BIG DATA ANALYSIS

THEORY AND PRACTICE FOR TAX INVESTIGATION

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

Thoughts and Opinions ...

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

Join at menti.com | use code 6981 8458

Short questions, Big Data

Please answer this quick survey about Big Data.

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.

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

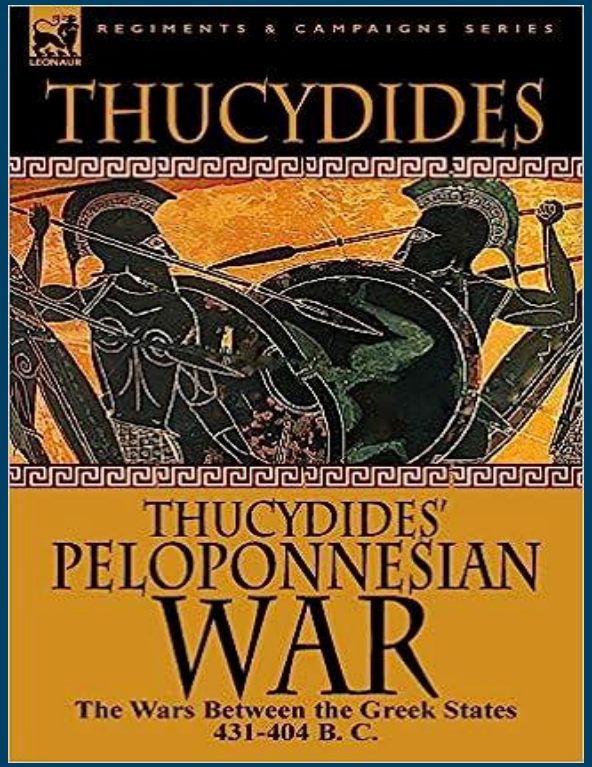
By the end of this module

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

A bit of history now

Data collection, data protection, cleaning the data, combining data, and finally analysing: helping decision making.



Data Analysis and Tax Administration

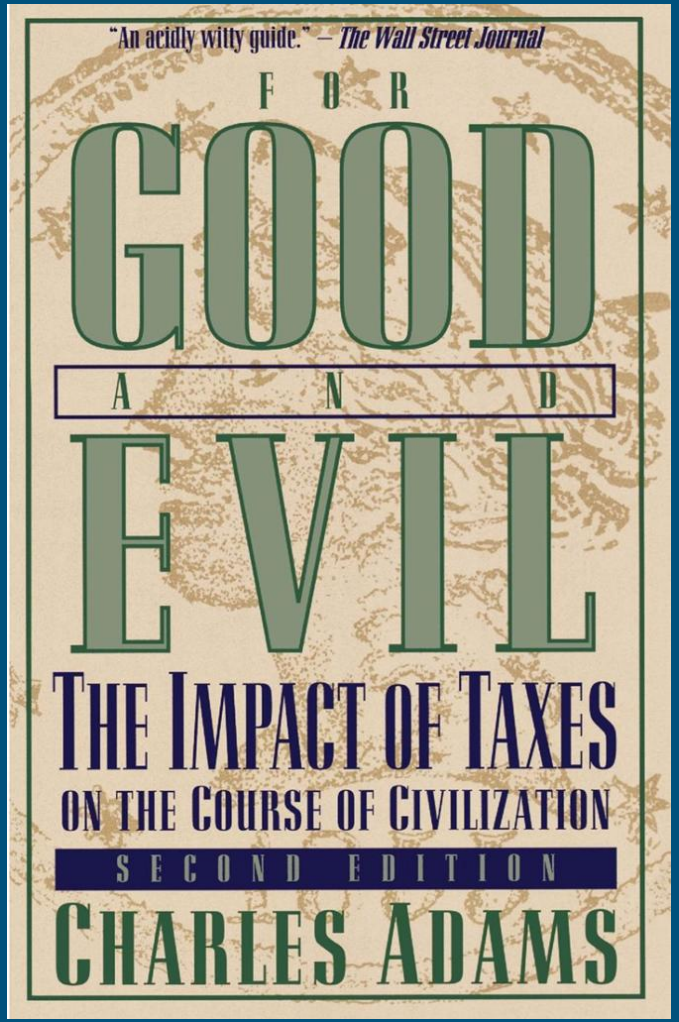
For Good and Evil: The Impact of Taxes on the Course of Civilization (Series..

The tax officials of the pharaoh—everywhere snooping, inspecting, recording, and arresting—even surveyed the nests of pigeons to count the eggs, making sure the pharaoh got his 20-percent cut.

For Good and Evil: The Impact of Taxes on the Course of Civilization (Series..

Painting from a Theban tomb of a "Scribe keeping account of the corn of Amun," from tax-immune temple lands of the priesthood. Here corn is being registered along with a flock of geese, as taxes for the priesthood. It

Recommendation!



DATA ANALYSIS

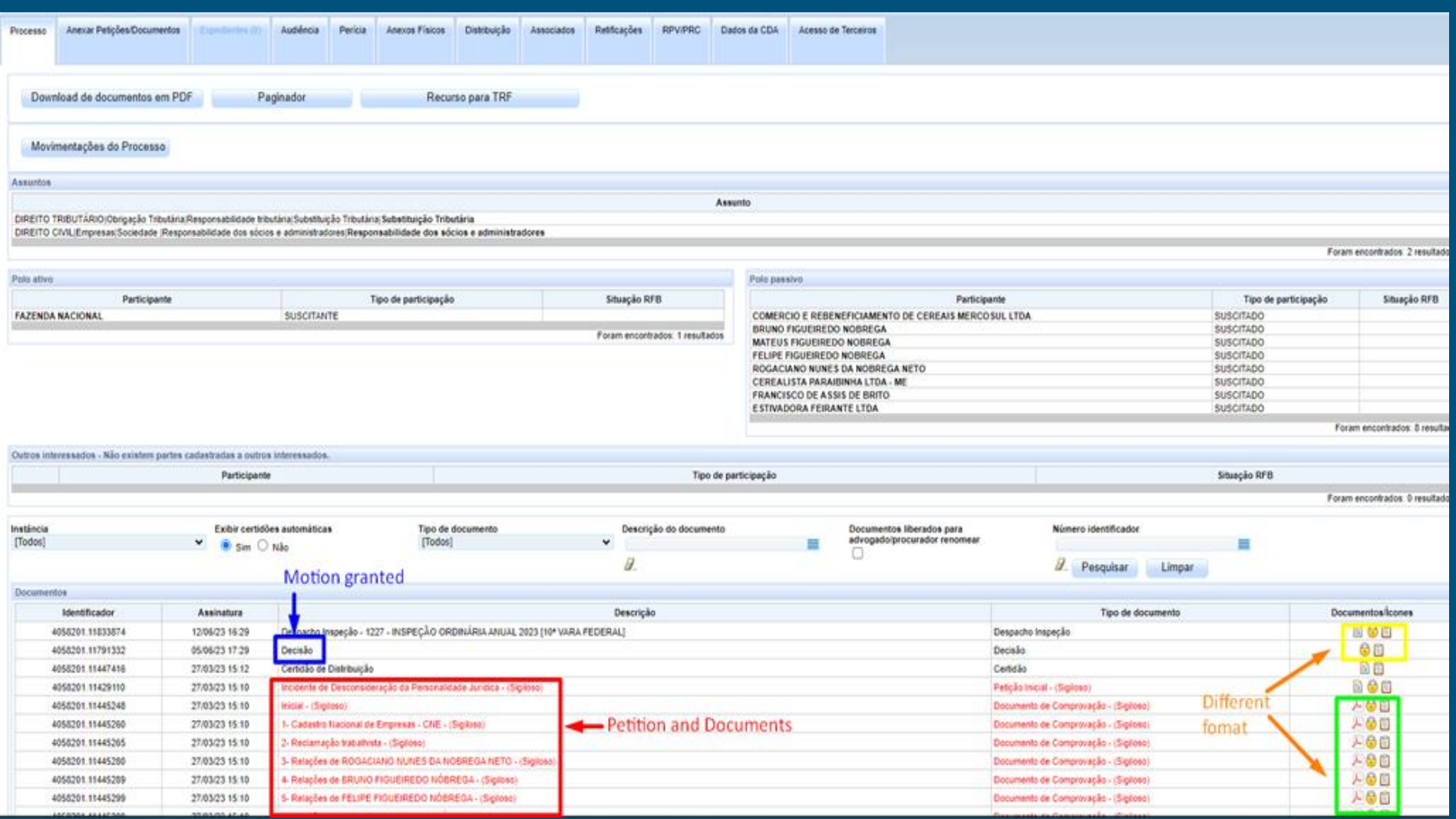
The process of collecting, cleaning, combining and formatting data in order to extract useful (valuable) information and insights to help decision making.

- **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.**

Why use Data Analysis?

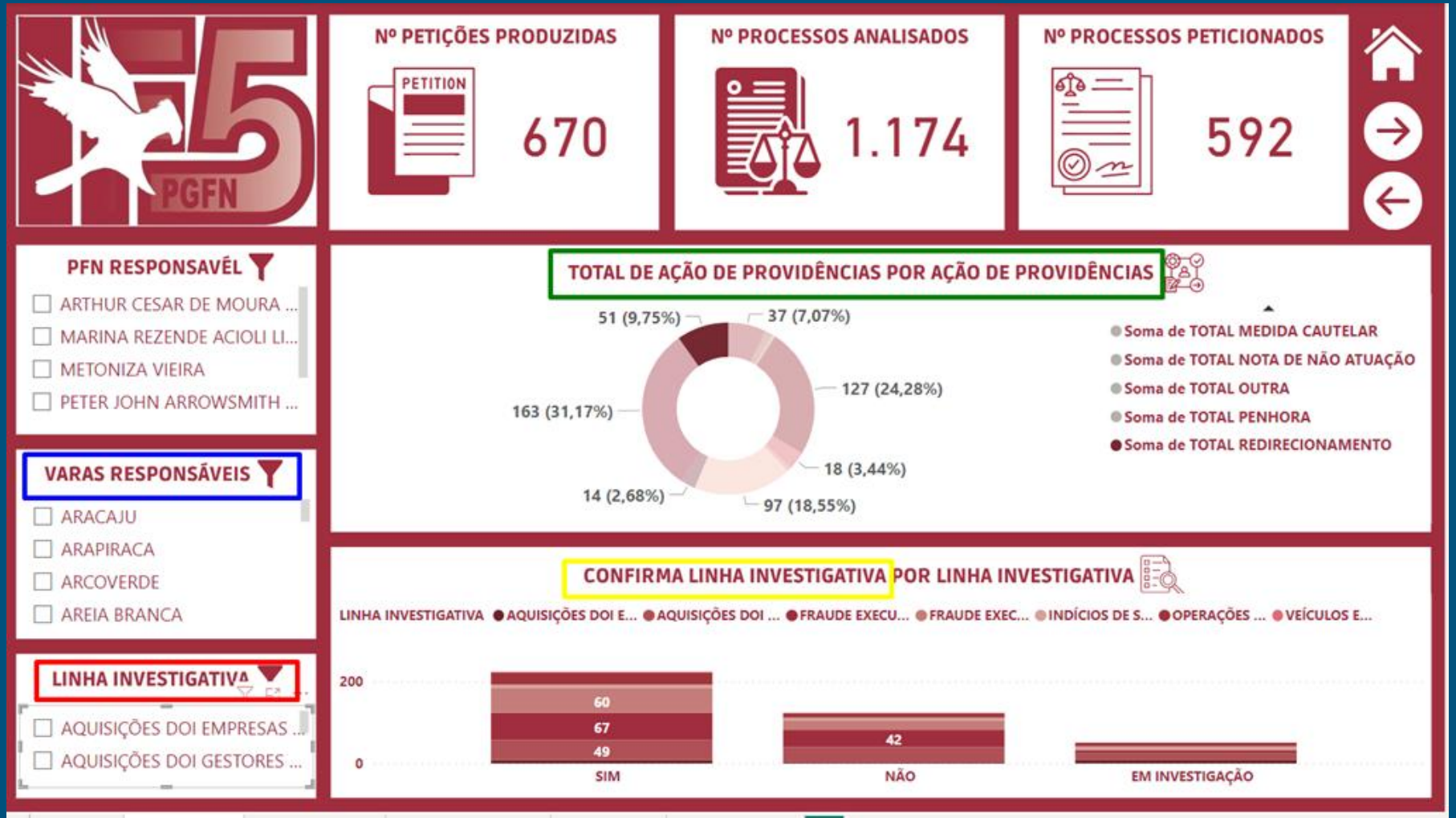
- Data analysis provides valuable insights and information that can help individuals and organizations make better decisions.
- Problem-solving: Data analysis helps identify problems.
- Cost-saving: Data analysis can help identify inefficiencies and waste in processes.
- Innovation: Data analysis can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

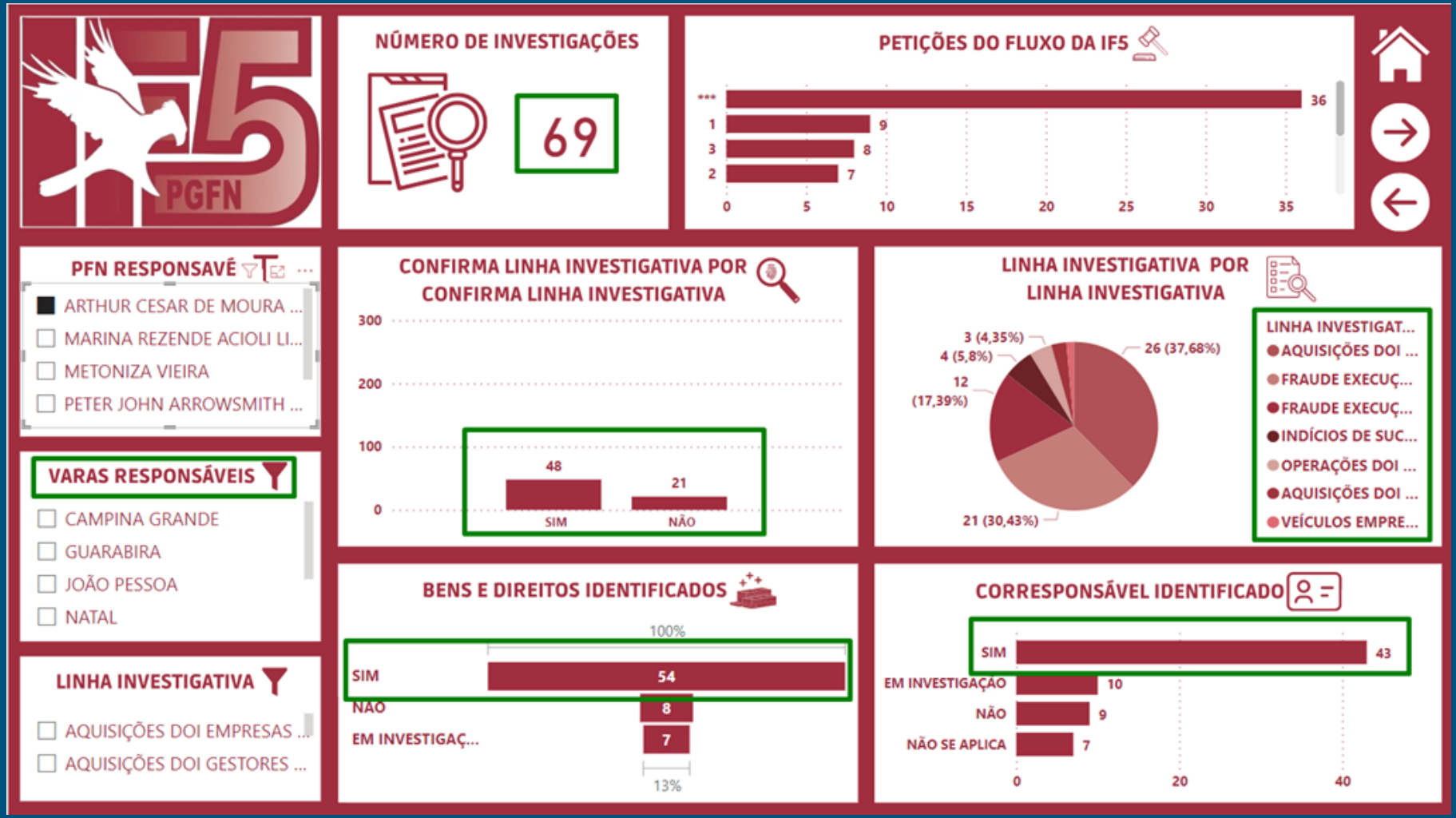
E	F	G	H	I	J	K	L	M	N	O	P
Origem do Devedor	Valor consolidado do débito	Nome do PFN Responsável	Linha Investigativa	Confirma linha investigativa?	Providências I	Providências II	Providências III	Corresponsável identificado?	Redirecionamento/IDPJ deferido?	Bens e Direitos Identificados	Bens e Direitos Penhorados?
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	
		Peter John Arrowsmith Cook	AQUISIÇÕES DOI GESTORES		NOTA DE NÃO						


Processo
Anexar Petições/Documentos
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
Motion granted
Documentos
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)
4058201.11445266
27/03/23 15:10
3- Relações de ROGACIANO NUNES DA NOBREGA NETO - (Sigiloso)
Documento de Comprovação - (Sigiloso)
4058201.11445269
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


Nº PETIÇÕES PRODUZIDAS
670
Nº PROCESSOS ANALISADOS
1.174
Nº PROCESSOS PETICIONADOS
592
PFN RESPONSAVÉL
ARTHUR CESAR DE MOURA ...
MARINA REZENDE ACIOLI LI...
METONIZA VIEIRA
PETER JOHN ARROWSMITH ...
VARAS RESPONSÁVEIS
ARACAJU
ARAPIRACA
ARCOVERDE
AREIA BRANCA
LINHA INVESTIGATIVA
AQUISIÇÕES DOI EMPRESAS ...
AQUISIÇÕES DOI GESTORES ...
TOTAL DE AÇÃO DE PROVIDÊNCIAS POR AÇÃO DE PROVIDÊNCIAS
51 (9,75%)
37 (7,07%)
127 (24,28%)
163 (31,17%)
18 (3,44%)
14 (2,68%)
97 (18,55%)
Soma de TOTAL MEDIDA CAUTELAR
Soma de TOTAL NOTA DE NÃO ATUAÇÃO
Soma de TOTAL OUTRA
Soma de TOTAL PENHORA
Soma de TOTAL REDIRECIONAMENTO
CONFIRMA LINHA INVESTIGATIVA POR LINHA INVESTIGATIVA
LINHA INVESTIGATIVA
AQUISIÇÕES DOI E...
AQUISIÇÕES DOI ...
FRAUDE EXECU...
FRAUDE EXEC...
INDÍCIOS DE S...
OPERAÇÕES ...
VEÍCULOS E...
200
0
60
67
49
42
SIM
NÃO
EM INVESTIGAÇÃO


NÚMERO DE INVESTIGAÇÕES
69
PETIÇÕES DO FLUXO DA IF5
***
1
3
2
36
9
8
7
0
5
10
15
20
25
30
35
PFN RESPONSAVÉ
ARTHUR CESAR DE MOURA ...
MARINA REZENDE ACIOLI LI...
METONIZA VIEIRA
PETER JOHN ARROWSMITH ...
CONFIRMA LINHA INVESTIGATIVA POR CONFIRMA LINHA INVESTIGATIVA
300
200
100
0
48
21
SIM
NÃO
LINHA INVESTIGATIVA POR LINHA INVESTIGATIVA
3 (4,35%)
4 (5,8%)
12 (17,39%)
26 (37,68%)
21 (30,43%)
LINHA INVESTIGAT...
AQUISIÇÕES DOI ...
FRAUDE EXECUÇ...
FRAUDE EXECUÇ...
INDÍCIOS DE SUC...
OPERAÇÕES DOI ...
AQUISIÇÕES DOI ...
VEÍCULOS EMPRE...
VARAS RESPONSÁVEIS
CAMPINA GRANDE
GUARABIRA
JOÃO PESSOA
NATAL
BENS E DIREITOS IDENTIFICADOS
100%
SIM
54
NÃO
8
EM INVESTIGAÇ...
7
13%
CORRESPONSÁVEL IDENTIFICADO
SIM
43
EM INVESTIGAÇÃO
10
NÃO
9
NÃO SE APLICA
7
0
20
40
LINHA INVESTIGATIVA
AQUISIÇÕES DOI EMPRESAS ...
AQUISIÇÕES DOI GESTORES ...

Why use Data Analysis?

- Problem-solving: Corporate group fraud 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%.

Outcome:

Interpretation and decision-making: 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.

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).

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.

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.

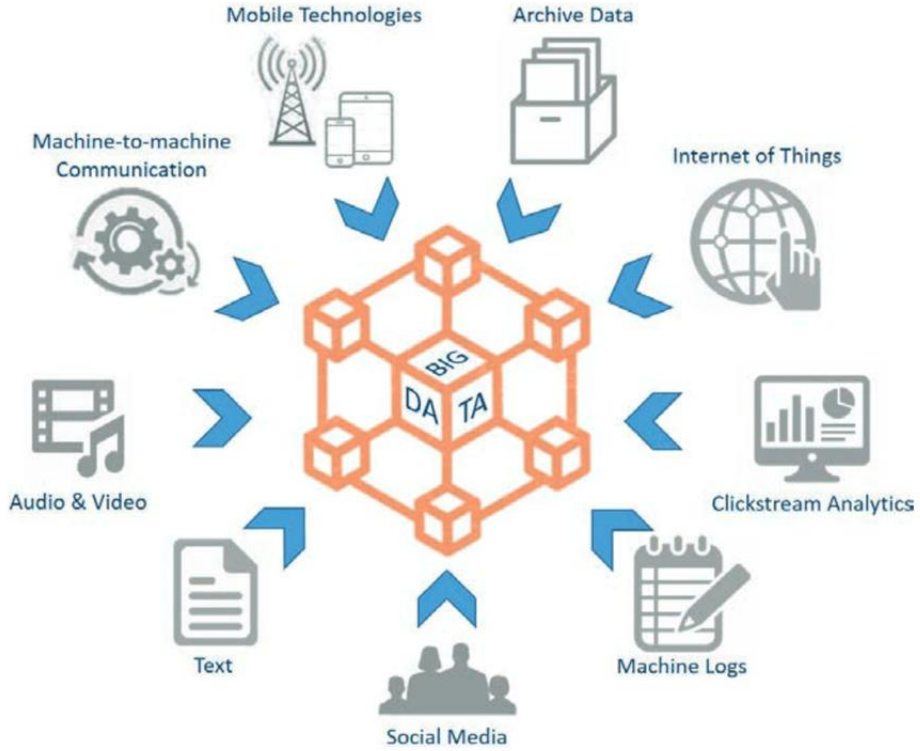
Structured, Semi-Structured and Unstructured Data

- 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.

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.

Figure 3.1. **Sources of Big Data**

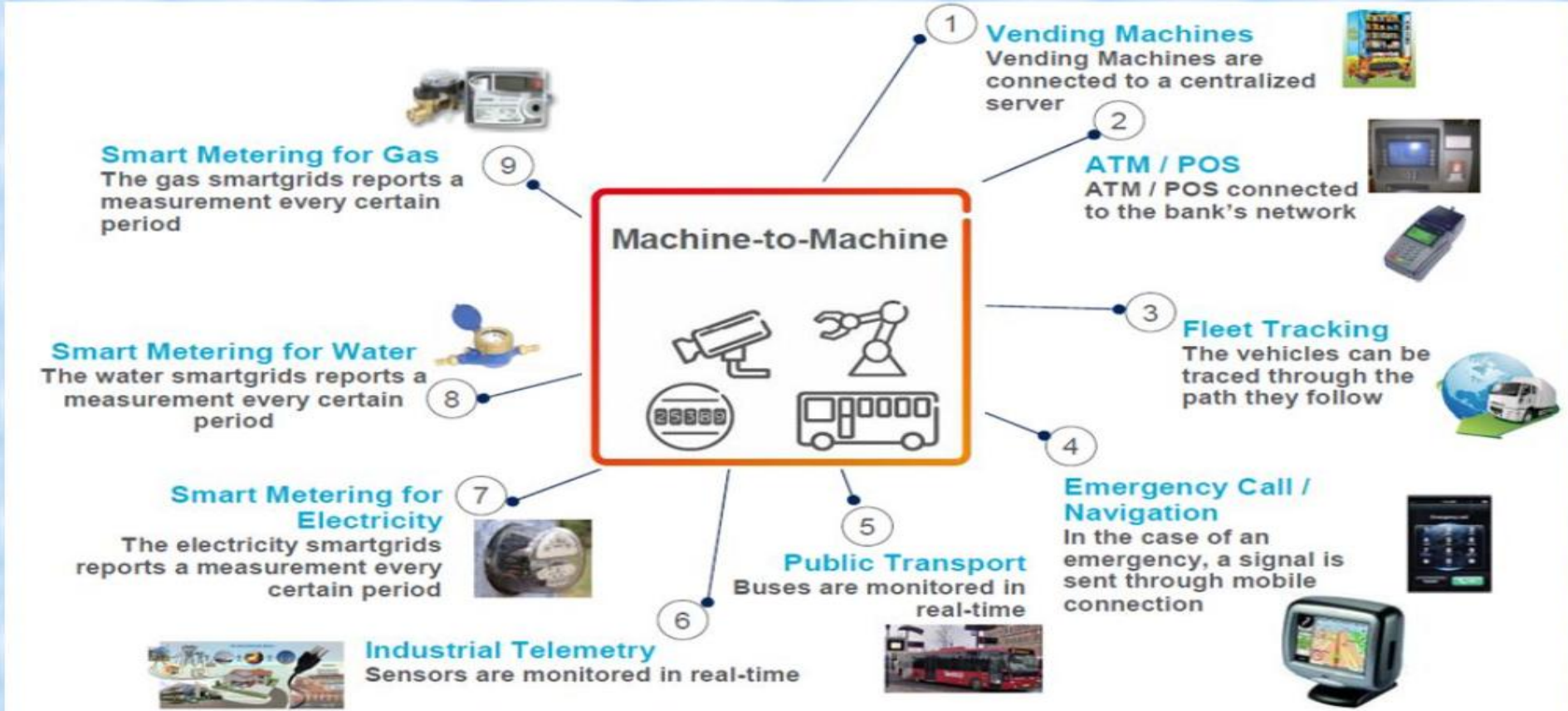


Source: Russian Federal Tax Service.

In Brazil: big numbers.

1. In 2022 alone, 3.838.063 new companies opened up, and 1.695.763 went bankrupt in Brazil. There are 20.191.290 companies regularly functioning in the country.
2. There are more than 30 million judicial cases (tax foreclosure, legal actions disputing taxation, 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 taxpayers to be investigated.

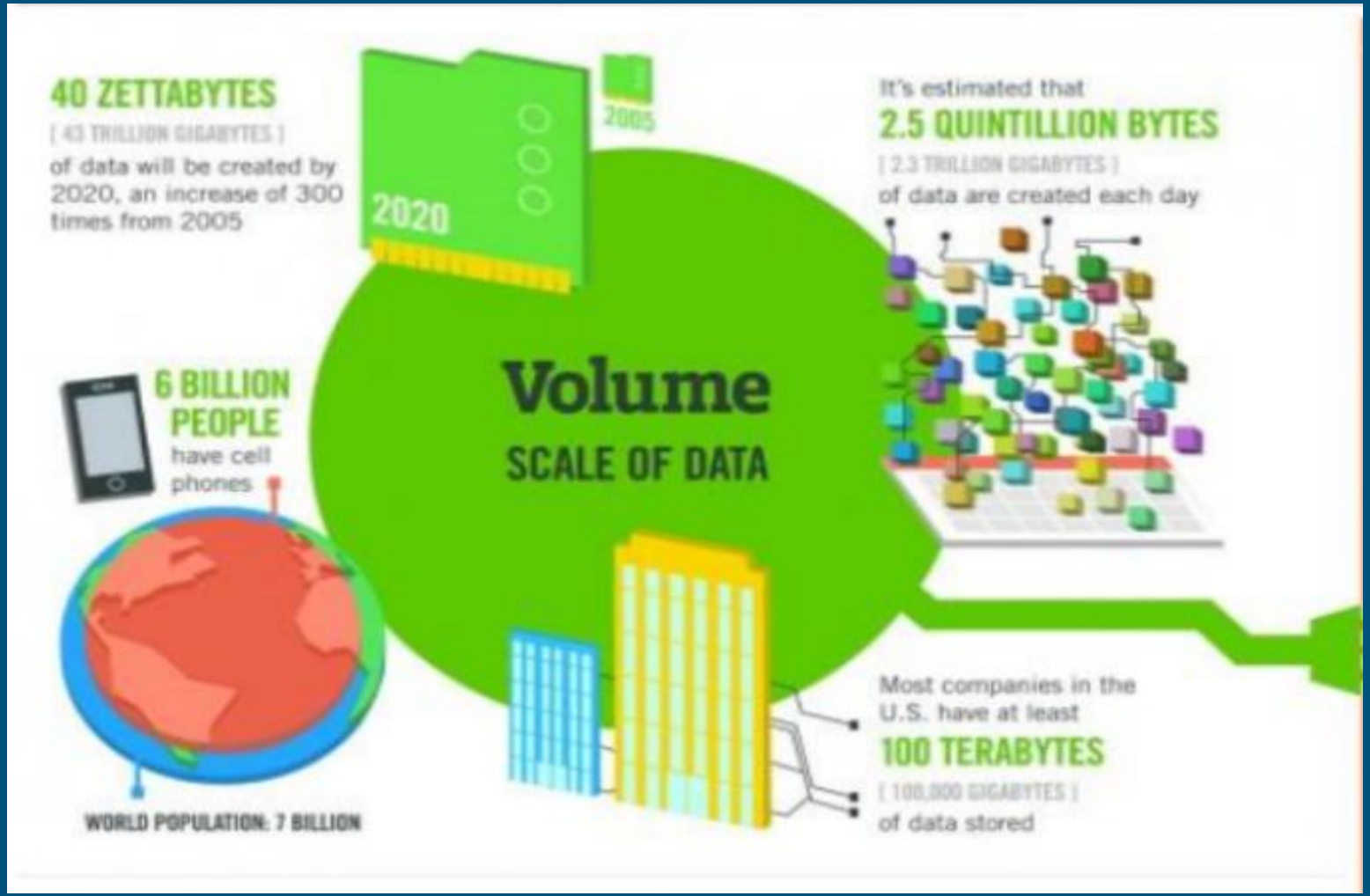
Potential M2M services



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.”


40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005
2020
2005
It's estimated that
2.5 QUINTILLION BYTES
[2.3 TRILLION GIGABYTES]
of data are created each day
Volume
SCALE OF DATA
6 BILLION PEOPLE
have cell phones
WORLD POPULATION: 7 BILLION
Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored

SI prefixes

V · T · E

Prefix		Base 10	Decimal	Adoption [nb 1]
Name	Symbol			
quetta	Q	10^{30}	1 000 000 000 000 000 000 000 000 000 000	2022[3]
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	1991
zetta	Z	10^{21}	1 000 000 000 000 000 000 000	
exa	E	10^{18}	1 000 000 000 000 000 000	1975[4]
peta	P	10^{15}	1 000 000 000 000 000	
tera	T	10^{12}	1 000 000 000 000	1960
giga	G	10^{9}	1 000 000 000	
mega	M	10^{6}	1 000 000	1873
kilo	k	10^{3}	1 000	1795
hecto	h	10^{2}	100	
deca	da	10^{1}	10	

The New York Stock Exchange captures

1 TB OF TRADE INFORMATION

during each trading session

Modern cars have close to

100 SENSORS

that monitor items such as fuel level and tire pressure

Velocity

ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

– almost 2.5 connections per person on earth

Variety

DIFFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES

[161 BILLION GIGABYTES]

30 BILLION PIECES OF CONTENT

are shared on Facebook every month

By 2014, it's anticipated there will be

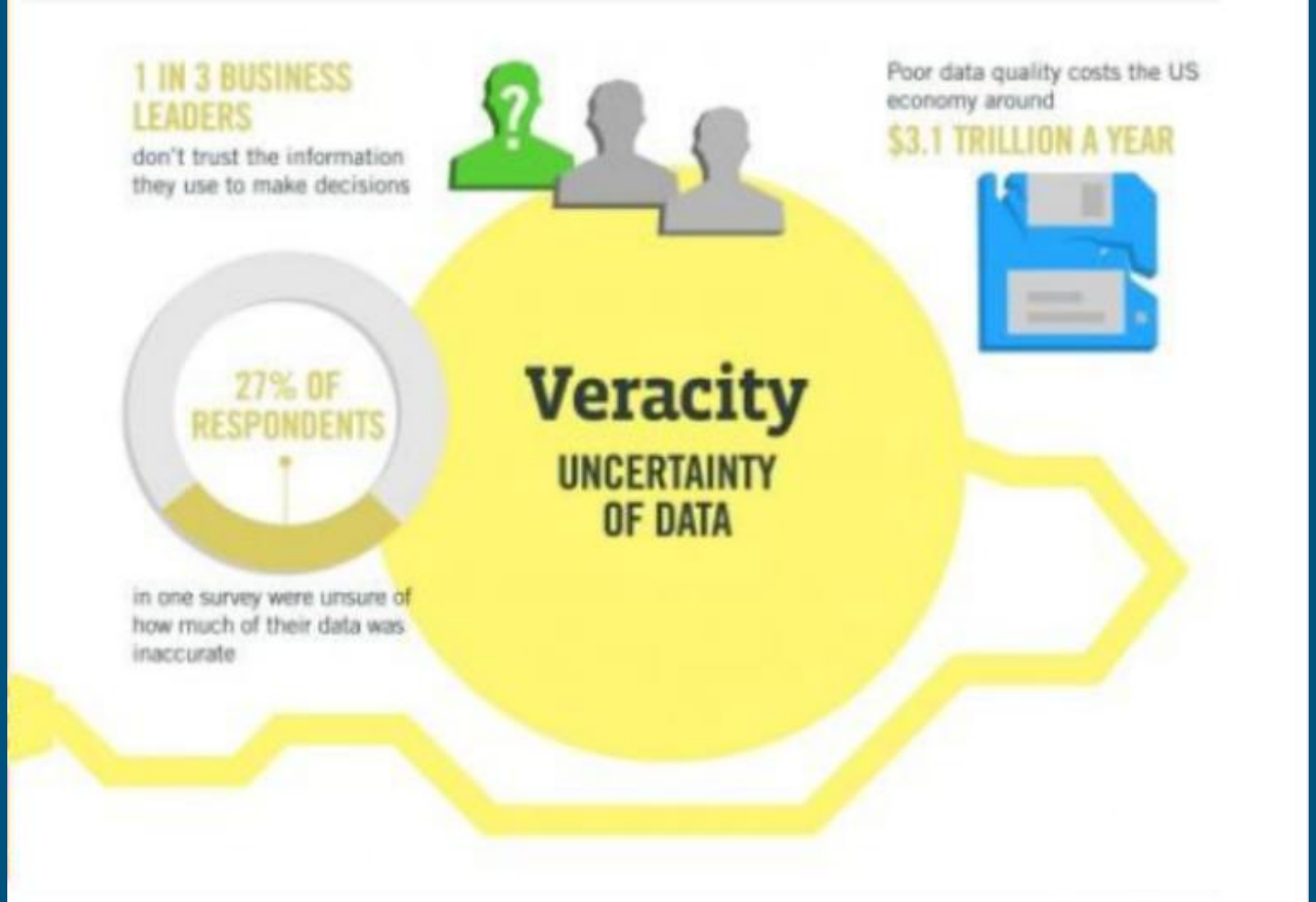
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO

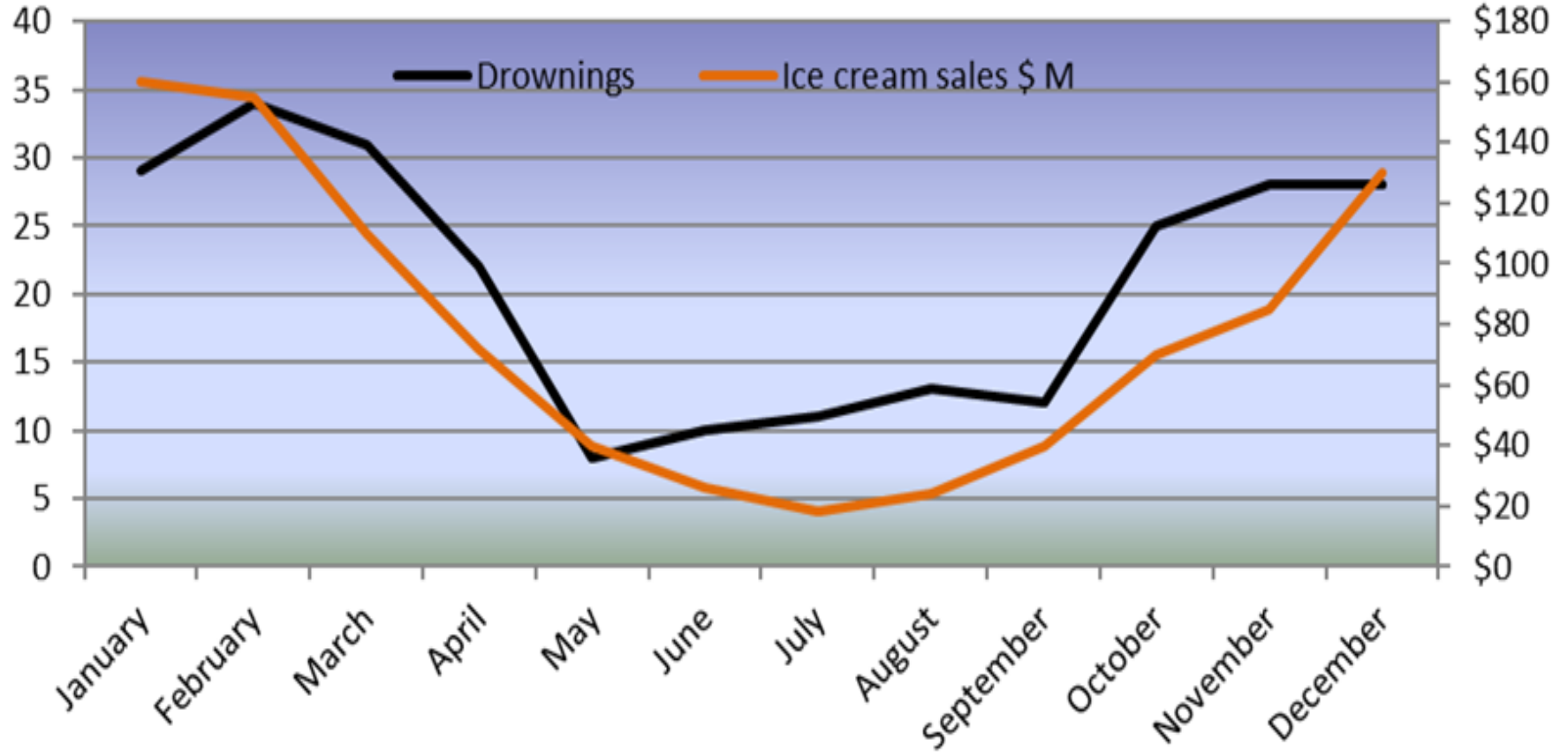
are watched on YouTube each month

400 MILLION TWEETS

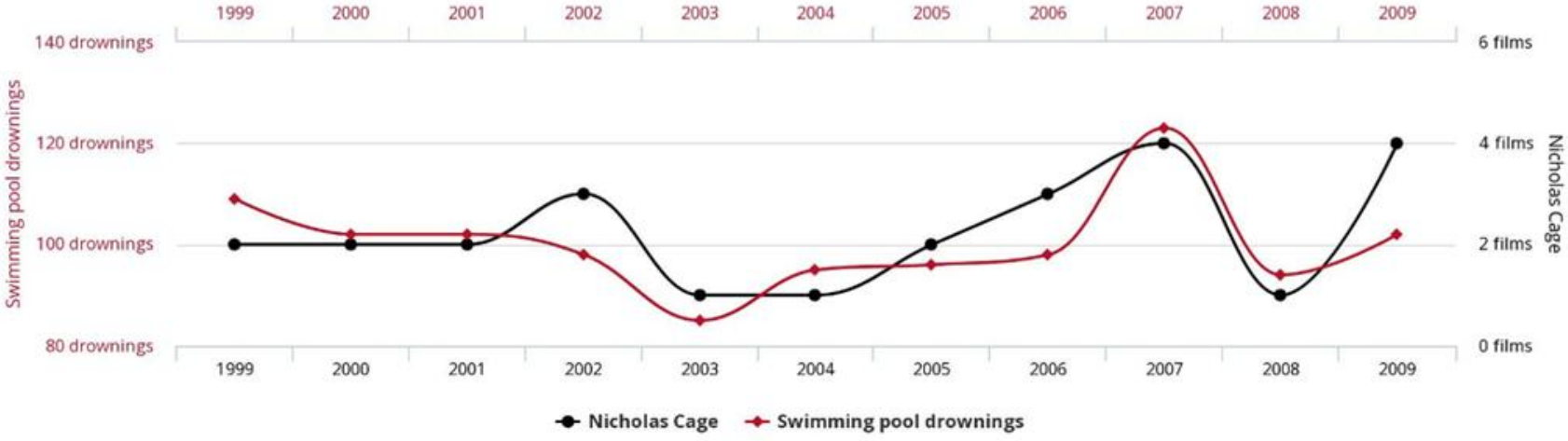
are sent per day by about 200 million monthly active users


1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions
Poor data quality costs the US economy around
$3.1 TRILLION A YEAR
27% OF RESPONDENTS
in one survey were unsure of how much of their data was inaccurate
Veracity
UNCERTAINTY OF DATA

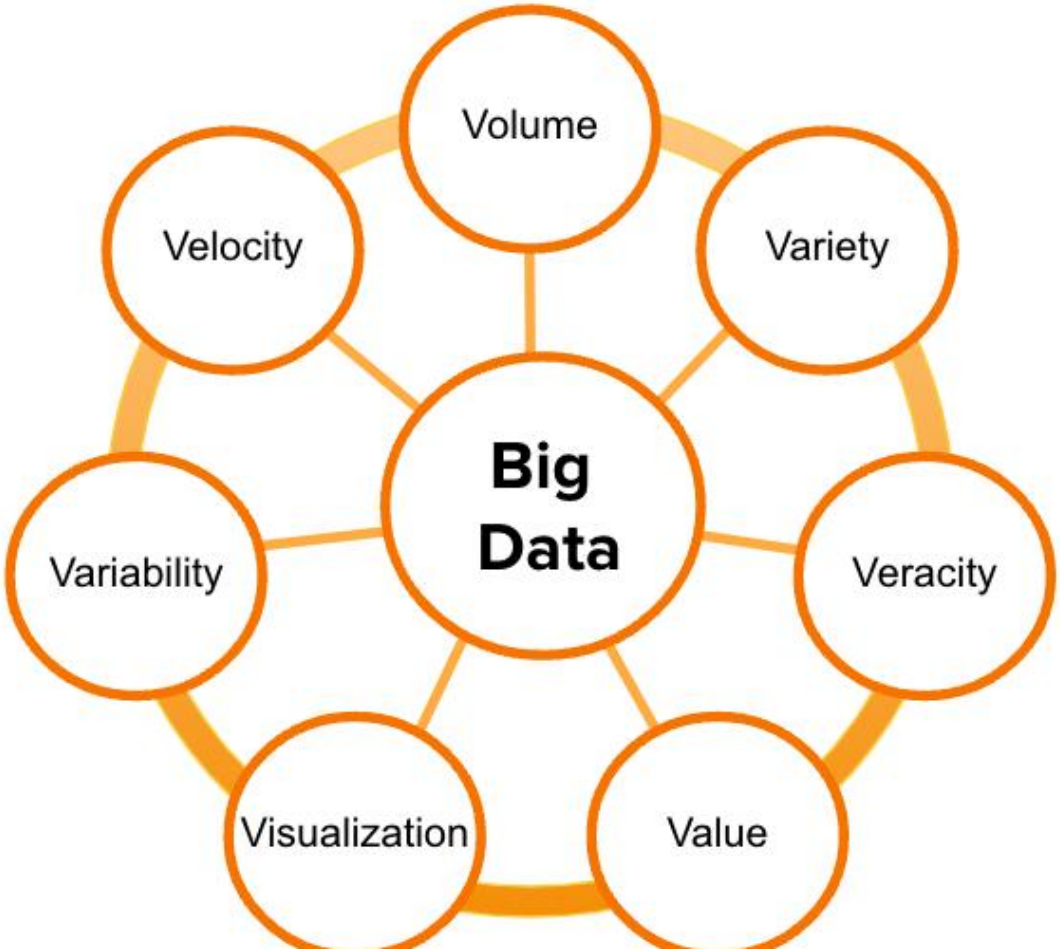
ICE CREAM SALES AND DROWNINGS




DANGER


1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
140 drownings
120 drownings
100 drownings
80 drownings
Swimming pool drownings
6 films
4 films
2 films
0 films
Nicholas Cage
Nicholas Cage
Swimming pool drownings

7 V'S OF BIG DATA



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

Big data and Tax and Crime Investigation: beyond the 7 Vs.

Can it be used as intelligence, as evidence, or both?

Is it legal according to your jurisdiction data protection act?

Big Data Analytics

Activity 1.

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

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

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

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

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

Benefits of using Analytics in tax administration

TAX FUNCTIONS :

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

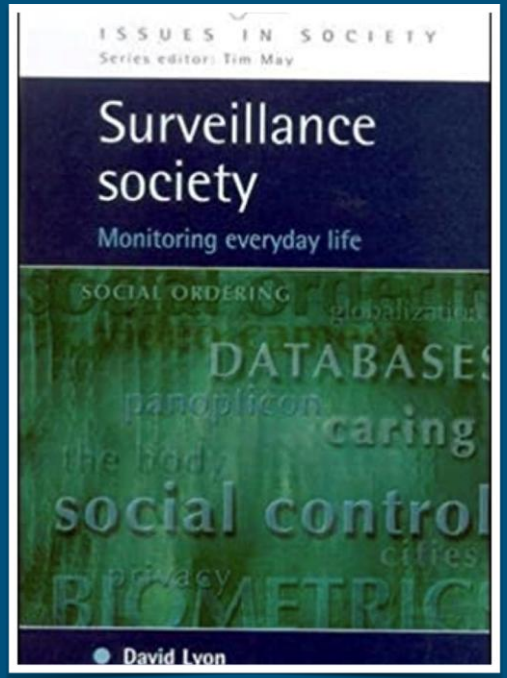
Examples in other areas:

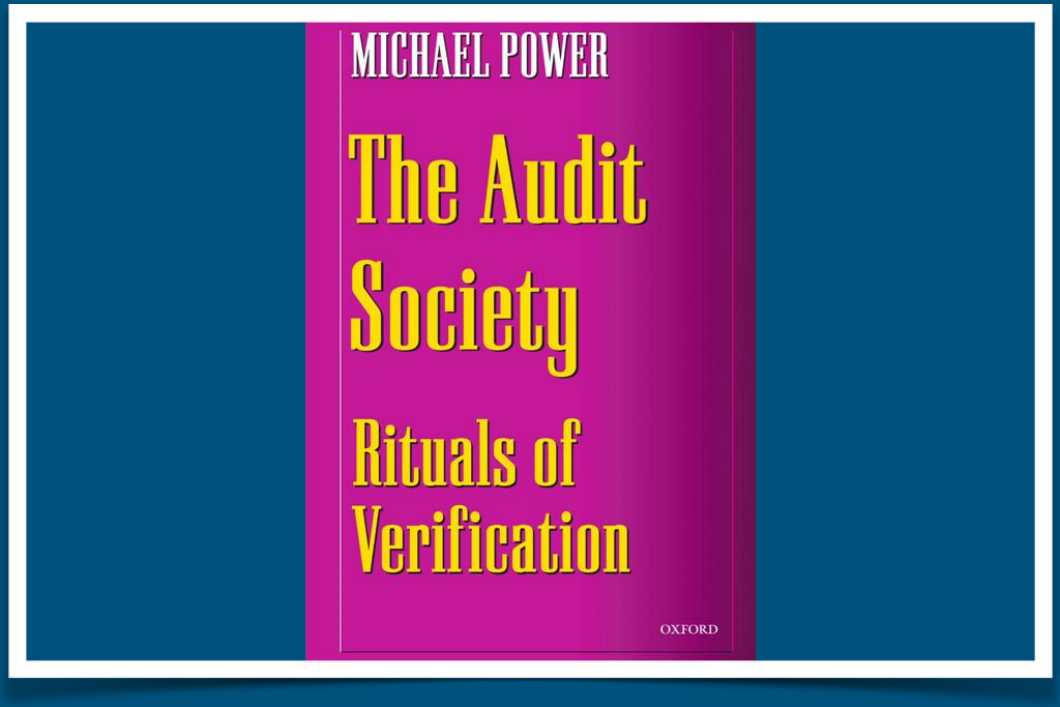
1. Health services: Pandemic control and vaccination.
2. Education: monitoring quality of schools and students' performance.
3. Social Services: tailored/customised services according to individual needs: migrants, homeless etc.
4. Procurement: competition and fairness.
5. Police: reuse of public open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

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.

STRONGLY RECOMMENDED





DISCUSSION:

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

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

Caption

Emerging risks from the use of Big Data Analysis

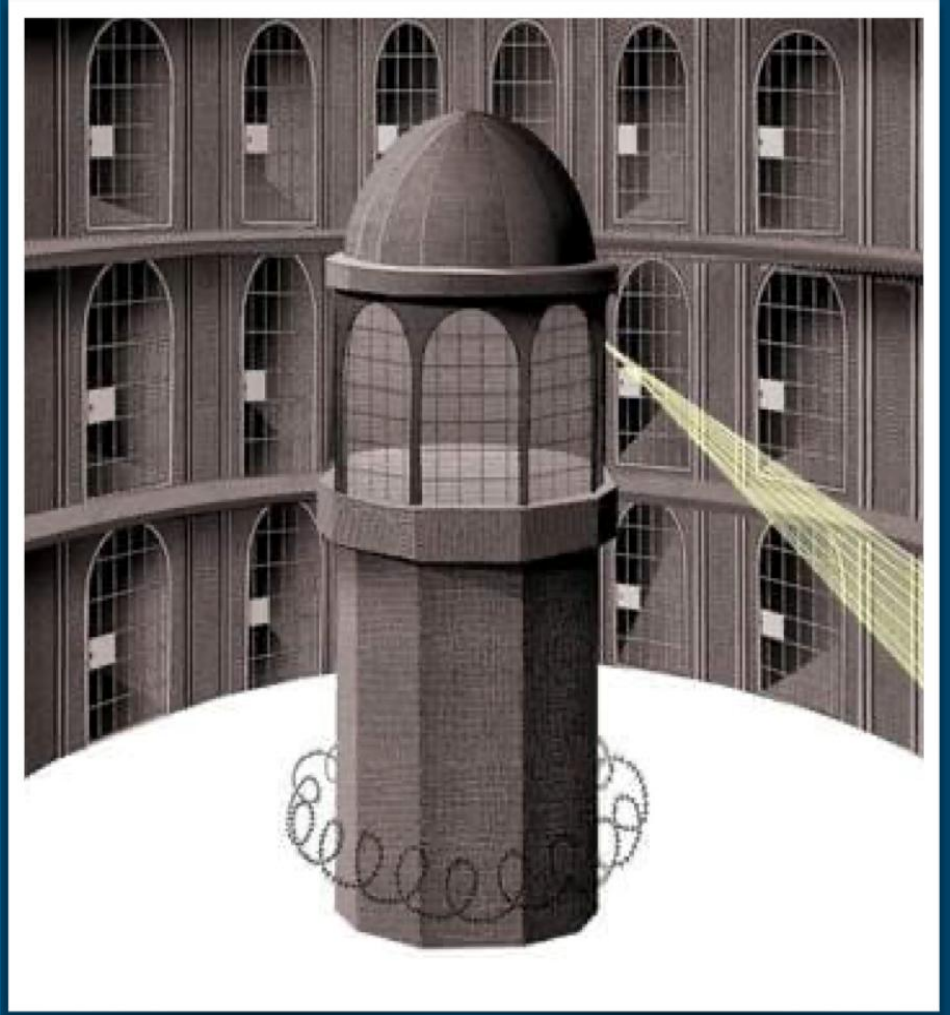
- Security protocols for access, etc.
- Representativeness and relevance of data: The former relates to whether data used provide an exhaustive representation of the population under study, with balanced representation of all relevant subpopulations. (avoiding bias). Data relevance involves the contribution 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 and accountability.

Emerging risks from the use of Big Data Analysis

- Disclosure and transparency (a critical issue when applying BDA in tax investigations: how to be transparent without revealing your investigative technique and strategy?).
- Outsourcing, third party providers.
- Inter agency exchange of information (especially between tax investigation teams and criminal prosecution services; for them tax data is classified and access depends on a court order).
- Regulatory considerations: investigative powers x taxpayers rights.
- Staff: training, ownership.
- Technical issues: storage, processing capacity, the right tool, keeping up with new technologies, etc.

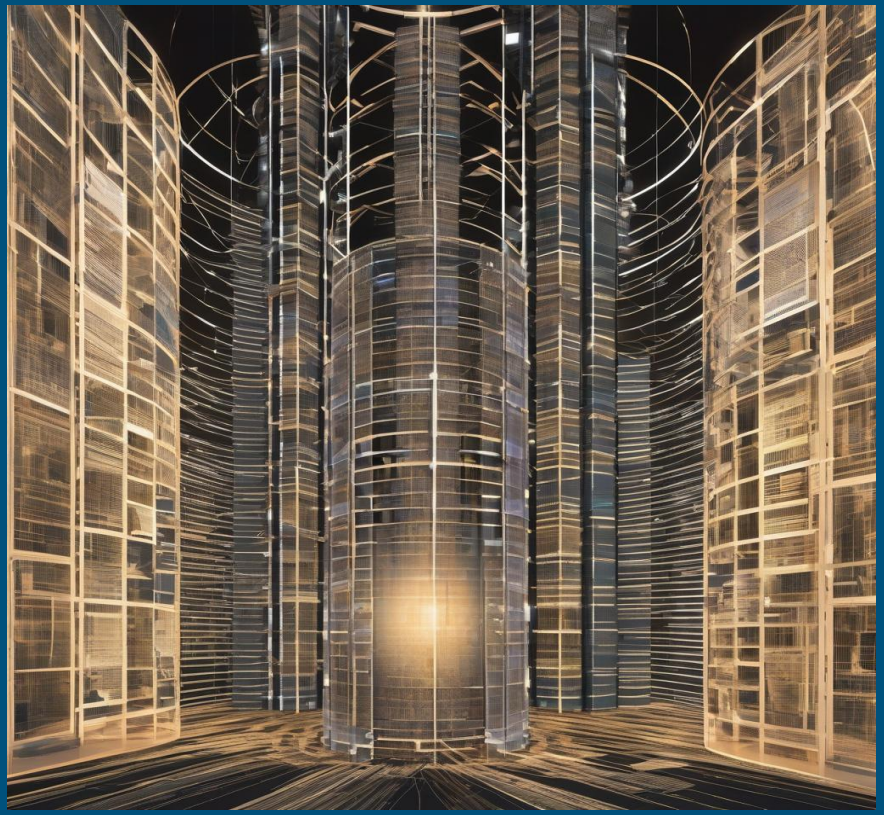
PANOPTICON

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



I - PANOPTICON

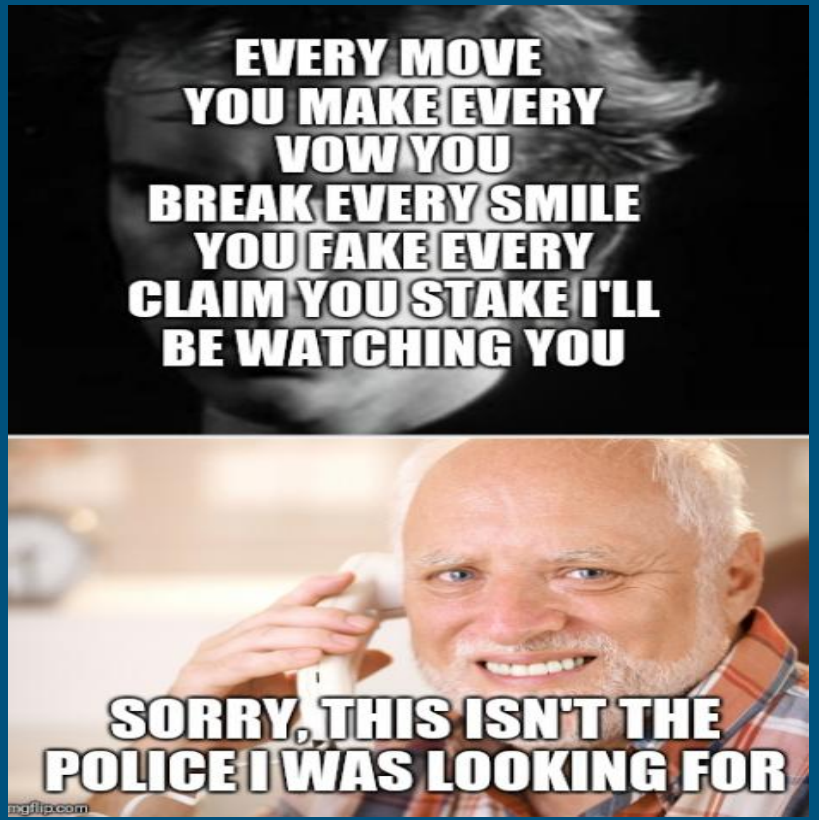
1. Doesn’t require physical presence.
2. Not just monitoring, but also collecting, processing, analysing, influencing.
3. Always on, everywhere.



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



The 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

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.

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**

GDPR - Recital 71

1The 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.

2Such 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.

3However, 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 (...)

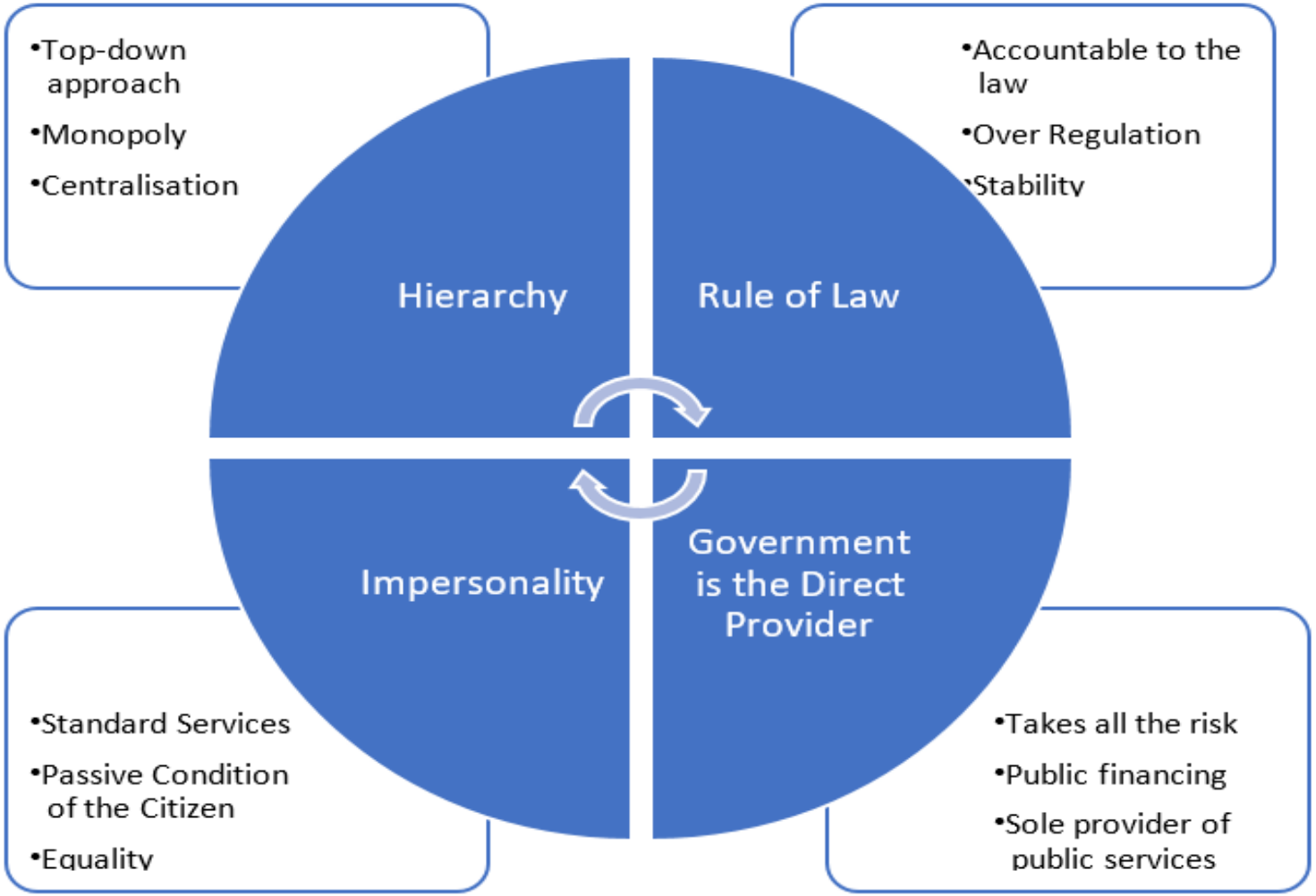
GDPR - Recital 112

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, (...)

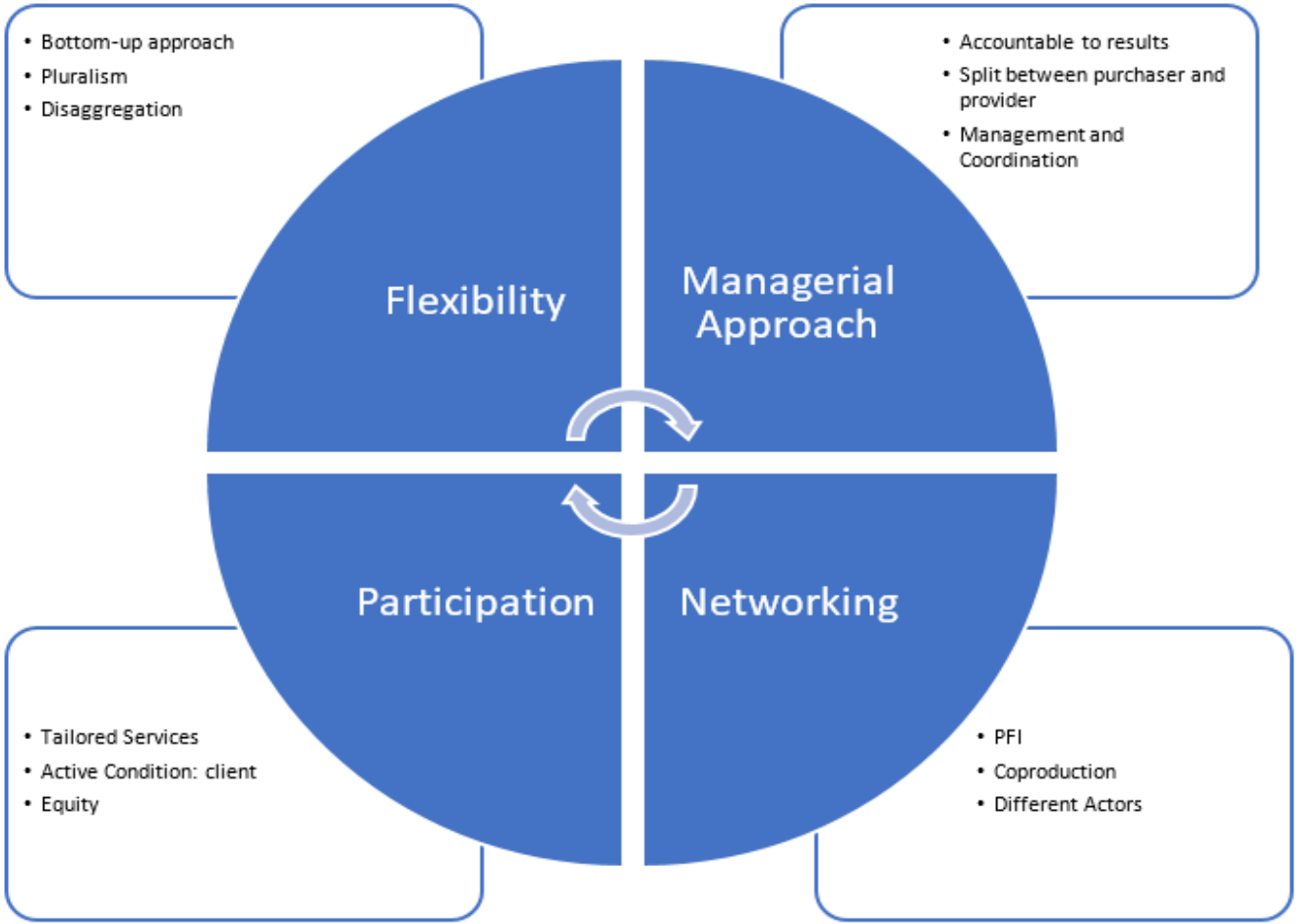
Management considerations

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

Two different models: Bureaucratic Administration x Mix Economy Management



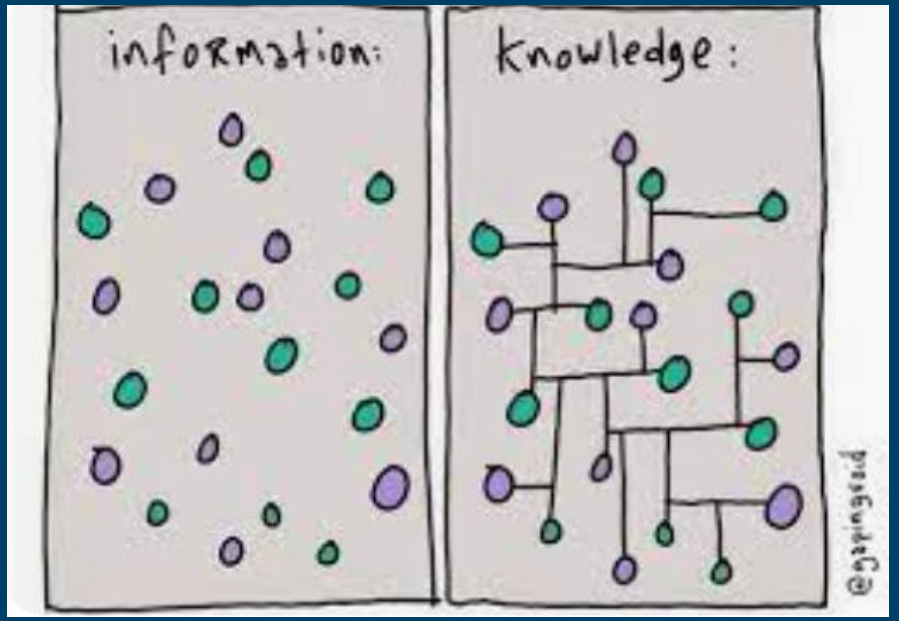
Source: own making.



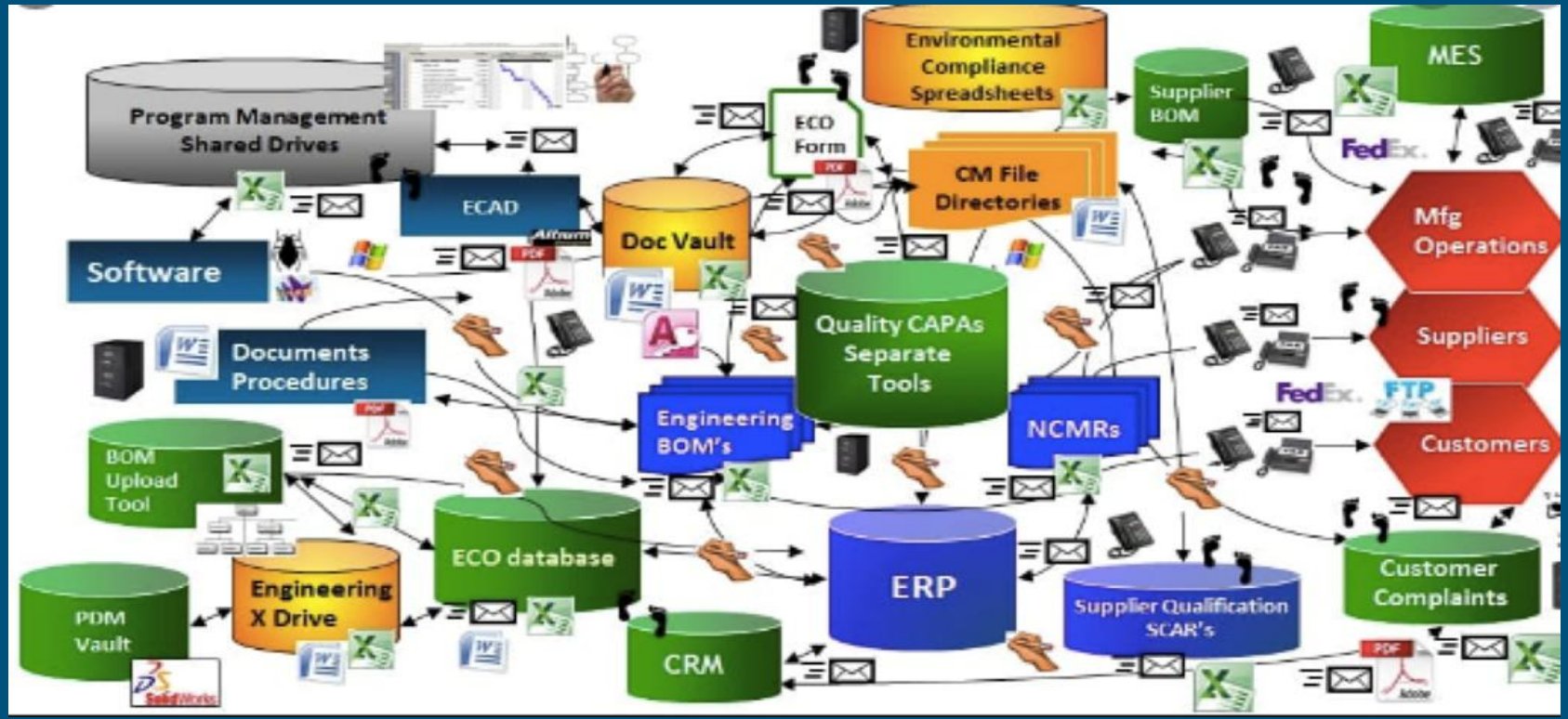
Source: own making.

Connecting the dots!

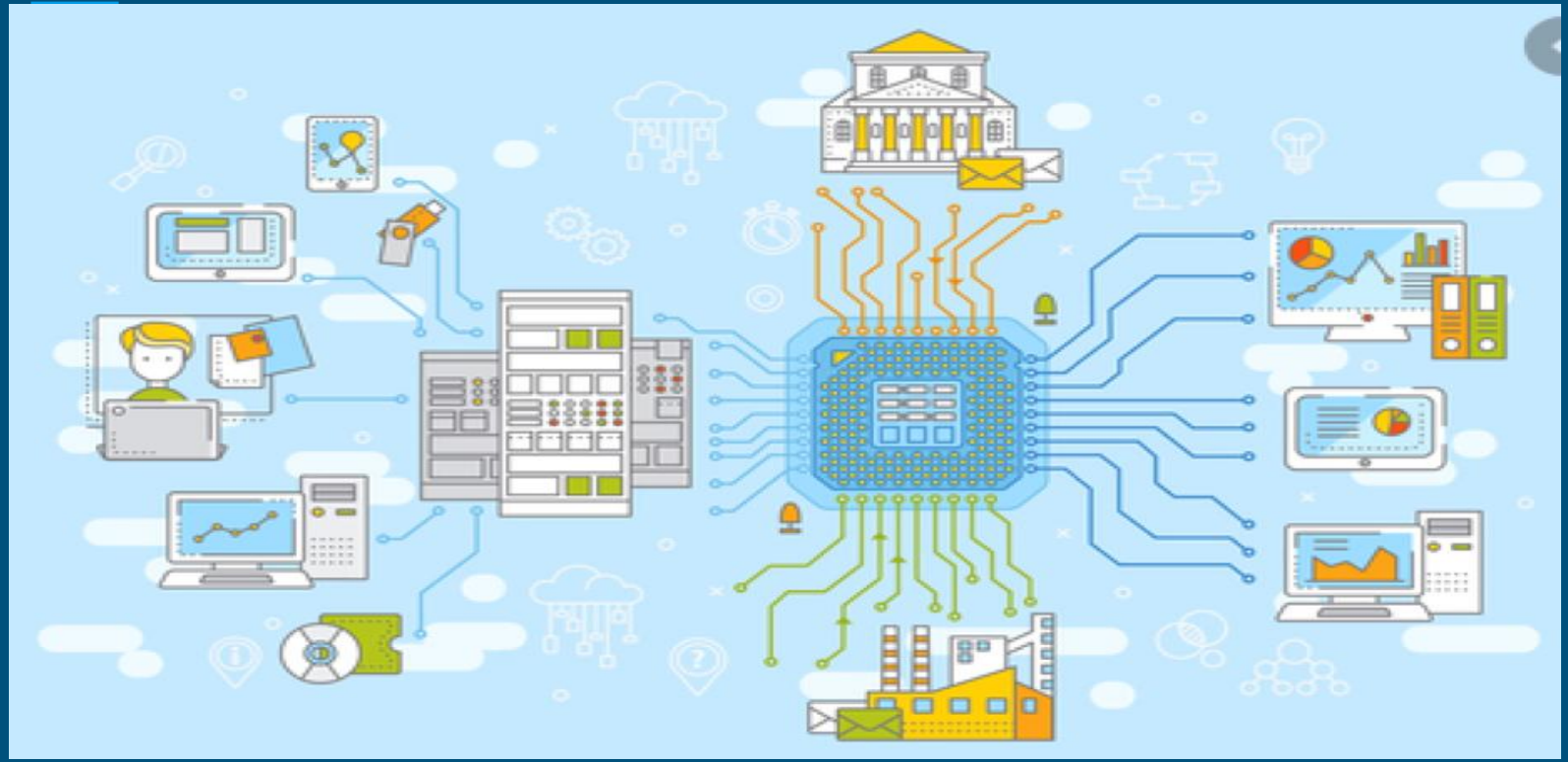
Example of massive use of Big Data Analytics: The PGFN case.



How it used to be



How it is now



DÍVIDA ABERTA APP

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

Check company's tax compliance using id number

Find local debtors (using georeferencing)

Whistleblower channel

Tax compliance services

PGFN news!

Big Data Analysis in Tax Investigation

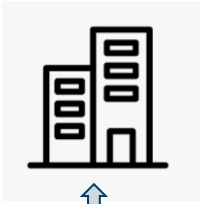
Mining data from 80+ different databases.

Structured and semi-structured data.

Writing the algorithm to search for patterns.

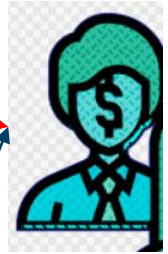
Combining and analysing data.

Expected result: Identification of shell companies + tax debt + assets transferred to a holding + beneficial owner = shell company shareholder.



TAX DEBT

Shareholder's WIFE



Company committing tax fraud

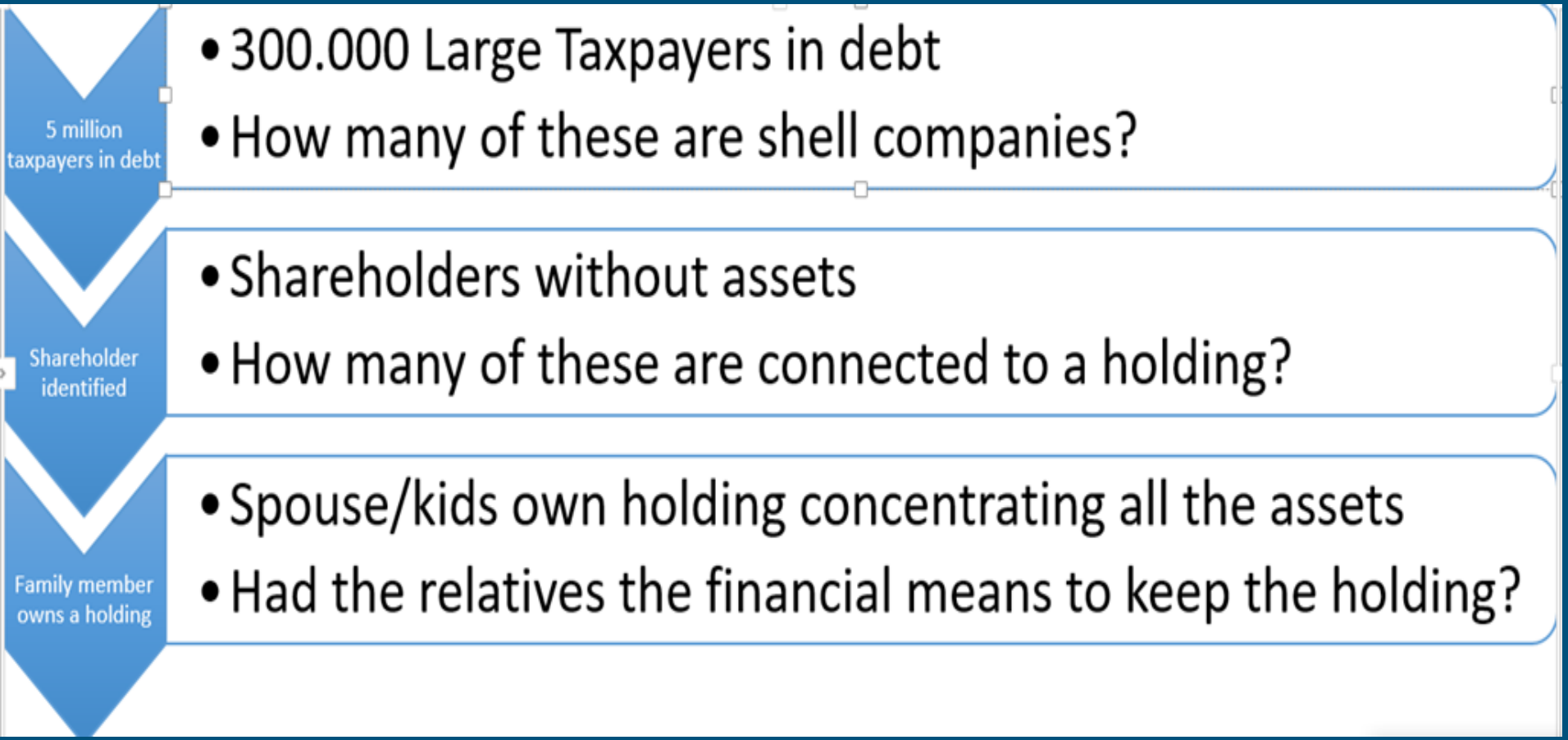
SHAREHOLDER

Shareholder

HOLDING

Transfer his assets to the wife´s Holding

A needle in a haystack!

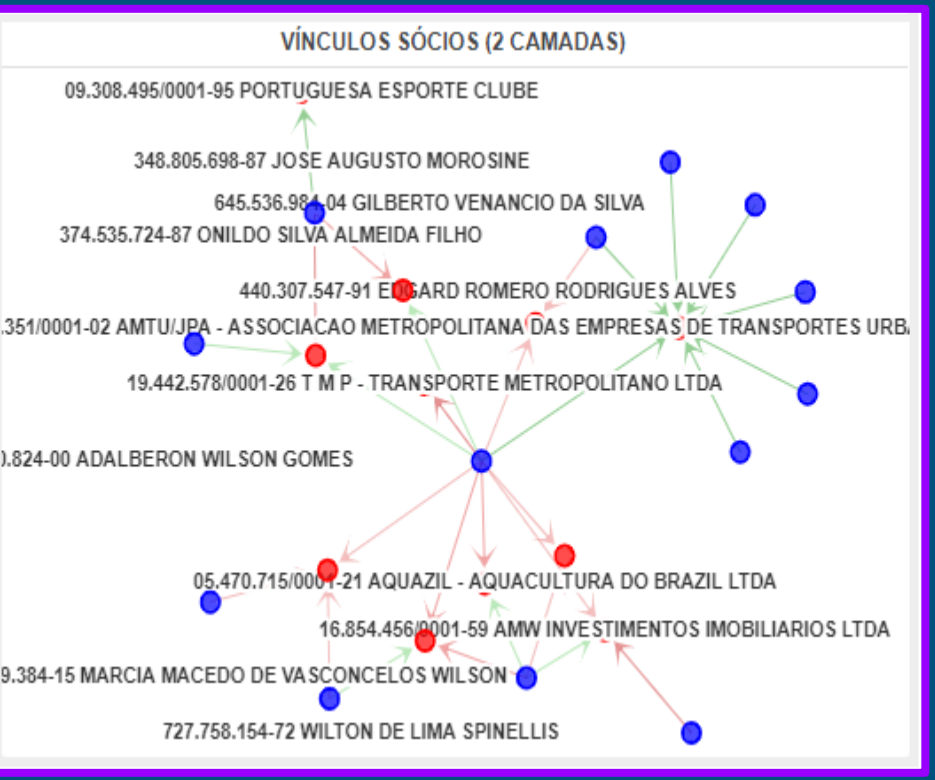


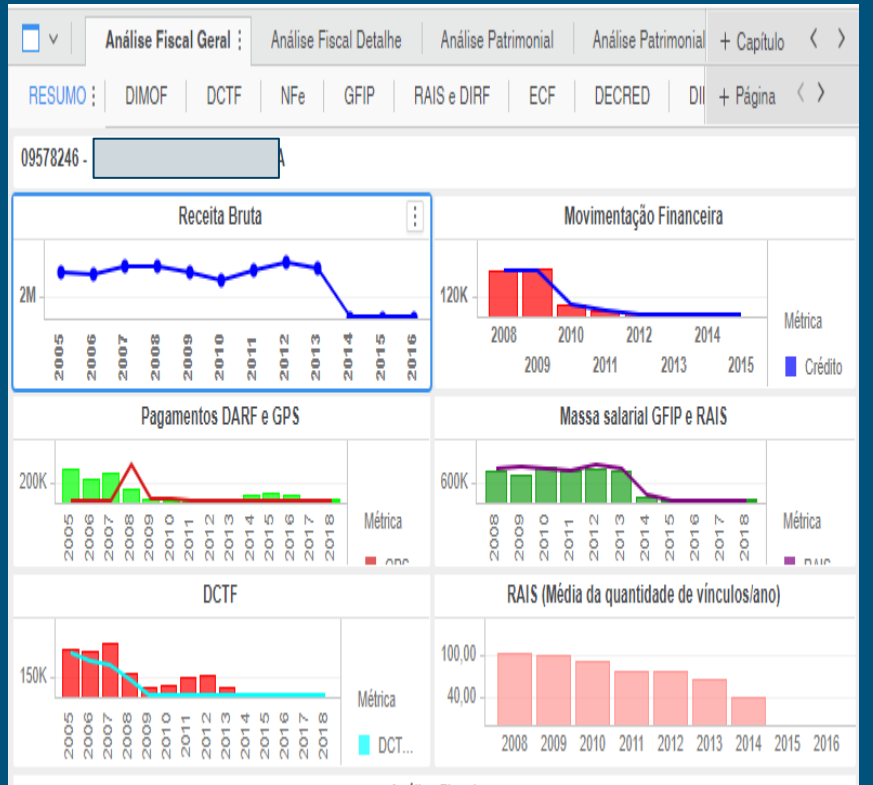
Explaining the Table

PATRIMONIAIS EM NOME DE PARENTES DE SÓCIO OU EX-SÓCIOS DE EMPRESAS DEVEDORAS

ocial	Federal State	City	Company ID Number	Company Name	Copany's current Situation	Cause of the situation	Total Tax Debt	Shareholder Capacity	Social Security Number	Shareholder's name	Shareholder entry	Shareholder's exit	Family Connection	Direcao Relacao	Connection's Social Security Number	Connection's name	Connection's capacity	Shareholder's situation	Shareholder entry	Shareholder exit	Holding's ID number
3ª REGIÃO	SP	PAULINIA	02270967	BOMM PETRO DISTRIB DE DERIVADOS DE PETROLEO E ALCOOL LTDA	BAIXADA	INEXISTENCIA DE FATO	372.769.409,20	SOCIO GERENTE	020114958 30	LUIZ CARLOS MONTEIRO	1998-09-01 00:00:00.0	1999-08-11 00:00:00.0	Companheiro(a) contrib. tenha filho, viva há mais de 5 anos ou cônjuge	<<<	04346325890	MARIA DE FATIMA MERINHO PAIS MONTEIRO	SOCIO ADMINISTRADOR	Ativado	1999-10-11 00:00:00.0	9999-01-01 00:00:00.0	03441931
3ª REGIÃO	SP	PAULINIA	02270967	BOMM PETRO DISTRIB DE DERIVADOS DE PETROLEO E ALCOOL LTDA	BAIXADA	INEXISTENCIA DE FATO	372.769.409,20	SOCIO GERENTE	020114958 30	LUIZ CARLOS MONTEIRO	1998-09-01 00:00:00.0	1999-08-11 00:00:00.0	Companheiro(a) contrib. tenha filho, viva há mais de 5 anos ou cônjuge	>>>	04346325890	MARIA DE FATIMA MERINHO PAIS MONTEIRO	SOCIO ADMINISTRADOR	Ativado	1999-10-11 00:00:00.0	9999-01-01 00:00:00.0	03441931
3ª REGIÃO	SP	PAULINIA	02270967	BOMM PETRO DISTRIB DE DERIVADOS DE PETROLEO E ALCOOL LTDA	BAIXADA	INEXISTENCIA DE FATO	372.769.409,20	SOCIO GERENTE	438247658 53	ROBERTO LUIZ BICUDO FERRARO	1997-12-08 00:00:00.0	1999-08-11 00:00:00.0	Companheiro(a) contrib. tenha filho, viva há mais de 5 anos ou cônjuge	<<<	02496015836	MARTA MAOLINE CHAVES	SOCIO	Ativado	2015-07-31 00:00:00.0	9999-01-01 00:00:00.0	03464098
3ª REGIÃO	SP	PAULINIA	02270967	BOMM PETRO DISTRIB DE DERIVADOS DE PETROLEO E ALCOOL LTDA	BAIXADA	INEXISTENCIA DE FATO	372.769.409,20	SOCIO GERENTE	438247658 53	ROBERTO LUIZ BICUDO FERRARO	1997-12-08 00:00:00.0	1999-08-11 00:00:00.0	Companheiro(a) contrib. tenha filho, viva há mais de 5 anos ou cônjuge	>>>	02496015836	MARTA MAOLINE CHAVES	SOCIO	Ativado	2015-07-31 00:00:00.0	9999-01-01 00:00:00.0	03464098

Making it visual: Visual Law.





A brief tour through some PGFN Big Data systems

- RELATÓRIOS DE APOIO
- ANALYTICS
- SNIPER / TRACKER/ RAPTOR
- DW

Is it 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.

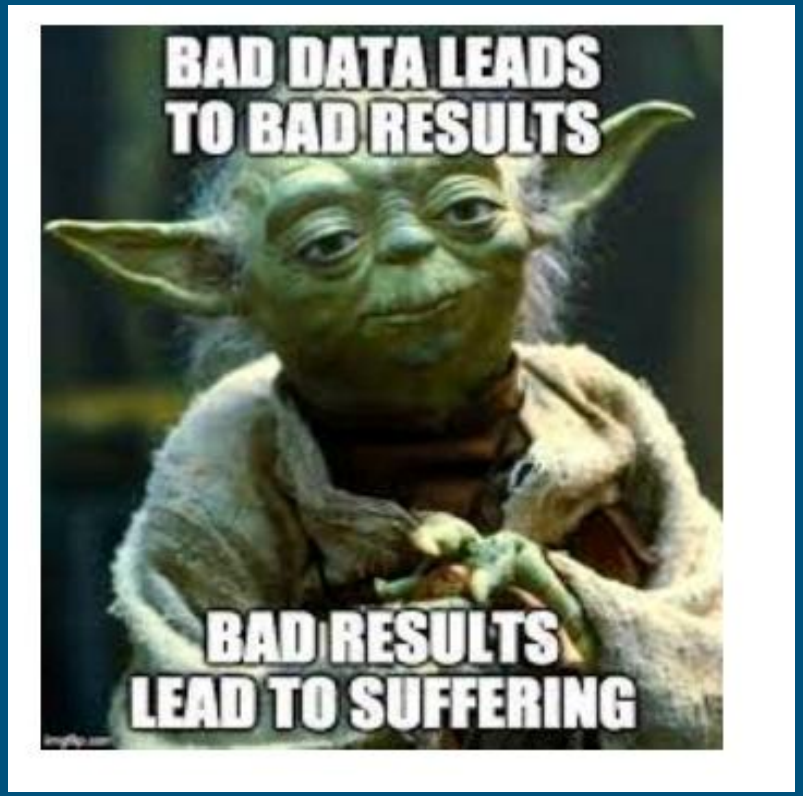
Becoming digital, Digital Governance.

Reaching for databases

Discussion: is it safe to outsource (storage, processing, cloud etc)?

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.



Fraud or 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 4 $2 million worth flats to wife.

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

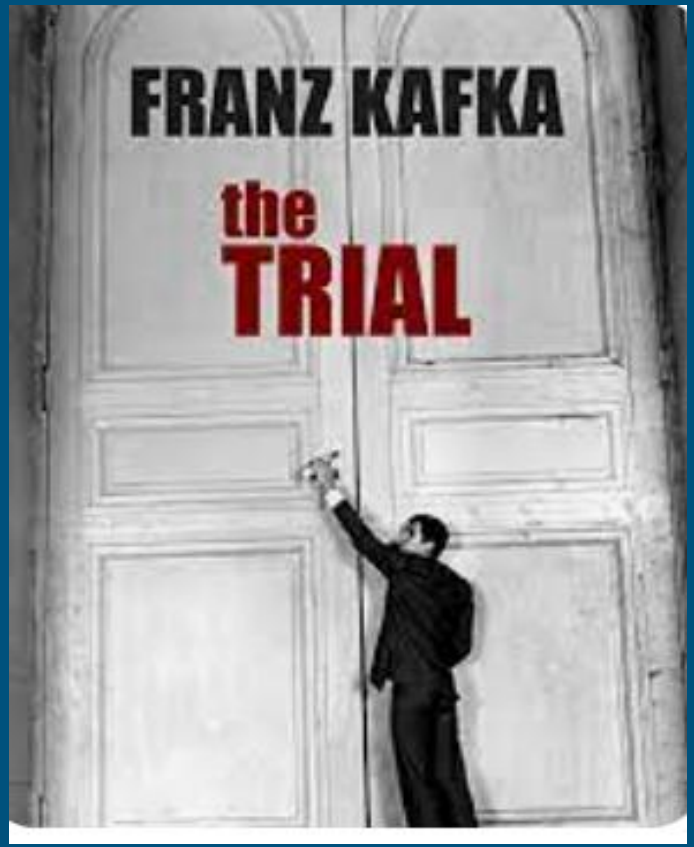
Fraudulent corporate group?

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 Commercial Bureau. Each company owned around $200,000. The accountant might be the enabler.

However….

The Post Office Scandal

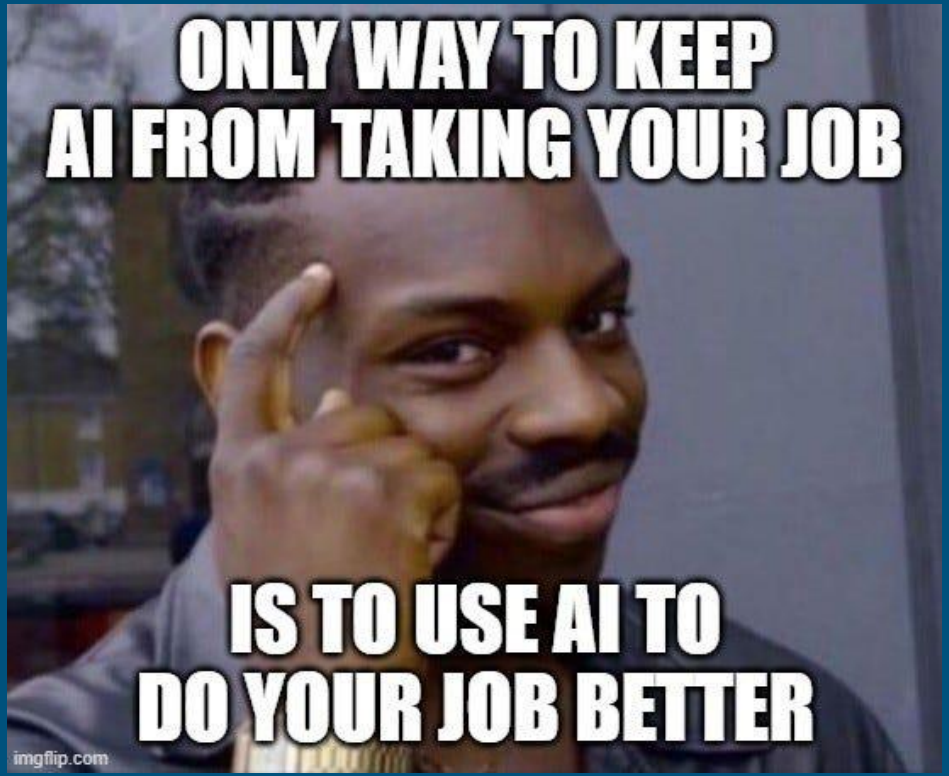
- A story of how overconfidence in data analysis ruined almost 900 lives: "The SYSTEM can do no wrong" !!
- Algorithm and accountability.
- The weight of electronic processed data as judicial evidence.
- Explainability: do judges and lawyers really understand big data?
- Big Data Analytics and credibility: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



Recommendation!

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 industries across various sectors.



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).

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:

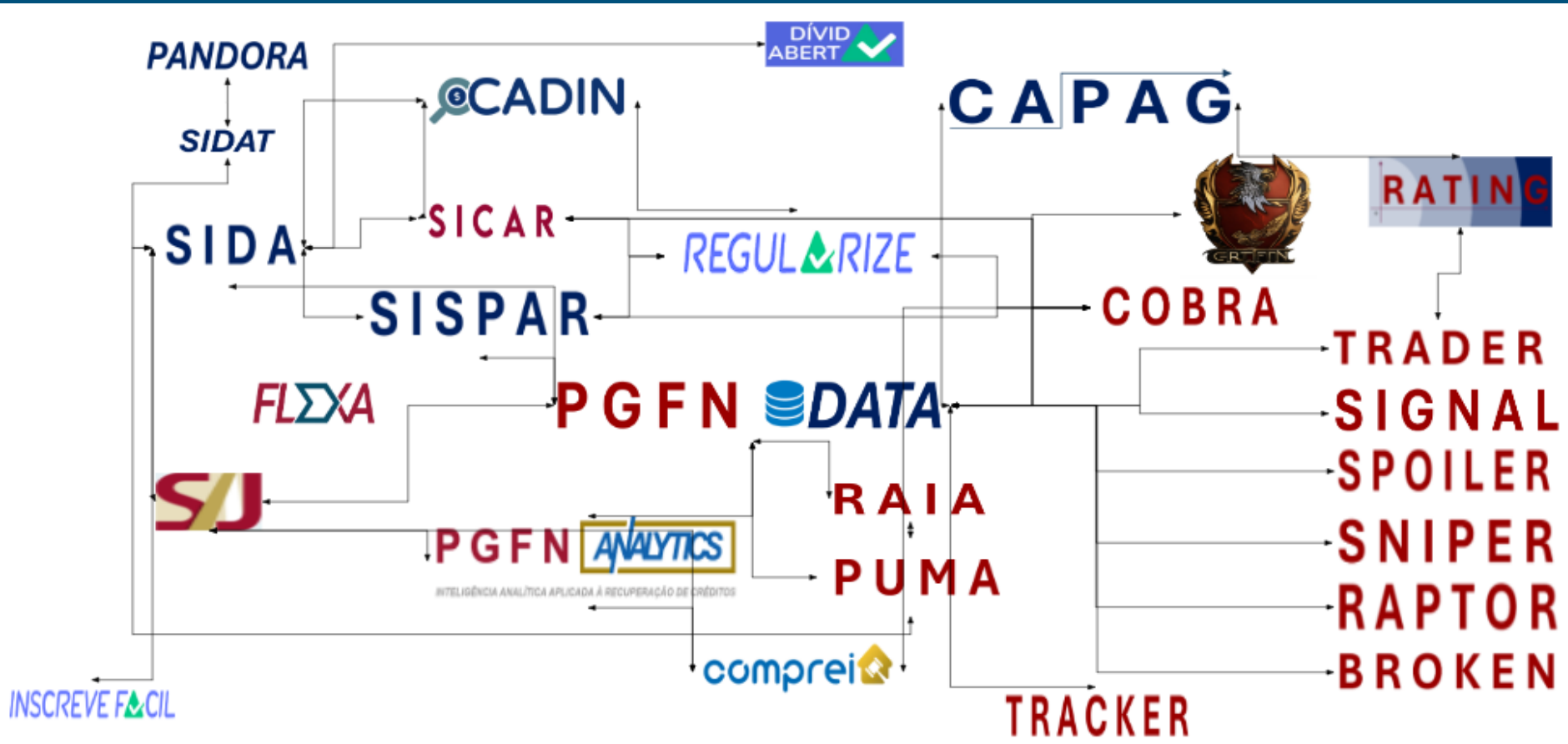
- Pattern Recognition: Identifying fraudulent schemes and patterns of behavior.
- Real-time Monitoring: Detecting fraudulent activities as they occur.

Money Laundering Investigation:

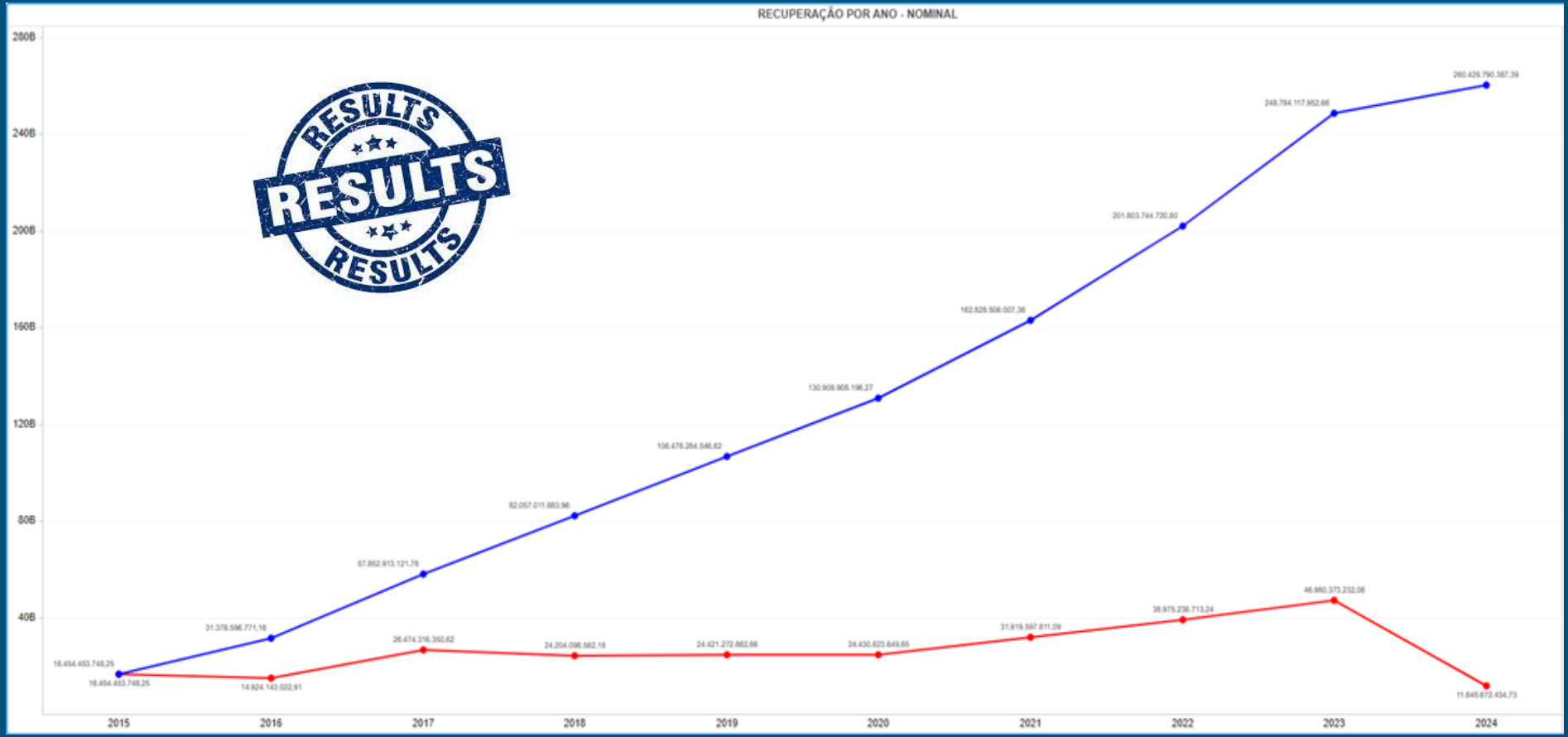
- Network Analysis: Analyzing financial transactions to identify suspicious networks.
- Behavioral Analysis: Identifying unusual spending patterns or asset transfers.

What PGFN has been doing with AI + Big Data

Systems and Apps “converse” with each other



Results!!!


RECUPERAÇÃO POR ANO - NOMINAL
280B
240B
200B
160B
120B
80B
40B
RESULTS
RESULTS
RESULTS
16.454.453.748,25
31.378.596.771,16
57.852.913.121,78
82.057.011.683,96
106.478.284.546,62
130.905.908.196,27
162.826.506.007,36
201.803.744.720,60
248.784.117.952,66
260.429.790.387,39
16.454.453.748,25
14.924.143.022,91
26.474.316.350,62
24.204.096.562,18
24.421.272.862,66
24.430.623.649,65
31.919.597.811,09
38.975.236.713,24
46.960.373.232,06
11.645.672.434,73
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024

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)

ChatPGFN

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

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

Activity 2

Advise the City Council on how Big Data 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.

Discussion and Wrap up

1. Big Data Analytics/AI and Tax Investigation.
2. Further reading.
3. Making connections.

Arthur Moura

E-mail: prof.acmoura@gmail.com
WhatsApp: 55 83 99350-4334