

AUDITORÍA Y GESTIÓN DE LOS FONDOS PÚBLICOS

Jorge Balaguer de la Riva
Técnico de Gestión de Auditoría
Área de Ayuntamientos. Audiencia de Cuentas de Canarias

“There will be no highlights on the eleven o'clock News(...)
The revolution will be no re-run brothers,
will be live”

Jill Scott-Heron.1971

La transformación digital no será televisada: el Big Data en la Auditoría Pública

RESUMEN/ABSTRACT:

El artículo aborda la necesidad de digitalizar el proceso auditor puesto que gran parte de nuestra realidad se ha “datificado”. Una vez transformados los datos es posible emplear técnicas de minería de datos, de textos o de procesos como herramientas para la auditoría. Incorporar éstas y otras prácticas del sector privado en la auditoría pública no está exento de dificultades. Los retos de la transformación digital en auditoría implican la adquisición de nuevas competencias y cambios estructurales en nuestras organizaciones.

This paper addresses on the need to digitize the auditing operations taking into account taht much of the information has been “datafied”. After a ETL process, it is possible to use data, text and process mining as tools for auditing.

But incorporating these, as well as private sector good practices into public auditing raises several difficulties. The challenges of digital transformation in auditing involve the acquisition of new skills and structural changes in public organizations.

AUDITORÍA PÚBLICA, DIGITALIZACIÓN, ANALÍTICA BIG DATA Y MINERÍA
PUBLIC AUDIT, DIGITALIZATION, BIG DATA ANALYTICS, MINING

PALABRAS CLAVE/KEYWORDS:

INTRODUCCIÓN. UNA DOBLE TRANSFORMACIÓN DIGITAL EN LA ADMINISTRACIÓN Y EN EL CONTROL EXTERNO

El cambio social asociado a la transformación digital obliga a hacer una reflexión estratégica sobre el papel de la Administración y las expectativas de una ciudadanía cada vez más digitalizada. Estas notas pretenden referirse a algunas consecuencias prácticas sobre la técnica fiscalizadora señalando algunos aspectos que pueden tanto impulsar como dificultar el proceso de adaptación de las ICEX al nuevo paradigma de *Big Data*.

No vamos a abordar el reto que supone para la Administración Pública aplicar sus políticas empleando tecnologías en evolución permanente y con una adecuada sistematización de los datos de gestión. Hablarle a la sociedad con sus herramientas es mejorar la función social de la Administración, pero tampoco vamos a abordar en estas notas cómo hay que articular estos servicios públicos desde la lógica de los receptores de los mismos.

La tecnología puede ser empleada por la Administración como palanca de participación en la gobernanza e intermediación con la ciudadanía. Tan sólo apuntar que para lograr estos objetivos, la ciudadanía espera que la Administración¹ ofrezca una experiencia de servicios digitales públicos parecida a sus experiencias de consumo, caracterizadas por la inmediatez, la exhaustividad, la automatización y la predicción².

Las campeonas de la intermediación con la ciudadanía (en este caso, consumidores) son las grandes tecnológicas, las denominadas GAFAs³. La alta productividad de estas compañías viene explicada por el aprovechamiento a fondo de las ventajas que ofrece la economía digital: unos bajos costes de transacción, un coste marginal de los productos que tienden a cero (es decir, que producir bienes adicionales no cuesta) y, por último, un traslado al consumidor de los costes de búsqueda, entrega, desplazamiento, decisión y “compromisos” asumidos a golpe de click. Estas empresas, ni van a resolver los conflictos sociales ni van a garantizar los derechos y libertades, pero están difuminando los límites entre los grandes sectores tradicionales en los que se organiza la economía desplazando incluso al sector público.

La Administración tampoco tiene costes para intermediar y convertirse en una plataforma de intercambio operativo con los agentes de nuestra sociedad, pero

tiene un problema para encontrar los estímulos para su adopción. Ignorar la transformación digital genera una frustración y alejamiento creciente en la ciudadanía que nos hace perder legitimidad especialmente entre las nuevas audiencias. Puede parecer exagerado que la Administración pueda llegar a dejar de cumplir con su función social y desaparecer en algunos ámbitos públicos sustituida por estas empresas, pero piense si alguien todavía consulta un callejero municipal oficial en la web del Ayuntamiento.

Incorporar la ciencia de datos a la Auditoría Pública

En el ámbito del control externo, de momento, no parece estar en cuestión la necesidad social de evaluaciones independientes y objetivas relativas al rendimiento de las políticas, programas u operaciones ejecutadas por las entidades públicas. Lo que debe reconsiderarse es la forma de comunicar la opinión para adaptarse a la lógica de los destinatarios, sean órganos legislativos y de supervisión o responsables de la gestión pública, ciudadanía, etc...

No parece estar en cuestión la rendición de cuentas, de resultados obtenidos o de uso de los recursos públicos. Pero señalemos que si los fiscalizados avanzan adecuadamente hacia la transparencia, es muy posible que acaben rindiendo por su cuenta y el intermediario, en nuestro caso la ICEX podría perder relevancia.

Tampoco parece estar en peligro que las fiscalizaciones sigan proporcionando opinión basada en evidencia suficiente y adecuada: el asunto es que la evidencia se ha digitalizado. En efecto, gran parte de nuestra realidad se ha “datificado” y para incorporar la variedad y volumen de datos que ello supone es necesario incorporar técnicas de la ciencia de datos en los procesos de auditoría. Aunque no hay una teoría única sobre cómo debe hacerse, si hay un consenso amplio en que en el éxito en la combinación de ambas disciplinas van a seguir siguiendo determinantes las decisiones que tome el auditor⁴.

Por último, la producción servicios públicos digitales tiene efecto jurídico y, por tanto, no sólo son fiscalizables desde la eficacia, eficiencia y economía sino también, desde la legalidad. Al fin y al cabo, estos servicios públicos los proporciona una programación informática con un código fuente que puede ser auditado en su definición, funcionamiento y sus errores⁵.

¹ 8 jornada d' auditoria del sector public, 2017. Els reptes de la transformació digital en el Sector Públic. Anna Fuster.

² https://retina.elpais.com/retina/2018/07/12/tendencias/1531403497_313941.html

³ Google, Amazon, Facebook y Apple.

⁴ Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors. 2016. Michael Alles, Glen L. Gray

⁵ Véase interesante artículo jurídico en: <http://www.lapaginadefinitiva.com/aboix/?p=1698> “Todo algoritmo usado por la Administración para adoptar decisiones no es sino un reglamento (...) De otro modo estaríamos aceptando una cierta “privatización” inadmisibles de algunas normas y de sus mecanismos de funcionamiento, (...) en este caso cuenta con el agravante de que la complejidad de este tipo de programación la hace especialmente opaca al escrutinio individual por cada ciudadano”.

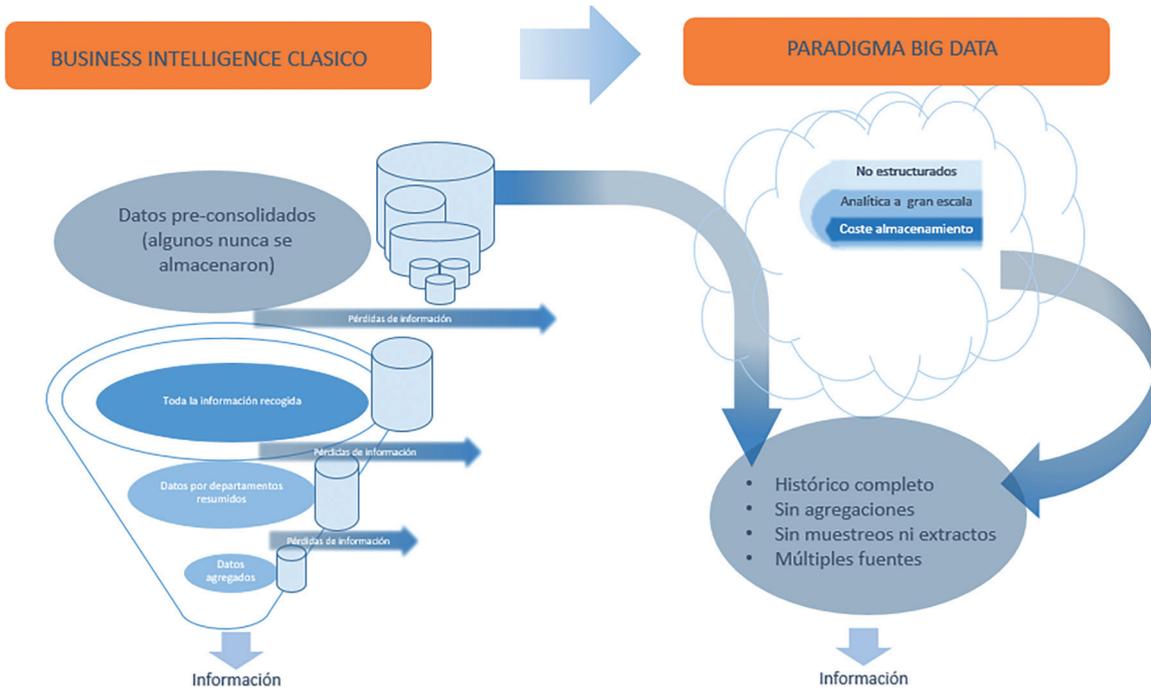
A nivel legislativo no existe una regulación al respecto, lo cierto es que el tema se está empezando a abordar por cuestiones de derechos fundamentales. <https://www.linkedin.com/pulse/sentencia-hist%C3%B3rica-del-tribunal-de-la-haya-anulando-manuela/>

1. EL ENTORNO BIG DATA: ANALÍTICA DE DATOS Y SU IMPACTO EN EL EJERCICIO DEL CONTROL

El desarrollo de la analítica de datos: Del BI clásico al Big Data

El *Big Data* está propiciando que los sistemas de información y los procesos tecnológicos generen datos de fuentes muy variadas y a gran velocidad⁶, también mejora y simplifica el procesado y análisis de estos datos mediante los procesos de analítica de datos. El análisis de esos datos revela conocimiento útil y criterios tanto para planificar una fiscalización como para validar nuevas pruebas que mejoren la opinión de auditoría.

La simplificación del uso de la tecnología permite analizar mayor cantidad y variedad de datos, superando un enfoque tradicional de la inteligencia de negocios, que se limitaba a datos exclusivamente financieros y contables con herramientas ad-hoc, hacia técnicas ampliadas de visualización y predicción. Para llegar a esta situación óptima, las organizaciones digitalizadas han modificado sus estructuras de almacenamiento, con o sin procesado previo, agregando los datos desde segmentos aislados e inconexos hacia grandes almacenes de datos (*datawarehouse*) o *data lakes*.



Los datos almacenados en los “sistemas *big data*”, tras un proceso de extracción, transformación y carga de datos (ETL), pueden ser convertidos en conocimiento. El cambio de paradigma facilita el empleo de datos dentro y fuera de la organización con un grado variable de agregación (llamado también granularidad), estableciendo métricas y alarmas automáticas (KPIs)⁷ y detectando patrones, tendencias y pronóstico. Los sistemas de analítica de datos comparten una serie de características:

VARIEDAD: muchas herramientas disponibles y muchas de ellas basadas en software libre y desarro-

lladas en colaboración por comunidades de desarrolladores.

FACILIDAD DE USO: una tendencia a la simplificación en el manejo y programación de analítica.

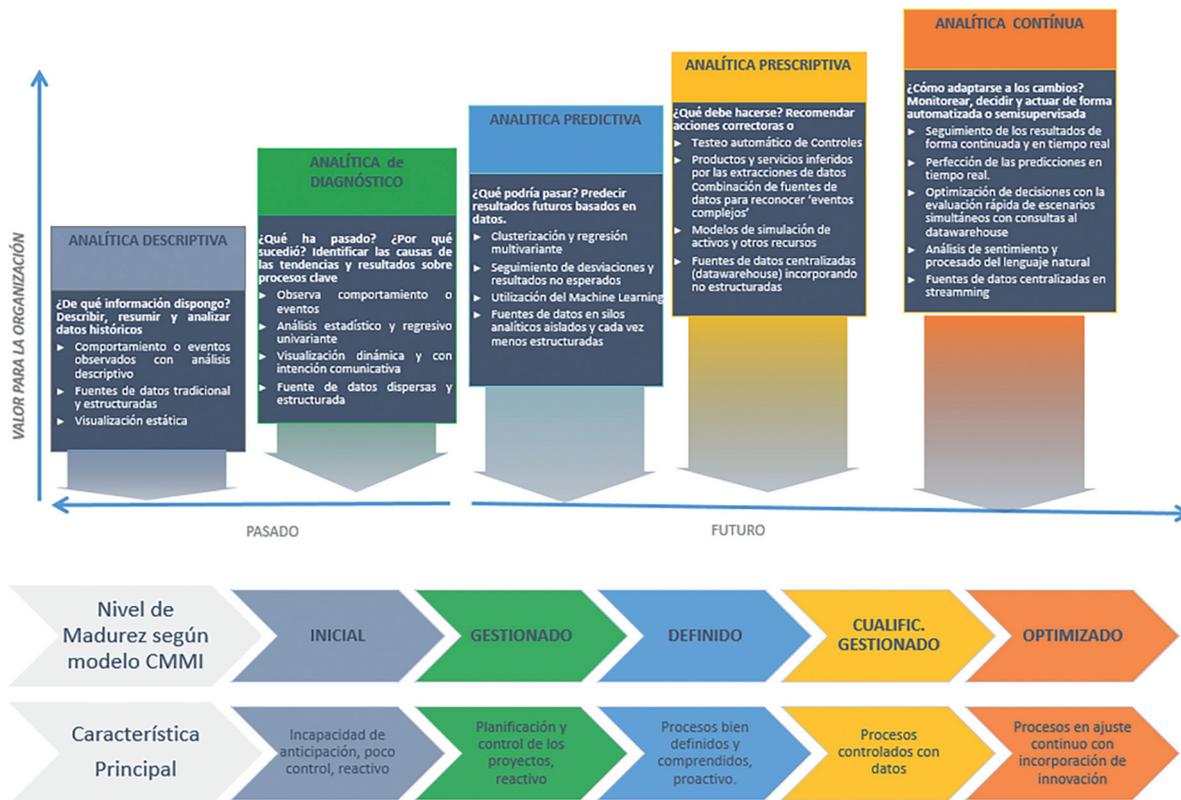
TRAZABILIDAD: Se evidencia con claridad de dónde viene el dato y cómo están siendo utilizados.

CONSISTENCIA: tanto en el formato visual como en los cálculos y transformación de los datos.

ADAPTABILIDAD: que se pueda adaptar a pequeños cambios sin tener que rehacerlo todo y que se pueda reciclar.

⁶Se puede comprobar la actividad en tiempo real tan sólo en internet en: <https://visual.ly/community/infographic/how/internet-real-time>.

⁷ KPI es acrónimo de *Key Performance Indicator*, es decir, indicadores de gestión. Durante el artículo se emplearán expresiones en inglés ampliamente utilizadas que devuelve resultados más refinados que empleando la terminología en castellano con la intención de facilitar las búsquedas de un/a lector/a interesado/a en profundizar en algún aspecto.



Existen varias clasificaciones y abundante literatura sobre el empleo de técnicas analíticas en organizaciones, si observamos los gráficos al Business intelligence pertenecen las tres primeras y business analytics (o analítica *big data*) son las otras dos. En definitiva, se trata de pasar del análisis basado en lo que ha sucedido a otro basado en la perspectiva, detección de patrones e incorporación de nuevas fuentes de datos que enriquezcan el análisis y, además, ofrezcan predicción y prescripción.

¿Cómo puede transformarse la Auditoría en un entorno Big Data?

Los sistemas de analítica en un entorno *Big Data* aplicados a la función fiscalizadora se caracterizan por:

Volumen: porque la facilidad para el procesado de grandes cantidades de datos incrementa la posibilidad de efectuar pruebas de verificación. No solo va a suponer una reformulación de la valoración del error y los criterios de importancia relativa sino que hasta el muestreo puede desaparecer.

Velocidad: la capacidad de anticipación es un valor en alza, así los destinatarios pasan de recibir un análisis ex post descriptivo a otro más prescriptivo y en tiempo real.

Variedad de los datos: incorporar en los análisis variedad de fuentes y de distinta naturaleza a las que recurrir, mejora las pruebas de comprobación y otorga mayor fiabilidad a las evidencias. Un buen sistema de analítica otorga mayor fiabilidad y ayuda a la robustez de las conclusiones de auditoría.

Veracidad de los datos: el desarrollo de la analítica va a necesitar de auditorías de sistemas de información que desarrolle pruebas sustantivas relacionadas con la integridad, exactitud y validez de la información para incrementar los niveles de seguridad significativa en la auditoría.

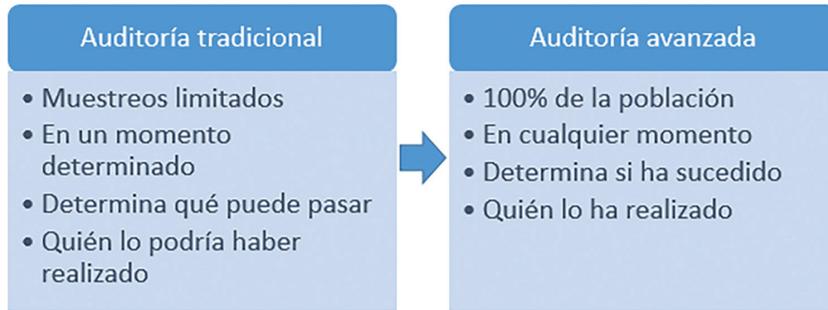
Viabilidad: la facilidad para encontrar fraude, datos atípicos junto con la automatización y reciclado de los procedimientos diseñados van a contribuir a encontrar la evidencia adecuada en las que se basa la opinión del auditor.

Visualización de datos aprovechando la variedad de formatos digitales en los que se pueden presentar y comunicar los resultados de hallazgos, mediciones, conclusiones y recomendaciones pone en cuestión el informe tipo "página de papel". También al control interno se le facilitan las acciones de seguimiento y de naturaleza correctiva.

Valor de los datos. Dentro del ciclo de vida de los datos sólo tienen valor para el negocio cuando se utilizan, o pueden ser útiles en el futuro. En nuestro sector, el dato adquiere el valor cuando se transforma en evidencia para basar una opinión.

En un entorno automatizable y con datos cada vez menos estructurados el reto es obtener evidencias su-

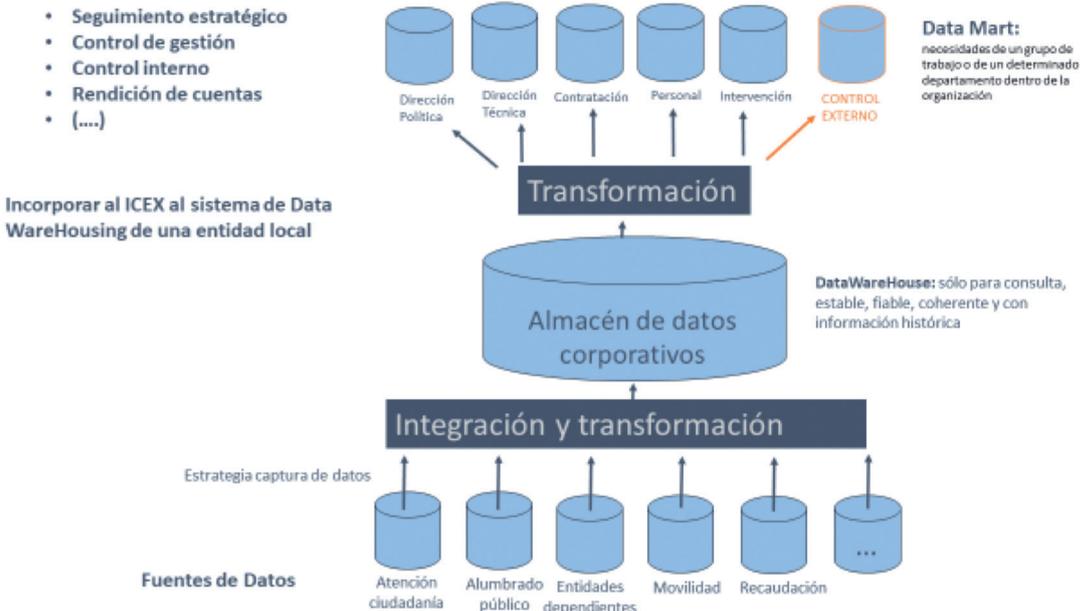
ficiente y adecuada de auditoría que reduzcan al mínimo el riesgo de auditoría. Una solución, no exenta de dificultades, pasa por emplear las nuevas herramientas que nos ofrece la tecnología *Big Data* de tal manera que, desde los almacenes de datos del fiscalizado, se extraigan, transformen y carguen en esquemas puestos a disposición de la institución fiscalizadora.



Muchas evidencias físicas, documentales y testimoniales se han digitalizado y su procesado para obtener evidencias de auditoría y control se ha simplificado con el desarrollo de la analítica *big data*. Además, estos procesos de análisis de datos no sólo pueden automatizarse (Artificial Intelligence) sino que pueden aprender y mejorar por sí mismos (Deep learning). En este contexto, es donde hay que situar las noticias sobre la robotización de actividades donde un lector informado conoce que el 45% de las actividades humanas podrían ser automatizadas⁸.

En lo que al control se refiere, la construcción de un buen sistema de analítica va a devolver las respuestas de auditoría suficientes para evaluar el nivel de eficacia, eficiencia y economía alcanzado en la utilización de los recursos disponibles, detectar sus posibles deficiencias y proponer a la dirección las recomendaciones oportunas en orden a la corrección de aquellas, de acuerdo con los principios generales de la buena gestión. Por tanto, desde el punto de vista del control:

Conexión al DataWarehouse del fiscalizado

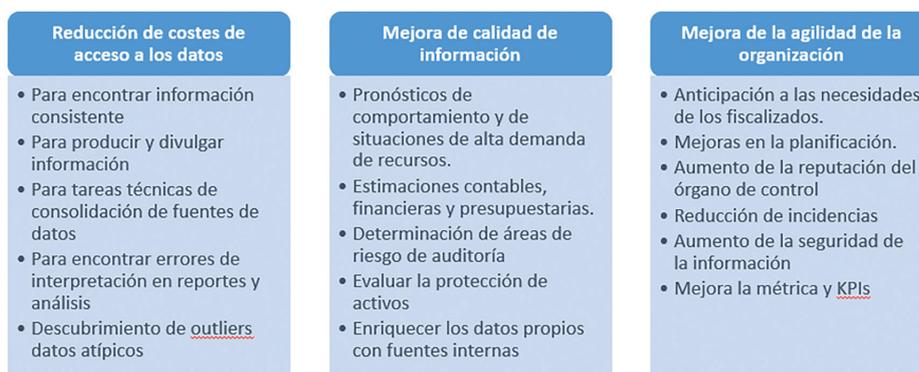


⁸ Desmitificando algunas cuestiones al respecto en: <https://www.raconteur.net/infographics/8-myths-about-ai-in-the-workplace>

Herramientas de analítica para el control interno: mejoran los reportings y permiten optimizar los recursos (incluso el mantenimiento predictivo de activos). Facilita la toma de decisiones a las personas con responsabilidades de vigilancia o capacidad de emprender acciones correctivas. Un buen sistema de analítica revelará en tiempo real si la entidad está adquiriendo, manteniendo y empleando recursos tales como personas, propiedades, instalaciones, etcétera, de forma económica y eficiente y las causas de ineficiencia, si las hubiera.

Herramientas de analítica para control externo: fundamental para la planificación y determinación del alcance, es decir, la naturaleza y extensión de los procedimientos que se aplicarán para obtener la evidencia de auditoría reduciendo los niveles de riesgo. En un nivel más avanzado, se aplicará en la obtención de evidencias: desde la identificación y valoración del riesgo pasando por los procedimientos analíticos, pruebas sustantivas de detalle y pruebas de control.

Impactos de la Analítica de datos en el ejercicio del control



2. NUEVA CAJA DE HERRAMIENTAS EN LOS PROCEDIMIENTOS DE AUDITORÍA

Para transformar los datos en evidencias de auditoría siguen siendo necesarias, la inspección, observación, indagación, confirmación, revisión de los cálculos, procedimientos analíticos y otras técnicas habituales de investigación: ahora con herramientas digitales. El software a emplear cada vez tiene más funcionalidades y, en ocasiones, su uso no es sencillo, pero tampoco especialmente complejo hasta tal punto que impidan que no-expertos puedan acceder a la utilización de la tecnología. El acceso a estas herramientas y a su aprendizaje está disponible y, frecuentemente es gratuito, por lo que optimizar el uso a las necesidades de auditoría está a nuestro alcance. En este apartado se comentan algunas técnicas que podemos aprender e incorporar al trabajo auditor.

2.1 Generalización de las Técnicas ETL (extracción, transformación y carga) de datos

Las técnicas ETL (extraction, transformation & load) se refieren a procesos mediante los cuales se extraen los datos de la fuente de datos, son depurados y homogeneizados conforme a los criterios que nos hubiésemos planteado y

se cargan en un almacén de datos para su utilización posterior. Las herramientas para construir un proceso ETL son cada vez más potentes y compatibles con distintas fuentes de datos, incluso incorporadas en las hojas de cálculo que manejamos a diario. Por el contrario, el software no va a impedir que el tratamiento de datos siga consumiendo el 80% del tiempo de trabajo de los analistas⁹.

Fase de Extracción de datos

En un primer momento de selección se recopilan las fuentes de datos existentes y, aunque la analítica de *Big Data*¹⁰ facilita el acceso a la misma, se requiere un gran esfuerzo en encontrar información y transferirla. Tecnológicamente, la generalización de los sistemas de almacenamiento con bases de datos no relacionales y sistemas distribuidos hacen que el sistema de extracción no resulte sencillo. En ocasiones, por motivos técnicos u operativos no se va a poder trabajar con toda la información disponible y, ya desde un primer momento, hay que aplicar técnicas de selección adecuada.

Una vez extraídos los datos se pasa a una fase de **exploración** descriptiva y gráfica¹¹ mediante técnicas de perfilado (*Data Profiling*), un proceso de examen de los datos

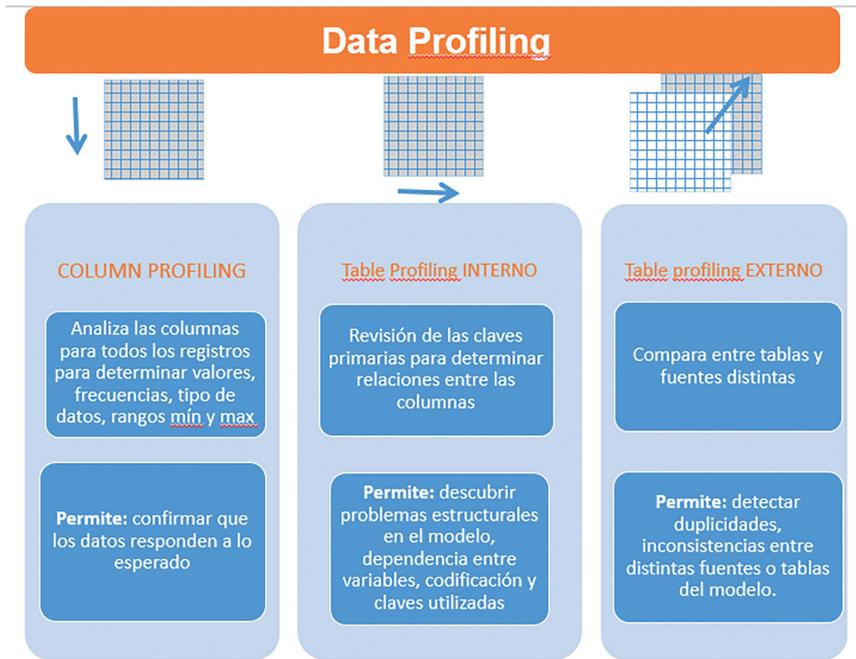
⁹ Distribución del tiempo de un científico de datos en un proyecto en: <https://medium.com/human-science-ai/how-i-spent-my-time-as-product-data-scientist-90e760044cd7>

¹⁰ Big Data en la Gestión de Registros de Auditoría. Pablo Castaño, Consultor de Ciberseguridad. Eulen seguridad. <https://es.slideshare.net/cprti/big-data-en-la-gestin-de-registros-de-auditora-big-data-en-la-gestin-de-registros-de-auditora>

¹¹ Puede encontrarse como EDA (*Exploratory Data Analysis*).

disponibles, se recolectan estadísticas e información sobre el comportamiento de cada variable. Un buen manejo de estadísticos elementales (mínimos, máximos, media,

moda, varianza,...) puede facilitarnos una orientación para detectar y seleccionar variables relevantes aunque se desconozcan las leyes que las relacionan entre sí.



En esta fase se detectan la presencia de información redundante, variables o rangos que acumulan la información más relevante o la existencia de valores atípicos (outliers) mediante gráficos de dispersión como el de caja-bigote. Si fuera necesario, hay que definir una política de imputación la información faltante o valores perdidos (*missing*). A diferencia de otras disciplinas, los outliers y los missing son una fuente de gran utilidad en auditoría dado que pueden indicar disfuncionalidades, fraudes o nuevas categorías de datos. También los datos erróneos o aparentemente irrelevantes tienen importancia cualitativa y es necesario definir una política para su tratamiento.

La utilización de software exploratorio va a requerir la definición de las características de cada variable:

- **Rango de los valores:** continuos, discretas o binarios.
- **Atributos:** numéricos, lógicos (v/false), categóricos (bajo, medio, alto), ordenados...
- **Escala variables:** nominal, ordinal o por intervalos.

Fase de transformación de datos

En esta fase se emplean distintas técnicas de combinación, transformación y preparación de los datos en función de los objetivos planteados. Las soluciones de software disponibles permiten realizar las transforma-

ciones sin alterar las fuentes originales de datos y que el proceso sea replicable para fuentes similares. También la transformación va dirigida a reducir el coste computacional de los procesos: su velocidad, almacenamiento o compresión, hay dos técnicas generalizadas:

- Reducir la dimensión del número de atributos o del número de casos que tenemos, es decir, encontrar un subconjunto de los atributos originales que permita obtener modelos de la misma calidad que los que se obtendrían utilizando todos los atributos.
- Categorizar las variables con distintas técnicas: agrupar o valores de un atributo en dos o más conjuntos distintos, agrupar las menos frecuentes bajo una etiqueta, discretizar variables que son continuas (binnig), crear variables dummies, ...

En fuentes de datos estructuradas hay que manejar con soltura funciones Query que nos permitan unir tablas, agrupar, seleccionar, establecer jerarquías de variables, extraer toda la información de registros de fechas (datastamp), etc...

Finalmente, con las técnicas de ETL es necesario apoyarse en sistemas de visualización que presenten los datos sin simplificaciones y permita el recorrido exploratorio permanente en un contexto "All data, All Time".

Una buena visualización logrará hacer simples contenidos complejos y son una herramienta imprescindible para comunicar los resultados de la auditoría.

2.2 Minería de procesos: Process mining

La digitalización facilita que las empresas parametrizen, cada vez, mayor cantidad de procesos y, por tanto, se crean y almacenan cada vez mayor información transaccional y operativa en soporte digital. La analítica *Big Data* puede ayudar tanto a recopilar y extraer de este tipo de información como a establecer métricas e indicadores útiles para la toma de decisiones. La minería de procesos o process mining, son técnicas para extraer información de los logs o metadatos de eventos generados en cualquier sistema de la organización (tramitadores de documentos, customer care, CRM, ERP, helpdesk, SCM, etc...).

Los logs son registros de la interacción de personas o cosas con los sistemas de la organización que, como mínimo, contienen: una identificación de la persona, una interacción con el sistema y un sello de tiempo (id-registro-timestamp). Con el despliegue del Internet de las Cosas (IoT) ni siquiera hace falta la intervención humana dado que sensores de ambiente, movimiento y otros elementos se les incorpora un chip con capacidad de emitir datos y, por tanto, incorporables al proceso analítico. Otra fuente importante de registros proviene de la web: clicks, interacción en redes sociales, fee-

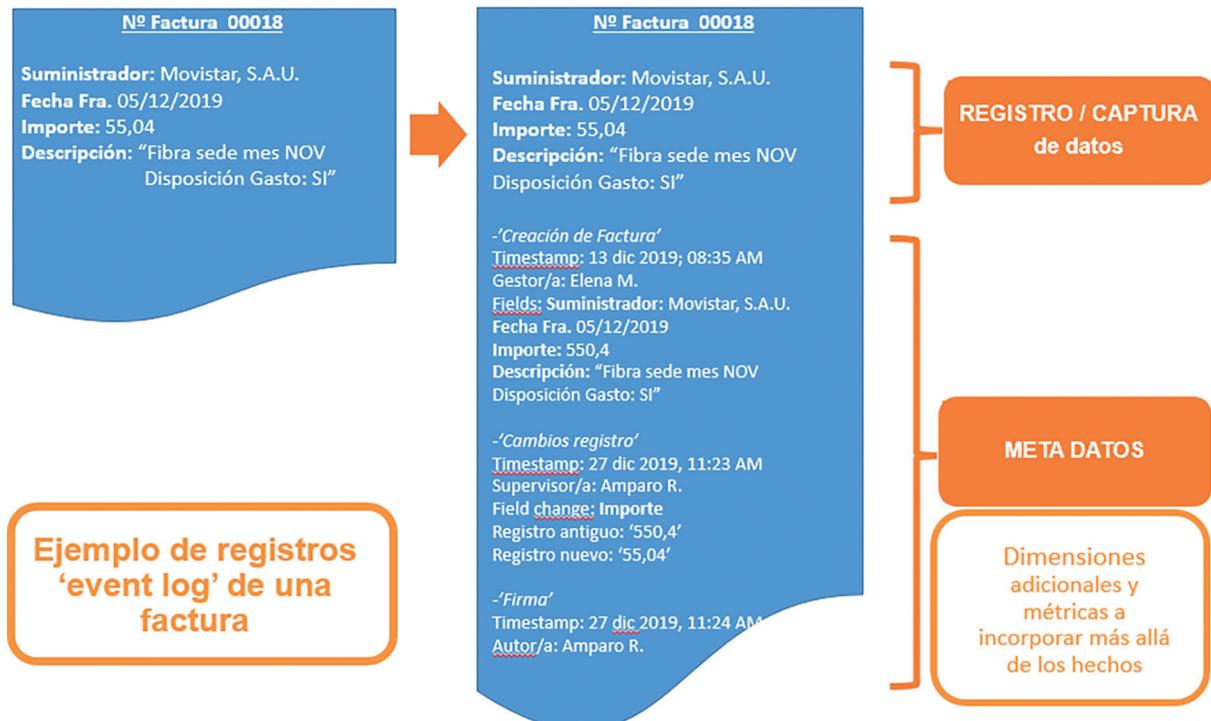
dbacks y otros registros generados por la actividad de los usuarios.

Los metadatos también alimentan la minería de procesos porque explican el contexto digital de una transacción/captura que se va registrando en los sistemas de información. En general, este contexto digital de los metadatos es de tres tipos:

- Técnicos: nombre de la fuente de datos, nombre de las tablas, columnas, tipo de dato, etc
- De Negocio: contexto que contiene al dato, glosario de nombres, definiciones, responsables, referencias de orígenes del dato, etc
- Operativos: información sobre el uso del dato, fecha y proceso de actualización, cantidad de accesos, última fecha de acceso, etc...

El ejemplo gráfico adjunto ilustra el registro de una factura que permite averiguar, por ejemplo, si alguien se ocupa siempre del mismo proveedor o registros fuera del horario laboral y otras cuestiones relevantes para el control interno y para la verificación de la segregación de funciones. El ejemplo se limita a una operación, pero si agregamos adecuadamente todo el proceso de registro de facturas permitiría definir patrones e irregularidades y las relaciones entre los individuos y la entidad fiscalizada.

Análisis de este tipo permiten descubrir (*process map y Discovery*), monitorizar (*conformance checking*) y optimizar (tiempos de los procesos y cuellos de botella).



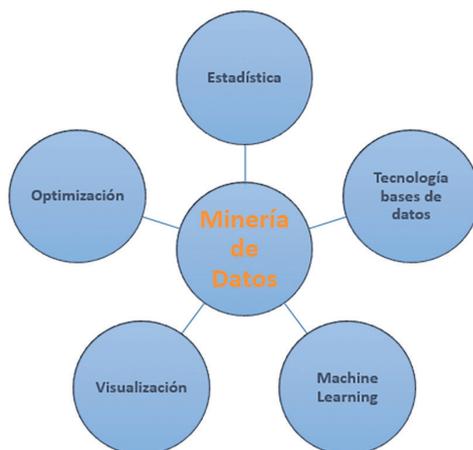
Ejemplo de registros 'event log' de una factura

Las herramientas de analítica van a poder tanto agregar y procesar el universo completo de registros granuleando y segmentando con un alto grado de precisión poniendo el foco en la actuación, no tanto en el contenido de esa actuación sino en el cómo se ha hecho. Siempre tenemos la posibilidad de añadir otras fuentes de datos por lo que se abre un mundo de pruebas de cumplimiento para las auditorías de sistemas y procedimientos.

Una analítica de procesos avanzada permite, en tiempo real, mejorar los procesos, la trazabilidad de las actividades, compliance y detectar el riesgo de fraude. Para el auditor externo, le ayuda a obtener el conocimiento necesario de los procesos de negocio y del organigrama real de la organización¹².

2.3 MINERÍA DE DATOS: *Data mining*

La analítica de *Big Data* ha impulsado la interrelación entre las técnicas que utiliza la minería de datos, muchas de ellas no son novedad puesto que provienen del análisis multivariante de la estadística descriptiva clásica. Con la capacidad y la relativa sencillez del procesado de las máquinas se pueden desarrollar métodos sencillos para abordar un tratamiento complejo de datos. El desarrollo de interfaces integradores de todas estas técnicas ha facilitado su combinación y presentarlas como un menú que facilita el descubrimiento automático de conocimiento. Gran parte de esta tecnología está basada en el software libre que incluyen comunidades abiertas de colaboración, por lo que es muy sencillo encontrar capacitación, ejemplos completos (*end-to-end*) y ayuda especializada.



En un gran porcentaje de los casos se desconocen las leyes que relacionan las inputs y las outputs, pero el objetivo de la minería de datos no es tratar de explicarlo todo, sino buscar cómo explicar los datos y concluir un MODELO que funcione, es decir, extraer patrones y relaciones no apreciables inicialmente para predecir comportamientos y en la toma de decisiones.

El desarrollo tecnológico ha permitido que los modelos sean replicables, automatizados, comparables y conectables entre sí y, que incluso, se perfeccionen de manera no supervisada. Vale la pena emplear recursos, pero hay que estar preparados para que una modelización no arroje resultados concluyentes. El desarrollo más conocido de la minería de datos es el *Machine Learning* cuyos algoritmos se configuran para dos objetivos:

- DESCRIPCIÓN: Técnicas que nos ayudan a entender la información masiva para su clasificación (clustering y segmentación), establecer reglas de asociación y dependencia, sistemas de recomendación, etc...

- PREDICCIÓN: combinando algoritmos con el uso de los modelos tradicionales de regresión y series temporales, análisis discriminante, métodos bayesianos, árbol de decisión y otros más avanzados como son las redes neuronales se pueden obtener patrones de comportamiento empleados en detección del fraude y crédito y caución. Los modelos pueden ser entrenados para el mantenimiento predictivo de los activos o la detección de cuellos de botella, pero también pueden pronosticar cuál es la siguiente tarea, cómo finalizan los procesos o qué recursos se van a necesitar.

Una vez seleccionados los datos tras un proceso ETL y tras seleccionar la variable objetivo, el conjunto de datos se divide en un grupo de entrenamiento y otro grupo de test, en general 70-80% de los datos para el primer grupo y el resto para el grupo de comprobación. Con el primer grupo se construye el modelo elegido y con el segundo grupo de datos se evalúan los errores y acier-

¹² Véase recientes trabajos de Tribunal Europeo de Cuentas: <https://medium.com/ecajournal/connecting-data-and-processes-in-audit-some-considerations-about-the-use-of-process-mining-231fe76f3f21>

tos del modelo con falsos positivos o falsos negativos (matriz de confusión) y otras técnicas de medición de la robustez de los modelos (curva ROC, etc..).

Hay que comparar entre varios modelos y transformando de datos iniciales empleando las técnicas de transformación (ETL) u otras más estadísticas (creación de *variables dummies*). El modelo más complicado, a veces no es el mejor y a pesar de todo habrá que estar preparados para que, con los datos seleccionados, no se pueda inferir nada. Con frecuencia, se producen sobreajustes en los modelos para los que también se emplean técnicas de modelado para tratar de evitarlo. En efecto, cuando se consideran muchos datos, hay que ajustar el modelo a más casos, con lo que nos perdemos detalles que pueden ser importantes o que pierdan capacidad predictiva. Cuando hemos solucionado estos problemas podemos tener un modelo más o menos aceptable al que poder introducirle datos nuevos con los que hacer clasificación o predicción.

2.4 minería de textos: *Text mining*

La minería de textos está mucho menos desarrollada que la minería de datos a pesar de que más del 90% de la información digital disponible está estructurada en textos y documentos tanto hablados como escritos. Esta disciplina se encarga del desarrollo de métodos, algoritmos y sistemas capaces de procesar datos textuales para su consulta, de tal manera que arroje tendencias, patrones, desviaciones y asociaciones en una colección de textos. La principal dificultad reside en que esta información suele estar sujeta a múltiples interpretaciones por lo que para poder aplicar cualquier desarrollo en esta área hay que analizar con precisión y despejar ambigüedades.

Para hacerlo posible se aplican técnicas de PLN (procesamiento del lenguaje natural)¹³ que necesitan de una colección (denominada corpus) de ejemplos y datos adecuadamente etiquetados de palabras y frases (tokenizado y lematización). El algoritmo puede ser entrenado para detectar patrones útiles, relaciones, clasificar documentos, resumir o predecir. Las técnicas más frecuentes son:

- Extracción de información: obtener las partes que interesan de un texto y pasarlas a un formato estructurado.
- Análisis de sentimientos: para entender la opinión general de un texto se utilizan colecciones de léxicos (lexicón) con un algoritmo clasificador. Los problemas a resolver son la subjetividad y tono, contexto, ironías y comparaciones. Pero no parecen relevantes cuando trabajamos con expe-

dientes administrativos, actas, memorias, etc...

- Clasificador de documentos: hace una asignación automática basada en sus contenidos., también tiene una versión manual en la que se categorizan los documentos.
- Generación automática de resúmenes: extraer y representar el contenido más importante, condensado y adaptado a las necesidades de la aplicación o del usuario final. Se realiza bien por extracción identificando los fragmentos clave o bien por abstracción, un parafraseo del texto original para hacerlo más claro y conciso.

La minería de textos tiene todavía un alto potencial por delante y a pesar de los esfuerzos públicos, especialmente en el ámbito científico e investigador con algún tímido avance en el ámbito sanitario¹⁴, lo cierto es que en idioma castellano queda mucho por hacer. La realidad es que tan sólo grandes empresas están dedicando recursos al asunto. En nuestro ámbito, sería necesaria alguna experiencia conjunta de los ICEX aprovechando la experiencia del sector privado.

3. ACELERADORES DIGITALES PARA LAS ICEX

3.1 Auditores deben adquirir competencias de analistas de datos

En muchas ICEX, asistidos por herramientas más o menos avanzadas, hemos incorporado la auditoría con datos basados en dos perfiles: por un lado, la persona experta con el conocimiento y la experiencia en auditoría y por otro, la persona que interactúa con el experto y que programa o extrae los datos. Ocasionalmente ambos perfiles podían concurrir en la misma persona, pero no parece que sea lo frecuente. Recientemente, algunas de nuestras instituciones están incorporando perfiles de auditoría de sistemas de información, con atribuciones en la evaluación de la integridad, fiabilidad y objetividad de los sistemas de información, para que se incorporen a la auditoría datos y fuentes digitales relevantes.

Limitar los puestos con alto grado de digitalización a estos perfiles podría dificultar la transformación digital de los ICEX puesto que, como hemos visto, la auditoría “tradicional” debe adquirir competencias en análisis de datos para no perder capacidad de escepticismo y juicio profesional. Las competencias en datos y bases de datos (eliminar de manera segura lo prescindible, realizar simulaciones, reaprovechar datos y las transformaciones ya construidas para otros trabajos, etc..) debe combinarse con la capacidad para identificar y manejar buenas fuentes, saber combinar información relevante

¹³ Más frecuente encontrarlo como: NLP (Natural Language Processing)

¹⁴ Véase: <https://www.plantl.gob.es/Paginas/index.aspx>

o documentar todo el proceso. Por tanto, el juicio, el escepticismo profesional o la experiencia de la auditoría tradicional siguen siendo imprescindibles, pero van necesitando de un implante biónico.

Una de las maneras de brindar una seguridad razonable de que la ICEX y su personal cumplen con las normas profesionales y los requerimientos aplicables¹⁵ exige no sólo que las personas que conformen los equipos de auditoría deben poseer en conjunto los conocimientos, las cualificaciones y la especialización necesarios sino que deben mantener su competencia profesional a través de un continuo desarrollo profesional.

El reto de ampliar nuestra competencias digitales debe plantearse de manera progresiva en varios niveles de madurez (básico, intermedio, avanzado) y escalando desde el análisis de datos ad-hoc hacia la monitorización continua de datos. Si tenemos una mirada ambiciosa, las competencias digitales no se limitarán al campo de la alfabetización de datos (*Data Literacy*) sino que se ocupará otras áreas tales como: comunicación y colaboración, creación de contenido digital, ciberseguridad y resolución de problemas¹⁶.

Para avanzar hacia organizaciones digitalizadas, lo deseable es la adquisición progresiva de otras competencias en otras áreas como:

- Habilidades de comunicación pública e interpersonal y mejorar en idiomas.
- Adquirir cultura de la innovación: atentos a los cambios y tendencias en el sector público y en ciencia y tecnología.
- Implantar metodologías de trabajo con datos, tipo scrum, usando técnicas de fraccionamiento de tareas que permitan escalar soluciones desde un mínimo viable a soluciones más avanzadas.
- Fomentar la cultura de aprendizaje: los recursos formativos son variados y accesibles desde nuestras máquinas individuales incorporando a nuestro trabajo el aprendizaje al modo “learning and test”.

3.2 Incorporar la gestión de datos como objetivo de auditoría

La auditoría de la gestión del dato puede ser una tarea inabarcable dado intervienen múltiples procesos, ciclo de vida del dato, metadatos, calidad del dato, soporte legal, organización, roles y responsabilidades, gestión de incidencias, documentación,....

Para la evaluación de la gestión de datos se pueden adoptar modelos y estándares de gestión de calidad internacionales¹⁷, así como nuestras normas técnicas y guías prácticas:



¹⁵ La ISSAI 30 - Código de Ética y la ISSAI 40 - Control de Calidad para las EFS

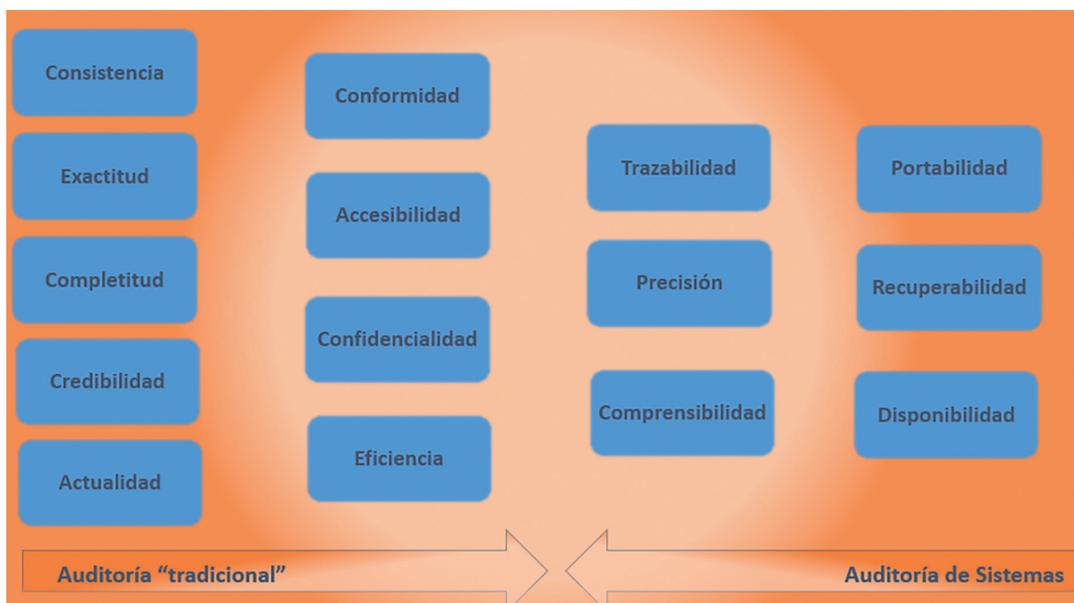
¹⁶ Ver la adaptación de los DIGCOMP que ha realizado la Comunidad Autónoma de Murcia definiendo las competencias digitales deseables del personal empleado público, <https://efiapmurcia.carm.es/web/pagina?IDCONTENIDO=41149&IDTIPO=11&RASTRO=cSm2813,5280,55544>

¹⁷ ISO8000, CMMI DMM, DAMA DMBOK, Manual de Intosai Auditoría de TI (2013)

Para la revisión de la arquitectura e infraestructuras de los sistemas no sólo es ocuparse del diseño sino de un procesamiento rápido y seguro. Es decir, por un lado revisar algoritmos, esquemas y código y, por otro, verificar la eficiencia de la estrategia empleada para la fragmentación, replicación y el almacenamiento distribuido y seguro.

También hay que ocuparse de una revisión limitada de la calidad del dato que podemos explicar conforme a sus criterios:

- Objetivas: integridad, exactitud, consistencia, existencia, completitud y validez
- Subjetivas: interpretabilidad, comprensibilidad, confianza, pertinencia, objetividad, oportunidad.
- Calidad de datos dependientes del sistema: disponibilidad, portabilidad, recuperabilidad.
- Calidad del dato inherente: exactitud, completitud, consistencia, credibilidad y actualidad



3.3 Adaptar técnicas de adquisición de relevancia digital de las organizaciones digitalizadas.

Se trata de emplear herramientas de visualización embebidas en los informes: que permitan la omnicanalidad y explorar los datos a diferentes grados de agregación de tal manera que contribuyan a reforzar o detallar la opinión del informe. Nunca ha habido tantas herramientas para difundir información y analizar datos como ahora, utilizarlas es una obligación si queremos hacernos un hueco en un entorno digital donde es muy sencillo crear y difundir información. El formato de informe como documento de papel ha sido desbordado por la tecnología y el consumo de la información va a condicionar el contenido y su forma de distribución. Si se quiere generar tráfico, ganar notoriedad y relevancia digital debemos incorporar técnicas de las organizaciones digitales que tratan de crear red de valor para los destinatarios y ganar a las nuevas audiencias.

Orientación al usuario/a: esmerarse en personalizar la experiencia de cliente, es decir, emplear diseños

orientados al usuario¹⁸, utilizar QRs, diseñar microsites, web vídeos,... utilizar buenas imágenes y otros recursos que retengan el interés. También cobran importancia los problemas tecnológicos tales como errores de navegación, rebajar los tiempos de carga de las páginas, etc...

Además, como organización digitalizada hay que estudiar la experiencia de usuario y medir los impactos utilizando las métricas e indicadores¹⁹ de la analítica de clientes que pueden incorporarse en la evaluación del impacto de nuestros informes: metodologías tipo embudo (Funnel), medir el número de interacciones, clicks y abandono de página, etc..

Por último, establecer canales de diálogo directo e instantáneo entre las personas destinatarias de los informes y las que gestionan la comunicación de los ICEX ayuda a mejora la personalización, proactividad y accesibilidad²⁰. No hay que olvidar que buena parte de las innovaciones y mejoras las proponen los usuarios.

¹⁸ UX design.

¹⁹ Para instalar google analytics para administraciones públicas consultar, Sergio jimenez.

²⁰ <https://analiticapublica.es/servicios-publicos-de-tres-estrellas-michelin/>

4. ALGUNAS DIFICULTADES PARA LA TRANSFORMACIÓN DIGITAL DE LAS ICEX

El principal reto de la digitalización de las ICEX no está en el uso de las nuevas herramientas, ni siquiera en la incorporación de técnicas modulares de trabajo, tipo scrum, que eviten caer en callejones sin salida o en eternizar los procesos de trabajo con datos. El desafío está en impulsar los cambios en el perfil profesional actual para adquirir el nivel de competencia digital adecuado. Es un asunto que no sólo afecta las ICEX sino a toda la Administración que tendrá que armonizar a las cuatro generaciones que estamos en edad laboral²¹ junto a las modalidades emergentes de teletrabajo.

Unas veces por falta de recursos y, otras por falta de comprensión de la transformación digital, entre otras causas, han originado la asimetría actual en el desarrollo tecnológico de nuestros fiscalizados. Dentro del ámbito subjetivo de actuación de las ICEX hay ejemplos avanzados de transformación, pero también es frecuente encontrar directivos que entienden la digitalización como “poner las cosas en internet”. Tampoco han ayudado los intentos regulatorios de las últimas décadas. El resultado es un universo de tecnificación muy desigual entre las organizaciones a fiscalizar.

Durante varias décadas de adopción de tecnologías se ha comprobado que hay cambios relevantes cada 5 años aproximadamente, y además, cada 2 o 3 años cambian los sistemas de procesamiento. Si a esto unimos la atomización de la toma de soluciones tecnológicas dentro de la Administración que incluye frecuentemente la adquisición de software cautivo y desarrollos ad-hoc, nos resulta un ecosistema con importantes problemas de interoperatividad.

5. CONCLUSIONES y RECOMENDACIONES

1. La Administración desarrolla sus servicios digitales públicos en un entorno marcado por las experiencias de consumo que ofrecen las gran-

des empresas tecnológicas caracterizadas por la inmediatez, la exhaustividad, la automatización y la predicción.

2. Gran parte de la realidad se ha “datificado” y para incorporar esta variedad y volumen de datos es necesario incorporar técnicas de la ciencia de datos en los procesos de auditoría tales como el process mining, *data mining* o *text mining*.
3. Las organizaciones digitalizadas han superado el esquema de Business Intelligence clásico para volcar en un gran almacén de datos la información generada por sus distintos sistemas de manera que pueda ser transformada para los distintos fines/departamentos. La auditoría externa debería aspirar a acceder a su propia transformación de datos.
4. La tecnología ha puesto al alcance de “no-técnicos” la posibilidad de extraer y transformar grandes cantidades de datos, de distintos orígenes y a tiempo real. Estas técnicas de ETL son imprescindibles para transformar datos del paradigma *Big Data* en evidencias.
5. El perfil del auditor “tradicional” debería completarse con competencias digitales avanzadas y la auditoría de sistemas y calidad del dato debería incorporarse de manera sistemática.
6. Las soluciones de visualización facilitan la exploración y difusión de información compleja por lo que tendrán que incorporarse o adaptarse a los informes de auditoría superando el formato “tipo papel”.
7. En un entorno digital con innovaciones constantes, las ICEX tienen la dificultad de impulsar un cambio cultural en sus organizaciones en un entorno de fiscalizados con un desigual grado de digitalización y con problemas de interoperatividad tecnológica entre sí.

BIBLIOGRAFÍA

Adrian Gepp, Martina K. Linnenluecke, Terrence J. O’Neill, Tom Smith (2018). *Big Data Techniques in auditing research and practice: current trends and future opportunities*. *Journal of accounting literature* nº 40.

Alles, M. & Gray, G. *Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors* (2016) *International Journal of Accounting Information Systems* nº 22.

Centre for the New Economy and Society. *World Economic Forum. The Future of Jobs Report* (2018). *Insight Report*

Cordero, Magdalena (2014). Un mundo digital, un reto para la auditoría.

Coyne, J.G., Coyne E.M. & Walker, K.B. *Accountants and Tech: a game changer?* (2017). *Strategic Fi-*

²¹ Discutiblemente etiquetados como: Baby boomers, generación x, millenials y generac z

nance Magazine.

Criado, I., Pastor V. y Villodre, Julián (2018). Un estudio sobre las comunidades digitales de los ayuntamientos españoles. Colección NovaGob Academia nº 3.

Diversos contenidos del Boletín semanal de datos.gob.es

Diversos contenidos de los cursos Executive de Transformación Digital (2018), Transformación Digital de Administraciones Públicas (2019) y del Master de *Big Data y Business Intelligence* (2020) de la Escuela de Organización Industrial.

Ferré, J. (2014) ¿Es posible la auditoría en un click? Presente y futuro de la auditoría IT

Forrester and Qlik. (2020). The human impact of Data Literacy: 5 steps to Driving Data Leadership in 2020.

Guthrie, R. Preparing Tomorrow's Auditors For The Future Of Tech-Driven Accounting. (2017). Forbes Community Voice.

Lucker, J., Hogan, S. & Bischoff T. *Predictably inaccurate: The prevalence and perils of bad big data* (2017). *Deloitte Review Issue* nº 21.

Manita, R., Elommal, N., Baudier, P. & Hikkerova, L. *The digital transformation of external audit and its impact on corporate governance* (2020). *Technological Forecasting & social change* nº 150.

Minguillón, A. (2014). La importancia de los sistemas de información en la Auditoría Pública.

Pérez Sarrión, Lorenzo (2020). Estrategia de implantación de la administración electrónica en un órgano de control externo (OCEX): No hablamos solo de Tecnología. *Revista de Auditoría Pública* nº 74.

Pérez García, I (2018), Tecnología y auditoría: principales retos. Instituto de Censores Jurados de cuentas de España. <https://www.icjce.es/tecnologia-auditoria-principales-retos>

Qlik, Accenture & Data Literacy Project. (2020). *The human impact of Data literacy.* Report

Sergio Jiménez (2019) Guía de Analytics para administraciones públicas.

Sergio Jiménez (2019). Herramientas digitales de rendición de cuentas.

Terraza Torra, Luis (2017). 2050 Odissea de la fiscalización. *Revista de Auditoría Pública* nº 69.

VV.AA. *Big Data & Digital Audit* (2020). *European Court of Auditors Journal* nº1

VV.AA. La transformación digital en el sector de la Auditoría (2019). Instituto de Censores Jurados de Cuentas de España.

VV.AA. Manual de la IDI y del WGITA sobre auditoría de TI para las EFS (2014), Intosai.