

El Dato

Unidad 2

Entendiendo la
importancia de los
datos y su impacto

Reflexión previa

“Yo he visto cosas que vosotros no creeríais. Atacar naves en llamas más allá de Orión. He visto rayos-C brillar en la oscuridad cerca de la Puerta de Tannhäuser. Todos esos momentos se perderán en el tiempo, como lágrimas en la lluvia. Es hora de morir....” Estas palabras que recordaréis de la película Blade Runner evocan el futuro, y sin duda alguna, el futuro hace tiempo que llegó, y para que las organizaciones crezcan, maduren o sobrevivan, necesitan *“conocerse”*.

¿Cómo se conoce a una organización? La respuesta es clara: mediante sus datos.

Las organizaciones capturan, crean, modifican, tratan datos sin parar. No estamos hablando solo de datos fiscales, financieros, contables o sanitarios, quizás los más conocidos. No. Estamos hablando del conjunto global de los datos de una organización. Es ese conjunto global el que describe a la compañía. Por eso es fundamental que todas las acciones de **Gobernanza del Dato** (Data Governance) estén patrocinadas por la máxima autoridad de la empresa. En varias ocasiones he leído alguna frase de William Edwards Demming (1900-1993) acerca del dato y su importancia, pero fueron estas dos las que más me gustaron:

“En Dios creo, el resto traedme datos”

“Sin datos, no eres más que una persona con opinión”.

Creo que cualquiera de las dos indica de un modo muy claro, lo importante que es el dato para una organización. Pero, ¿por qué es tan importante? Piensa en una construcción, ¿qué elemento se repite en todas? El ladrillo. Pues justamente por eso es tan importante, porque **para cualquier construcción de inteligencia empresarial, la unidad básica es el dato.**

El dato produce información y ésta, a su vez, conocimiento. Por tanto, podemos afirmar que **el dato es un activo clave estratégico para cualquier organización.** Ya sabemos que el dato es una pieza básica en el buen funcionamiento de una compañía....ahora ¿qué?

Calidad del dato

Según lo que plantea la norma ISO 9000: 2000, la calidad se podría definir como **"el grado en el que un conjunto de características inherentes cumple con los requisitos, esto es, con la necesidad o expectativa establecida, generalmente implícita u obligatoria"**.

Aplicado a los datos, se podría decir que **para que un dato sea de calidad debe ser:**



100%
Quality

Completo
Fiable
Actualizado
Accesible
Exacto
Consistente
Relevante
Legal

La falta de calidad de los datos es uno de los principales problemas a los que se enfrentan las personas responsables de sistemas de información y las empresas en general, pues representa claramente uno de **los problemas "ocultos" más graves y persistentes** en cualquier organización.



¿Por qué es tan importante la calidad de los datos?



Los crecientes avances de la tecnología en cuanto a gestión de datos e información han llevado a empresas y a las propias administraciones públicas a enfrentarse día a día a un **aumento en la cantidad y diversidad de los datos que deben gestionar.**

Como nos podemos imaginar, este aumento exponencial del volumen de información dificulta cada vez más el manejo de los datos, afectando directamente al desempeño de la organización y a la toma de decisiones y, por consiguiente, **dificultando la gestión empresarial.**

Dicho esto, parece evidente que se debe asegurar una buena calidad de los datos.



¿Qué podemos hacer?

Reconocer la calidad del dato como una prioridad de la organización.

Asegurar la calidad del dato debe ser una **actividad estratégica** de cualquier organización para garantizar que los datos se ajustan a las necesidades y objetivos de cada negocio.

¿Qué necesitamos?



Es una actividad que incluye, entre otros, aplicaciones informáticas, funciones, roles, estructura organizativa, procesos de verificación, supervisión y monitorización, métricas, indicadores y procesos de corrección y mejora de la calidad del dato.

Consiste, en definitiva, en una **pieza fundamental de la gobernanza del dato** de una organización que se desarrolla e implementa a través de políticas específicas y que implica a todas las personas.

¿Para qué?

Para que los datos se adecúen al uso que se les pretende dar y éste, al ser utilizado, **satisfaga las expectativas** de las personas usuarias.

Conclusión

La organización debe aplicar el **aseguramiento de la calidad de los datos** en toda la empresa y en todo el ciclo de vida de los datos para que estos se ajusten a los fines del negocio y cumplan todos los requisitos de:

- | | |
|-----------------|-------------------------------------|
| · Exactitud | · Relevancia |
| · Completitud | · Accesibilidad |
| · Unicidad | · Prontitud |
| · Integridad | · Seguridad, privacidad y legalidad |
| · Fiabilidad | · Metadatado |
| · Consistencia | · Georreferenciación |
| · Actualización | · Publicación y reutilización |



Es precisamente cuando dichas dimensiones o atributos son ambiguas o no están claramente definidas, cuando las empresas caen en errores en los datos.

¿Qué son los metadatos?

La consultora Gartner define los metadatos como "la información que describe varias facetas de un activo de información para mejorar su utilidad a lo largo de su ciclo de vida. **Son los metadatos los que convierten la información en un activo**".

En pocas palabras, los metadatos son datos sobre datos. Se generan cada vez que los datos se capturan de una fuente, las personas usuarias acceden a ellos, se mueven a través de una organización, se integran o aumentan con otros datos de otras fuentes, se perfilan, se limpian y se analizan. **Para que un dato sea de calidad, tiene que tener metadatos.**

Reflexión: si el 90% de la información es no-estructurada y no es posible procesarla por máquinas fácilmente, ¿cómo encontrar esa información o saber qué puede sernos útil? La respuesta está en los metadatos.

Un lenguaje común: Do you speak "Data"?

Todas las organizaciones precisan de un vocabulario o lenguaje común en torno a los datos para que sus departamentos se puedan interrelacionar adecuadamente. Imaginaros que cada departamento de una organización hablara en un idioma diferente. Lo difícil que sería interrelacionarse e intercambiar información. Esto que se ve muy fácil cuando hablamos de idiomas, es trasladable a los datos.

La gobernanza del dato persigue el establecimiento de ese lenguaje común que permita el acceso e intercambio de datos entre diferentes departamentos y áreas, y posibilite el cruce y el análisis de todos esos datos de diferentes fuentes sin enormes esfuerzos y costes de limpieza y conversión.

Los catálogos de datos, los glosarios de negocio y los diccionarios de datos son herramientas que ayudan en la elaboración de ese lenguaje común tan necesario en toda empresa. Saber dónde están los datos, qué significan exactamente y conocer sus características y atributos es crucial para que todo el sistema de información funcione correctamente y todas las personas hablen el mismo idioma.

La importancia de la calidad del dato

Basado en hechos reales:

Siguiendo la normativa del Banco Central Europeo (ECB), la asignación del riesgo en las entidades financieras a las operaciones crediticias está asociado al tipo de operación (Préstamos, Hipoteca, etc...) al CNAE, esto es, a la profesión de la persona beneficiaria, al importe y plazo y **al nivel de exposición del riesgo de la propia beneficiaria**.

La mala calidad del dato puede tener consecuencias inesperadas y de gran impacto en una organización. No se salva ninguna. Ni las más grandes.

Las consecuencias pueden suponer pérdidas, quejas, insatisfacción de la clientela, ineficiencias, incremento de costes...

Erase una vez un "torero" que fue a un banco...

[ver vídeo "La importancia del dato"](#)



En cierta ocasión, una entidad financiera lanzó una campaña de préstamos personales a profesionales liberales y autónomas. Los sistemas de la entidad no obligaban a las personas empleadas de las sucursales a determinar la profesión de las beneficiarias. Podían pasar de pantalla, formalizar la operación, sin marcar la profesión correcta de la misma.

Además, la parte desarrolladora de sistemas habían incluido el campo de profesión en un campo desplegable, cuyas opciones no estaban ordenadas, y donde la profesión que aparecía por defecto era "torero".

Ese mes las provisiones subieron en más de 500 Millones de Euros. Esto afectó a la cuenta de resultados directamente, ya que el importe provisionado se resta a los beneficios declarados de la entidad y, por tanto, altera el valor bursátil de la misma.

Desde Calidad de Dato, se analizó las causas posibles. Buscando en los datos que impactan en el cálculo de provisiones, se detectó que había un aumento de aproximadamente 60.000 Toreros en la base de datos. Esto llevó a investigar las operaciones en las que estaban implicados y se detectó que eran siempre en la contratación del nuevo préstamo personal a profesionales liberales y autónomos.



Conclusiones:

- **La calidad del dato en origen es primordial, evitar que entren datos con mala calidad evitan impactos económicos y de riesgo. Además, evitan trabajo posterior de limpieza de datos con coste para las empresas.**
- **La calidad es cosa de todas las personas, de las que determinan qué datos son clave, de las que introducen los datos, y de las que lo implantan en el sistema. La calidad hay que mantenerla a lo largo de todo el ciclo de vida del dato.**



Bizkaia