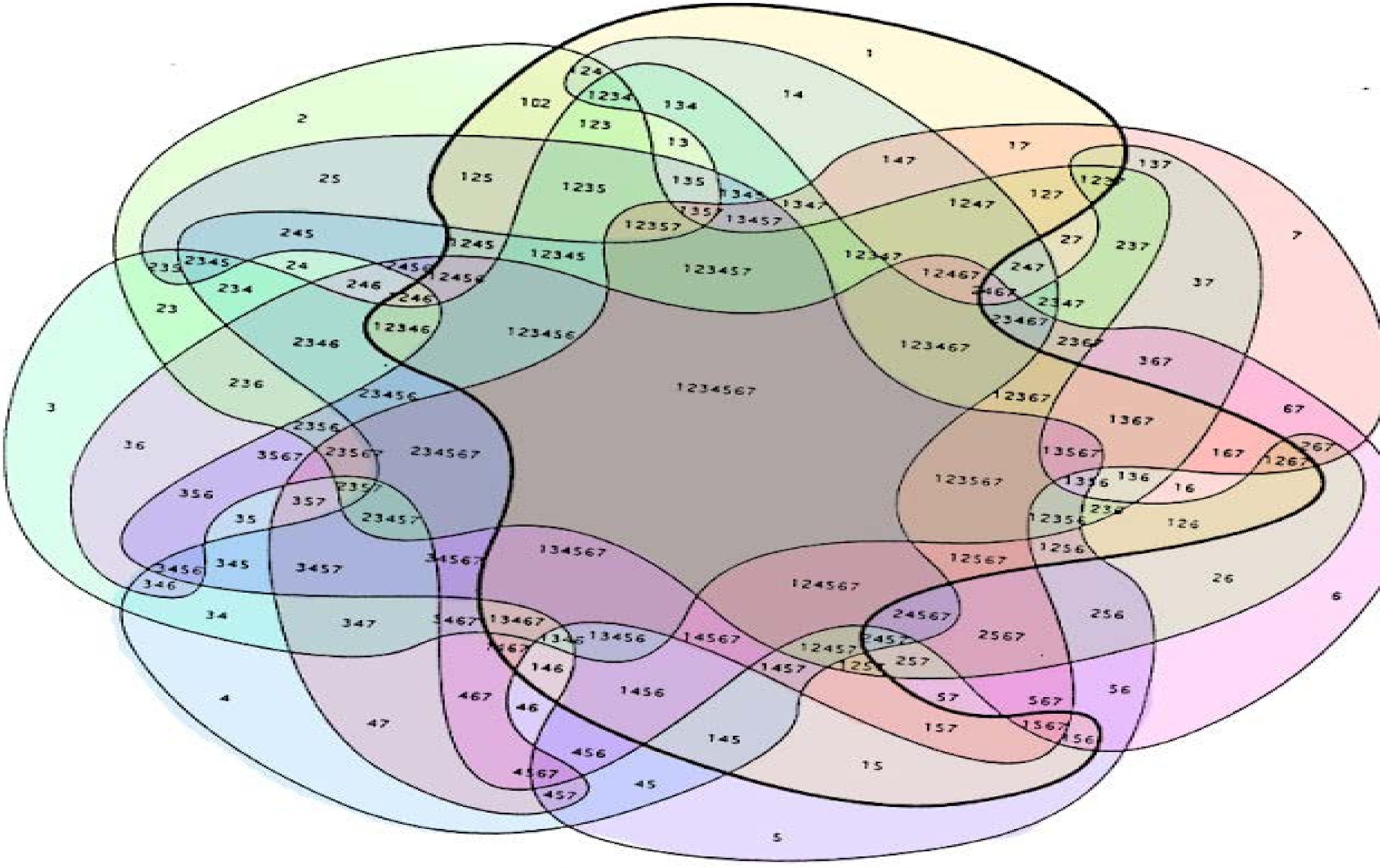
PETER WILL uses a computer to find oil. As director of systems science at Schlumberger Research in Connecticut, a division of the Schlumberger Corporation, he heads a team of experts who use artificial-intelligence systems to find oil throughout the world.

"We drop sensors into a well so the data from the sensors can be transmitted to a centralized site where A.I. software is used to interpret the signals," Mr. Will explains, using the abbreviation for artificial intelligence. "The raw signals measure physical phenomena. Inferences are then made from the physical phenomena in order to give clients answers. As a result of this information, we know where to drill, how much we'll get out of the ground and whether hydrocarbons are present."



Source: New York Times, 24 March 1985

Still tackling the buzzwords and definitions



Shifting Paradigms for Big Data Analysis

Traditional data analysis	New data analysis
Some of the data	All of the data
Clean Data	Messy data
Deterministic relationships	Complex coupling relationships
Interrogation of data to test hypotheses	Discovery of insight
Lag-time analysis of data	Real-time analysis of data

Source: Government of Australia (2015), Australia Better Practice Guide or Big Data, Version 2.0, https://www.finance.gov.au/sites/default/files/APS-Better-Practice-Guide-for-Big-Data.pdf

Many external challenges leading to strategic reorientation and new approaches among SAIs

External challenges…

…leading to internal change.

Source: OECD analysis of 2015 annual reports of select SAIs

SAIs for oversight, insight and foresight

Role	Purpose	Examples of SAI Activities
Oversight	Assurance and compliance	Assessing the compliance, as well as effectiveness and efficiency of, internal audit in supporting a more robust system of internal control and risk management (e.g. Brazil's Tribunal de Contas da União)
Insight	Identifying systemic issues, cross-cutting challenges and trends	Ranking high-risk institutions in public administration to encourage improvements in management before risks materialise and investigating overlapping goals (E.G., High-Risk Lists of U.S. GAO)
Foresight	Forecasting policy implications and predicting risks	Assessing government's preparedness to address population ageing and workforce management, or assessing progress against international commitments to project whether goals would be reached (e.g. UK Forecasting in Government)

Data & analytics for hindsight, insight and foresight

Role	Purpose	Examples
Hindsight	What happened?	Data matching and data mining to assess improper payments for government services Benford's Law to identify anomalies
Insight	What's going on and why?	Applying statistical models to understand extent of existing problem Text mining and text analysis to gain insights from unstructured data Advanced visualtisations and dashboards to inform management decision-making
Foresight	What could happen in the future?	Using historical data to predict high-risk claims in medical programmes, and identify patterns that can inform improvements to existing control activities Applying statistical mode to understand trends and explore future scenarios

2. 5,000 METERS – INSTITUTIONAL PERSPECTIVES

Key Elements for Effective Analytics at an Institutional Level

- Governance – vision, strategy, policies, roles and responsibilities, controls, data agreements, M&E, etc.
- Culture – commitment of leadership, awareness of value, collaboration internally and externally, embedded in audit proceesses
- People – technical skills and knowledge, subject matter expertise, legal advice
- Processes – processes for analyzing as well as identifying, collecting and assessments reliability and validity, continuous M&E
- Technology – infrastructure, tools and software, based on objectives

Culture is key (again)

Category	Lagging	Basic	Advanced	Leading
Leadership buy-in	Leaders do not know how data analytics can help advance the entities' objectives and inform decision making.	Leaders want to use data analytics for specific objectives but do not know how.	Leaders have a clear idea about how data analytics can be used for specific objectives and communicate its value.	Leaders demand that decisions are driven by data analytics.
Staff buy-in	Staff do not understand the value of data analytics, which they view as a hindrance.	Staff are aware of the value of data analytics but use it seldomly, sporadically or incompletely.	Staff use data analytics for decision making because they are required to do so.	Staff use data analytics for decision making because they want to do so.

Source: OECD, adapted from the Center for Data Science and Public Policy, the University of Chicago. (2018). Data Maturity Framework - Center for Data Science and Public Policy, website, http://dsapp.uchicago.edu/resources/datamaturity/ (Accessed 21 February 2018).

A few leading practices for developing a culture/buy-in for data analytics

- Build support within the program with small wins
- Tailor the output to the intended audience to ensure the results are usable.
- Support across the institution and, but in particular, from program managers.
- Develop skills, but avoid siloes
- Others?

How technology-ready are you?

Category	Lagging	Basic	Advanced	Leading
Accessibility	Only accessible with the application where it is collected.	Data can be accessed outside the application, but exist in a proprietary format that requires specialized analysis software.	Data are machine readable in standard open formats	Data are machine readable in standard open format and available through an application program interface.
Integration	Data sit in the source systems.	Data are exported occasionally and integrated in an ad-hoc manner.	Data are maintained in a central data warehouse that allows for automatic, realtime aggregation and linking.	External data are also integrated.

Source: OECD, adapted from the Center for Data Science and Public Policy, the University of Chicago. (2018). Data Maturity Framework - Center for Data Science and Public Policy, website, http://dsapp.uchicago.edu/resources/datamaturity/ (Accessed 21 February 2018).

How's your data looking?

Category	Lagging	Basic	Advanced	Leading
Relevance and Sufficiency	The data collected are irrelevant to the objective or the entity does not collect relevant data.	Some of the data the entity has are relevant, but it is insufficient because key fields are missing.	Data exist that are helpful and relevant for addressing the objective, but are not sufficient to address it well.	All relevant data about the individuals or entities to be analyzed exists within the entity and it is sufficient to address the objective.
Quality	Missing rows (individuals/entities missing in the data).	Missing columns (variables missing).	No missing data, but errors in data such as typos.	No missing data and no errors in data collection.

Source: OECD, adapted from the Center for Data Science and Public Policy, the University of Chicago. (2018). Data Maturity Framework - Center for Data Science and Public Policy, website, http://dsapp.uchicago.edu/resources/datamaturity/ (Accessed 21 February 2018).

Data readiness tailored to asset declarations

Lagging

- The entity does not collect data on all public officials required to disclose their assets and does not have access to data sources necessary to verify the accuracy of data that are collected.

Basic

- The entity collects data for all public officials required to disclose their assets, but may lack information on some assets required to be disclosed. The entity has minimal access to external data sources necessary to verify the accuracy of the data.

Advanced

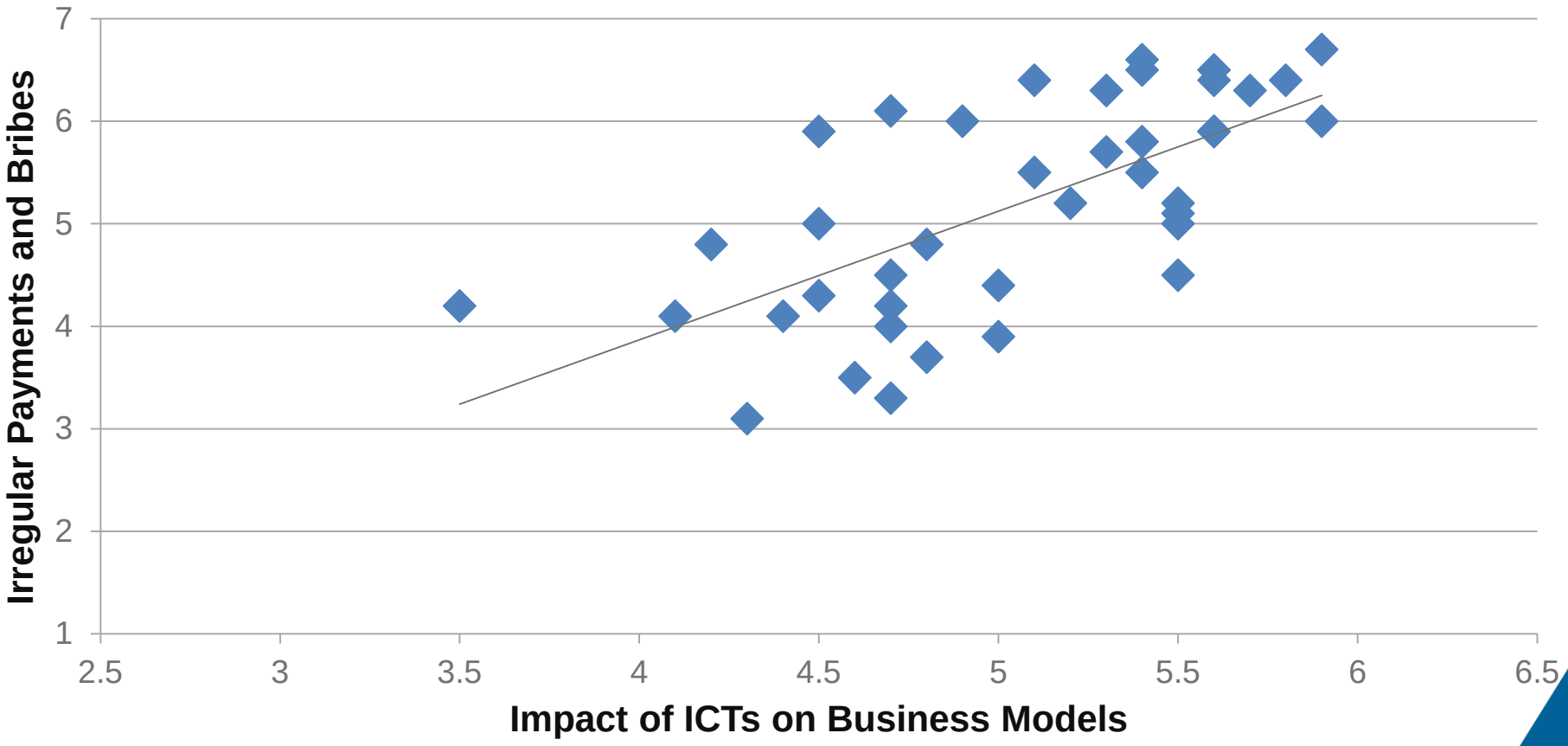
- The entity collects data on income and all relevant assets required to be disclosed by law for all public officials required to disclose their assets; however, data formats and information may not be standardized, resulting in typos or other errors in data collection. The entity has access to some external sources.

Leading

- The entity collects data on income and all relevant assets required to be disclosed by law for all public officials required to disclose their assets. Data collection methods incorporate basic checks to ensure that only valid data are accepted into the system. The entity has access to all external data sources necessary to verify the accuracy of asset-disclosure information. the information disclosed by public officials.

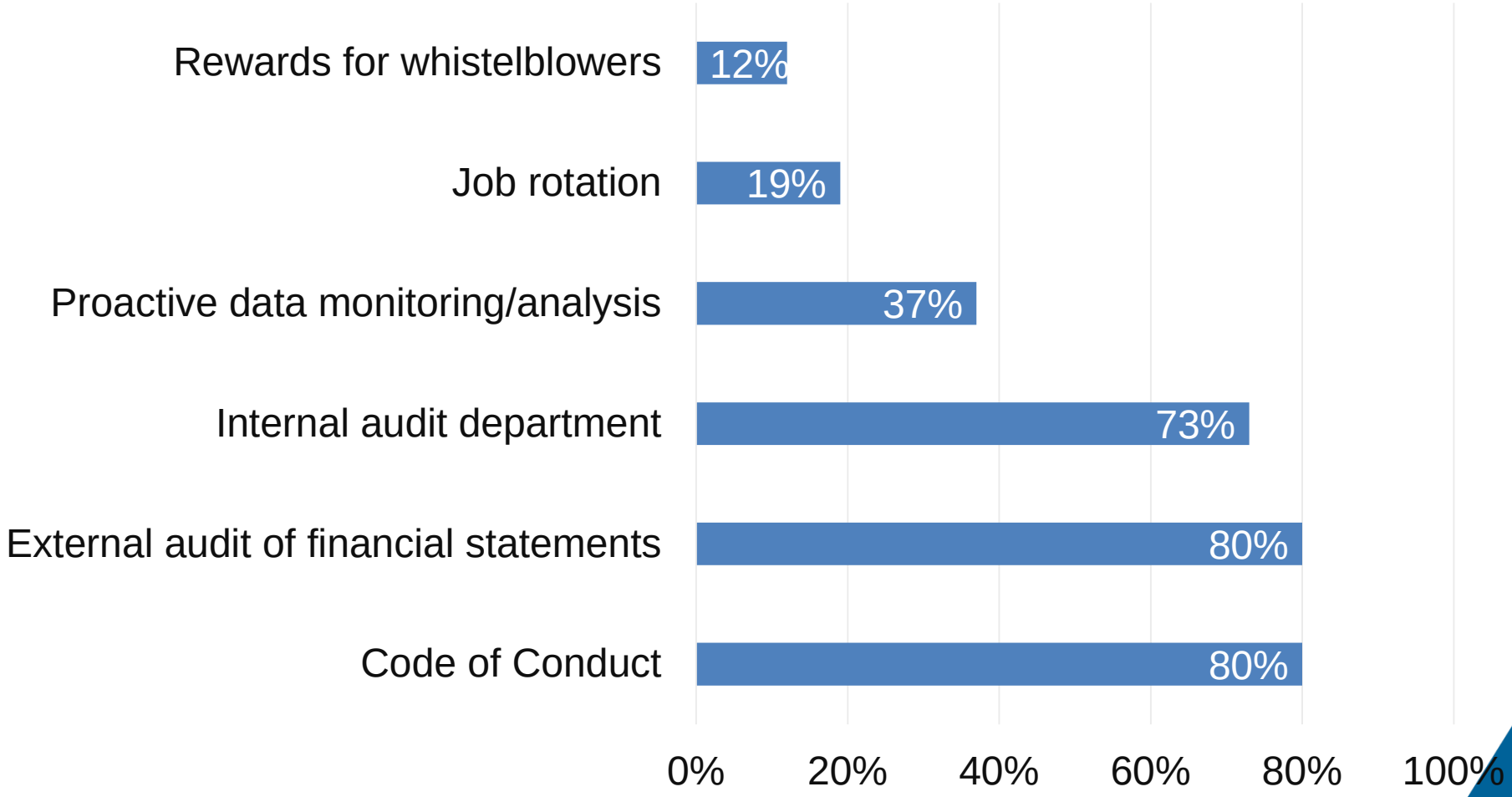
3. 100 METERS - DATA ANALYTICS FOR INTEGRITY

Impact of ICT on Fraud in OECD Countries



Source: Data taken from World Economic Forum, 'The Global Information Technology Report 2016' http://www3.weforum.org/docs/GITR2016/GITR_2016_full%20report_final.pdf ;
And the 'The Global Competitiveness Report 2015-16 ' http://www3.weforum.org/docs/gcr/2015-2016/Global_Competitiveness_Report_2015-2016.pdf

Which anti-fraud controls are most and least common?



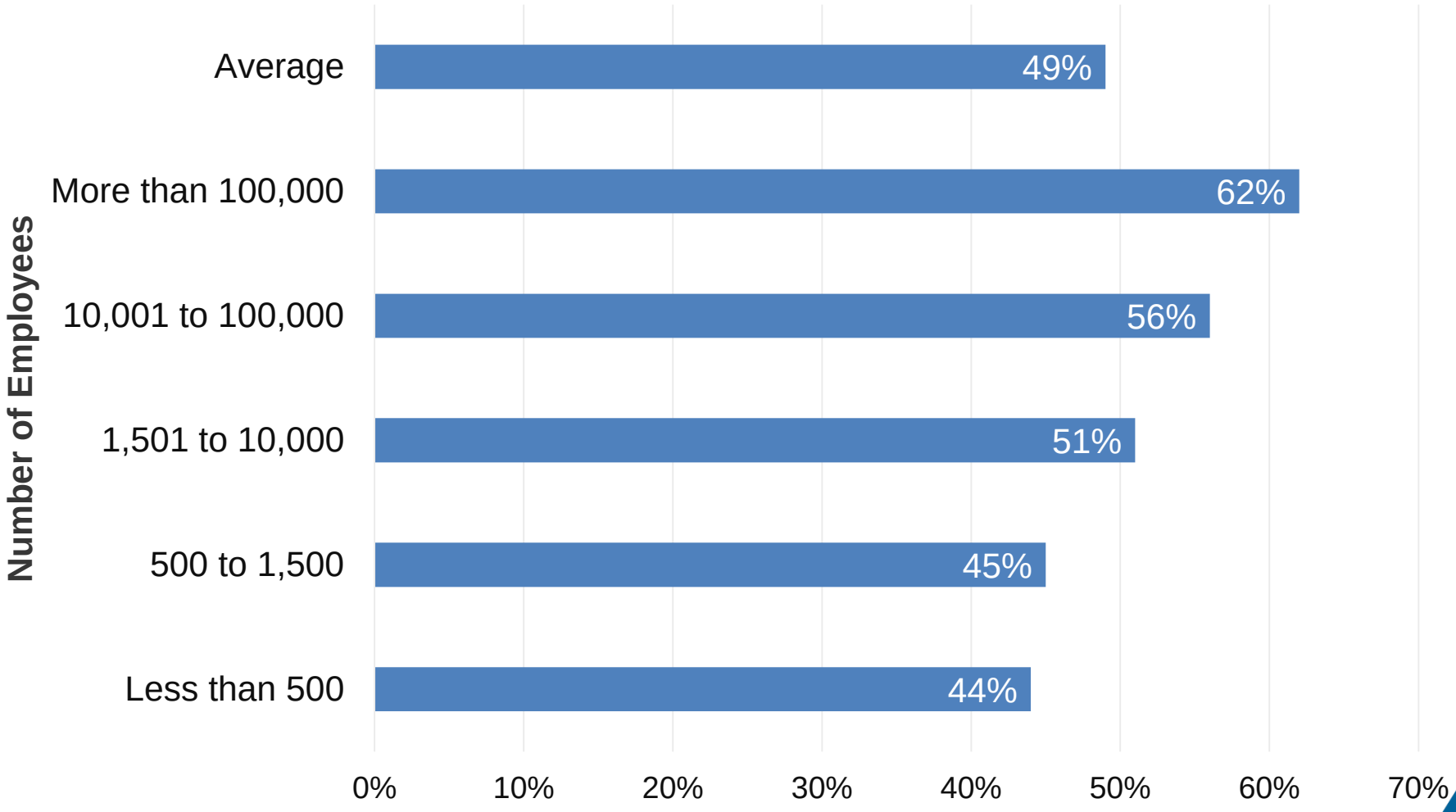
Source: ACFE, 2018 Report to the Nations

An effort at measuring the business case (top 3 controls)

Control	Percent of cases	Control in Place	Control not in place	Percent Reduction
Code of Conduct	80	110,000	250,000	56
Proactive data monitoring/analysis	37	80,000	165,000	52
Surprise audits	37	75,000	152,000	51

Source: ACFE, 2018 Report to the Nations

Data analytics used to identify potential fraud, the IA perspective

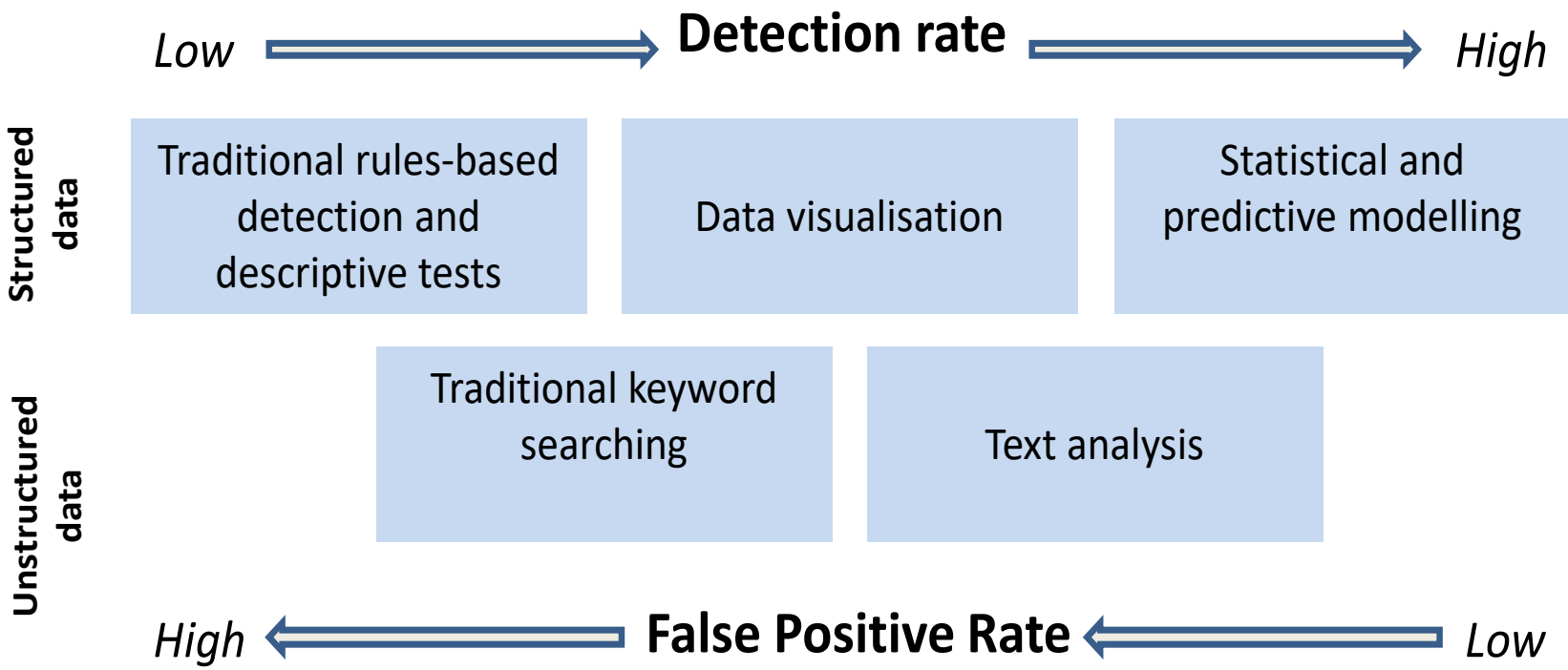


Source: Araj, Farah G. (2015), Responding to Fraud Risk: Exploring Where Internal Auditing Stands, the IIA Research Foundation, CBOK 2015 Practitioners Survey *Note:* Q96: Does your internal audit department use data mining or data analytics for the following activities? Exhibit shows respondents who chose "identification of possible frauds." *n* = 11,101.

Data analytics has applications across prevention, detection and response

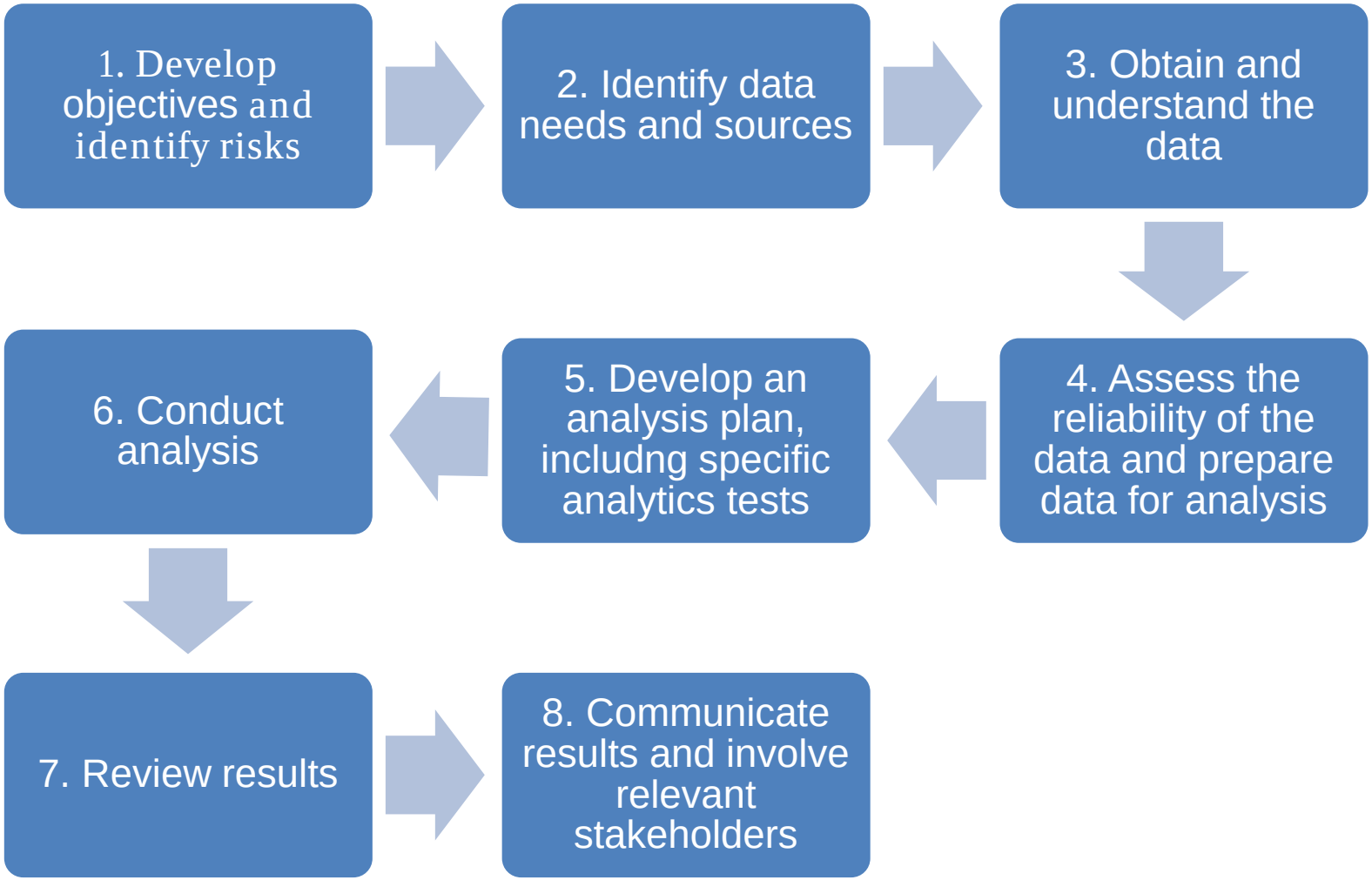
- Activities to prevent, detect and respond to fraud or corruption are interdependent and mutually reinforcing
- Data analytics for prevention
  - Data matching to verify eligibility
  - Predictive analytics
  - System edit checks
  - Informing risk assessments
- Data analytics for detection
  - Data matching after payment disbursement
  - Data mining
- Data analytics for response
  - Automating detection algorithms based on results of investigations
  - Providing evidence for substantiating cases

Not all methodologies are created equal

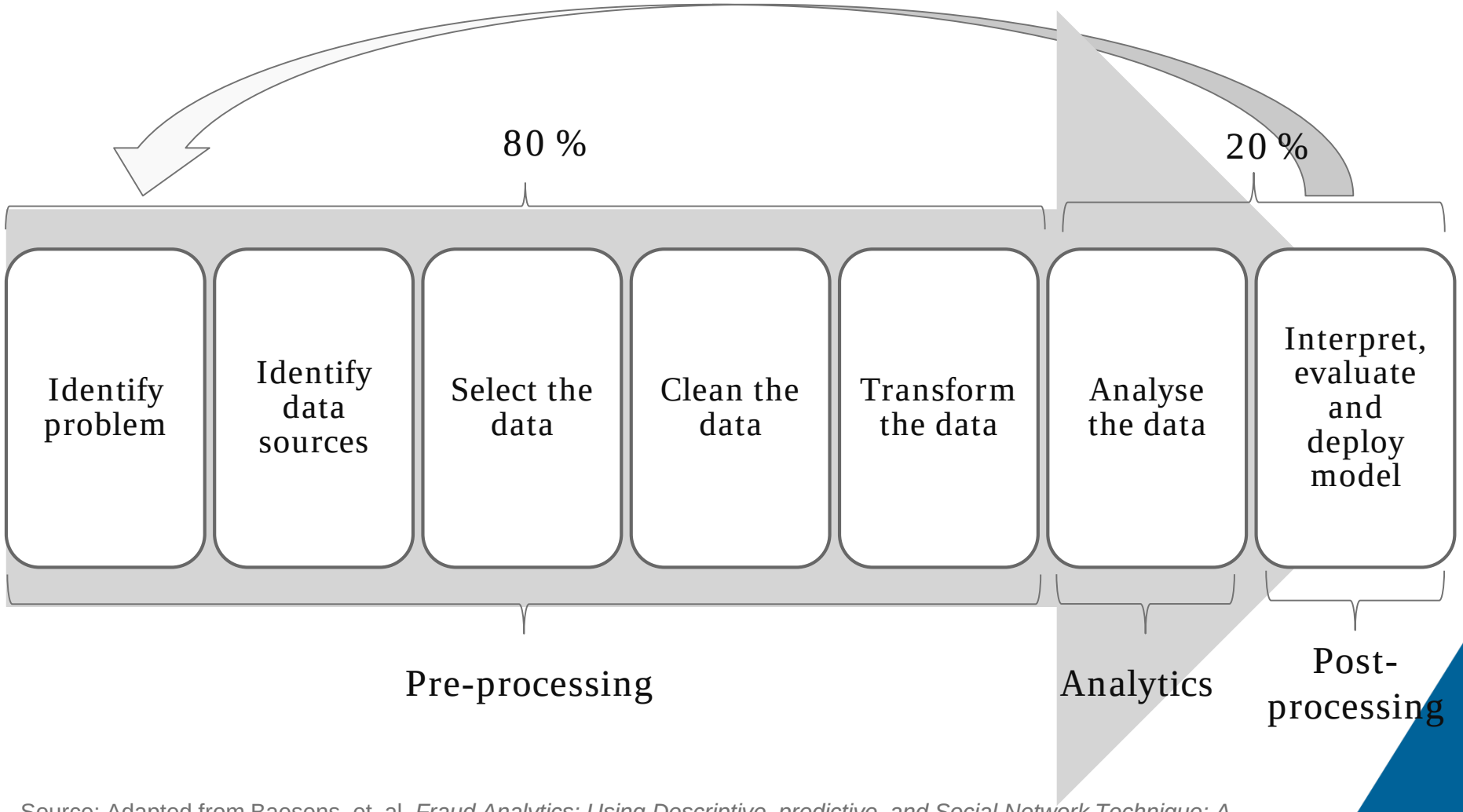


Source: Adapted from EY (2016), *Trends in Data Analytics: Fraud Detection*. Presentation at the New England Intergovernmental Audit Form, presentation delivered to the New England Intergovernmental Audit Forum, 3 November 2016, https://www.regonline.com/custImages/300000/300624/2016/11-3_Feinstein.pdf.

Planning and execution of data analytics plans



Fraud Analytics Process Model emphasizes upstream and feedback loop



Source: Adapted from Baesens, et. al, *Fraud Analytics: Using Descriptive, predictive, and Social Network Technique: A guide to data science for fraud detection,* John Wiley and Sons, Inc. 2015

Example 1: The GAO uses analytics to identify improper activities, including potential fraud

Telling Report Title: “Pervasive Passport Fraud Not Identified, but Cases of Potentially Fraudulent and High-Risk Issuances Are under Review”

Summary of GAO’s Matching Analysis and Nongeneralizable Samples

GAO’s population of passport applicants	Matching criteria			Total passport issuances[a]	Sample size	Sample cases with potential fraud or high-risk indicators
	Social Security number (SSN)	Name	Date of birth			
Deceased individuals	✓	✓	✗	181	15	1
State prisoners (from 11 states)	✓	✓	✓	68	14	7
Federal prisoners	✓	✓	✓	206	15	0
Individuals with active warrants	✓	✓	✓	486	15	3
				Total:	**59**	**11**

✓ Matched the Department of State’s passport data

✗ Not used or available for matching analysis

Source: GAO analysis of data provided by the Department of State (State), the Social Security Administration (SSA), state departments of correction, the federal Bureau of Prisons (BOP), and the U.S. Marshals Service.

Example 2: the U.K. HM Revenue and Customs and big data

Predictive analytics allows HMRC allows it to integrate and analyse information from separate data sources to identify personal tax fraud

Links taxpayers to more than one billion pieces of information held in the Connect system, which is fed from 28 data sources from government entities, onshore and offshore banks

Compares data with taxpayer self-assessment, and produces a spider diagram to visualise connections to addresses, trusts, businesses, etc.

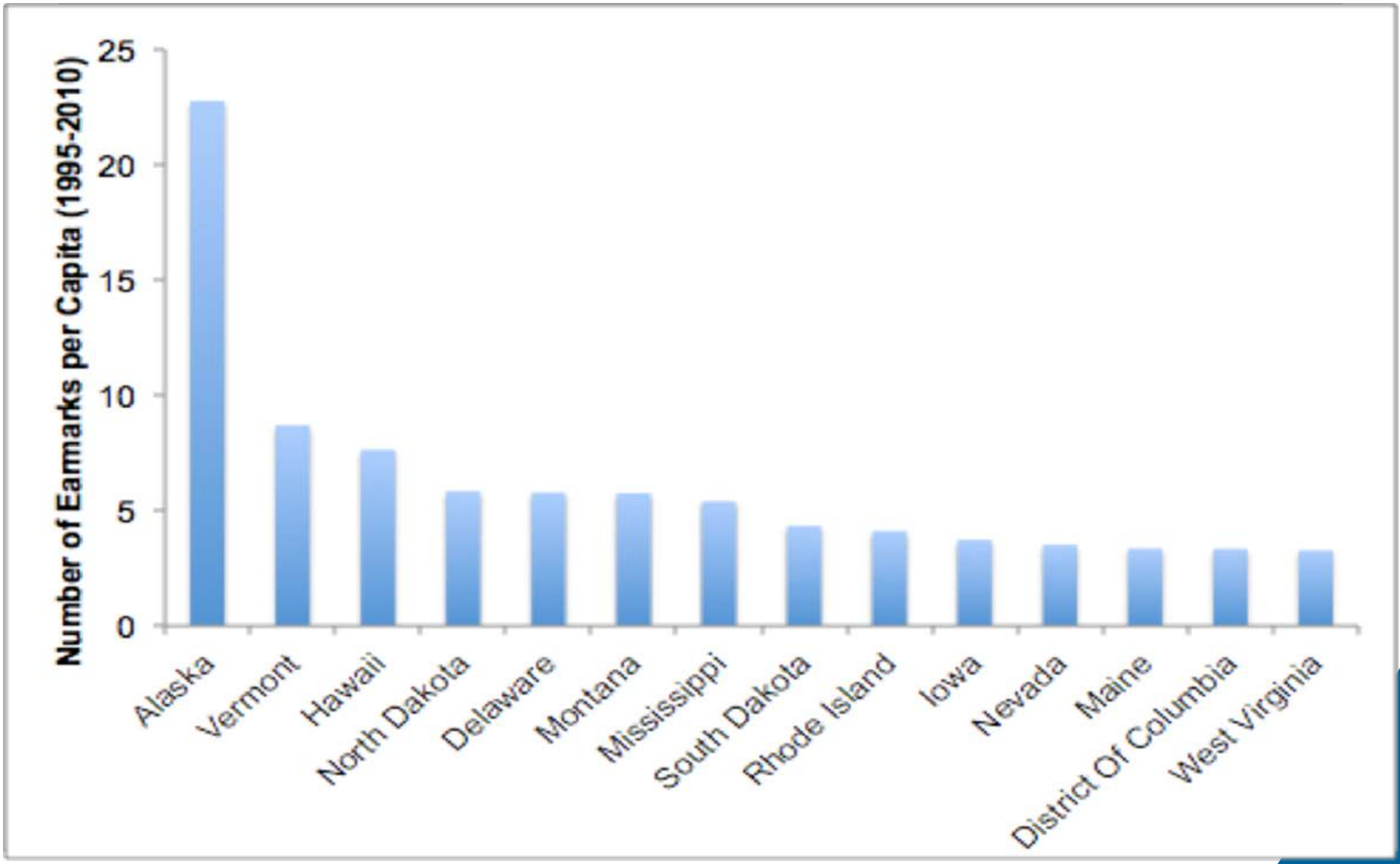
2008-2009, implementation costs for the system were around 18.8 million GBP

The system resulted in increased tax yields of 572 million GBP, a 30:1 return-on-investment

In addition to identifying potentially fraudulent claims, the system also enables HMRC to process legitimate claims faster

Source: JMP Statistical Discovery, “The Benefits of Analytics in the Public Sector: The Smart Way to Cut Costs, Optimise Performance and Deliver Reform in the Public Sector.”

Example 3: Concerned citizens using text analytics to spot earmarks



Example 4: Behavioral insights to mine keyword terms linked to the Fraud Triangle

Rationalisation	Incentive/pressure	Opportunity
…I deserve it	…make the number	…special fees
…nobody will find out	…don't leave a trail	…client-side storage
…they owe it to me	…not comfortable	…off the books
…everybody does it	…why are we doing this?	…cash advance
…fix it later	…pull out all the stops	…side commission
…not hurting anyone	…want no part of this	…backdate
…won't miss it	…only a timing difference	…no inspection
…don't get paid enough	…not ethical	…no receipt
		…smooth earnings
		…pull earnings forward

Source: EY (2016), Trends in Data Analytics: Fraud Detection. Presentation at the New England Intergovernmental Audit Form, presentation delivered to the New England Intergovernmental Audit Forum, 3 November 2016, https://www.regonline.com/custImages/300000/300624/2016/11-3_Feinstein.pdf.

Of course, there are challenges

- Auditors don’t own the data – need for sharing and success of D&A is based on this (as well as legal savvy)
- Time and resources constraints
- False positives
- How will D&A affect the 3/4 Lines of Defense? Maturity mismatches between auditor and auditees
- Return on investment and cost/benefit analyses
- Risk of non-compliance with audit standards, particularly when mixed with investigative techniques
- Data security and privacy
- Falling down the rabbit hole
- Etc., etc.

But thankfully there also known common elements of success and “good enough practices” (a recap)

- High-level commitment and communication is key
- But so is ground-up initiative and giving space for this
- Given the importance of the “pre-processing” phase, understand and invest in data readiness
- Have clear objectives
- Start with small wins and demystify
- Then enable audit staff to build experience and confidence through a structured roll-out programme
- Experiment, network and share experiences!
- Others?

Data analytics complements, but does not replace, professional skepticism

"Auditors should take care that they are not over relying on data analytics. As powerful as these tools are, or are expected to become, they nonetheless are **not substitutes** for the auditor's knowledge, judgment, and exercise of professional skepticism."

~ Public Company Accounting Oversight Board Member Steven B. Harris, April 2017

BONUS SLIDES - A FEW NOTES ON OECD

OECD is a learning organisation, creating and sharing evidence for improving SAIs and government

- **Building** knowledge, such as:
  - Safeguarding integrity through risk management/control
  - Contributing to the evidence base on anti-corruption measures and effective auditing (e.g. comparative analyses, surveys and leading practices)
  - Closing implementation gaps through improved risk mapping and data analytics
- **Communicating** evidence through key products, such as Integrity Reviews, Government at a Glance and international standards and recommendations (e.g. 2017 OECD Recommendation on Public Sector Integrity)
- **Learning** through dialogue, workshops, forums and bridging gaps between experts

More comparative data, analyses & country reviews

No Shortage of Other Collective Efforts

- Standards and Guidelines (International, national and subnational)
- Working Groups and task forces
- Conferences and seminars
- Specialised networks
- OECD Recommendations:
  - 2017 Recommendation of the Council on Public Integrity
  - 2014 Recommendation of the Council on Digital Government Strategies
  - 2014 Recommendation of the Council on Public Procurement


OECD