BIG DATA ANALYSIS

THEORY AND PRACTICE FOR TAX INVESTIGATION

Restricted Use - À usage restreint

YOUR TUTOR TODAY

ARTHUR MOURA:

Federal Prosecutor for the National Treasury since 2003

Former Head of the Tax Investigation Unit (5th Region)

Chevening Scholar 2017-2018

Master in Public Administration - University of Nottingham (UK)

Writer and Professor of Tax and Procedure Law

Restricted Use - À usage restreint

Thoughts and Opinions …

Are mine and not necessarily coincide with those of my Agency (PGFN) or the Academy

Restricted Use - /

MODULE PLANNING

Focus on Management, rather than statistics or computer science.

First Part: Data Analysis, Big Data, and Management.

Exercise 1: applying Big Data Analysis to the public sector.

Second Part: Practical example of Big Data Analysis in Tax Investigation. AI and Tax Investigation.

Exercise 2: applying Big Data Analysis to tax and crime investigation.

Restricted Use - À usage restreint

By the end of this module

You will:

a. have a better understanding of Data Analysis and Big Data and their uses in managing tax investigations
b. be familiar with the potential of those tools for you and your agency
c. be aware of some of the risks involved in Big Data
d. get a grasp on AI related to tax crime investigation.

Restricted Use - À usage restreint

A bit of history

- Data collection: counting bricks
  Data protection: covering the wall
- Cleaning the data: spotting outliers
- Combining data: numbers of bricks x size of each brick
- Data analysis: Precise height of the ladder

Restricted Use

SUMERIAN IPAD

Data collection: Census and Satraps

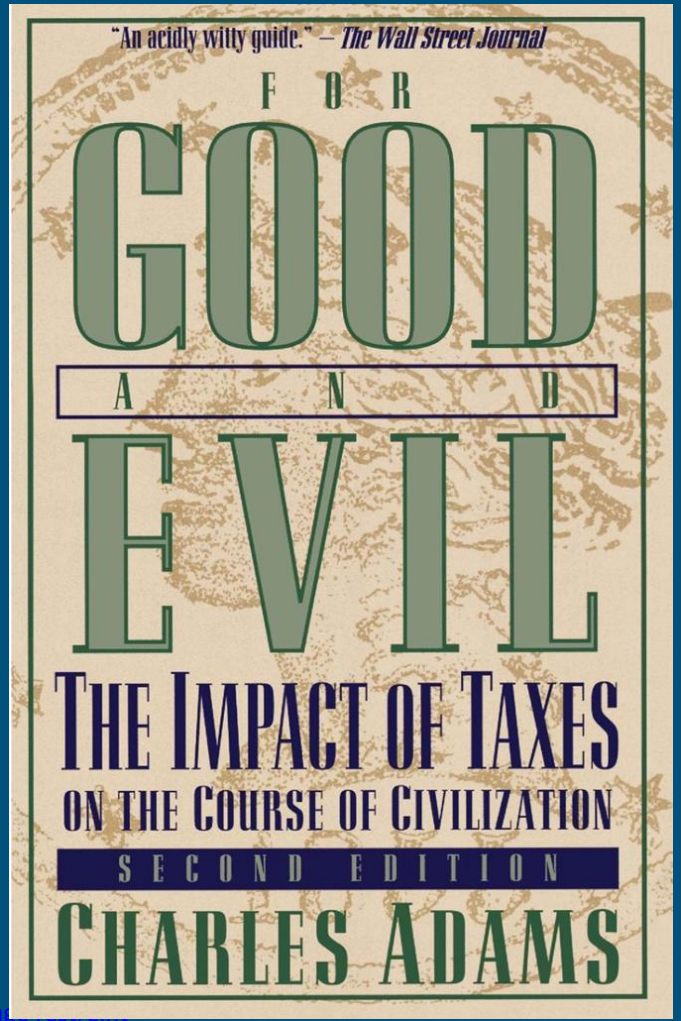
Data Storage: Clay Tablets

Combining the data: size of population and sales records

Data Analysis: food distribution and tax collection.

Restricted Use - À usage restreint

Recommendation!



Restricted Use - À usage

DATA ANALYSIS

The process of collecting, cleaning, combining and formatting data in order to extract useful (valuable) information and insights to help decision making*.

- Collecting: diversity of sources, formats, languages.
- Mining data: too much noise for you to pick a note.
- Cleaning: involves removing duplicate data, correcting errors, and dealing with missing values.*
- Combining: data fusion, establishing correlations to better understand the data and identify patterns, trends, and outliers.
- Formatting: visual communications, dashboards etc.

Restricted Use - À usage restreint

Why use Data Analysis*?

- Data analysis provides valuable insights and information that can help individuals and organizations make better decisions.
- Problem-solving: Data analysis helps identify problems.
- Cost-saving: Data analysis can help identify inefficiencies and waste in processes.
- Innovation: Data analysis can help identify new opportunities.
- Improved performance: Data analysis can help individuals to deliver more.

Restricted Use - À usage restreint

E	F	G	H	I	J	K	L	M	N	O	P
Origem do Devedor	Valor consolidado do débito	Nome do PFN Responsável	Linha Investigativa	Confirma linha investigativa?	Providências I	Providências II	Providências III	Corresponsável identificado?	Redirecionamento/IDPJ deferido?	Bens e Direitos Identificados	Bens e Direitos Penhorados?
NATAL	375.213,17	Peter John Arrowsmith Cook Junior	AQUISIÇÕES DOI GESTORES EMPRESAS BAIXADAS	NÃO	NOTA DE NÃO ATUAÇÃO			NÃO SE APLICA	NÃO SE APLICA	NÃO	NÃO
NATAL	395.743,78	Peter John Arrowsmith Cook Junior	AQUISIÇÕES DOI GESTORES EMPRESAS BAIXADAS	NÃO	REDIRECIONAMENTO	PENHORA		SIM	NÃO	SIM	NÃO
NATAL	672.826,95	Peter John Arrowsmith Cook Junior	AQUISIÇÕES DOI GESTORES EMPRESAS BAIXADAS	NÃO	AJUIZAMENTO	PENHORA		NÃO	NÃO	SIM	NÃO
NATAL	1.033.491,89	Peter John Arrowsmith Cook Junior	AQUISIÇÕES DOI GESTORES EMPRESAS BAIXADAS	NÃO	NOTA DE NÃO ATUAÇÃO			SIM	SIM	NÃO	NÃO
AREIA BRANCA	168.333,14	Peter John Arrowsmith Cook Junior	AQUISIÇÕES DOI GESTORES EMPRESAS BAIXADAS	NÃO	REDIRECIONAMENTO	PENHORA		SIM	SIM	SIM	NÃO
ASSU	210.916,73	Peter John Arrowsmith Cook Junior	AQUISIÇÕES DOI GESTORES EMPRESAS BAIXADAS	NÃO	NOTA DE NÃO ATUAÇÃO			NÃO SE APLICA	NÃO SE APLICA	NÃO	NÃO
ASSU	933.629,81	Peter John Arrowsmith Cook Junior	AQUISIÇÕES DOI GESTORES EMPRESAS BAIXADAS	NÃO	NOTA DE NÃO ATUAÇÃO			NÃO SE APLICA	NÃO SE APLICA	NÃO	NÃO
ASSU	300.895,32	Peter John Arrowsmith Cook Junior	AQUISIÇÕES DOI GESTORES EMPRESAS BAIXADAS	NÃO	AJUIZAMENTO	IDPJ		SIM		SIM	
CEARÁ-MIRIM	458.858,68	Peter John Arrowsmith Cook Junior	AQUISIÇÕES DOI GESTORES EMPRESAS BAIXADAS	NÃO	REDIRECIONAMENTO	PENHORA		SIM	SIM	SIM	
[illegible]	[illegible]	Peter John Arrowsmith Cook [illegible]	AQUISIÇÕES DOI GESTORES [illegible]	[illegible]	NOTA DE NÃO [illegible]			[illegible]	[illegible]	[illegible]	[illegible]

Processo | Anexar Petições/Documentos | Expedientes (0) | Audiência | Perícia | Anexos Físicos | Distribuição | Associados | Retificações | RPV/PRC | Dados da CDA | Acesso de Terceiros

Download de documentos em PDF | Paginador | Recurso para TRF

Movimentações do Processo

Assuntos

Assunto
DIREITO TRIBUTÁRIO|Obrigação Tributária|Responsabilidade tributária|Substituição Tributária|**Substituição Tributária**
DIREITO CIVIL|Empresas|Sociedade |Responsabilidade dos sócios e administradores|**Responsabilidade dos sócios e administradores**

Foram encontrados: 2 resultado

Polo ativo

Participante	Tipo de participação	Situação RFB
FAZENDA NACIONAL	SUSCITANTE	

Foram encontrados: 1 resultados

Polo passivo

Participante	Tipo de participação	Situação RFB
COMERCIO E REBENEFICIAMENTO DE CEREAIS MERCOSUL LTDA	SUSCITADO	
BRUNO FIGUEIREDO NOBREGA	SUSCITADO	
MATEUS FIGUEIREDO NOBREGA	SUSCITADO	
FELIPE FIGUEIREDO NOBREGA	SUSCITADO	
ROGACIANO NUNES DA NOBREGA NETO	SUSCITADO	
CEREALISTA PARAIBINHA LTDA - ME	SUSCITADO	
FRANCISCO DE ASSIS DE BRITO	SUSCITADO	
ESTIVADORA FEIRANTE LTDA	SUSCITADO	

Foram encontrados: 8 resulta

Outros interessados - Não existem partes cadastradas a outros interessados.

Participante	Tipo de participação	Situação RFB

Foram encontrados: 0 resultado

Instância: [Todos] | Exibir certidões automáticas: Sim / Não | Tipo de documento: [Todos] | Descrição do documento | Documentos liberados para advogado/procurador renomear | Número identificador | Pesquisar | Limpar

Documentos

Motion granted

Identificador	Assinatura	Descrição	Tipo de documento	Documentos/Ícones
4058201.11833874	12/06/23 16:29	Despacho Inspeção - 1227 - INSPEÇÃO ORDINÁRIA ANUAL 2023 [10ª VARA FEDERAL]	Despacho Inspeção	
4058201.11791332	05/06/23 17:29	Decisão	Decisão	
4058201.11447416	27/03/23 15:12	Certidão de Distribuição	Certidão	
4058201.11429110	27/03/23 15:10	Incidente de Desconsideração da Personalidade Jurídica - (Sigiloso)	Petição Inicial - (Sigiloso)	
4058201.11445248	27/03/23 15:10	Inicial - (Sigiloso)	Documento de Comprovação - (Sigiloso)	
4058201.11445260	27/03/23 15:10	1- Cadastro Nacional de Empresas - CNE - (Sigiloso)	Documento de Comprovação - (Sigiloso)	
4058201.11445265	27/03/23 15:10	2- Reclamação trabalhista - (Sigiloso)	Documento de Comprovação - (Sigiloso)	
4058201.11445280	27/03/23 15:10	3- Relações de ROGACIANO NUNES DA NOBREGA NETO - (Sigiloso)	Documento de Comprovação - (Sigiloso)	
4058201.11445289	27/03/23 15:10	4- Relações de BRUNO FIGUEIREDO NÓBREGA - (Sigiloso)	Documento de Comprovação - (Sigiloso)	
4058201.11445299	27/03/23 15:10	5- Relações de FELIPE FIGUEIREDO NÓBREGA - (Sigiloso)	Documento de Comprovação - (Sigiloso)	

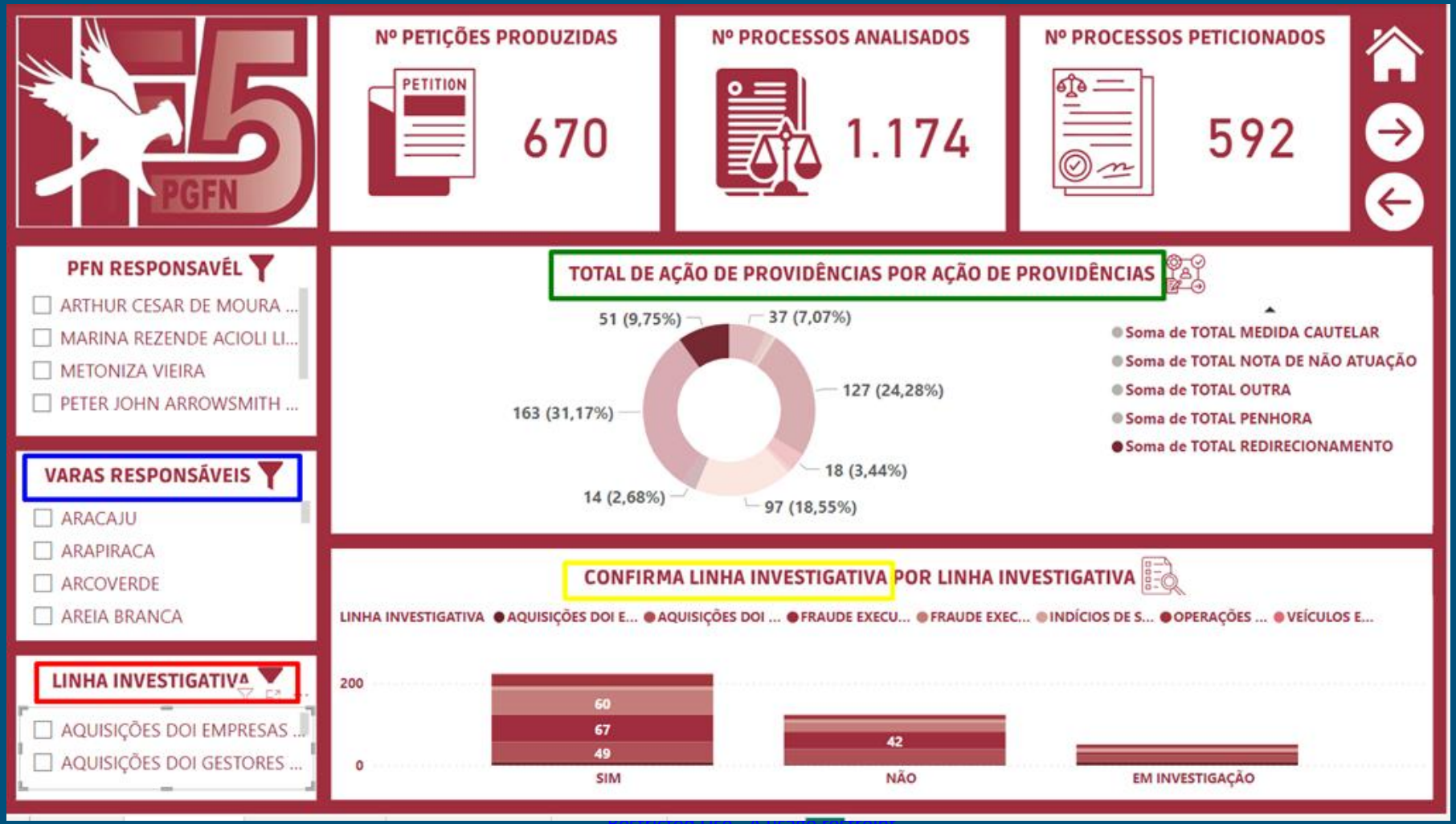
Petition and Documents

Different fomat

RESULTADOS DO IF5 2022
PGFN

- Relatório Geral ← TEAM RESULTS
- Relatório Individual ← INDIVIDUAL RESULTS
- Regularização da Dívida ← PAYMENT AND OTHER OUTCOMES
- Valores e Bens ← ASSET RECOVERY
- Árvore Hierárquica

Visualizações
Criar visual
Filtros
Valores
Adicionar os campos de
Drill-through
Relatório cruzado
Manter todos os filtros
Adicionar os campos de

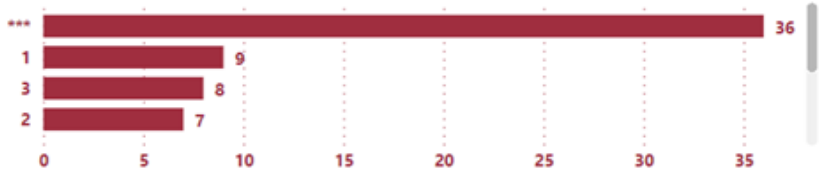


Restricted Use - A usage restreint

NÚMERO DE INVESTIGAÇÕES

69

PETIÇÕES DO FLUXO DA IF5



PFN RESPONSAVÉ

- ARTHUR CESAR DE MOURA ...
- MARINA REZENDE ACIOLI LI...
- METONIZA VIEIRA
- PETER JOHN ARROWSMITH ...

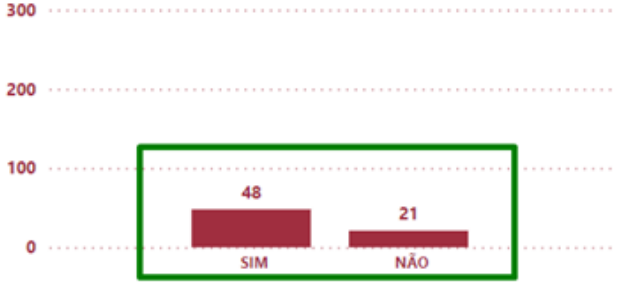
VARAS RESPONSÁVEIS

- CAMPINA GRANDE
- GUARABIRA
- JOÃO PESSOA
- NATAL

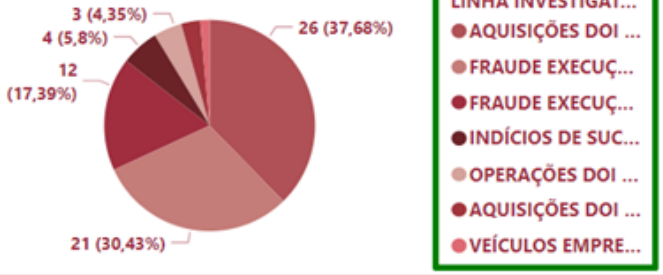
LINHA INVESTIGATIVA

- AQUISIÇÕES DOI EMPRESAS ...
- AQUISIÇÕES DOI GESTORES ...

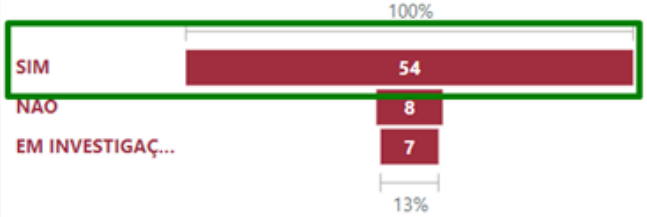
CONFIRMA LINHA INVESTIGATIVA POR CONFIRMA LINHA INVESTIGATIVA



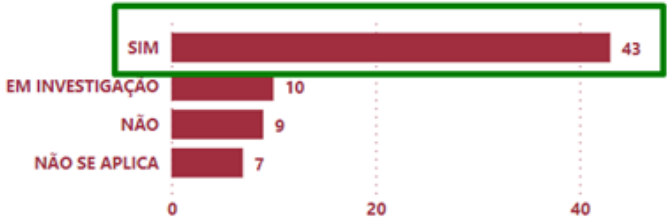
LINHA INVESTIGATIVA POR LINHA INVESTIGATIVA



BENS E DIREITOS IDENTIFICADOS



CORRESPONSÁVEL IDENTIFICADO



Restricted Use - A usage restreint

Why use Data Analysis?

- Problem-solving: Corporate group fraud typology failed in more than 50% of the cases.
- Cost-saving: The 5th Unit did more with less.
- Innovation: Jurimetrics in tax investigation judicial litigation.
- Improved performance: 50%.

Restricted Use - À usage restreint

Outcome:

Optimising operation by focusing on successful typologies, distributing the cases in accordance with each one’s abilities and having a better understanding of how each judge/court operates.

Restricted Use - À usage restreint

Was that Analysis Big Data Analysis?

Big Data: the use of large scale computing power (...) to collect, process and analyse data characterised by a large volume, velocity, variety and value (OECD, 2016).

Restricted Use - A

Massive volume of data

Big data originally referred to the tremendous amount of data digitally generated in the www. That massive amount includes email, photos, texts, videos, podcasts… but also metadata, such as the data produced every time you access your bank account, watch netflix or buy a product in an online store.

9GAG

Big data - 9GAG

Restricted Use - A

Tax administration generates and uses loads of data

- Registry and documentation of transactions, operations, sales, services and assets for tax purposes.
- Financial, commercial and labour registries.
- Tax returns and other tax duties and obligations.
- Audits and inspections.
- Administrative opinions, procedures and decisions.

Restricted Use - À usage restreint

Structured, Semi-Structured and Unstructured Data

- Structured and tabulated data is relatively easy to manage and is amenable to statistical analysis.
- In contrast, unstructured data is not so easily categorized and includes photos, videos, tweets, and word-processing documents.
- Dealing with unstructured data is challenging: since it cannot be stored in traditional databases or spreadsheets, special tools have had to be developed to extract useful information.

Restricted Use - À usage restreint

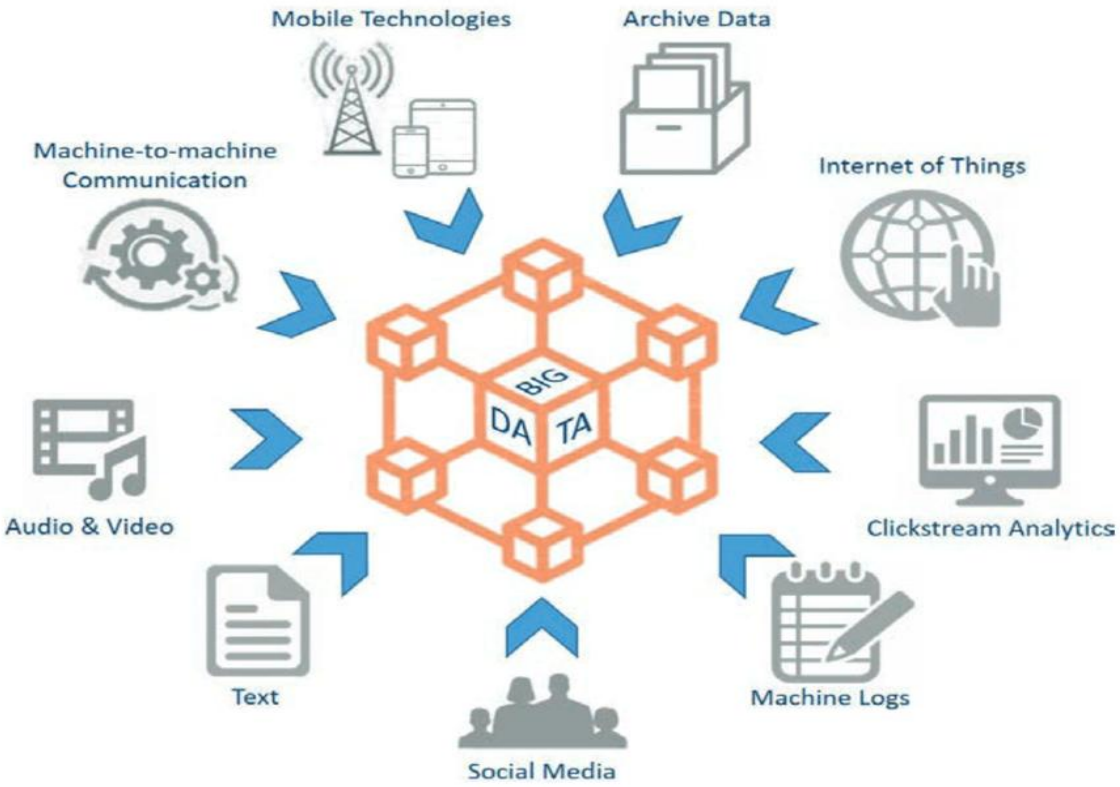
Challenges!

- Storing absurdly massive amounts of data in an organized way.
- 90% of data is unstructured.
- When the number of variables becomes large, the number of spurious correlations also increases.
- Extraordinary processing capacity: expensive equipment, development of new softwares.

Restricted Use - À usage restreint

SOURCES OF BIG DATA: another challenge!

Figure 3.1. **Sources of Big Data**



Source: Russian Federal Tax Service.

In Brazil: big numbers.

1. In 2024 alone, 4.254.903 new companies opened up, and 2.436.190 went bankrupt in Brazil. There are 21.095.654 companies regularly functioning in the country.
2. There are more than 30 million tax legal cases (tax foreclosure, legal actions disputing taxation, tax fraud etc).
3. 40 million individuals to fill a tax return each year.
4. Federal tax arrears: 500 billion Dollars.
5. 20.000 whistleblowing cases in 3 years.
6. 25.000 large delinquent taxpayers to be investigated.

Restricted Use - À usage restreint

The Vs of Big Data

Gartner's big data definition is not much longer than a tweet:

"Big data" is high-volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."

Restricted Use - À usage restreint

VOLUME*: SCALE OF DATA

- 147 zettabytes of data generated per year, which is equivalent to 402.74 million terabytes of data per day.
- Over 8.58 billion mobile subscriptions in use worldwide
- There are 18 billion IoT connected devices worldwide
- A big dataset usually has at least a million rows.
- More than 97.2% of companies invest in big data solutions
- The global payments industry handled 3.4 trillion transactions in 2023, with a value of $1.8 quadrillion and a revenue pool of $2.4 trillion.

Restricted Use - À usag

SI prefixes

v·

Prefix		Base 10	Decimal	Adopt [nb 1
Name	Symbol			
quetta	Q	10^{30}	1 000 000 000 000 000 000 000 000 000 000	2022
ronna	R	10^{27}	1 000 000 000 000 000 000 000 000 000	
yotta	Y	10^{24}	1 000 000 000 000 000 000 000 000	199
zetta	Z	10^{21}	1 000 000 000 000 000 000 000	
exa	E	10^{18}	1 000 000 000 000 000 000	1975
peta	P	10^{15}	1 000 000 000 000 000	
tera	T	10^{12}	1 000 000 000 000	196
giga	G	10^{9}	1 000 000 000	
mega	M	10^{6}	1 000 000	187
kilo	k	10^{3}	1 000	
hecto	h	10^{2}	100	179
deca	da	10^{1}	10	

VELOCITY*: SPEED AT WHICH DATA IS GENERATED

- Approximately 30.7 million online stores, or e-commerce sites, worldwide
- More than 300 million photos get uploaded per day. Every minute there are 510,000 comments posted and 293,000 statuses updated.
- OECD Digital Government Policy Framework (DGPF): Digital by design; Data-driven public sector; Government as a platform; Open by default; User-driven; Proactiveness (2020)**

Restricted Use - À usage restreint

VARIETY*: DIFFERENT FORMS OF DATA

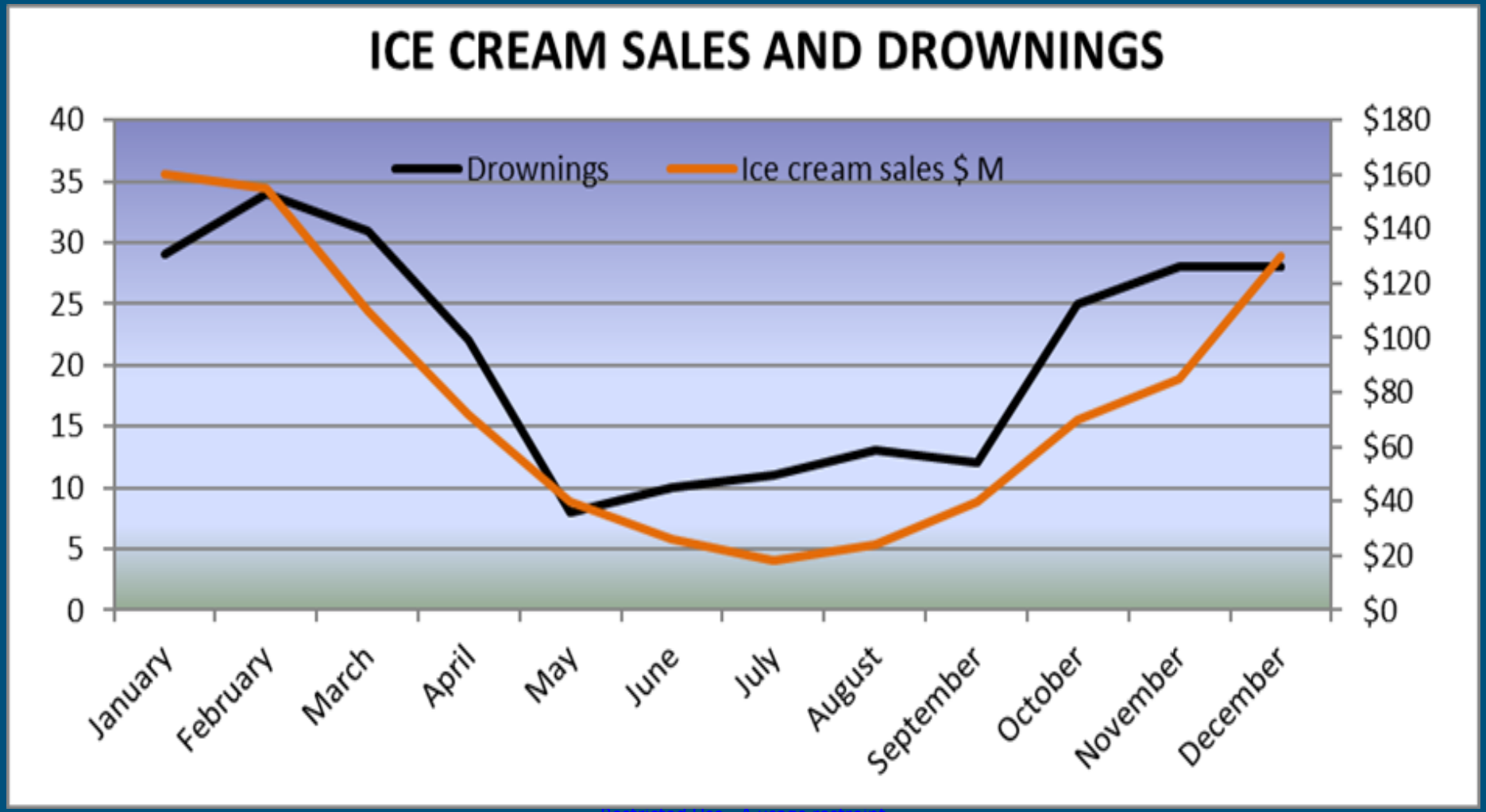
- Social media: photos, videos, audios, tweets, posts, comments etc.
- Online transactions
- Geospatial data
- Biodata
- Forms, reports, tables, spreadsheets, essays etc.
- Operational data, event data, logs and metrics, scripts and algorithms, searches and prompts, etc.

Restricted Use - À usage restreint

VERACITY: DATA ACCURACY

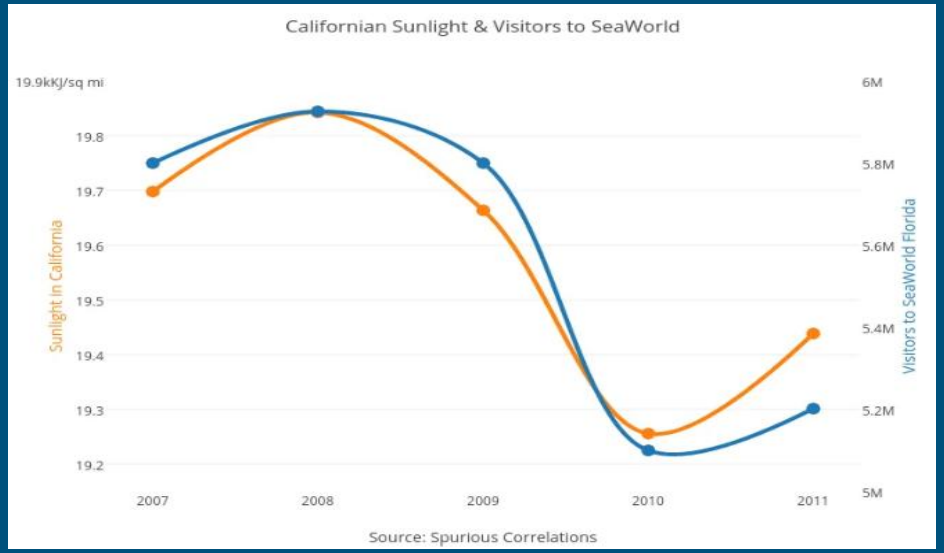
- The degree to which data is precise, correct, and free of errors.
- It's a key component of data quality and is essential for making reliable and valid decisions and insights from data.
- What is the truth and who’s to tell?
- Who watches the Watchmen?

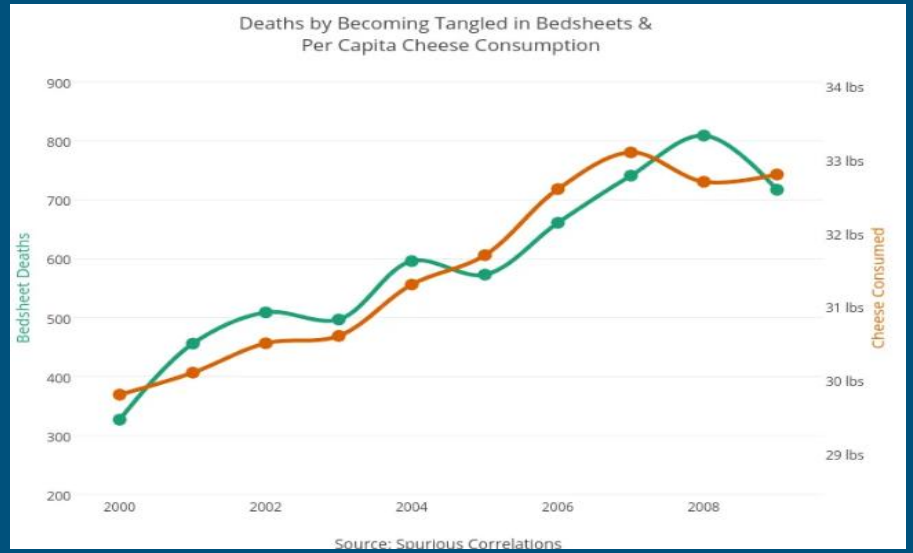
Restricted Use - À usage restreint

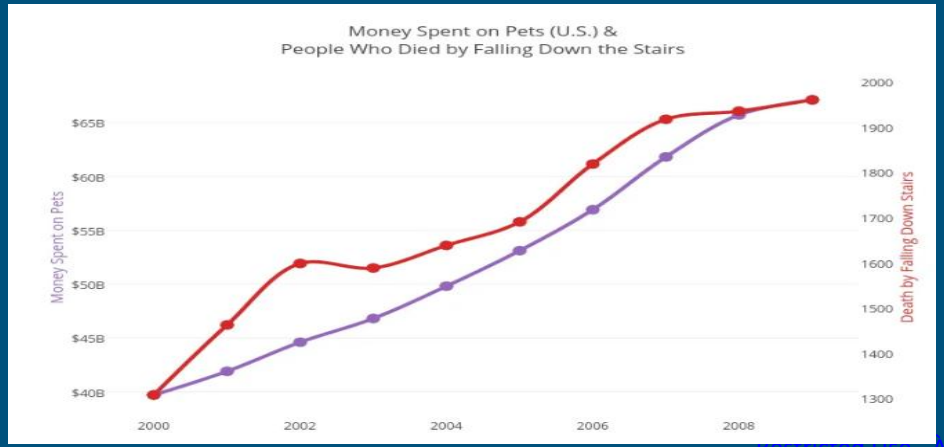


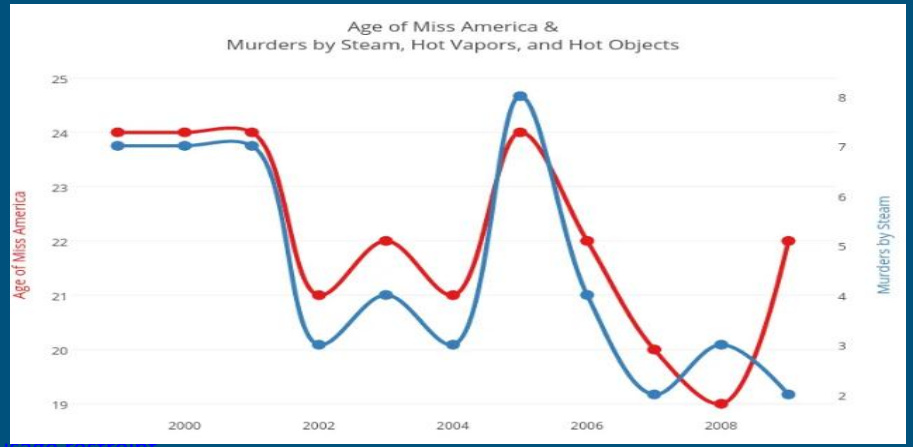
Restricted Use - A usage restreint

Restreint - À usage restreint









Restricted Use - A usage restraint

https://www.advantecs.co

Restricted Use - À usage restreint

VALIDATION*: my personal contribution to the “V - trend”

<u>Big data and Tax and Crime Investigation: yet another V</u>.

- Can it be used as intelligence, as evidence, or both?
- Is it legal according to your jurisdiction data protection act?

Restricted Use - À usage restreint

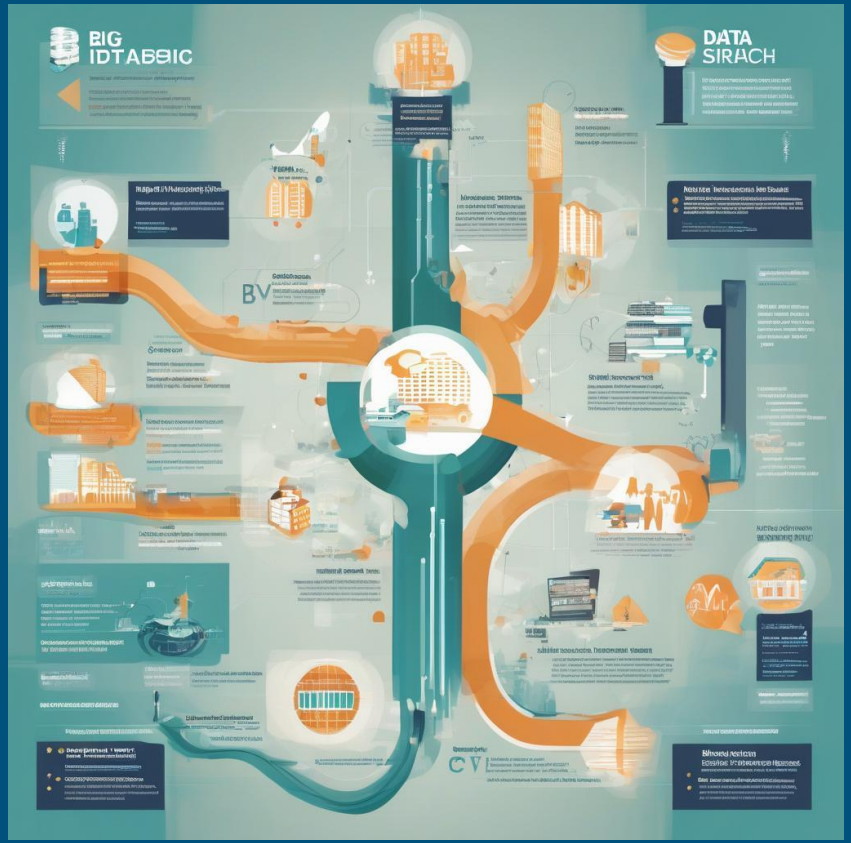
Big Data Analytics

Restricted Use - À usage restreint

Activity 1.

Discuss in which areas of the public sector Big Data Analysis could improve public service delivery within tax administration.

- Breakout room.
- Duration: 15 minutes.
- Choose a spokesperson to tell us the main ideas of the group.



Restricted Use - À usage restreint

Benefits of using Analytics in tax administration

SOURCE: FTA SECRETARIAT.

MANAGEMENT :

- Produce more accurate revenue estimates for the jurisdiction, to better plan government spending
- Support tax administration management, finance ministry and parliament in evaluation of potential policy changes through forecasting and what-if-analysis, including possible introduction of new revenue sources
- Uncover opportunities for increasing effectiveness and efficiency within the administration

Restricted Use - À usage restreint

Benefits of using Analytics in tax administration

TAXPAYER SERVICES :

- Uncover opportunities for reducing taxpayer burden
- Improve the accuracy and efficiency of taxpayer information campaigns
- Improve the quality of chatbots
- Discover tax administration reputation through sentiment analysis of social media

Restricted Use - À usage restreint

Benefits of using Analytics in tax administration

COMPLIANCE :

- Uncover fraud attempts that involve a range of factors that otherwise would not be considered together
- Uncover fraud attempts through network analysis and integration of data sources
- Discover trends in tax fraud through linking social media data with taxpayer data where legislation allows
- Uncover risk of taxpayer bankruptcy in order to take pre-emptive measures
- Improve the accuracy of audit case selection
- Improve the efficiency and effectiveness of compliance management through automated auditing
- Produce more accurate taxpayer risk segmentation, in order to target taxpayer measures according to risk of and expected reasons for noncompliance
- Use patterns in tax administration data sources to reduce the probability of future delays or lack of filing or payment

Restricted Use - À usage restreint

Benefits of using Analytics in tax administration

TAX FUNCTIONS :

- Uncover lack of registration and improve registration completeness
- Improve debt management

Restricted Use - À usage restreint

Examples in other areas:

1. Health services: Pandemic control and vaccination.
2. Education: monitoring quality of schools and students' performance.
3. Social Services: tailored/customised services according to individual needs: migrants, homeless etc.
4. Procurement: competition and fairness. M2M purchases.
5. Police: reuse of public open data to provide predictive policing
6. Emergency response: use of data from cars' sensors and social media to provide quick response.

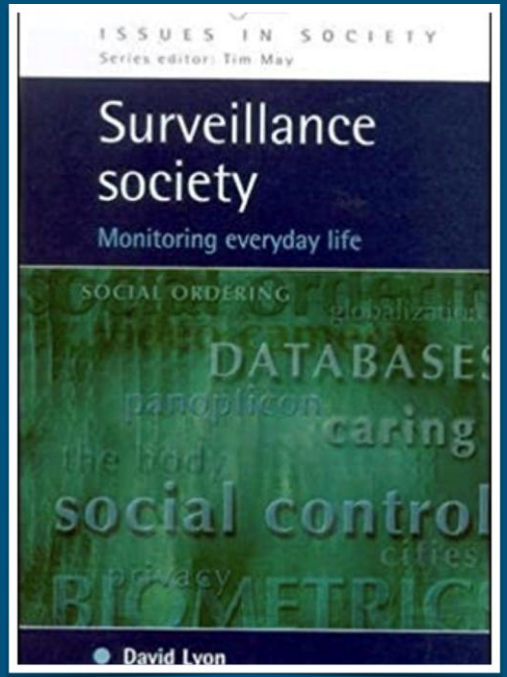
Restricted Use - À usage restreint

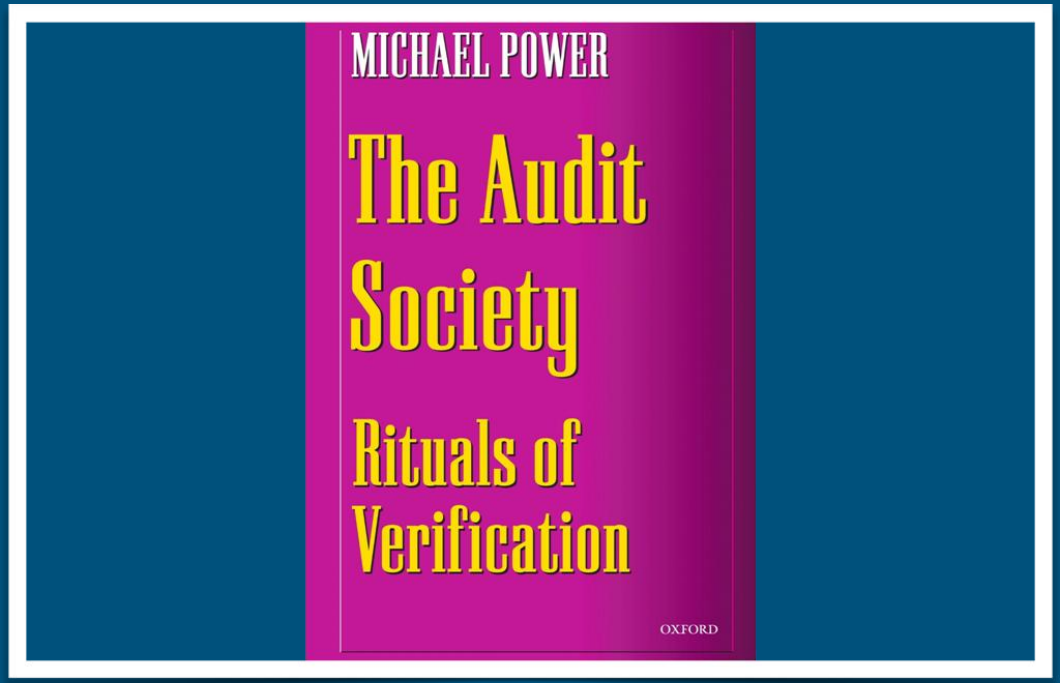
BIG DATA ANALYTICS AND INVESTIGATION

1. Connectivity creates a paradox of modern times: "The more we know or learn through connected networks, the more is known and learned about us" (Schmidt et al. 2013).
2. Surveillance is part of every bureaucratic organisation. It entails collection, processing, analysing and sharing data.
3. Surveillance is related to control and governance.
4. Big Data Analytics took Surveillance to a whole new level.

Restricted Use - À usage restreint

STRONGLY RECOMMENDED





Restricted Use - À usage restreint

DISCUSSION:

What are the risks of using Big Data Analysis, considering:

a. **your agency**
b. **tax crime investigation**
c. **the use of Big Data Analytics as evidence before the courts**

Caption

Restricted Use - À usage restreint

Emerging risks from the use of Big Data Analysis

- Security protocols for access, etc.
- Representativeness relates to whether data used provide an exhaustive representation of the population under study, with balanced representation of all relevant subpopulations.
- Data relevance involves the contribution of data used to describe the phenomenon at hand without including exogenous (misleading) information.
- Data privacy and confidentiality.
- Bias and discrimination.
- Explainability (to the general public, to taxpayers and to audit courts).
- Auditability of the algorithms. Accountability.

Restricted Use - À usage restreint

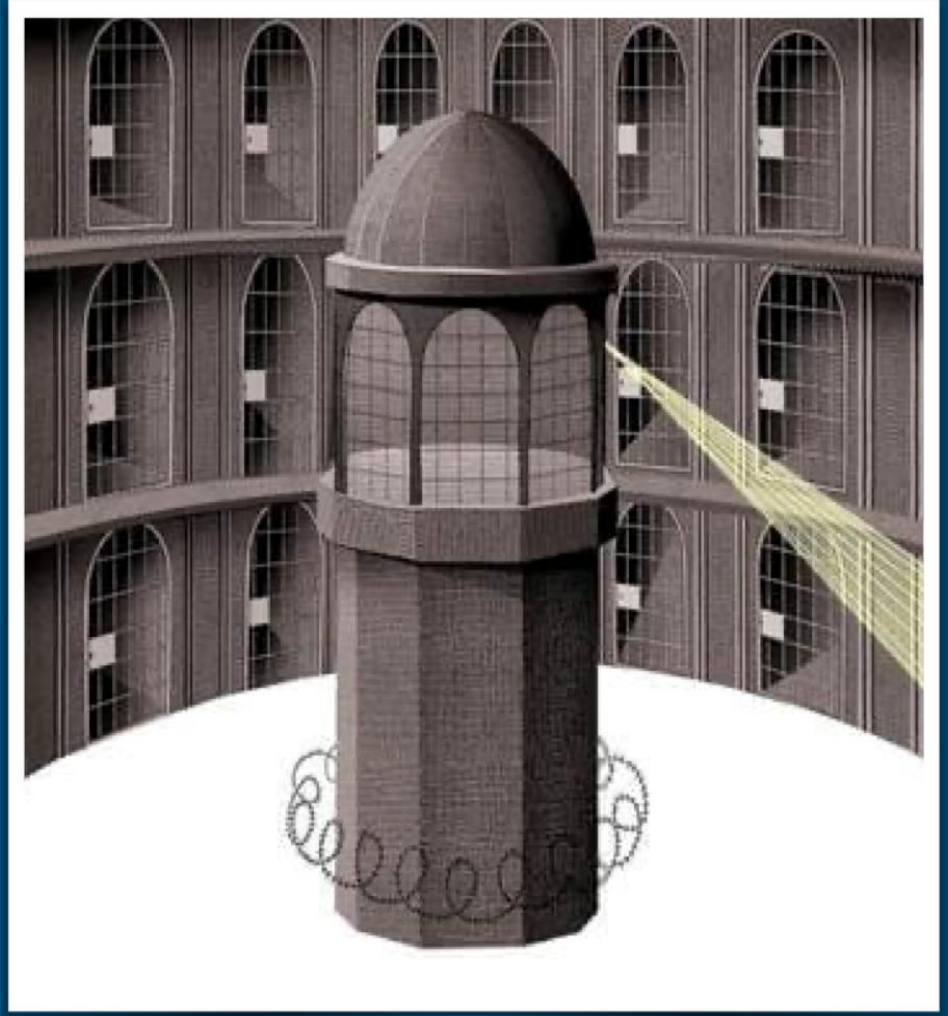
Emerging risks from the use of Big Data Analysis

- Disclosure and transparency* .
- Outsourcing, third party providers.
- Interagency exchange of information*
- Regulatory considerations: investigative powers x taxpayers rights.
- Staff: training, ownership.
- Technical issues: storage, processing capacity, the right tool, keeping up with new technologies, etc.

Restricted Use - À usage restreint

PANOPTICON

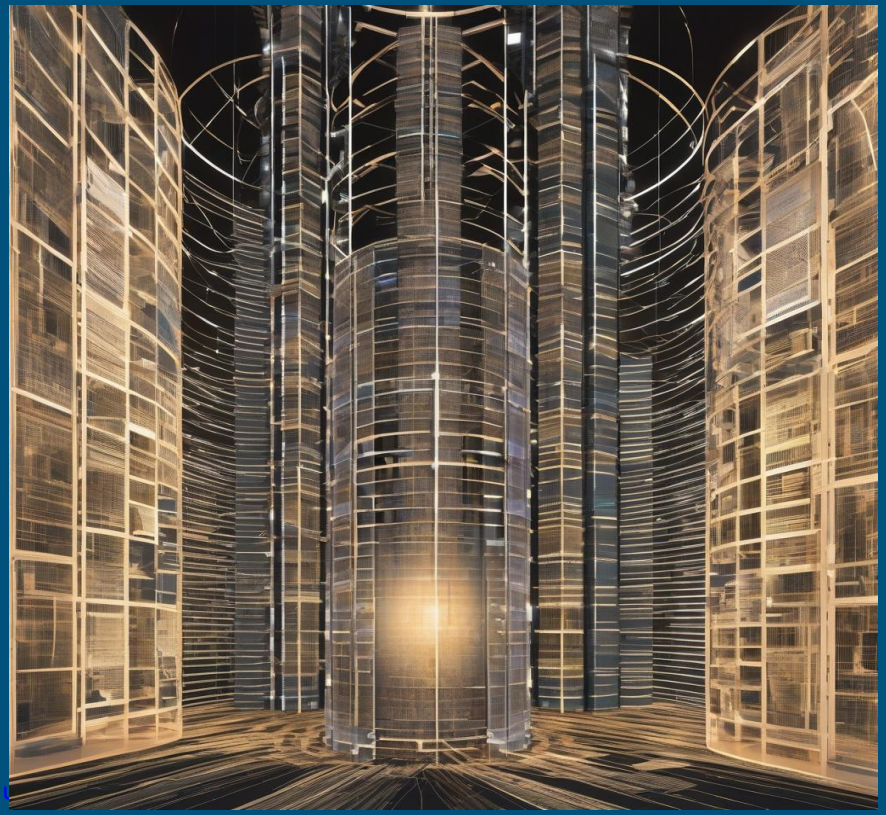
1. Bentham / Foucault
2. 24/7 Surveillance.
3. You never know when and who is monitoring, only that you're being watched.



Restricted Use - A

I - PANOPTICON

1. Doesn’t require physical presence.
2. Not just monitoring, but also collecting, processing, analysing, influencing.
3. Always on, everywhere.

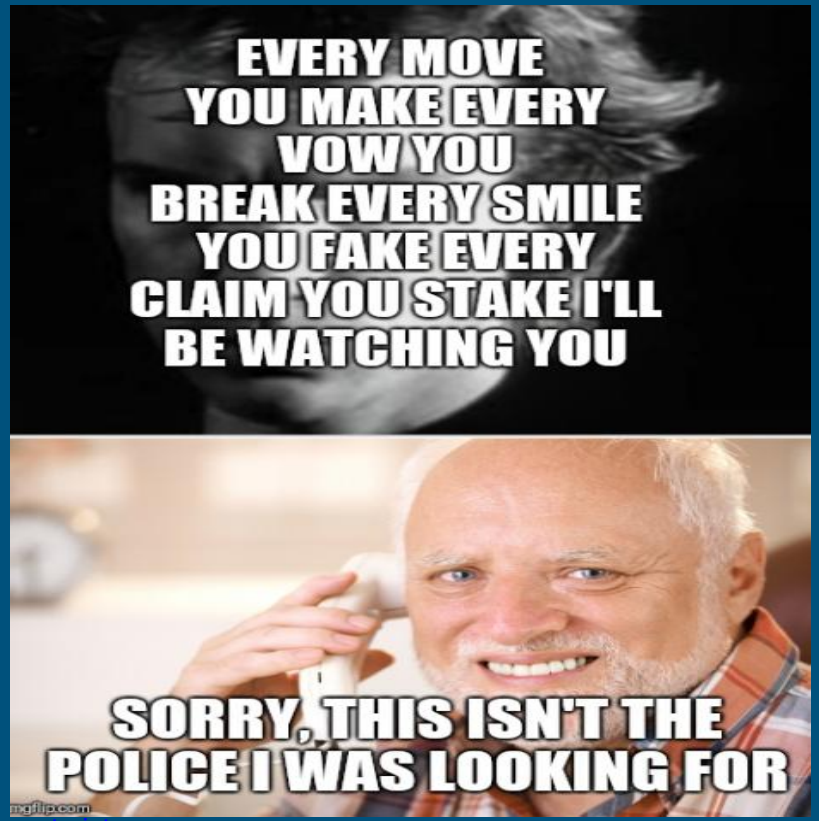


Restricted Use - À

I-PANOPTICON AND VALIDATION

The use of Big Data for surveillance must observe the legal procedures and respect taxpayers' rights if the operation is to be considered legal.

A good example is the EU GDPR - General Data Protection Regulation



Restricted Use - À usage restreint

The EU GDPR includes the following rights (Information Commissioner's Office, 2018):

- Right to be informed about surveillance
- Right of access
- Right to rectification
- Right to erasure
- Right to restrict processing
- Right to data portability
- Right to object
- Rights related to automated decision making including profiling

Restricted Use - À usage restreint

SURVEILLANCE X INDIVIDUAL RIGHTS

- Regulation : limitation surveillance or restricting civil rights?
- Taking away governmental powers of monitoring and auditing isn’t feasible, since both are inherent to government.
- These powers can be restricted in two ways: either the scope of government action is reduced, or limits are imposed on the collection, processing and sharing of data.
- The EU GDPR chose the second alternative.

Restricted Use - À usage restreint

GDPR - Recital 31

1.Public authorities to which personal data are disclosed in accordance with a legal obligation for the exercise of their official mission, such as **tax and customs authorities, financial investigation units**, independent administrative authorities, or financial market authorities responsible for the regulation and supervision of securities markets should not be regarded as recipients if they receive personal data which are necessary to carry out a particular inquiry in the general interest, in accordance with Union or Member State law.

2. The requests for disclosure sent by the public authorities should always be in writing, reasoned and occasional and should not concern the entirety of a filing system or lead to the interconnection of filing systems.

3.**The processing of personal data by those public authorities should comply with the applicable data-protection rules according to the purposes of the processing**

Restricted Use - A usage restreint

GDPR - Recital 71

1 The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention.

2 Such processing includes 'profiling' that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her.

3 However, decision-making based on such processing, including profiling, should be allowed where expressly **authorised by Union or Member State law** to which the controller is subject, including for fraud and tax-evasion monitoring (...)

GDPR - Recital 112

Those derogations should in particular apply to data transfers required and necessary for important reasons of public interest, for example in cases of international data exchange between competition authorities, tax or customs administrations, (...)

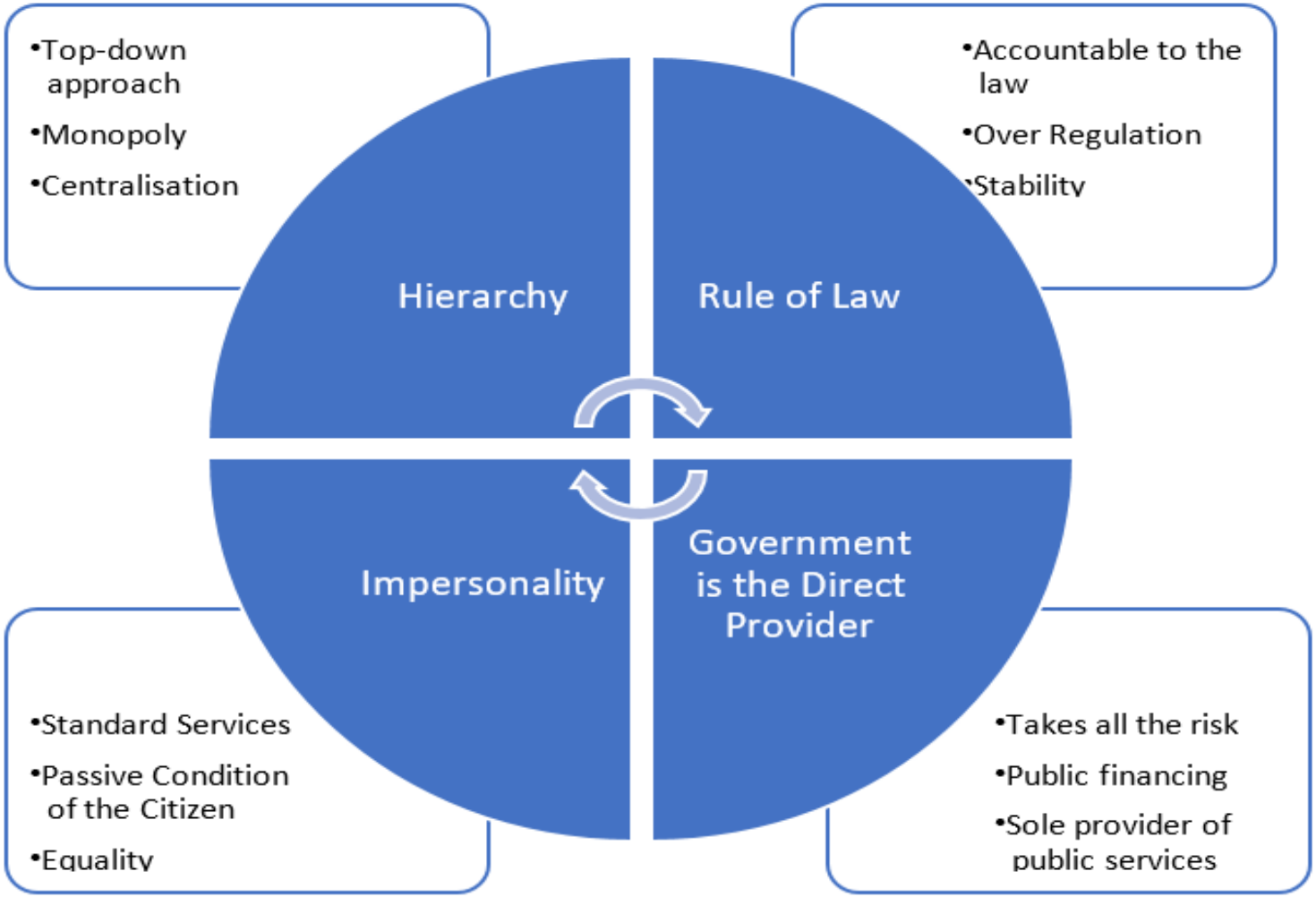
Restricted Use - À usage restreint

Management considerations

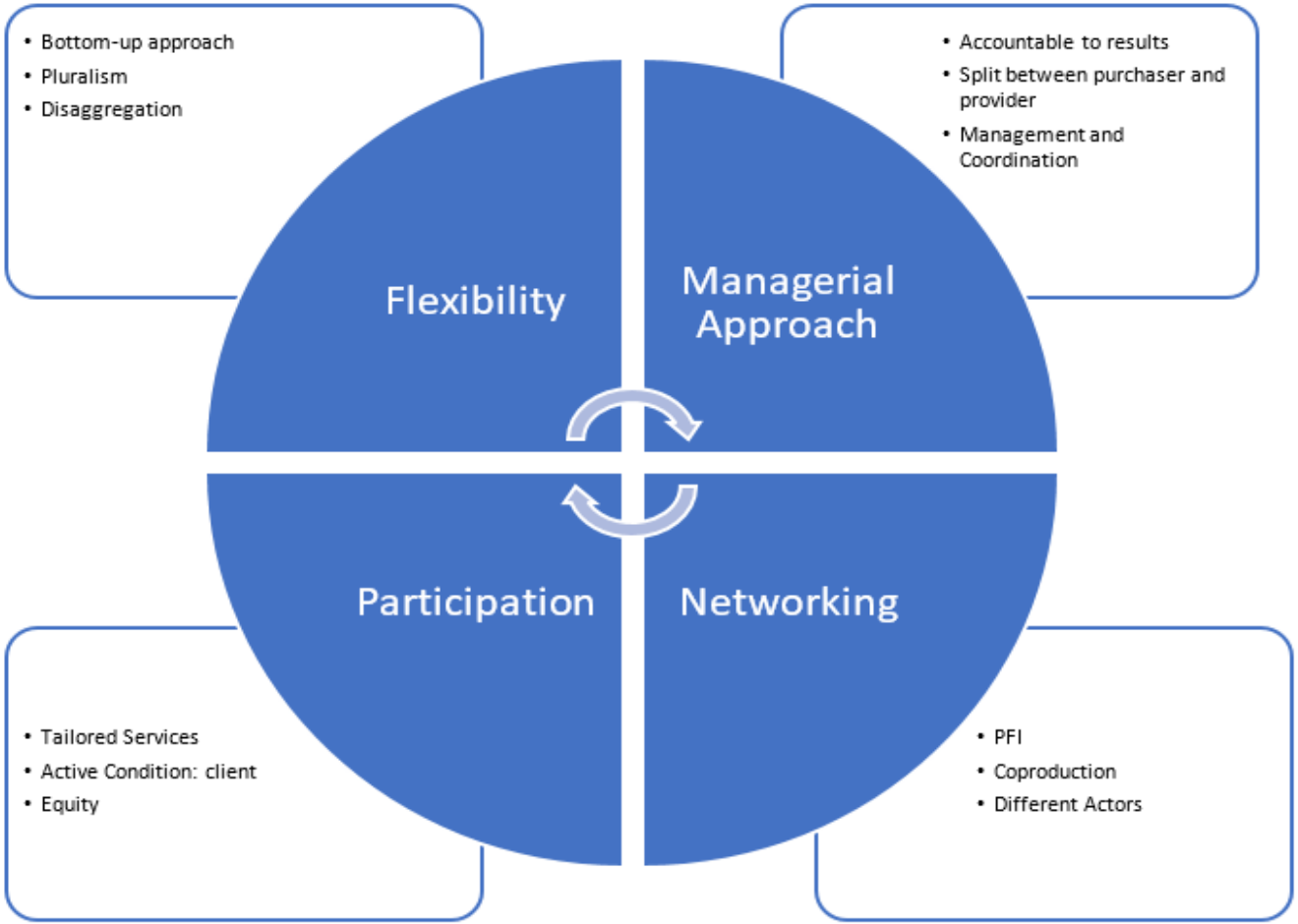
Besides the legal issues and expected costs, is your Agency ready to implement Big Data Analytics in tax investigation?

Two different models: Bureaucratic Administration model x Mix Economy Management model

Restricted Use - À usage restreint



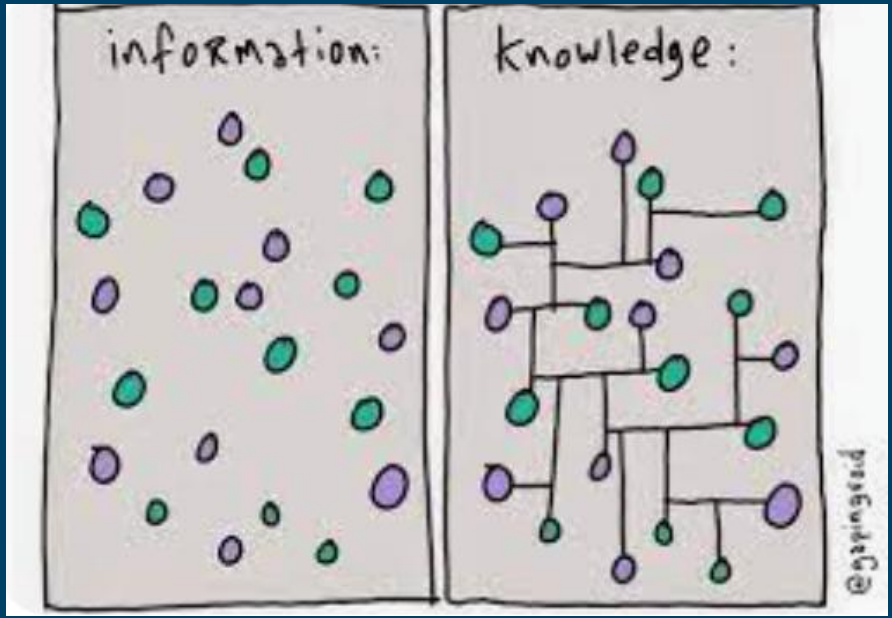
Source: own making.



Source: own making.

Connecting the dots!

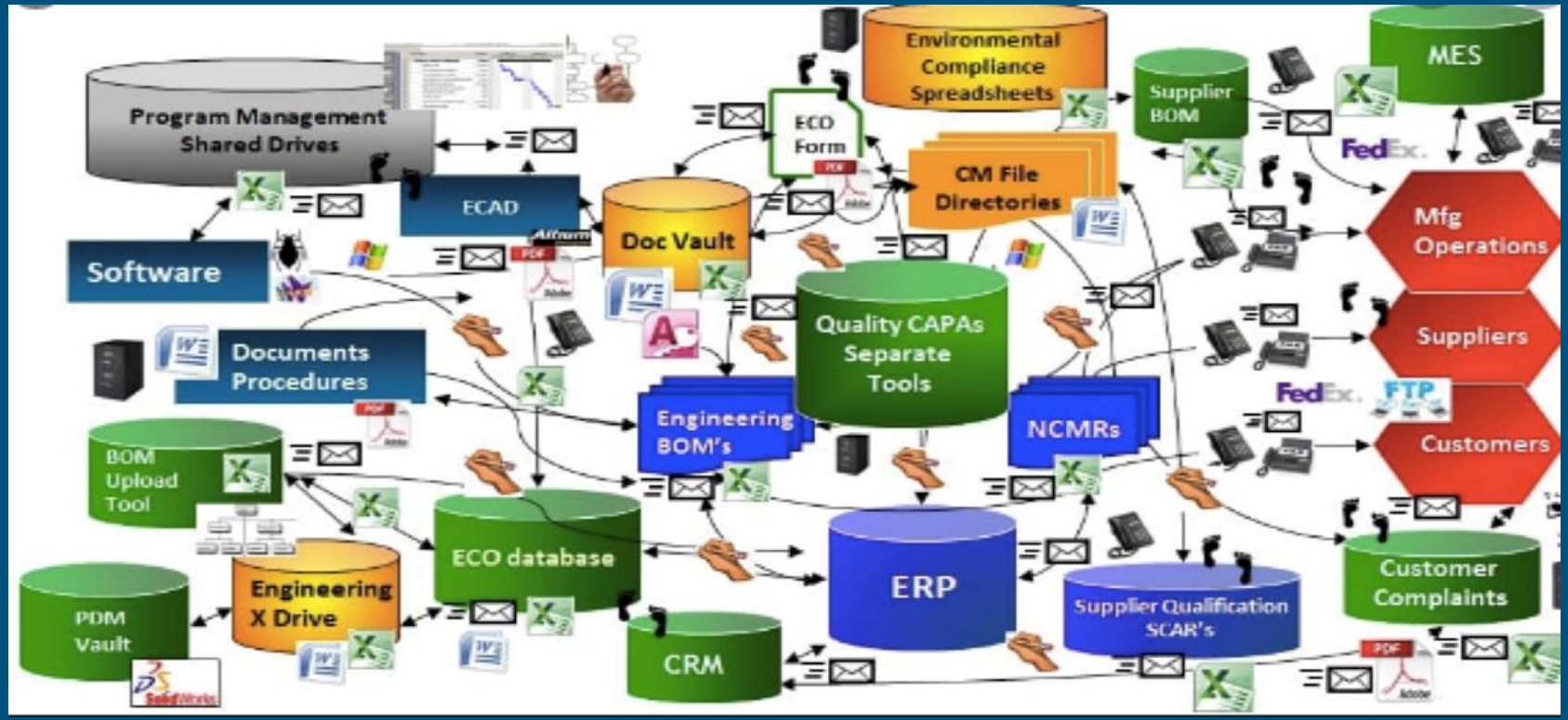
Example of massive use of Big Data Analytics: The PGFN case.



Restricted Use - A

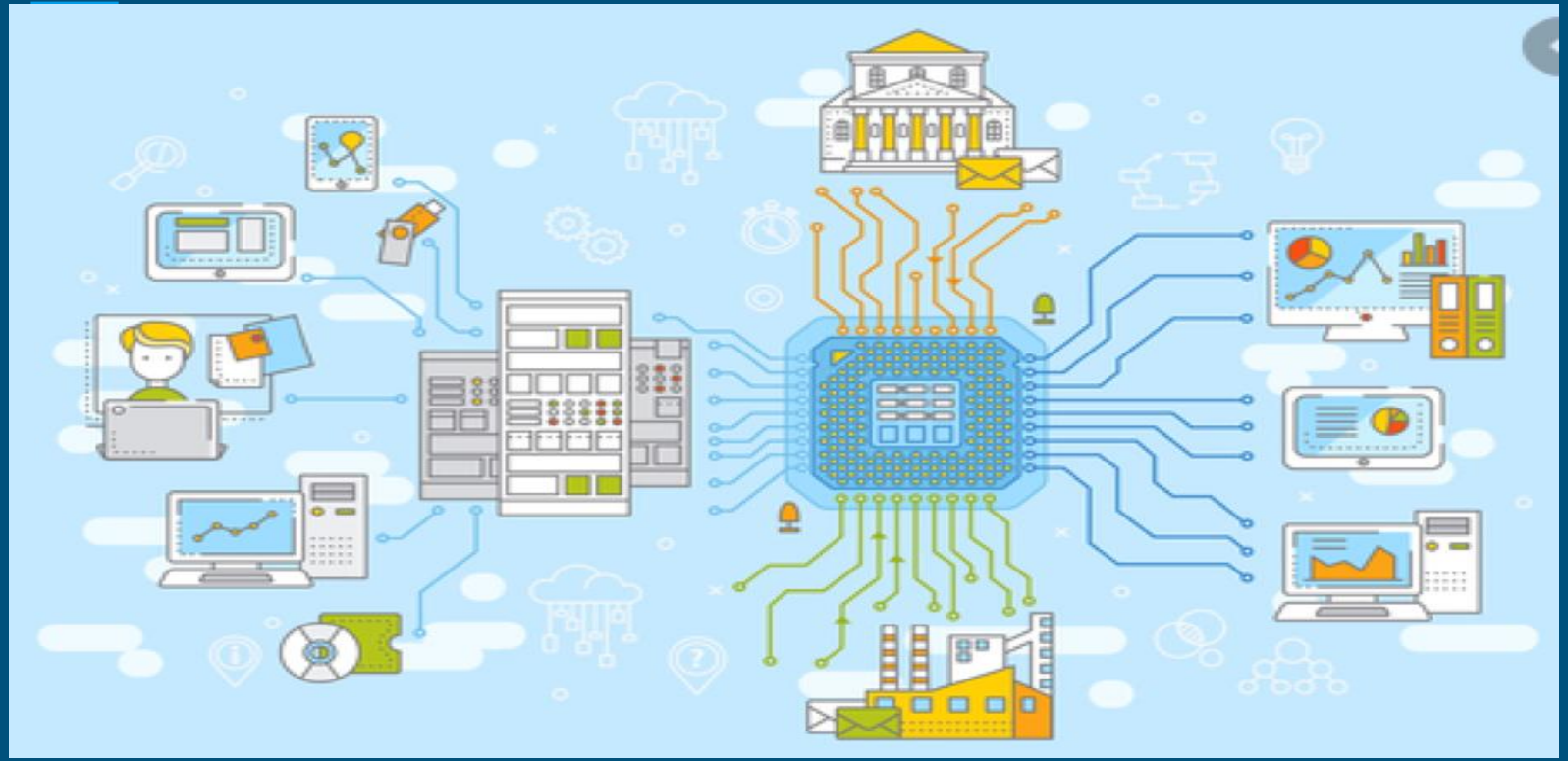
Restricted Use - À usage restreint

How it used to be



Restricted Use - À usage restreint

How it is now



Restricted Use - À usage restreint

DÍVIDA ABERTA APP

QR CODE SCAN to check e-invoices (is the company owing taxes?)

Check company's tax compliance using id number

Find local tax debtors (using georeferencing)

Whistleblower channel

Tax compliance services

PGFN news!

Restricted Use - À usage restreint

Big Data Analysis in Tax Investigation

Mining data from 80+ different databases.

Structured and semi-structured data.

Writing the algorithm to search for patterns.

Combining and analysing data to spot delinquent taxpayers.

Restricted Use - À usage restreint

TAX DEBT

Beneficial Owner's WIFE

Company committing tax fraud

Beneficial Owner

Shareholder

HOLDING

Transfers his assets to wife´s Holding

Restricted Use -

A needle in a haystack!

5 million taxpayers in debt

- 300.000 Large Taxpayers in debt
- How many of these are shell companies?

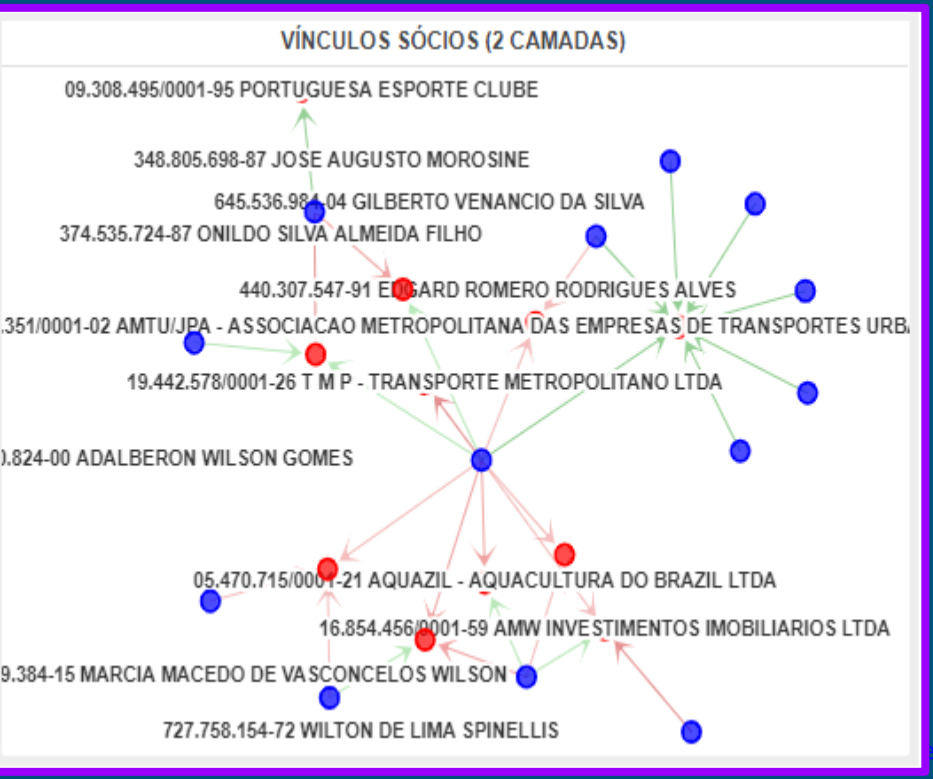
Shareholder identified

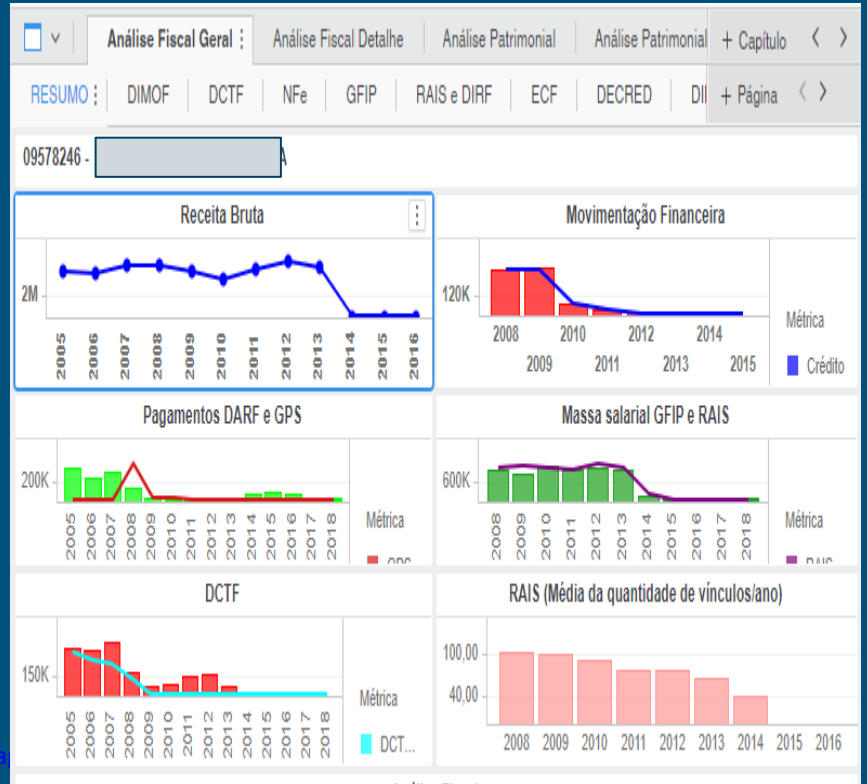
- Shareholders without assets
- How many of these are connected to a holding?

Family member owns a holding

- Spouse/kids own holding concentrating all the assets
- Had the relatives the financial means to keep the holding?

Making it visual: Visual Law.





BACK OFFICE

6 TAX INVESTIGATION UNITS
100+ PROSECUTORS

Restricted Use - À usage restreint

Getting to know PGFN Analytics

Restricted Use - À usage restreint

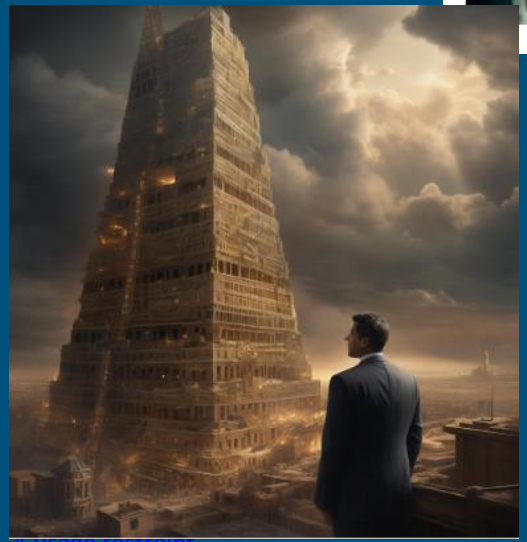
Is Big Data expensive?

- Implementation: hardware, storage system, high speed internet, processing capacity.
- Training the staff: post-graduation on Big Data Analysis, Data Science, courses on AI etc.
- Costs of maintenance.
- Becoming digital, Digital Governance.
- Reaching for databases

Restricted Use - À usage restreint

What if the "system" got it wrong?

1. Can the algorithm be held accountable?
2. Using digital evidence as supporting evidence, never as the sole basis for conviction.
3. Bad Data: error when collecting, sampling, storing.
4. Spurious relations: poor interpretation.
5. Babel Tower infatuation.



Restricted Use - A usage restreint

Fraud or just divorce?

System programmed to find *asset transfer fraud* : to simulate a legal transfer of assets to a trusted party (usually family member) to avoid lien, confiscation etc.

Target located: Mister Love, owing $500,000 in taxes, transferred $2 million worth flats to wife.

However, further investigation revealed that he got divorced and his former wife got the flats in the deal.

Restricted Use - À usage restreint

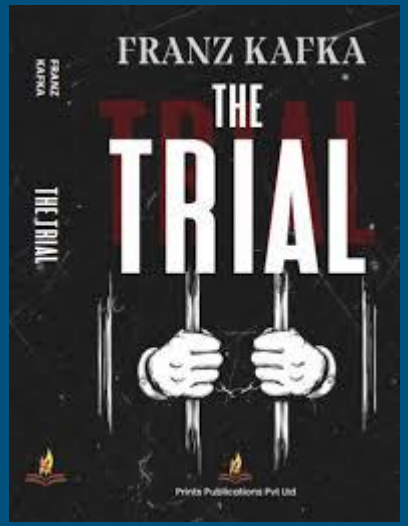
Fraudulent corporate group?

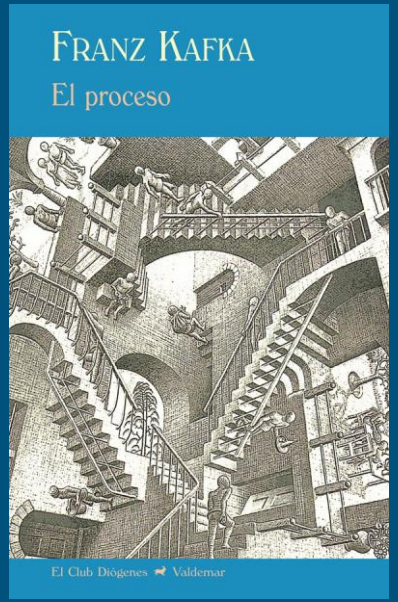
According to BDA report, 5 different agribusiness companies were operating in the same place (small room in an old building in town). They all had the same accountant, who registered the companies statutes in the local Commercial Bureau. Each company owed around $200,000 in taxes. The accountant might be the enabler of a fraudulent scheme. 5 big companies functioning in the same small room? And they have the same account? Highly suspicious…

Restricted Use - À usage restreint

The Post Office Scandal

- A story of how overconfidence in data analysis ruined almost 900 lives: "The SYSTEM can do no wrong" !!
- Algorithm and accountability.
- The weight of electronic processed data as judicial evidence.
- Explainability: do judges and lawyers really understand big data?
- Big Data Analytics and credibility: transparency, open data, taxpayers rights.





Restricted Use - À usage restreint

BBC

Home News Sport Business Innovation Culture Arts Travel Earth Audio Video Live

What is the Post Office Horizon scandal?

More than 900 sub-postmasters were prosecuted for stealing because of incorrect information from the Horizon computer system. It has been called the UK's most widespread miscarriage of justice.

The Post Office itself took many cases to court, prosecuting 700 people between 1999 and 2015. Another 283 cases were brought by other bodies, including the Crown Prosecution Service (CPS).

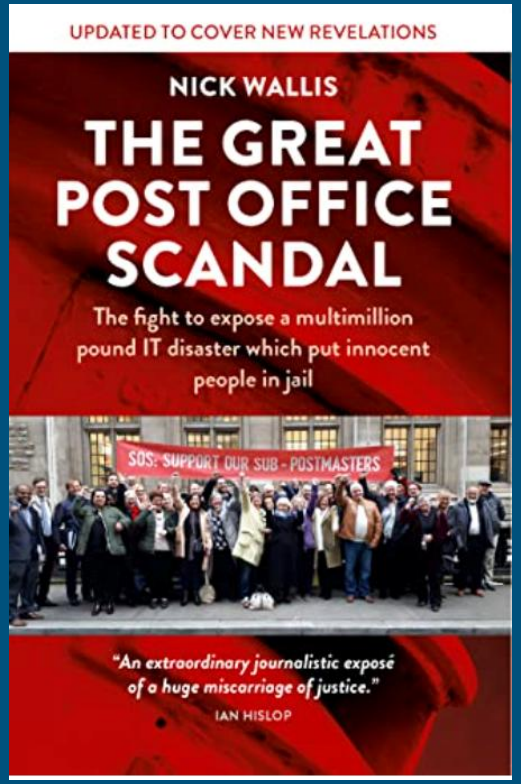
Many sub-postmasters went to prison for false accounting and theft, and many were financially ruined.

In 2017, a group of 555 sub-postmasters took legal action against the Post Office. In 2019, **it agreed to pay them £58m in compensation**, but much of the money went on legal fees.

The Metropolitan Police **is investigating the Post Office over potential fraud offences.**

Restricted Use - A usage restreint

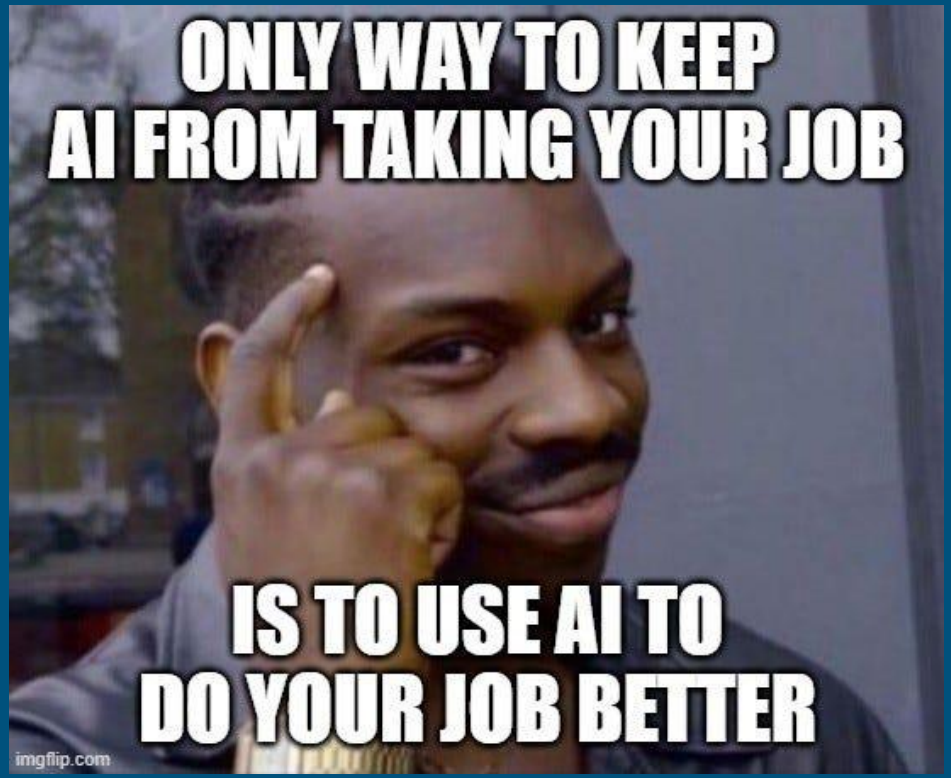
Recommendation!



Restricted Use - À usage restreint

What about AI ?

1. AI and big data are complementary technologies that work together to unlock the potential of data.
2. AI provides the intelligence to analyze and interpret big data, while big data provides the fuel for AI to learn and improve.
3. The combination of AI and big data is driving innovation and transforming both private and public sector.



Restricted Use - À usage restreint

Figure 9

Top Technologies to Accelerate Investigations

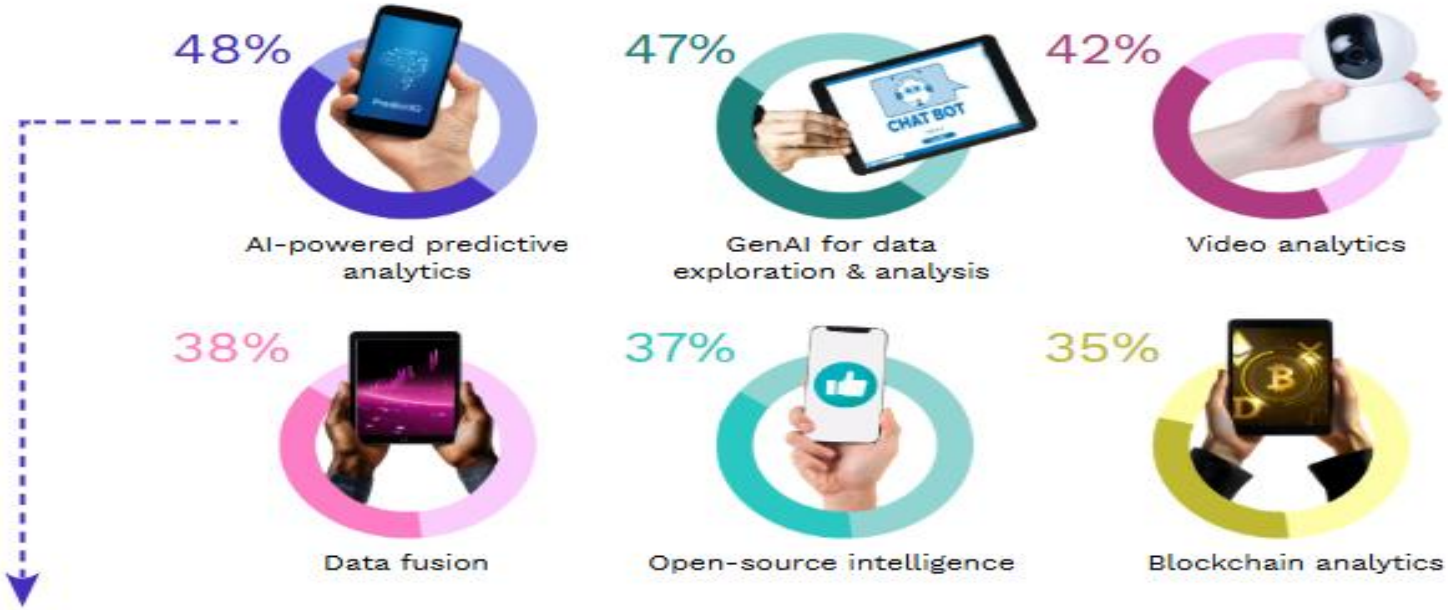


Figure 10

2025 Law Enforcement Outlook:
https://engage.cognyte.com/s/e3b95dd0/?page=14

Restricted Use - À usage restreint

AI Techniques for Tax Crime Investigation

Machine Learning:

- Supervised Learning: Training models on labeled data to predict outcomes (e.g., identifying fraudulent tax returns).
- Unsupervised Learning: Discovering patterns in unlabeled data (e.g., clustering taxpayers with similar behaviors).

Deep Learning:

- Neural Networks: Complex algorithms that can learn from large datasets (e.g., analyzing unstructured data like emails or documents).

Natural Language Processing (NLP):

- Text Analysis: Extracting information from textual data (e.g., identifying keywords related to tax evasion in documents).

Restricted Use - À usage restreint

AI Applications in Tax Crime Investigation

Risk Assessment:

- Predictive Modeling: Identifying high-risk taxpayers based on historical data and behavioral patterns.
- Anomaly Detection: Identifying unusual activities or transactions that may indicate fraudulent behavior.

Data Analysis and Visualization:

- Data Mining: Discovering hidden patterns and relationships within large datasets.
- Data Visualization: Presenting complex data in a clear and understandable manner.

Audit Selection and Prioritization:

- Risk-Based Auditing: Focusing on taxpayers with the highest likelihood of non-compliance.
- Intelligent Audit Scheduling: Optimizing audit resources based on risk and workload.

Fraud Detection:

- Pattern Recognition: Identifying fraudulent schemes and patterns of behavior.
- Real-time Monitoring: Detecting fraudulent activities as they occur.

Money Laundering Investigation:

- Network Analysis: Analyzing financial transactions to identify suspicious networks.
- Behavioral Analysis: Identifying unusual spending patterns or asset transfers.

Restricted Use - À usage restreint

What PGFN has been doing with AI + Big Data

RAPTOR

FERRAMENTAS X DAU

TRADER

SNIPER

COBRA

comprei

RAIA

PUMA

SISPAR

PANDORA

SIDAT

CADIN

PGFN ANALYTICS

INTELIGÊNCIA ANALÍTICA APLICADA À RECUPERAÇÃO DE CRÉDITOS

PGFN DATA

DÍVID ABERT

REGULARIZE

FERRAMENTAS X FRAUDE

BROKEN

SPOILER

RATING

CAPAG

INSCREVE FÁCIL

FLEXA

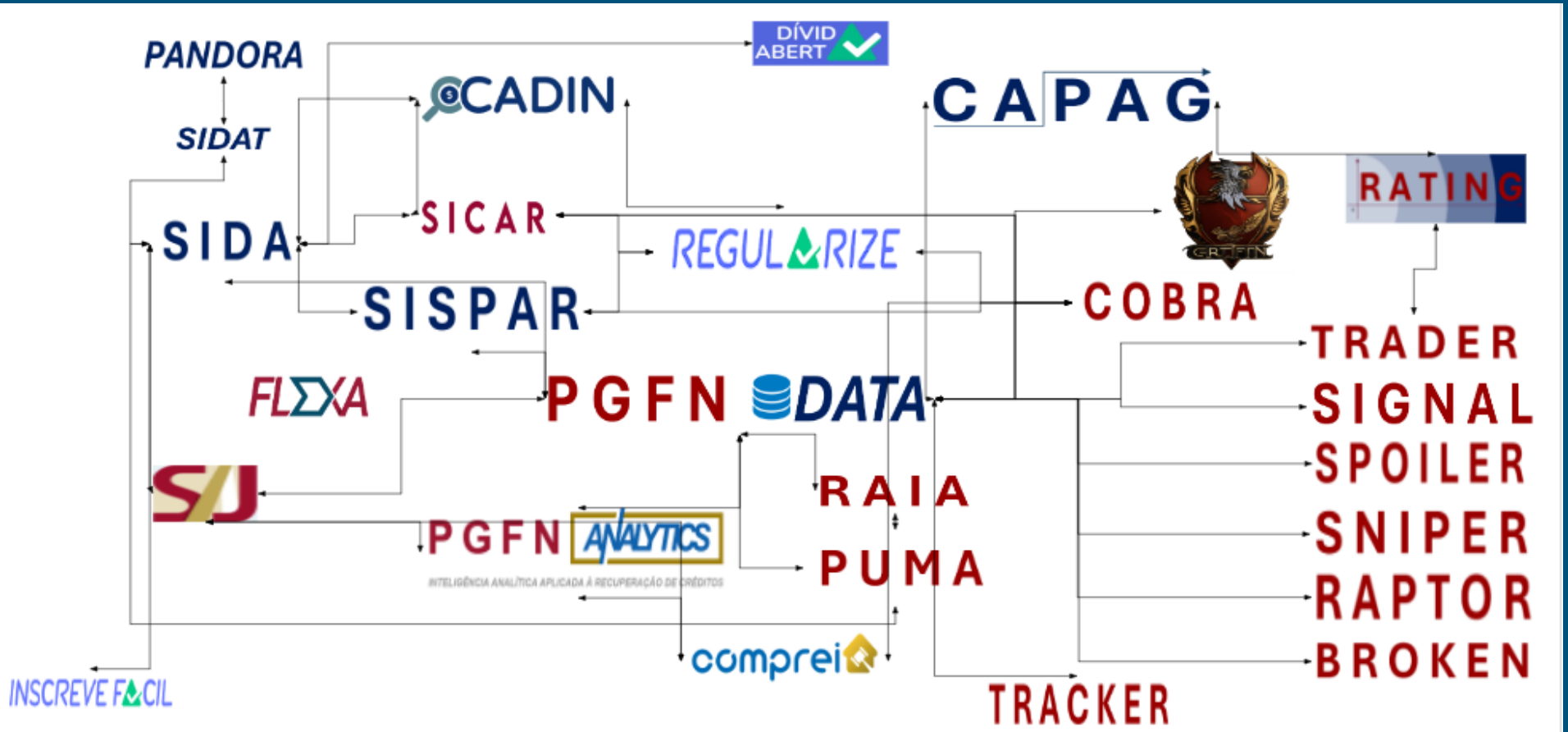
TRACKER

SICAR

SIGNAL

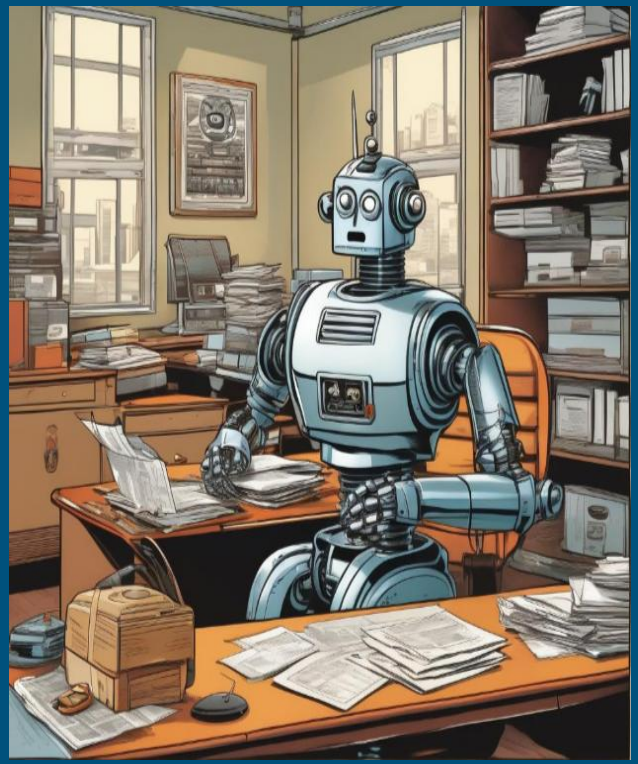
SIDA

Systems and Apps “converse” with each other



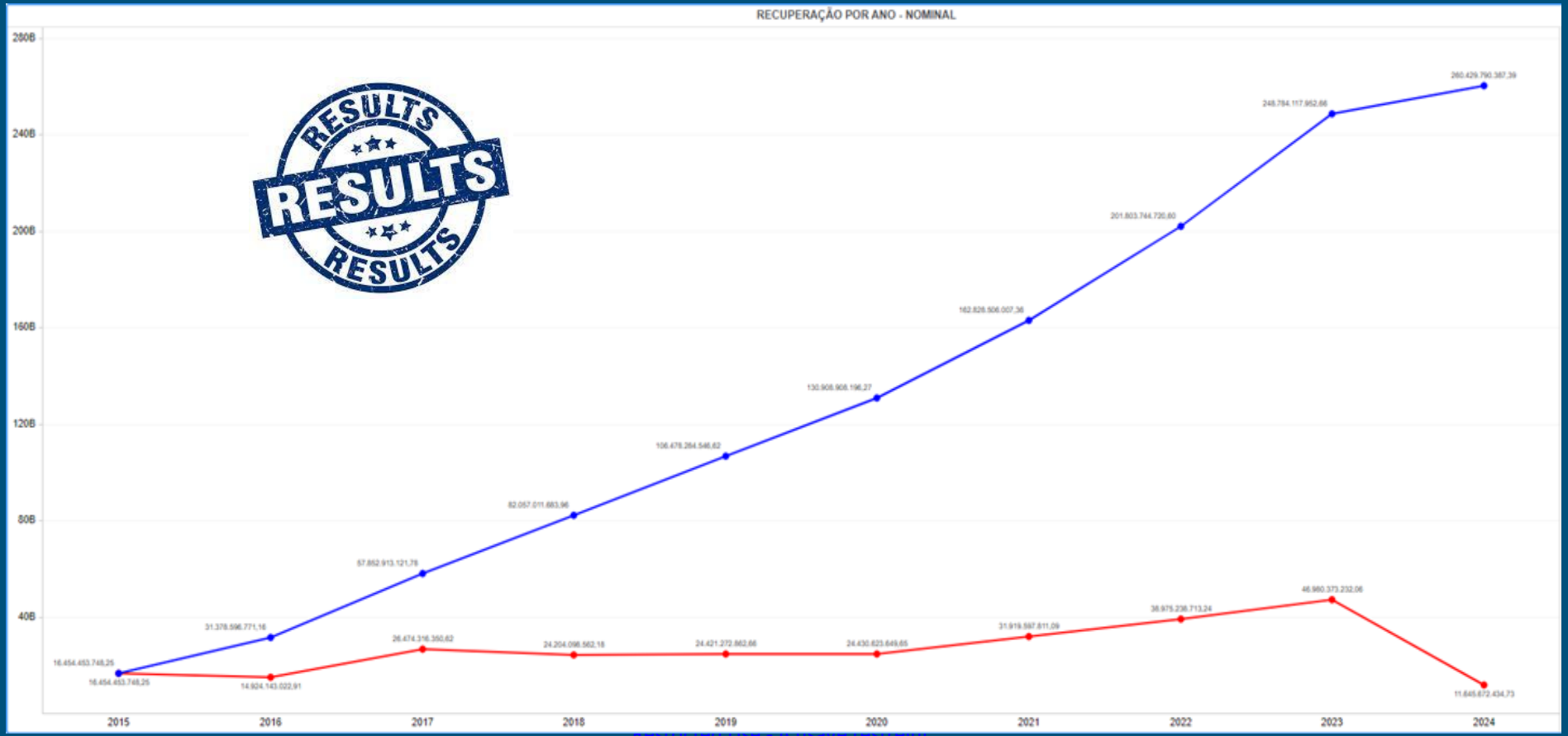
INTEGRATION AND ASSET RECOVERY

- AI linked PGFN Data, FLEXA, PGFN Analytics, Integra Service e SAJ, boosting SPOILER and PGFN Analytics. (April, 2025)
- PGFN Analytics searches for assets (of delinquent taxpayers) 24/7
- Whenever located, the system read all the judicial cases (tax foreclosure*) and chooses the best one to file a motion detailing the asset and asking court order to freeze.
- In the first day it was fully operational: 310 motions filed to freeze thousands of bank accounts; 347 motions filed to freeze hundreds of vehicles.



Restricted Use - À usage restreint

Results!!!



Restricted Use - A usage restreint

The Palantír Problem

- When using a Palantír to gaze further abroad, you don’t know who else is looking (privacy issues)
- The Palantír may show only what you want to see or what someone wants you to see (bias / interference/ behavioural engineering)
- You may see things and situations that do not exist but are presented as real (hallucination)

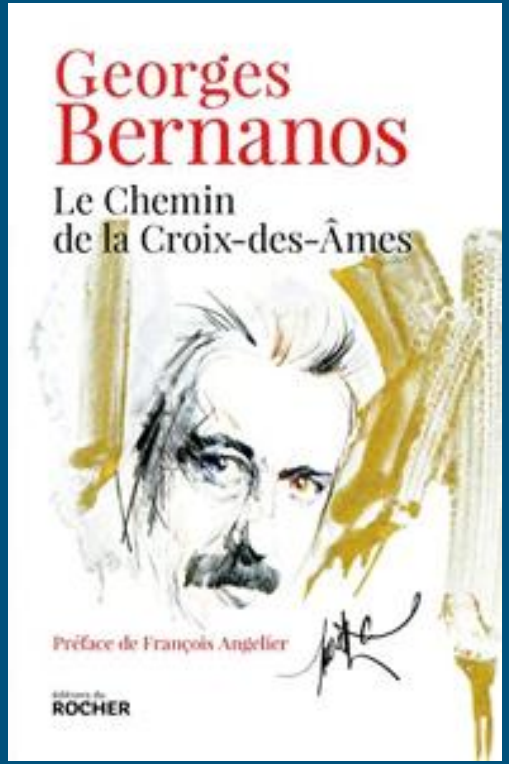
Restricted Use - À usage restreint

ChatPGFN

Restricted Use - À usage restreint

Tax crime investigators and AI

"The danger is not in the multiplication of machines, but in the ever-increasing number of men accustomed from their childhood to desire only what machines can give."



Restricted Use - À usage restreint

Activity 2.

You were hired as Tax and Big Data consultant to advise Digitopolis City Council on how to improve data collection of City Council tax and VAT, to spot potential tax offences concerning evasion and to improve asset recovery. In the city of Digitopolis e-government is a trend, so the following areas within the government are 100% digital:

- Citizens, business, and properties (e.g. base registries, transactions)
- transactions between taxpayers and the government, including invoices

Restricted Use - À usage restreint

Activity 2

- Public procurement and public expenses
- Public bodies and employees
- Registry of companies and employees
- Geographical data mainly related to cadastral
- Geospatial data
- Content related to traffic and vehicles ownership
- Real estate ownership and other rights over immovable property

Restricted Use - À usage restreint

Activity 2

Advise the City Council on how Big Data/ AI could be used to:

1.Improve City Council Tax collection:

2. Improve VAT collection:

3. Spot tax offences:

4. Improve asset recovery:

Breakout room. Duration: 15 minutes. Groups of 5 or 10 participants. Choose a spokesperson to tell us the main ideas of the group.

Restricted Use - À usage restreint

Discussion and Wrap up

1. Big Data Analytics/AI and Tax Investigation.

1. Making connections: prof.acmoura@gmail.com / Instagram: @arthurleeds

Restricted Use - À usage restreint

Arthur Moura

Restricted Use - A