



ANÁLISIS DE BIG DATA



TEORÍA Y PRÁCTICA PARA LA
INVESTIGACIÓN TRIBUTARIA



IMPARTE LA PONENCIA HOY:

ARTHUR MOURA

Fiscal Federal del Tesoro Nacional desde 2003

Antiguo jefe del Departamento de Investigación Tributaria (5.ª Región)

Beneficiario del Programa de Becas Chevening 2017-2018

Máster en Administración Pública - Universidad de Nottingham (Reino Unido)

Escritor y profesor de Derecho Fiscal y Procesal

Estas reflexiones y opiniones...

son mías y no coinciden necesariamente con las del organismo en el que trabajo (la PGFN) o con las de la Academia



Short questions, Big Data

Please answer this quick survey about Big Data.



PLANIFICACIÓN DEL MÓDULO

Enfoque centrado en la Administración, más que en la estadística o la informática.

Primera parte: Análisis de datos, big data y gestión o administración (*management*).

Ejercicio 1: aplicación del análisis de big data al sector público.

Segunda parte: Ejemplo práctico del análisis de big data en investigación tributaria.

Ejercicio 2: aplicación del análisis de big data a la investigación tributaria y penal.

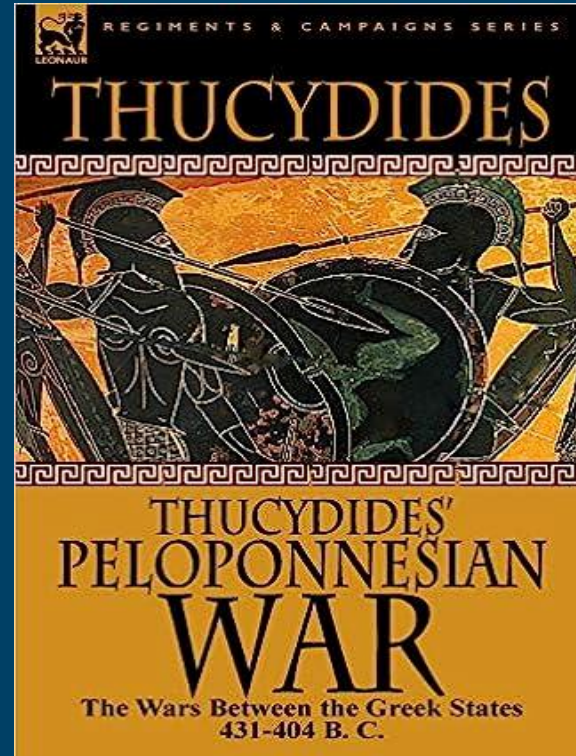
Al final de este módulo

Usted:

- a. Comprenderá mejor el análisis de datos y big data y sus usos en la gestión de investigaciones tributarias.
- b. Se habrá familiarizado con el potencial de esas herramientas tanto a nivel personal como en general para el organismo en el que trabaja.
- c. Será consciente de algunos de los riesgos que entraña el uso de big data.

Un poco de historia

Recopilación de datos,
protección de datos, limpieza de
datos, combinación de datos y,
por último, análisis: ayuda a la
toma de decisiones.



Análisis de datos y Administración tributaria

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



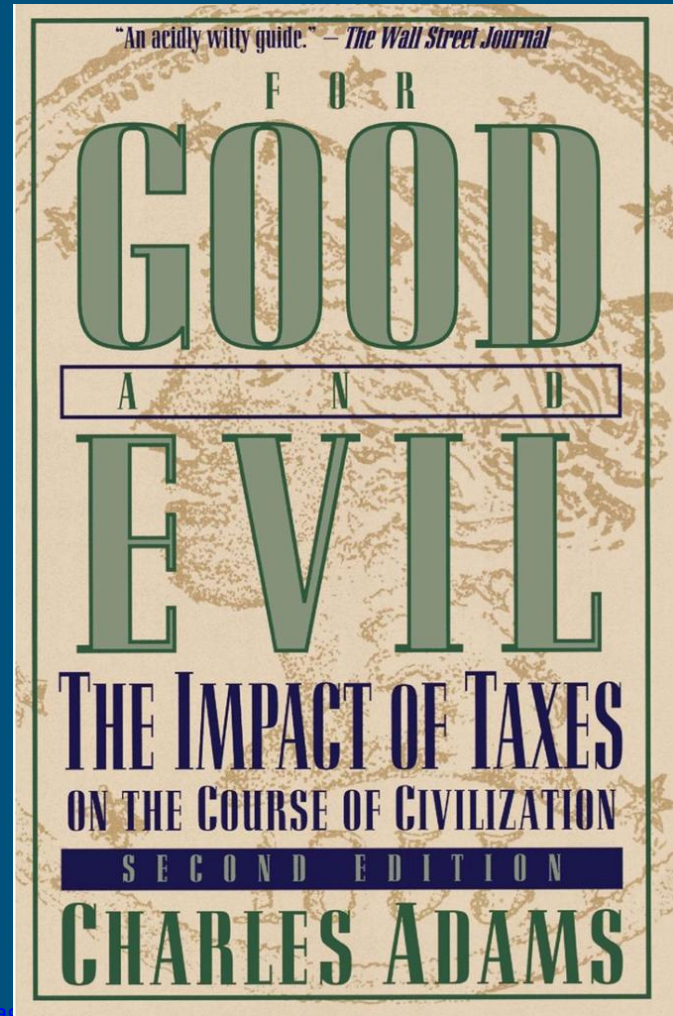
The tax officials of the pharaoh—everywhere snooping, inspecting, recording, and arresting—even surveyed the nests of pigeons to count the eggs, making sure the pharaoh got his 20-percent cut.

ph 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

¡Recomendación!



ANÁLISIS DE DATOS

El proceso de recopilación, limpieza, combinación y formateo de datos con el fin de extraer información útil (valiosa) y conocimientos que ayuden a la toma de decisiones.

- **Recopilación:** diversidad de fuentes, formatos, lenguas.
- **Minería de datos:** demasiado ruido como para que sea posible identificar una nota.
- **Limpieza:** consiste en eliminar datos duplicados, corregir errores y solucionar los valores que faltan.
- **Combinación:** fusión de datos, establecimiento de correlaciones para comprender mejor los datos e identificar patrones, tendencias y valores atípicos.
- **Formateo:** comunicaciones visuales, cuadros de mando, etc.



¿Por qué utilizar el análisis de datos?

- El análisis de datos proporciona valiosos conocimientos e información que pueden ayudar a individuos y organizaciones a tomar mejores decisiones.
- Resolución de problemas: El análisis de datos ayuda a identificar problemas.
- Ahorro de costos: El análisis de datos puede ayudar a detectar ineficiencias y desaprovechamientos en los procesos.
- Innovación: El análisis de datos puede ayudar a identificar nuevas oportunidades.
- Mejora del rendimiento: El análisis de datos puede ayudar a las personas a rendir más.

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	933.829,81	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	
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

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

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 resultados

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 CEREJAS 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 resultados

Outros interessados - Não existem partes cadastradas a outros interessados.

Participante	Tipo de participação	Situação RFB
--------------	----------------------	--------------

Foram encontrados: 0 resultados

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

Identificador	Assinatura	Descrição	Tipo de documento	Documentos/Ícones
4056201.11633874	12/04/23 16:29	Despacho Inspeção - 1227 - INSPEÇÃO ORDINÁRIA ANUAL 2023 (10ª VARA FEDERAL)	Despacho Inspeção	
4056201.11791332	05/06/23 17:29	Decisão	Decisão	
4056201.11447416	27/03/23 15:12	Certidão de Distribuição	Certidão	
4056201.11429110	27/03/23 15:10	Incidente de Desconsideração da Personalidade Jurídica - (Sigiloso)	Petição Inicial - (Sigiloso)	
4056201.11445248	27/03/23 15:10	Inicial - (Sigiloso)	Documento de Comprovação - (Sigiloso)	
4056201.11445260	27/03/23 15:10	1- Cadastro Nacional de Empresas - CNE - (Sigiloso)	Documento de Comprovação - (Sigiloso)	
4056201.11445265	27/03/23 15:10	2- Reclamação trabalhista - (Sigiloso)	Documento de Comprovação - (Sigiloso)	
4056201.11445280	27/03/23 15:10	3- Relações de ROGACIANO NUNES DA NOBREGA NETO - (Sigiloso)	Documento de Comprovação - (Sigiloso)	
4056201.11445289	27/03/23 15:10	4- Relações de BRUNO FIGUEIREDO NOBREGA - (Sigiloso)	Documento de Comprovação - (Sigiloso)	
4056201.11445299	27/03/23 15:10	5- Relações de FELIPE FIGUEIREDO NOBREGA - (Sigiloso)	Documento de Comprovação - (Sigiloso)	
4056201.11445300	27/03/23 15:10		Documento de Comprovação - (Sigiloso)	

Motion granted

Petition and Documents

Different format

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

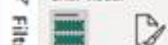


Árvore Hierárquica



Visualizações

Criar visual



Filtros



Valores

Adicionar os campos de

Drill-through

Relatório cruzado

Manter todos os filtros

Adicionar os campos de



Nº PETIÇÕES PRODUZIDAS



670

Nº PROCESSOS ANALISADOS



1.174

Nº PROCESSOS PETICIONADOS



592



PFN RESPONSÁVEL

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

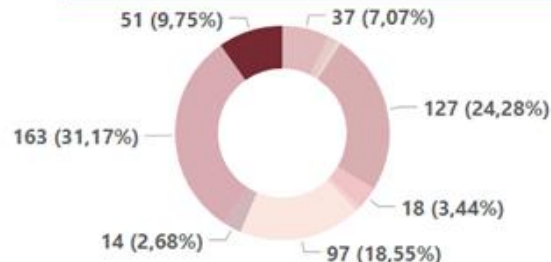
VARAS RESPONSÁVEIS

- ☐ ARACAJU
- ☐ ARAPIRACA
- ☐ ARCOVERDE
- ☐ AREIA BRANCA

LINHA INVESTIGATIVA

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

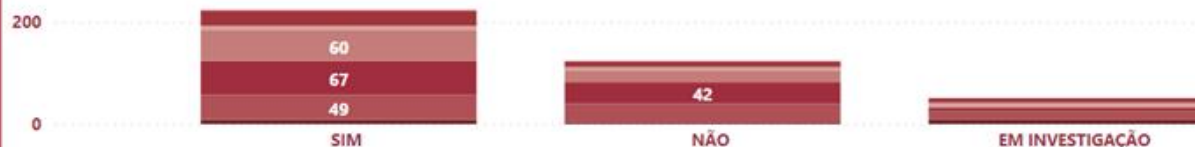
TOTAL DE AÇÃO DE PROVIDÊNCIAS POR AÇÃO DE PROVIDÊNCIAS



- 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...





NÚMERO DE INVESTIGAÇÕES



69

PETIÇÕES DO FLUXO DA IF5



PFN RESPONSÁVEL

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

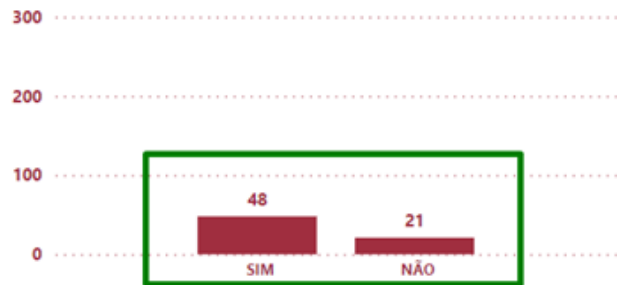
VARAS RESPONSÁVEIS

- ☐ CAMPINA GRANDE
- ☐ GUARABIRA
- ☐ JOÃO PESSOA
- ☐ NATAL

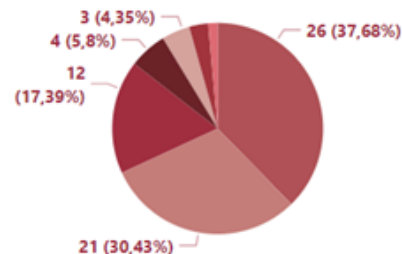
LINHA INVESTIGATIVA

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

CONFIRMA LINHA INVESTIGATIVA POR CONFIRMA LINHA INVESTIGATIVA



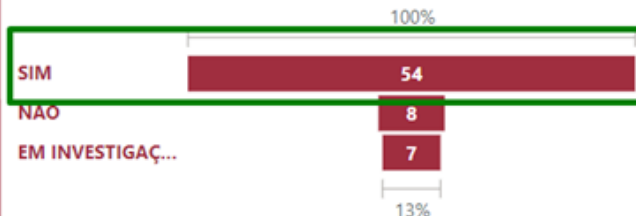
LINHA INVESTIGATIVA POR LINHA INVESTIGATIVA



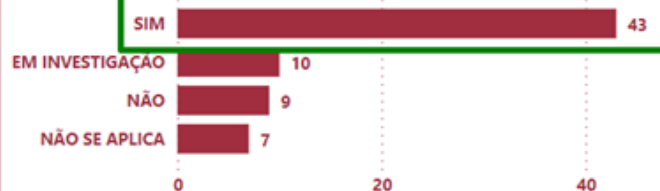
LINHA INVESTIGAT...

- AQUISIÇÕES DOI ...
- FRAUDE EXECUÇ...
- FRAUDE EXECUÇ...
- INDÍCIOS DE SUC...
- OPERAÇÕES DOI ...
- AQUISIÇÕES DOI ...
- VEÍCULOS EMPRE...

BENS E DIREITOS IDENTIFICADOS



CORRESPONSÁVEL IDENTIFICADO



¿Por qué utilizar el análisis de datos?

- Resolución de problemas: El fraude en grupos empresariales fracasó en más del 50% de los casos.
- Ahorro de costos: El Departamento hizo más con menos.
- Innovación: La jurimetría en los litigios judiciales de investigación tributaria.
- Mejora del rendimiento: un 50%

Resultado:

Interpretación y toma de decisiones: optimizar el funcionamiento centrándose en las tipologías de éxito, distribuyendo los asuntos en función de las capacidades de cada uno y disponiendo de un mejor conocimiento del funcionamiento de cada juez/tribunal.

¿Constituye eso un análisis de big data?

Big data: es el uso de potencia informática a gran escala [...] para recopilar, procesar y analizar datos que se caracterizan por su gran volumen, velocidad, variedad y valor (OCDE, 2016).



Volumen masivo de datos

En un principio, el término «big data» hacía referencia a la enorme cantidad de datos generados digitalmente en la red. Esa ingente cantidad incluye correos electrónicos, fotos, textos, vídeos, podcasts... pero también metadatos, como los datos que se producen cada vez que accede a su cuenta bancaria, ve Netflix o compra cualquier producto en una tienda online.



La Administración tributaria genera y utiliza montones de datos

- Registro y documentación de transacciones, operaciones, ventas, servicios y activos con fines fiscales.
- Registros de datos financieros, comerciales y laborales.
- Declaraciones fiscales y otros deberes y obligaciones de carácter tributario.
- Auditorías e inspecciones.
- Dictámenes, procedimientos y decisiones administrativas.



Datos estructurados, semiestructurados y no estructurados

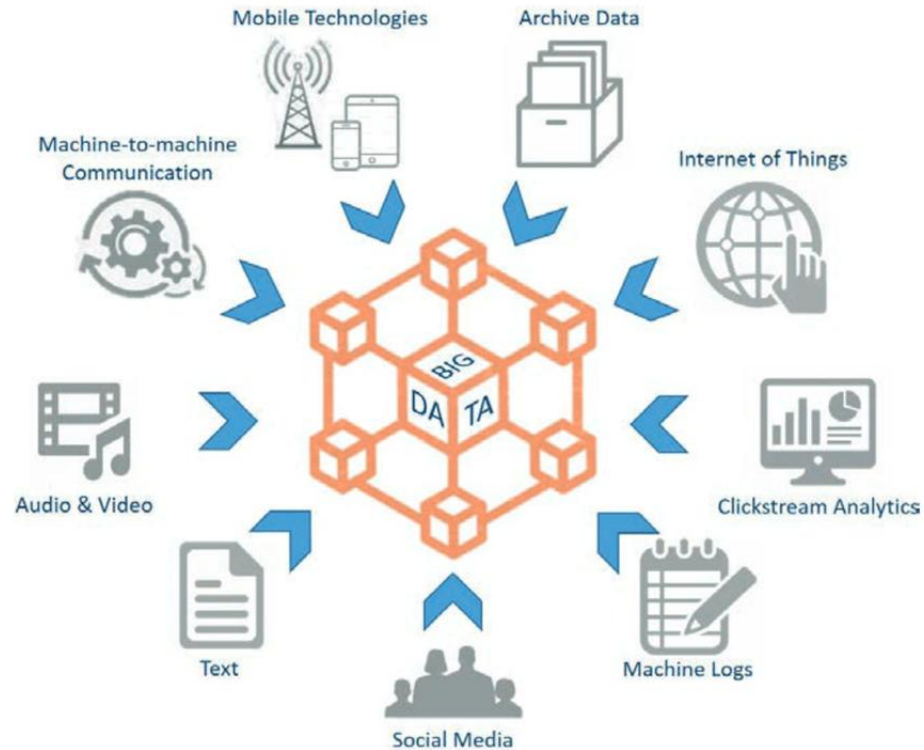
- Los datos estructurados y tabulados son relativamente fáciles de gestionar y son aptos para el análisis estadístico.
- En cambio, los datos no estructurados no son tan fáciles de clasificar e incluyen fotos, vídeos, tuits y documentos de procesamiento de textos.
- Tratar con datos no estructurados plantea ciertas dificultades: dado que estos no pueden almacenarse en bases de datos u hojas de cálculo tradicionales, ha habido que desarrollar herramientas especiales para extraer información útil.

Desafíos:

- Almacenar cantidades absurdamente masivas de datos de forma organizada.
- El 90% de los datos no están estructurados.
- Cuando el número de variables es elevado, también aumenta el número de correlaciones espurias.
- Capacidad de procesamiento extraordinaria: equipos de coste muy elevado, desarrollo de nuevos softwares.



Figure 3.1. Sources of Big Data

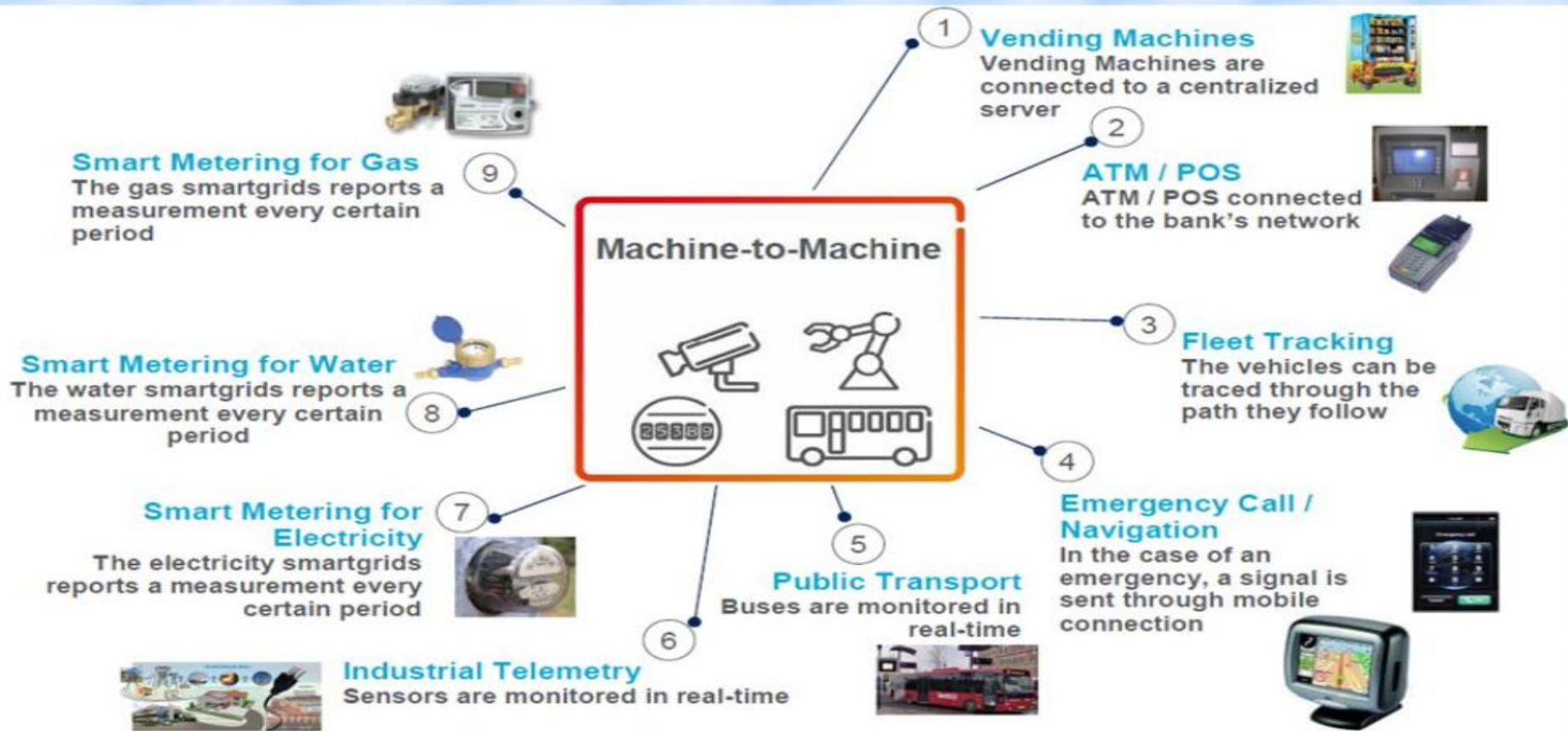


Source: Russian Federal Tax Service.

En Brasil: grandes cifras.

1. Solo en 2022, en Brasil se crearon 3.838.063 nuevas empresas y quebraron 1.695.763. Actualmente hay 20.191.290 empresas funcionando regularmente en el país.
2. Hay más de 30 millones de causas judiciales (ejecuciones tributarias, acciones judiciales de impugnación de impuestos, etc.).
3. Cada año, 40 millones de personas físicas están obligadas a presentar una declaración fiscal.
4. Impuestos federales en apremio: 500.000 millones de USD.
5. En tres años, se han denunciado 20.000 casos de irregularidades.
6. El número de grandes contribuyentes a investigar asciende a 25.000.

Potential M2M services



Las «v» de los big data

La definición de «big data» de Gartner no es mucho más larga que un tuit:

«El término “big data” hace referencia a activos de información de gran volumen, velocidad y variedad que exigen formas rentables e innovadoras de procesamiento de la información para mejorar el conocimiento y la toma de decisiones.»



40 ZETTABYTES

[40 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

[2.5 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	1 000 000 000 000 000 000 000 000 000 000 000	2022 ^[3]
ronna	R	10 ²⁷	1 000 000 000 000 000 000 000 000 000 000 000	
yotta	Y	10 ²⁴	1 000 000 000 000 000 000 000 000 000 000 000	1991
zetta	Z	10 ²¹	1 000 000 000 000 000 000 000 000 000 000 000	
exa	E	10 ¹⁸	1 000 000 000 000 000 000 000 000 000 000 000	1975 ^[4]
peta	P	10 ¹⁵	1 000 000 000 000 000 000 000 000 000 000 000	
tera	T	10 ¹²	1 000 000 000 000 000 000 000 000 000 000 000	1960
giga	G	10 ⁹	1 000 000 000 000 000 000 000 000 000 000 000	
mega	M	10 ⁶	1 000 000 000 000 000 000 000 000 000 000 000	1873
kilo	k	10 ³	1 000 000 000 000 000 000 000 000 000 000 000	1795
hecto	h	10 ²	100 000 000 000 000 000 000 000 000 000 000	
deca	da	10 ¹	10 000 000 000 000 000 000 000 000 000 000 000	

The New York Stock Exchange captures

1 TB OF TRADE INFORMATION

during each trading session



By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

— almost 2.5 connections per person on earth



Velocity

ANALYSIS OF STREAMING DATA



Modern cars have close to

100 SENSORS

that monitor items such as fuel level and tire pressure



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

150 EXABYTES

[181 BILLION GIGABYTES]



By 2014, it's anticipated there will be

**420 MILLION
WEARABLE, WIRELESS
HEALTH MONITORS**

Variety

DIFFERENT FORMS OF DATA

**4 BILLION+
HOURS OF VIDEO**

are watched on
YouTube each month



**30 BILLION
PIECES OF CONTENT**

are shared on Facebook
every month



400 MILLION TWEETS

are sent per day by about 200
million monthly active users



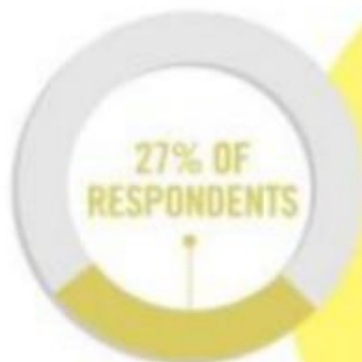
1 IN 3 BUSINESS LEADERS

don't trust the information they use to make decisions



Poor data quality costs the US economy around

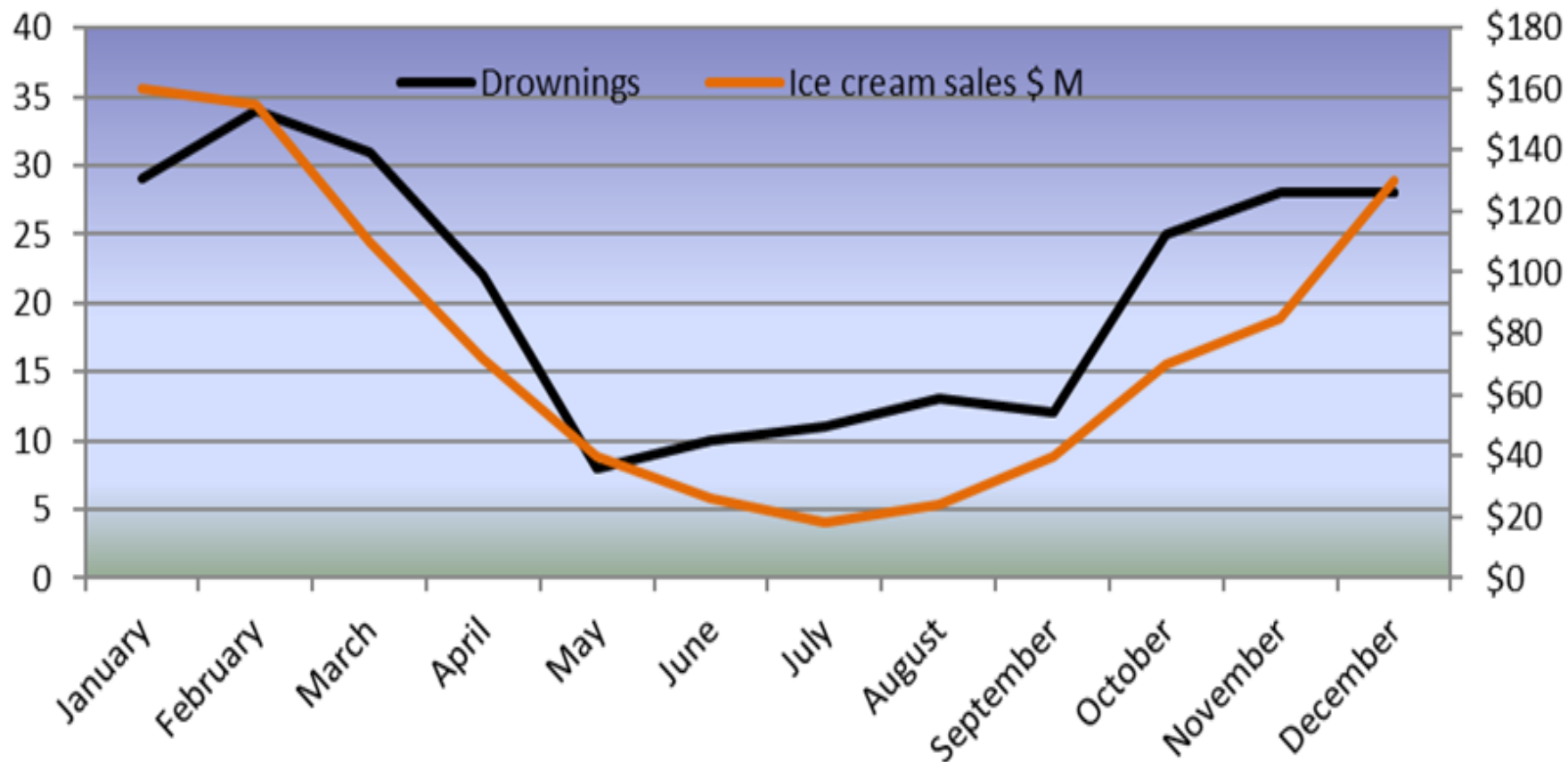
\$3.1 TRILLION A YEAR



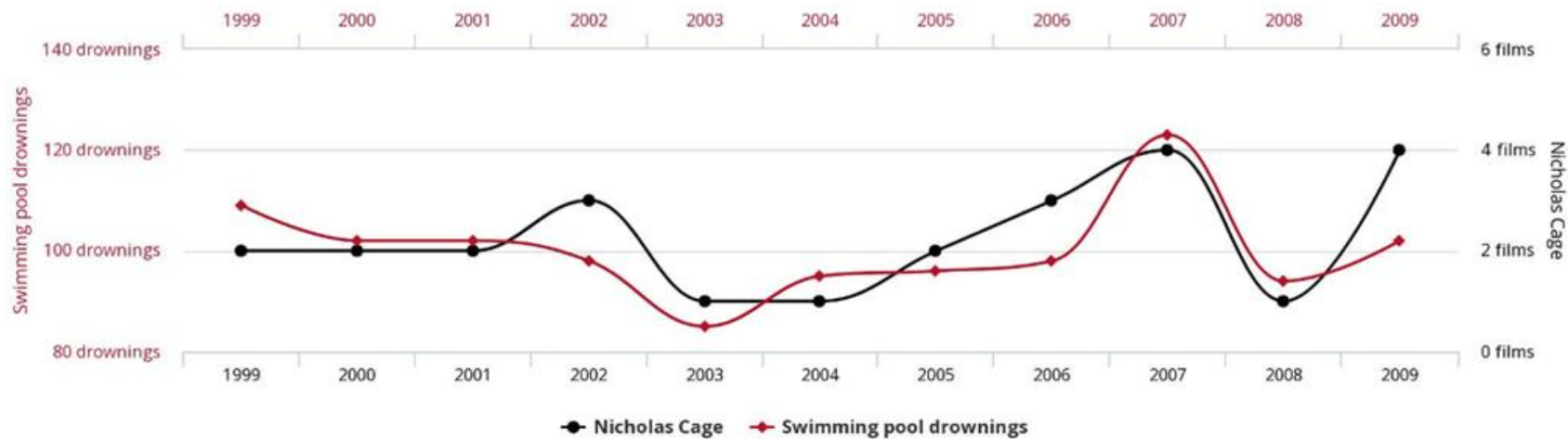
in one survey were unsure of how much of their data was inaccurate

Veracity
UNCERTAINTY OF DATA

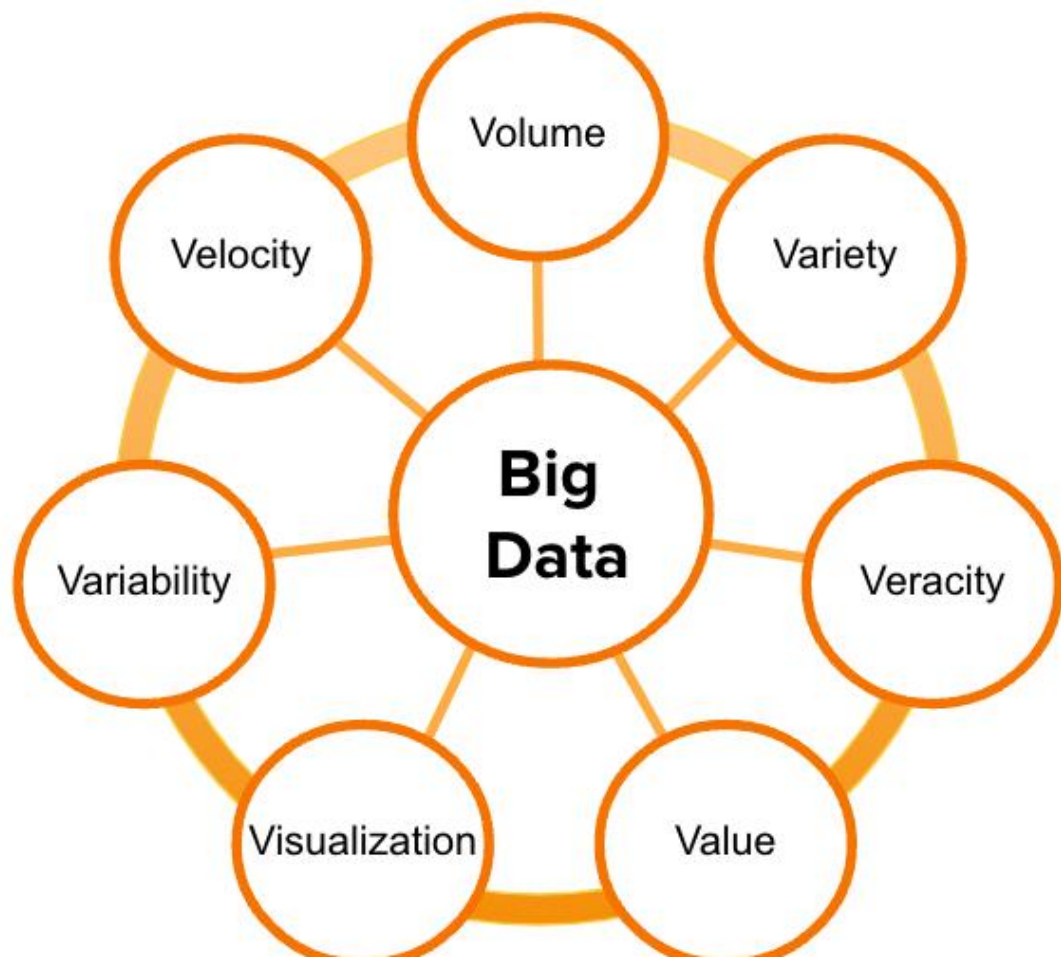
ICE CREAM SALES AND DROWNINGS







7 V'S OF BIG DATA



VALIDACIÓN: mi contribución personal a la «v - tendencia»

Big data e investigación tributaria y penal: más allá de las 7 uves.

¿Puede utilizarse como información, como prueba o como ambas cosas?

¿Es legal según la ley de protección de datos de su jurisdicción?



Análisis de big data



Actividad 1.

Debata en qué áreas del sector público el análisis de big data podría contribuir a mejorar la prestación de servicios públicos en el ámbito de la Administración tributaria.

- Sala de debate.
- Duración: 5 minutos.
- Grupos de 5 participantes. Elija a un portavoz para que exponga las ideas principales del grupo.

Ventajas de utilizar la analítica en la Administración tributaria

FUENTE: SECRETARÍA DEL FTA

GESTIÓN O ADMINISTRACIÓN (*MANAGEMENT*):

- Generar estimaciones de ingresos más precisas para la jurisdicción, con el fin de planificar mejor el gasto público.
- Apoyar a la dirección de la Administración tributaria, al Ministerio de Hacienda y al Parlamento a la hora de evaluar posibles reformas políticas mediante previsiones y análisis hipotéticos, incluida una eventual introducción de nuevas fuentes de ingresos.
- Poner de manifiesto oportunidades que permitan aumentar la eficacia y la eficiencia en el seno de la Administración.

Ventajas de utilizar la analítica en la Administración tributaria

SERVICIOS AL CONTRIBUYENTE :

- Explorar nuevas vías para reducir la carga del contribuyente.
- Mejorar la precisión y la eficacia de las campañas de información al contribuyente.
- Mejorar la calidad de los chatbots.
- Profundizar en la percepción pública sobre la Administración tributaria mediante el análisis de sentimientos expresados en redes sociales.

Ventajas de utilizar la analítica en la Administración tributaria

CUMPLIMIENTO:

- Detectar intentos de fraude mediante la identificación de una serie de factores interrelacionados que, de otro modo, no se considerarían conjuntamente.
- Detectar intentos de fraude a través del análisis de redes y la integración de diversas fuentes de datos.
- Detectar tendencias en el fraude fiscal mediante la vinculación de datos de las redes sociales con la información de los contribuyentes, siempre que la legislación lo permita.
- Evaluar el riesgo de quiebra de los contribuyentes para implementar medidas preventivas adecuadas.
- Optimizar la precisión de la selección de casos de inspección.
- Aumentar la eficiencia y eficacia en la gestión del cumplimiento mediante inspecciones automatizadas.
- Realizar una segmentación más precisa del riesgo de los contribuyentes, con el fin de orientar las medidas de control en función del riesgo de incumplimiento y sus causas subyacentes.
- Utilizar patrones en las fuentes de datos de la Administración tributaria para minimizar la probabilidad de retrasos o incumplimientos en futuras presentaciones o pagos.

Ventajas de utilizar la analítica en la Administración tributaria

FUNCIONES DE LA ADMINISTRACIÓN TRIBUTARIA:

- Detectar la falta de registro y mejorar la exhaustividad del registro.
- Mejorar la gestión de la deuda tributaria.

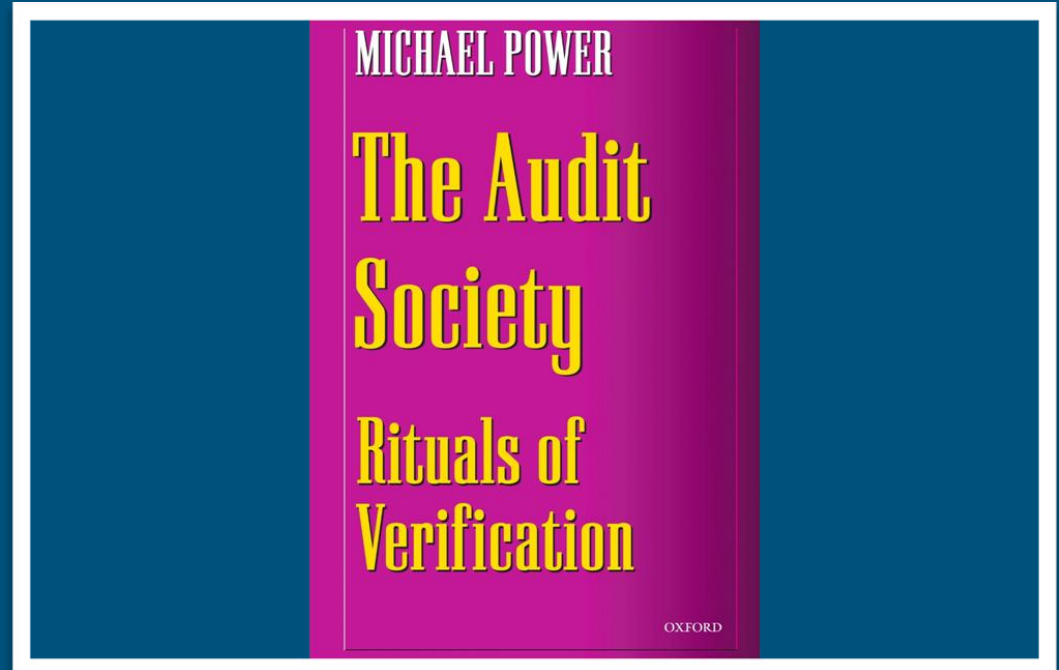
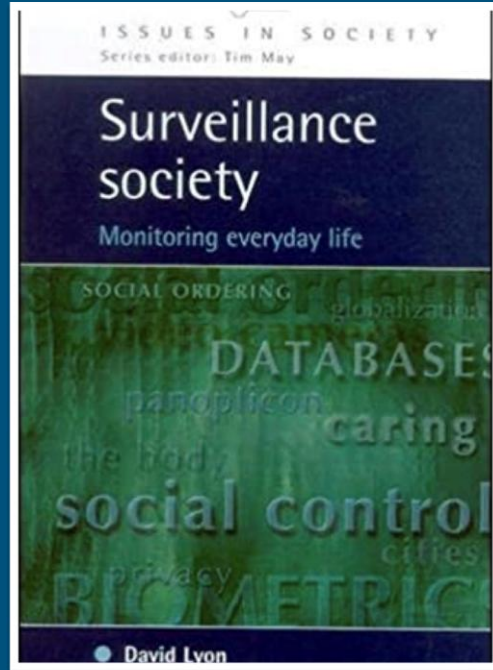
Ejemplos en otros ámbitos:

1. Servicios sanitarios: control de pandemias y vacunación.
2. Educación: control de calidad de los centros de enseñanza y del rendimiento de los alumnos.
3. Servicios sociales: ofrecer servicios adaptados/personalizados según las necesidades individuales: como las de los inmigrantes, las personas sin hogar, etc.
4. Contratación pública: fomentar la competencia y la equidad.
5. Policía: reutilizar datos de dominio público para realizar una labor policial predictiva.
6. Respuesta a emergencias: utilizar datos de sensores en vehículos y procedentes de las redes sociales para proporcionar una respuesta rápida.

INVESTIGACIÓN Y ANÁLISIS DE BIG DATA

1. La conectividad crea una paradoja de los tiempos modernos: «Cuanto más sabemos o aprendemos a través de las redes conectadas, más sabemos y aprendemos sobre nosotros mismos» (Schmidt et al. 2013).
2. La vigilancia forma parte de toda organización burocrática. Implica recopilar, procesar, analizar y compartir datos.
3. La vigilancia está relacionada con el control y la gobernanza.
4. El análisis de big data ha llevado la vigilancia a un nivel completamente nuevo.

ALTAMENTE RECOMENDABLES



DEBATE:

¿Cuáles son los riesgos de utilizar el análisis de big data, teniendo en cuenta:

- a. el organismo en el que usted trabaja
- b. la investigación de delitos fiscales
- c. el uso del análisis de big data como prueba ante los tribunales



Caption

Riesgos emergentes del uso del análisis de big data

- Protocolos de seguridad de acceso, etc.
- Representatividad y relevancia de los datos: La representatividad se refiere a si los datos utilizados ofrecen una representación exhaustiva de la población objeto de estudio, con una representación equilibrada de todas las subpoblaciones pertinentes (evitando sesgos). La relevancia implica que los datos utilizados contribuyan a describir el fenómeno en cuestión sin incluir información exógena (equívoca).
- Privacidad y confidencialidad de los datos.
- Sesgo y discriminación.
- Explicabilidad (al público en general, a los contribuyentes y a los tribunales de auditoría).
- Auditabilidad de los algoritmos y rendición de cuentas.

Riesgos emergentes del uso del análisis de big data

- Divulgación y transparencia (cuestión crítica a la hora de aplicar el análisis de big data en investigaciones tributarias: ¿cómo ser transparente sin revelar las técnicas y estrategias de investigación?)
- Externalización, terceros proveedores.
- Intercambio de información entre distintos organismos públicos (especialmente entre los equipos de investigación tributaria y los servicios de procesamiento penal; estos últimos no pueden acceder a los datos fiscales sin una orden judicial).
- Consideraciones reglamentarias: poderes de investigación vs derechos de los contribuyentes.
- Personal: formación, responsabilización.
- Cuestiones técnicas: almacenamiento, capacidad de procesamiento, herramientas adecuadas y mantenerse actualizado con las nuevas tecnologías.

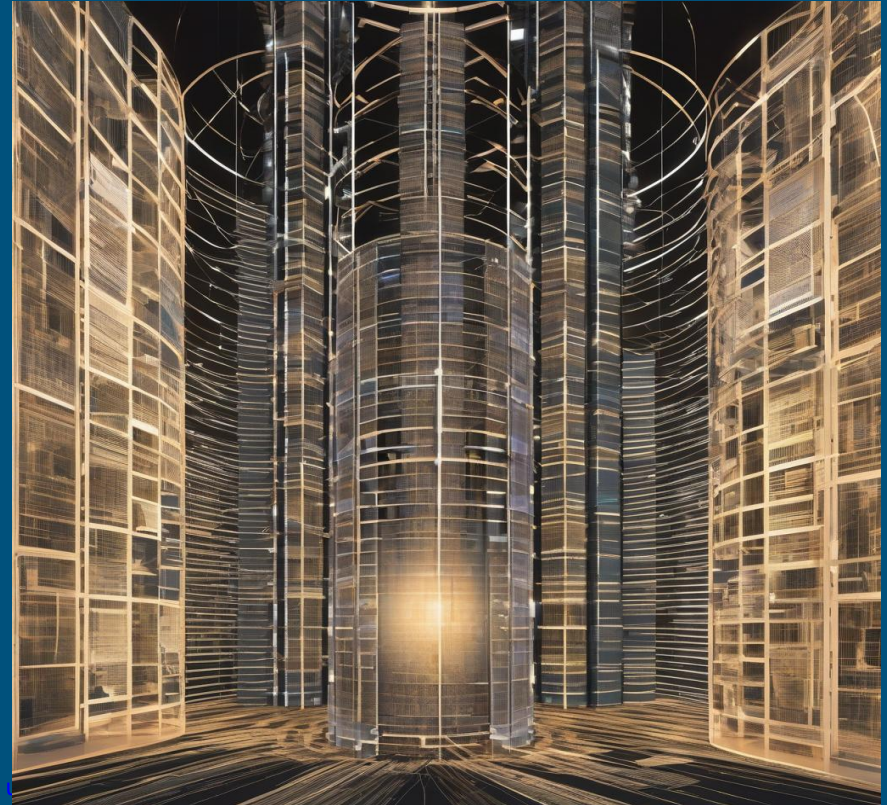
PANÓPTICO

1. Bentham / Foucault
2. Vigilancia 24/7.
3. Nunca sabes cuándo y quién te vigila, solo que te están vigilando.



I - PANÓPTICO

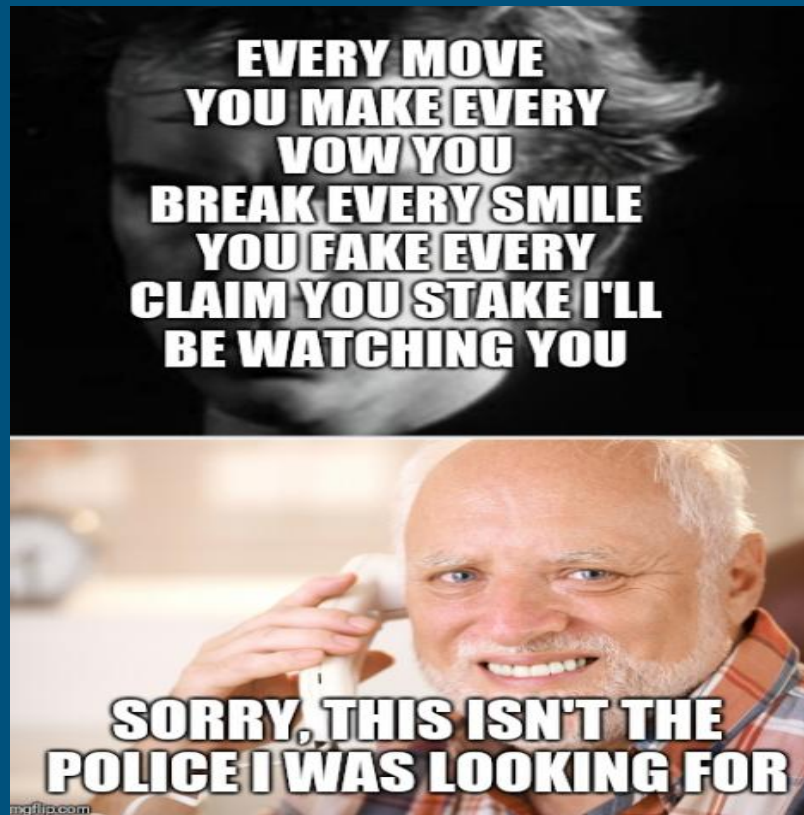
1. No requiere presencia física.
2. No solo vigilar, sino también recopilar, procesar, analizar e influir.
3. Siempre activo, en todas partes.



I-PANÓPTICO Y VALIDACIÓN

El uso de big data con fines de vigilancia debe respetar los procedimientos legales y los derechos de los contribuyentes si se desea que la operación sea considerada legal.

Un buen ejemplo de este enfoque es el Reglamento General de Protección de Datos (RGPD) de la UE.



El RGPD otorga los siguientes derechos (Information Commissioner's Office, 2018):

- Derecho a ser informado sobre la vigilancia
- Derecho de acceso
- Derecho de rectificación
- Derecho de supresión
- Derecho a restringir el tratamiento
- Derecho a la portabilidad de los datos
- Derecho de oposición
- Derechos relacionados con la toma de decisiones automatizada, incluida la elaboración de perfiles

VIGILANCIA VS DERECHOS INDIVIDUALES

Regulación: ¿limitar la vigilancia o restringir los derechos civiles?

Eliminar los poderes gubernamentales de control e inspección no es factible, ya que ambos son inherentes al gobierno.

Sin embargo, estos poderes pueden restringirse de dos maneras: reduciendo el alcance de la acción gubernamental o imponiendo límites a la recopilación, procesamiento y compartición de datos.

El RGPD de la UE optó por la segunda alternativa.



RGPD - Considerando 31

1. Las autoridades públicas a las que se comunican datos personales en virtud de una obligación legal para el ejercicio de su misión oficial, como las autoridades fiscales y aduaneras, las unidades de investigación financiera, las autoridades administrativas independientes o los organismos de supervisión de los mercados financieros encargados de la reglamentación y supervisión de los mercados de valores, no deben considerarse destinatarios si reciben datos personales que son necesarios para llevar a cabo una investigación concreta en interés general, de conformidad con el Derecho de la Unión o de los Estados miembros.

2. Las solicitudes de comunicación de las autoridades públicas siempre deben presentarse por escrito, de forma motivada y con carácter ocasional, y no deben referirse a la totalidad de un fichero ni dar lugar a la interconexión de varios ficheros.

3. El tratamiento de datos personales por dichas autoridades públicas debe ser conforme con la normativa en materia de protección de datos que sea de aplicación en función de la finalidad del tratamiento.

RGPD - Considerando 71

1. El interesado debe tener derecho a no ser objeto de una decisión, que puede incluir una medida, que evalúe aspectos personales relativos a él, y que se base únicamente en el tratamiento automatizado y produzca efectos jurídicos en él o le afecte significativamente de modo similar, como la denegación automática de una solicitud de crédito en línea o los servicios de contratación en red en los que no medie intervención humana alguna.
2. Este tipo de tratamiento incluye la elaboración de perfiles consistente en cualquier forma de tratamiento de los datos personales que evalúe aspectos personales relativos a una persona física, en particular para analizar o predecir aspectos relacionados con el rendimiento en el trabajo, la situación económica, la salud, las preferencias o intereses personales, la fiabilidad o el comportamiento, la situación o los movimientos del interesado, en la medida en que produzca efectos jurídicos en él o le afecte significativamente de modo similar.
3. Sin embargo, se deben permitir las decisiones basadas en tal tratamiento, incluida la elaboración de perfiles, si lo autoriza expresamente el Derecho de la Unión o de los Estados miembros aplicable al responsable del tratamiento, incluso con fines de control y prevención del fraude y la evasión fiscal [...].

RGPD - Considerando 112

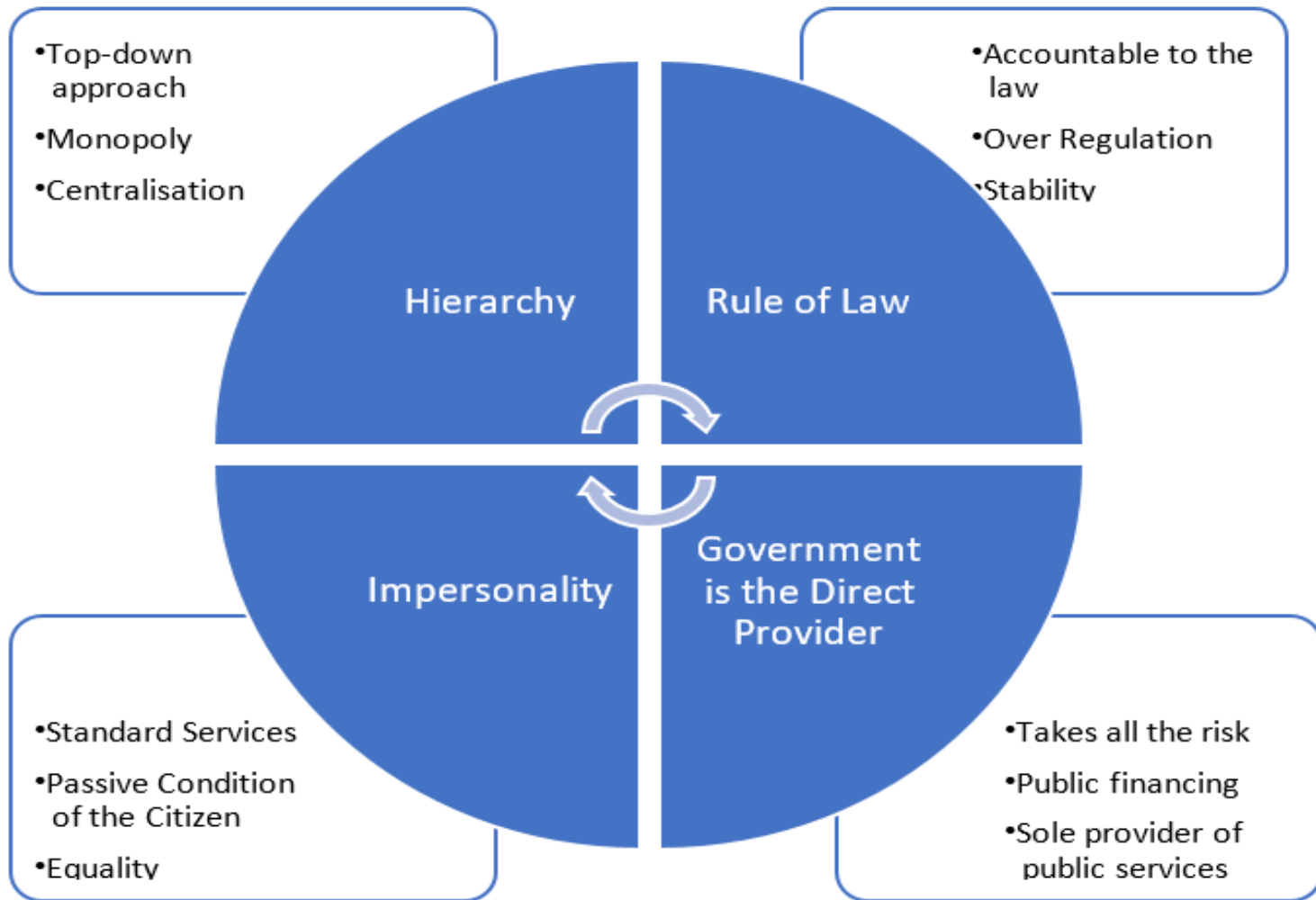
Dichas excepciones deben aplicarse en particular a las transferencias de datos requeridas y necesarias por razones importantes de interés público, por ejemplo en caso de intercambios internacionales de datos entre autoridades en el ámbito de la competencia, **administraciones fiscales o aduaneras**, [...]

Consideraciones en materia de gestión/administración (*management*)

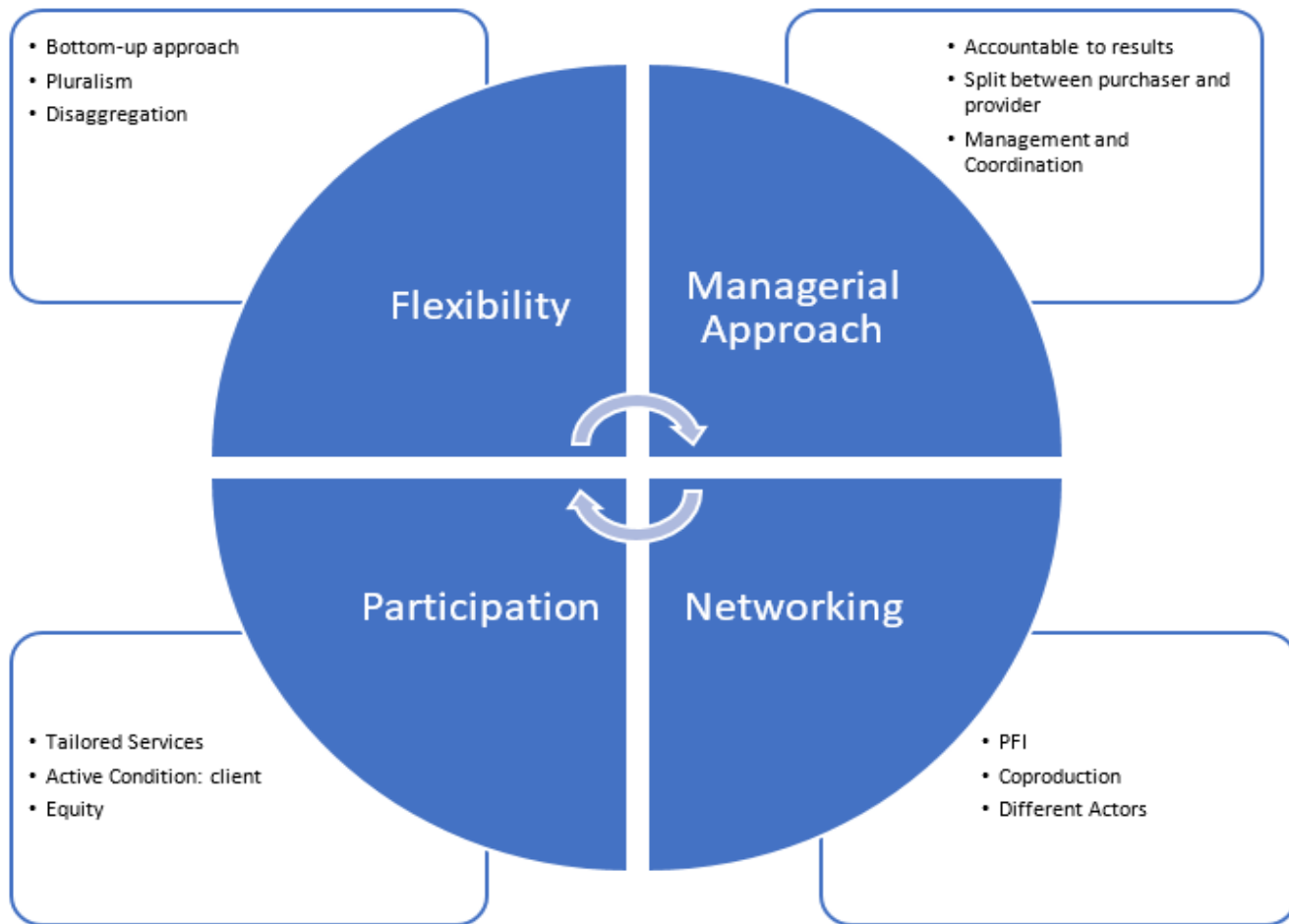
Además de las cuestiones legales y los costos previstos, ¿está su organismo preparado para implantar el análisis de big data en la investigación tributaria?

Dos modelos diferentes:
Administración burocrática vs Gestión de economía mixta





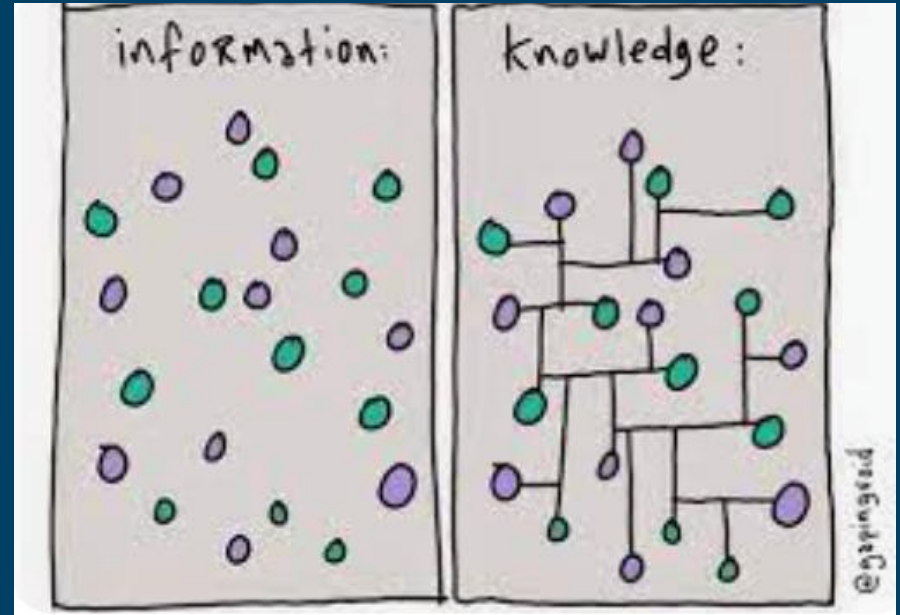
Source: own making.



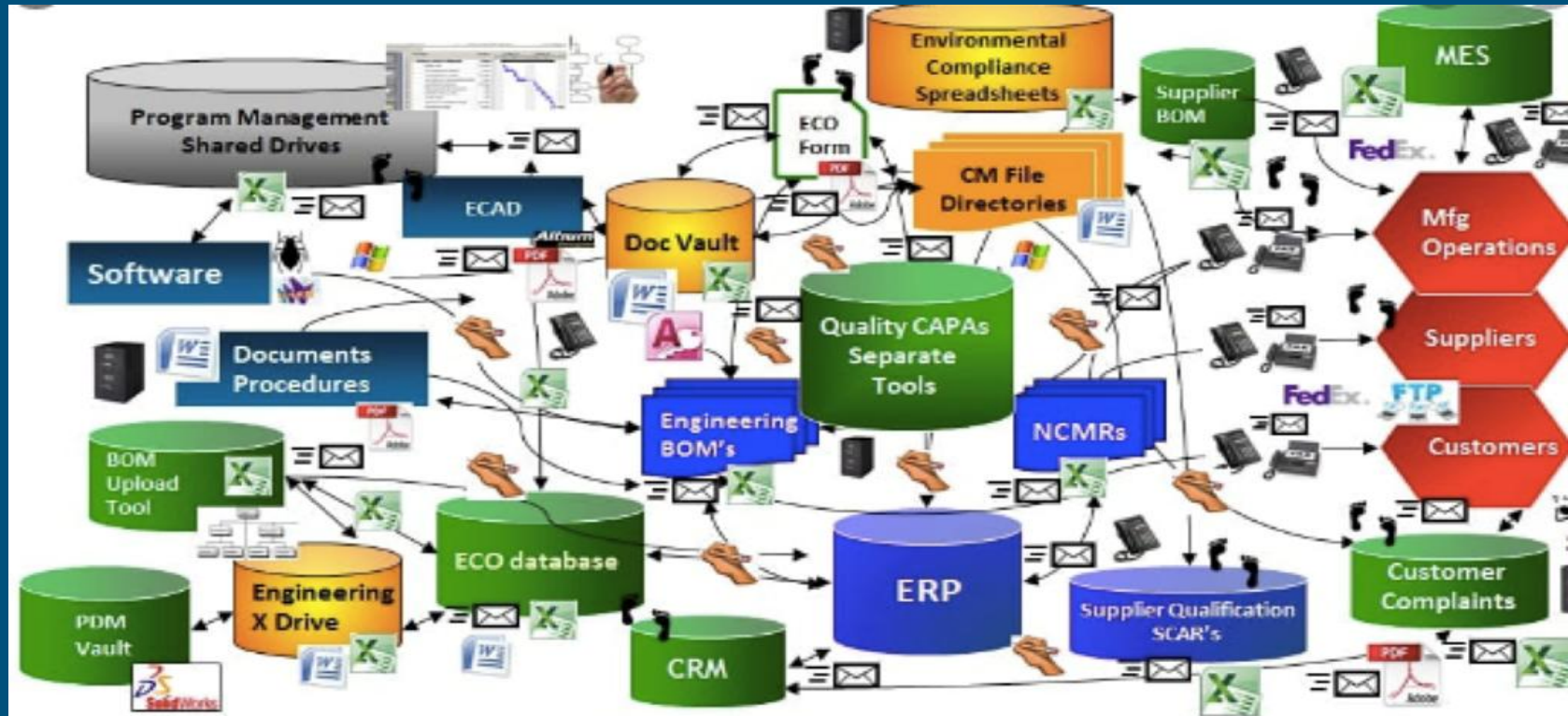
Source: own making.

Conectando los puntos

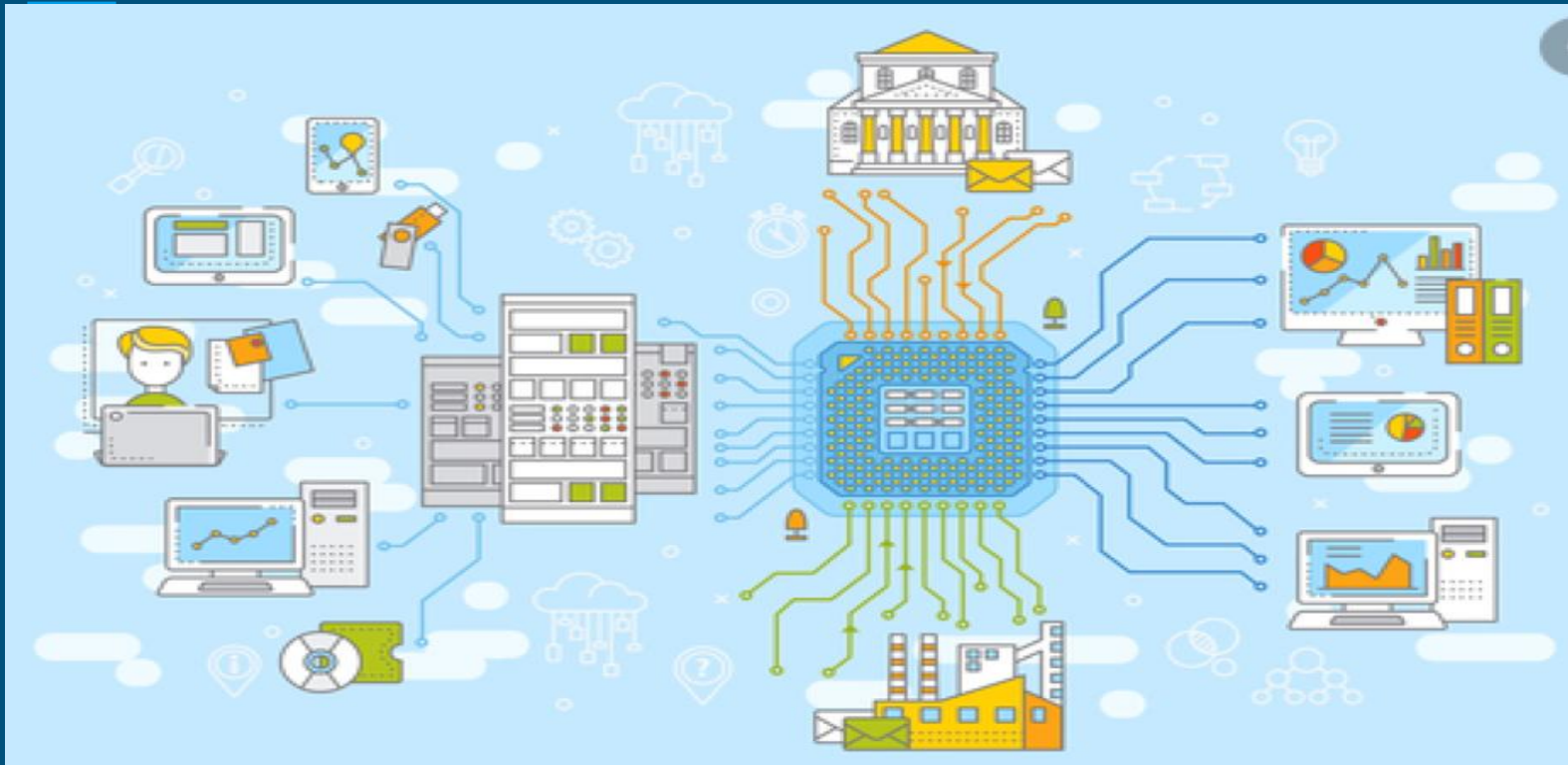
Ejemplo de uso masivo del análisis de big data: El caso de la PGFN.



Cómo era antes



Cómo es ahora



APP DÍVIDA ABERTA



ESCÁNER DE CÓDIGOS QR para comprobar facturas (¿debe impuestos la empresa?)

Comprobar el cumplimiento de las obligaciones tributarias de la empresa mediante el número de identificación

Encontrar deudores locales (mediante georreferenciación)

Canal de denuncia

Servicios para el cumplimiento de las obligaciones tributarias

¡Noticias de la PGFN!

Análisis de big data en investigaciones tributarias

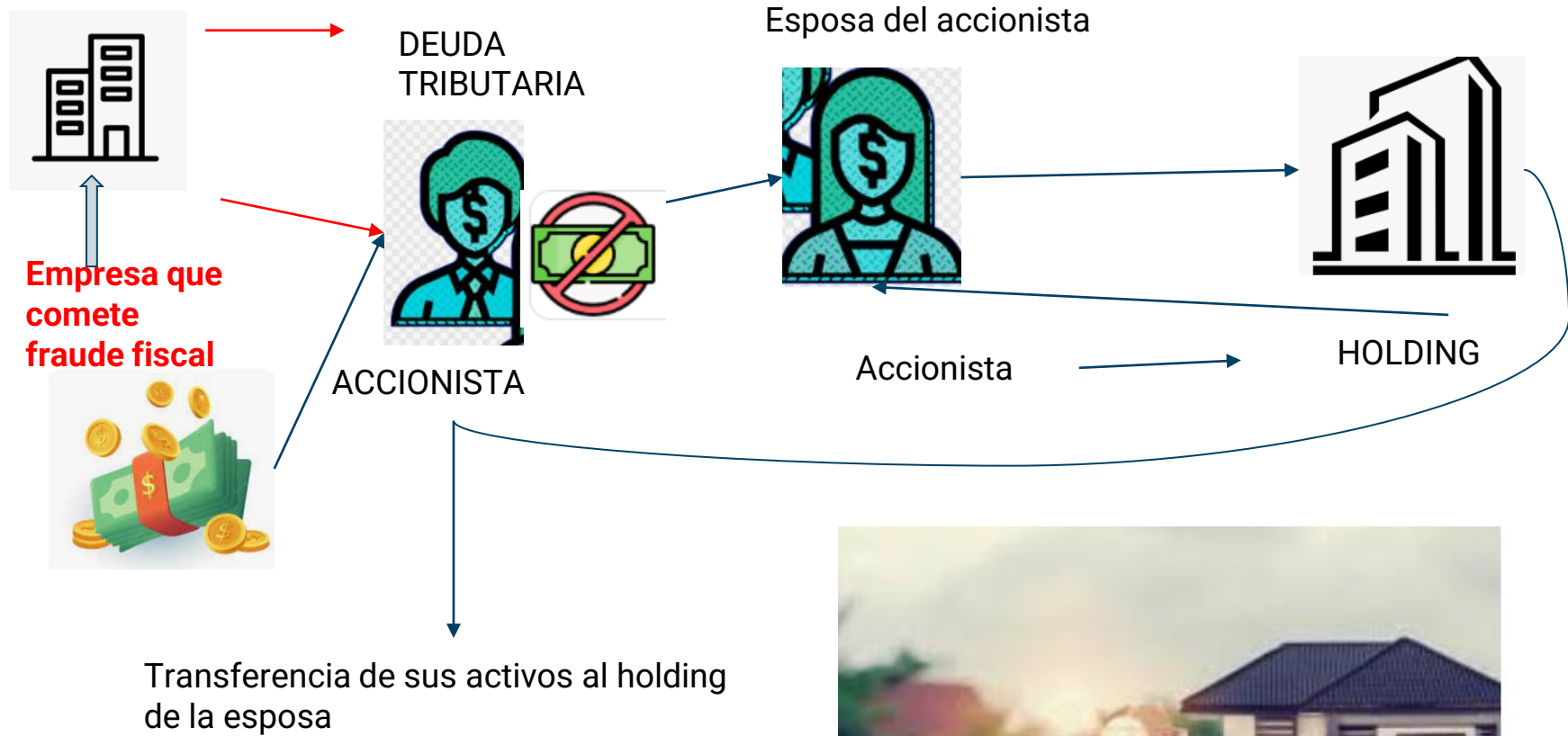
Extracción de datos de más de 80 bases de datos diferentes.

Datos estructurados y semiestructurados.

Escritura del algoritmo de búsqueda de patrones.

Combinación y análisis de datos.

Resultado esperado: Identificación de sociedades fantasma o ficticias + deuda tributaria + activos transferidos a un holding + beneficiario efectivo = accionista de sociedad ficticia.



¡Una aguja en un pajar!

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

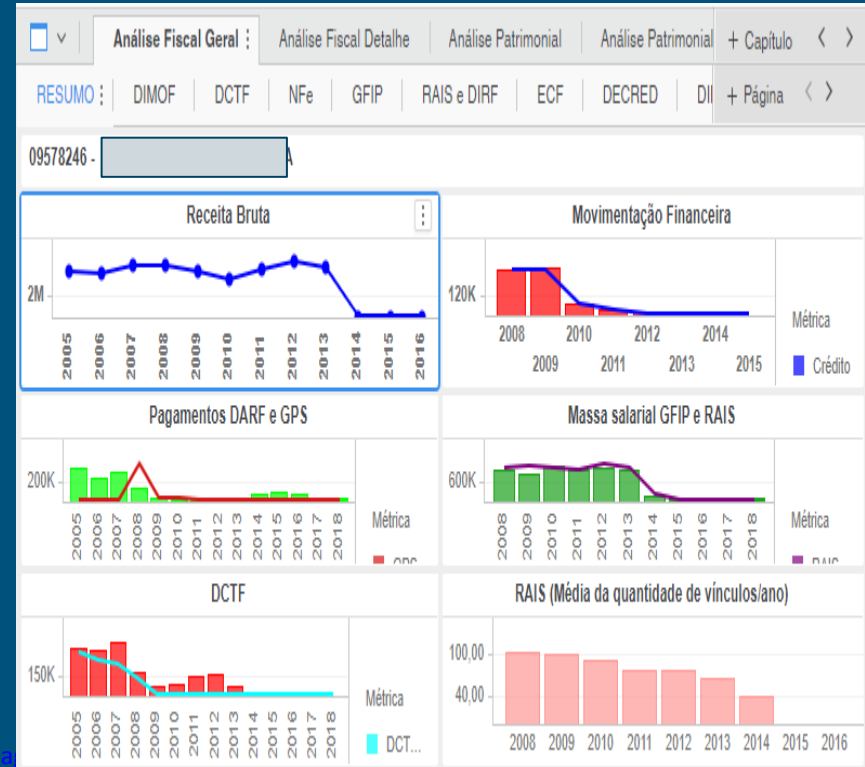
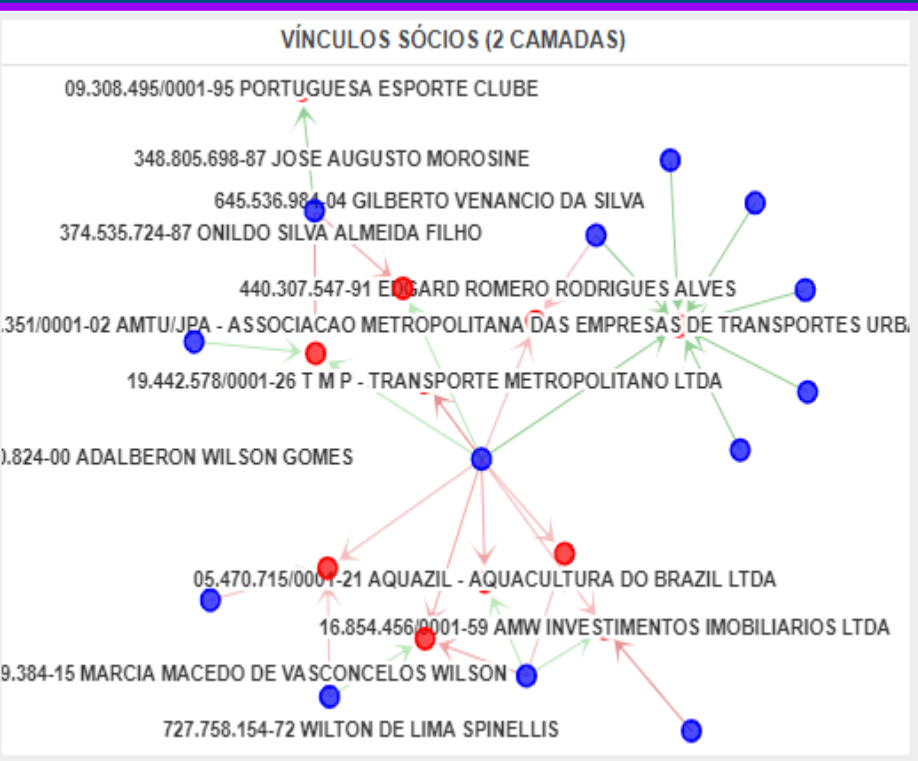
- Spouse/kids own holding concentrating all the assets
- Had the relatives the financial means to keep the holding?

Explicação de la tabla

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

Local	Federal State	City	Company ID Number	Company Name	Company'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	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	PAULINA	02270967	BOMM PETRO DISTRIB DE DERIVADOS DE PETROLEO E ALCOOL LTDA	BAXADA	INEXISTENCIA DE FATO	372.769.409,20	SÓCIO GERENTE	020114958 30	LUZ CARLOS MONTEIRO	1998-09-01 00:00:00.0	1999-08-11 00:00:00.0	Companheiro(a) contribuinte, tenha filho, viva há mais de 5 anos ou cônjuge	<<<	04346325690	MARIA DE FATIMA MOURÃO PAIS MONTEIRO	SÓCIO ADMINISTRADOR	Ativo	1999-10-11 00:00:00.0	9999-01-01 00:00:00.0	03441931
3ª REGIÃO	SP	PAULINA	02270967	BOMM PETRO DISTRIB DE DERIVADOS DE PETROLEO E ALCOOL LTDA	BAXADA	INEXISTENCIA DE FATO	372.769.409,20	SÓCIO GERENTE	020114958 30	LUZ CARLOS MONTEIRO	1998-09-01 00:00:00.0	1999-08-11 00:00:00.0	Companheiro(a) contribuinte, tenha filho, viva há mais de 5 anos ou cônjuge	>>>	04346325690	MARIA DE FATIMA MOURÃO PAIS MONTEIRO	SÓCIO ADMINISTRADOR	Ativo	1999-10-11 00:00:00.0	9999-01-01 00:00:00.0	03441931
3ª REGIÃO	SP	PAULINA	02270967	BOMM PETRO DISTRIB DE DERIVADOS DE PETROLEO E ALCOOL LTDA	BAXADA	INEXISTENCIA DE FATO	372.769.409,20	SÓCIO GERENTE	438247658 53	ROBERTO LUZ BICUDO FERRARO	1997-12-08 00:00:00.0	1999-08-11 00:00:00.0	Companheiro(a) contribuinte, tenha filho, viva há mais de 5 anos ou cônjuge	<<<	02496015836	MARTA MAOLINE CHAVES	SÓCIO	Ativo	2015-07-31 00:00:00.0	9999-01-01 00:00:00.0	03464098
3ª REGIÃO	SP	PAULINA	02270967	BOMM PETRO DISTRIB DE DERIVADOS DE PETROLEO E ALCOOL LTDA	BAXADA	INEXISTENCIA DE FATO	372.769.409,20	SÓCIO GERENTE	438247658 53	ROBERTO LUZ BICUDO FERRARO	1997-12-08 00:00:00.0	1999-08-11 00:00:00.0	Companheiro(a) contribuinte, tenha filho, viva há mais de 5 anos ou cônjuge	>>>	02496015836	MARTA MAOLINE CHAVES	SÓCIO	Ativo	2015-07-31 00:00:00.0	9999-01-01 00:00:00.0	03464098

Hacerlo visual: Derecho visual.



Breve recorrido por algunos sistemas de big data de la PGFN

- RELATÓRIOS DE APOIO
- ANÁLISIS
- FRANCO TIRADOR / RASTREADOR / RAPTOR
- DW



¿Es caro?

Implementación: hardware, sistema de almacenamiento, Internet de alta velocidad, capacidad de procesamiento

Formación del personal: posgrado en análisis de big data, ciencia de datos, cursos sobre IA, etc.

Costos de mantenimiento

Digitalización, gobernanza digital

Búsqueda de bases de datos

Debate: ¿Es seguro subcontratar (almacenamiento, procesamiento, nube, etc.)?

¿Y si el «sistema» se equivoca?

1. ¿Se puede responsabilizar al algoritmo?
2. Utilizar las pruebas digitales como pruebas de apoyo, nunca como única base para la condena.
3. Datos erróneos: error en la recopilación, el muestreo, el almacenamiento.
4. Relaciones espurias: mala interpretación.
5. Fascinación de la Torre de Babel.



¿Fraude o divorcio?

Sistema programado para encontrar *fraude de transferencia de activos*: simular una transferencia legal de activos a una persona de confianza (normalmente un familiar) para evitar embargos, confiscaciones, etc.

Objetivo localizado: Mister Love, que debía 500.000 USD en impuestos, transfirió a su esposa 4 pisos valorador en 2 millones.

Sin embargo, investigaciones posteriores revelaron que se había divorciado y que en el acuerdo a su exmujer se le habían adjudicado dichos pisos.

¿Grupo empresarial fraudulento?

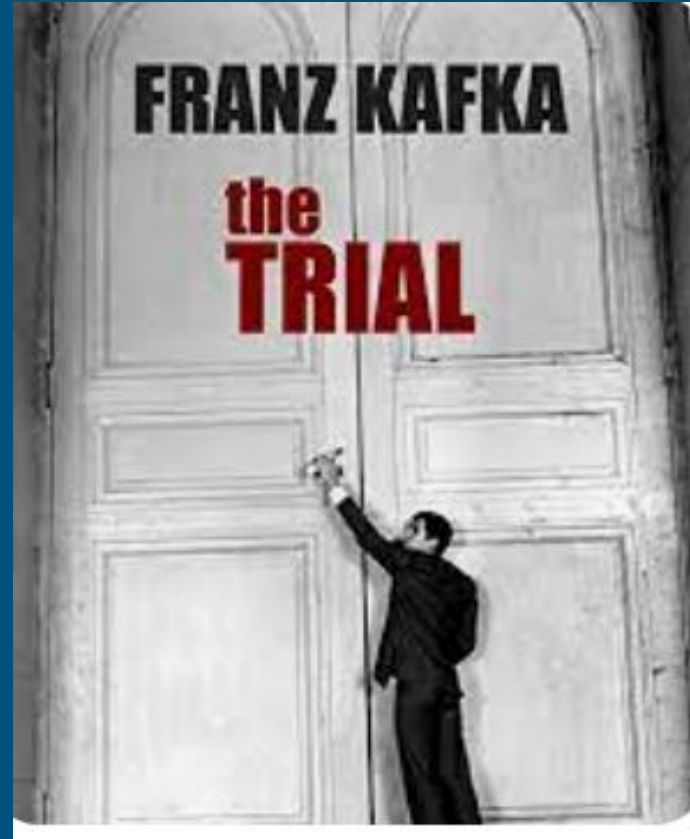
Según los resultados del análisis de big data, 5 empresas agroalimentarias diferentes operaban en el mismo lugar (una pequeña habitación en un edificio antiguo de la ciudad). Todas tenían el mismo contable, quien registró los estatutos de las empresas en el Registro Mercantil. Cada empresa poseía unos 200.000 USD. El contable podría ser el facilitador.

Sin embargo....

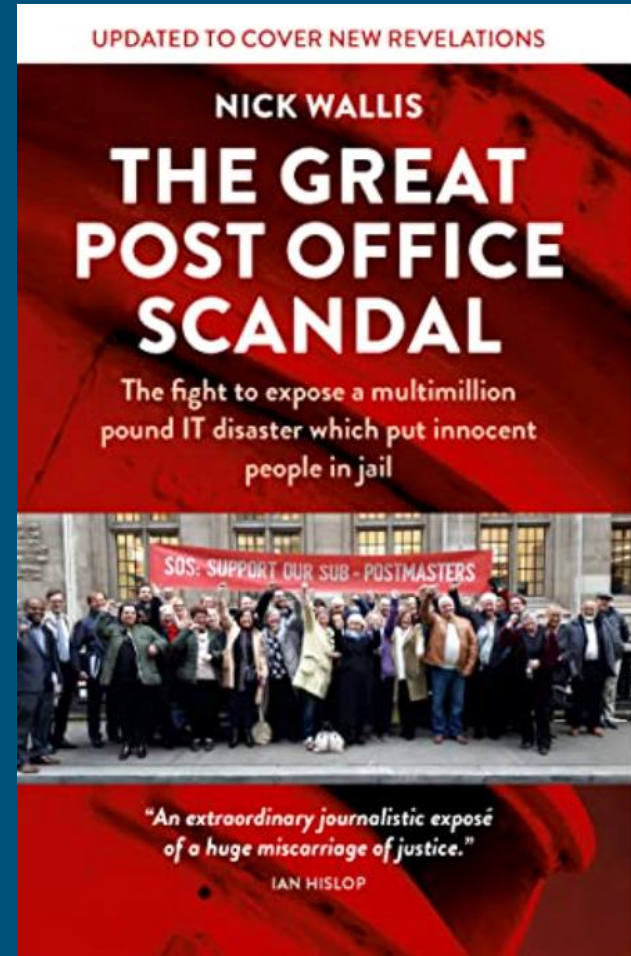


El escándalo de Correos

- Una historia de cómo el exceso de confianza en el análisis de datos arruinó casi 900 vidas: ¡«El SISTEMA no puede equivocarse»!
- Algoritmo y responsabilidad.
- El peso como prueba judicial de los datos procesados electrónicamente.
- Explicabilidad: ¿entienden realmente los jueces y abogados los big data?
- Análisis de big data y credibilidad: transparencia, datos de dominio público, derechos de los contribuyentes.

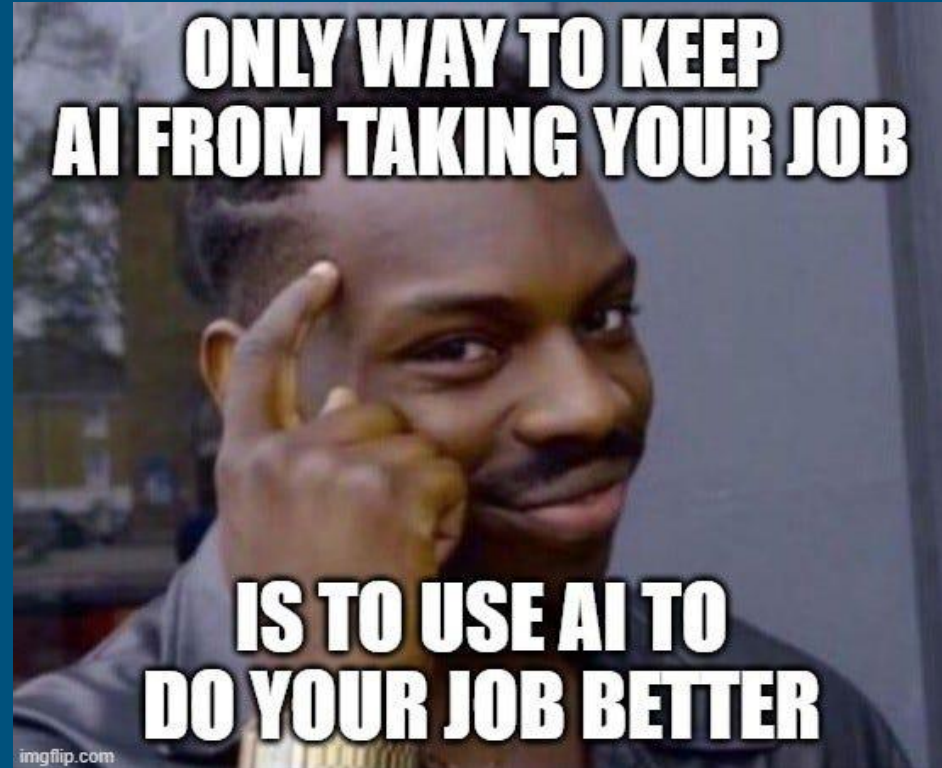


¡Recomendación!



¿Y qué pasa con la IA?

1. La IA y los big data son tecnologías complementarias que se combinan para aprovechar el potencial de los datos.
2. La IA proporciona la inteligencia necesaria para analizar e interpretar los big data, mientras que estos últimos aportan el combustible para que la IA aprenda y mejore.
3. La combinación de IA y big data está impulsando la innovación y transformando industrias de diversos sectores.



Técnicas de IA para la investigación de delitos fiscales

Aprendizaje automático (*Machine Learning*):

- Aprendizaje supervisado: Entrenamiento de modelos con datos etiquetados para predecir resultados (por ejemplo, identificar declaraciones de impuestos defraudatorias).
- Aprendizaje no supervisado: Descubrimiento de patrones en datos no etiquetados (por ejemplo, agrupar a los contribuyentes con comportamientos similares).

Aprendizaje profundo (*Deep Learning*):

- Redes neuronales: Algoritmos complejos capaces de aprender de grandes conjuntos de datos (por ejemplo, analizar datos no estructurados como correos electrónicos o documentos).

Procesamiento del lenguaje natural (*Natural Language Processing* o «NLP»):

- Análisis de textos: Extracción de información de datos textuales (por ejemplo, identificar palabras clave relacionadas con la evasión fiscal en documentos).



Aplicaciones de la IA en la investigación de delitos fiscales

Evaluación de riesgos:

- Modelado predictivo: Identificar a los contribuyentes de alto riesgo basándose en datos históricos y patrones de comportamiento.
- Detección de anomalías: Identificar actividades o transacciones inusuales que puedan indicar un comportamiento sospechoso de fraude.

Análisis y visualización de datos:

- Minería de datos: Descubrir patrones y relaciones ocultos en grandes conjuntos de datos.
- Visualización de datos: Presentar datos complejos de forma clara y comprensible.

Selección y priorización de casos que inspeccionar:

- Labor de inspección basada en el riesgo: Centrarse en los contribuyentes que presentan una mayor probabilidad de incumplimiento.
- Programación inteligente de las inspecciones: Optimizar los recursos de inspección en función del riesgo y la carga de trabajo.

Detección del fraude:

- Reconocimiento de patrones: Identificar tramas y patrones de comportamiento sospechosos de fraude.
- Control en tiempo real: Detectar actividades fraudulentas en el mismo momento en que se producen.

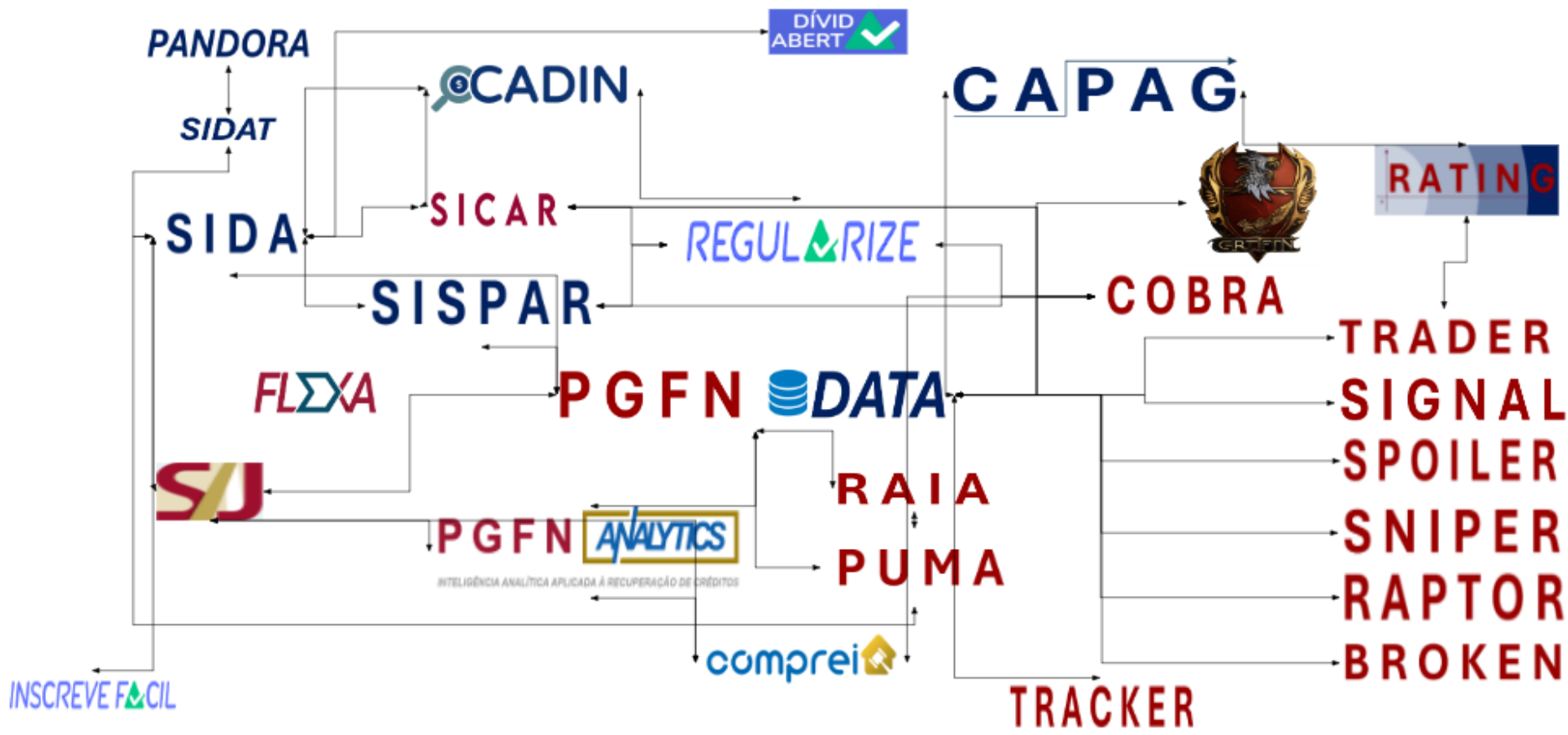
Investigación sobre lavado de activos:

- Análisis de redes: Analizar transacciones financieras a fin de identificar posibles conexiones sospechosas.
- Análisis del comportamiento: Identificar patrones de gasto o transferencias de activos inusuales.

Qué ha estado haciendo la PGFN con IA + big data



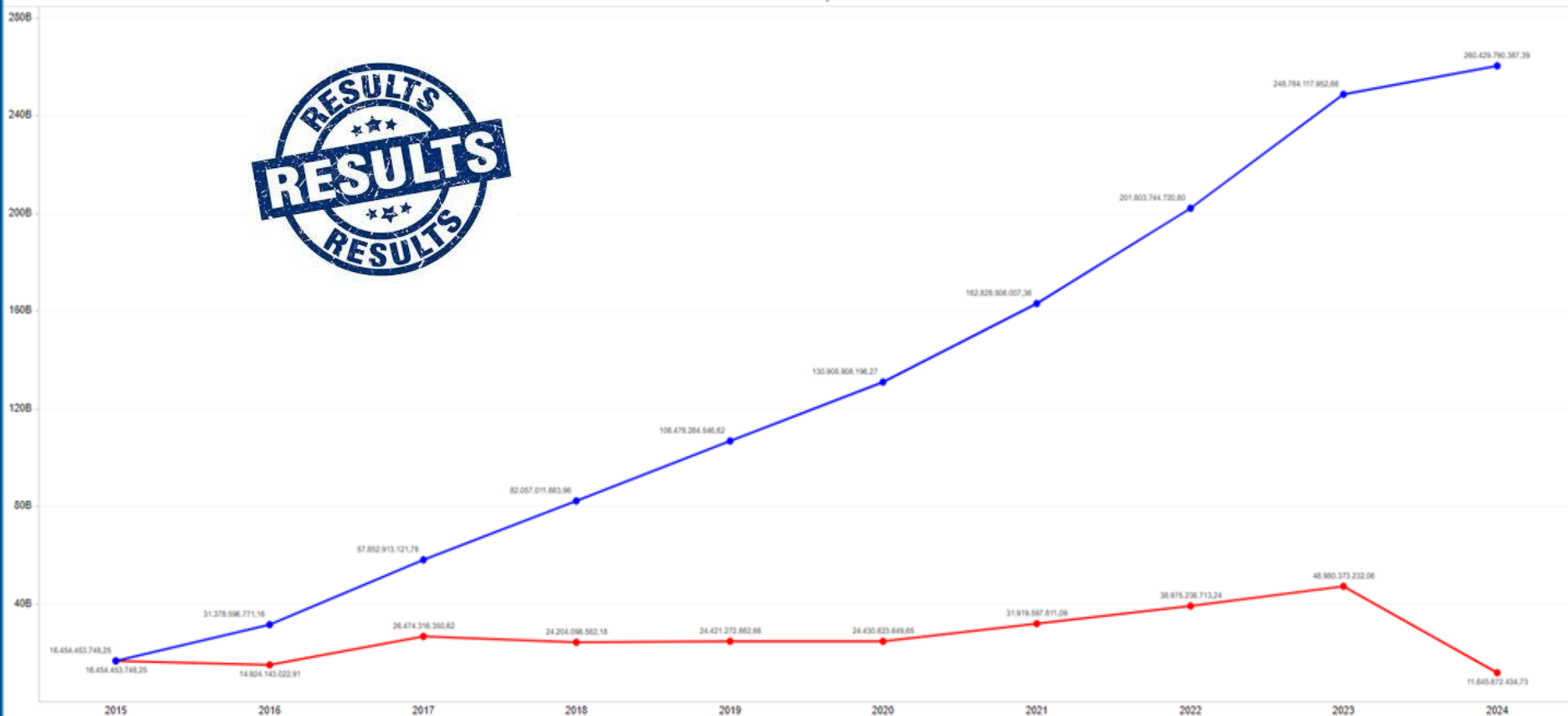
Sistemas y aplicaciones «conversan» entre sí



!!!Resultados!!!



RECUPERAÇÃO POR ANO - NOMINAL



El problema de Palantir

- Cuando se utiliza un Palantír para mirar más allá, no se sabe quién más está mirando... (cuestiones de privacidad).
- El Palantir puede mostrar solo lo que quieres ver o lo que alguien quiere que veas (sesgo / interferencia / ingeniería del comportamiento).
- Puede que veas cosas y situaciones que no existen pero que se presentan como reales (alucinación).



HEY

**THERE'S SOMETHING
IN HERE CALLED "AI"**

imgflip.com

 **ChatPGFN**

restricted Use - A usage restraint

Actividad 2.

Usted ha sido contratado como consultor fiscal y de big data para asesorar al Ayuntamiento de Digitópolis sobre cómo mejorar la recopilación de datos de los impuestos municipales y del IVA, detectar posibles delitos fiscales derivados de la evasión y optimizar la recuperación de activos. En la ciudad de Digitópolis la administración electrónica marca tendencia, así que las siguientes áreas dentro del gobierno son 100% digitales:

- Ciudadanos, empresas y propiedades (por ejemplo, registros base, transacciones)
- Las transacciones entre los contribuyentes y la Administración, incluidas las facturas

A usage restraint



Actividad 2

- Contratación pública y gasto público
- Organismos públicos y empleados de la Administración
- Registro de empresas y trabajadores
- Datos geográficos relacionados principalmente con el catastro
- Datos geoespaciales
- Contenidos relacionados con el tráfico y la titularidad de vehículos
- Propiedad inmobiliaria y otros derechos sobre bienes inmuebles

Actividad 2

Asesorar al Ayuntamiento sobre cómo podrían utilizarse los big data para:

1. Mejorar la recaudación de los impuestos municipales:
2. Mejorar la recaudación del IVA:
3. Detectar infracciones fiscales:
4. Optimizar la recuperación de activos:

Sala de debate. Duración: 15 minutos. Grupos de 5 a 10 participantes. Elija a un portavoz para que exponga las ideas principales del grupo.

Debate y recapitulación

1. Análisis de big data/IA e investigaciones tributarias.
2. Más información.
3. Establecer relaciones.

Arthur Moura

E-mail: prof.acmoura@gmail.com

WhatsApp: +55 83 99350-4334

