
El ciclo de vida del dato

PID_00246836

Marcos Pérez González

Tiempo mínimo de dedicación recomendado: 3 horas



Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea este eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares del copyright.

Índice

Introducción	5
Objetivos	6
1. El dato como activo de valor para la organización	7
1.1. Estrategias competitivas basadas en activos	7
1.2. El dato como activo de valor	8
2. El ciclo de vida de los datos	11
2.1. ¿Qué es el ciclo de vida de un activo?	11
2.2. ¿Cómo medir el valor del dato?	12
2.3. Fases del ciclo de vida del dato	15
2.4. ¿Cómo es posible detectar los cambios en un dato?	17
3. Gobierno del dato	20
3.1. Orígenes del <i>data governance</i>	20
3.2. Definición de <i>data governance</i>	21
3.3. Objetivos del <i>data governance</i>	22
3.4. Fases del <i>data governance</i>	25
4. Componentes y madurez del <i>data governance</i>	29
4.1. Componentes	29
4.1.1. Normas, estándares y estrategia	31
4.1.2. Calidad del dato	31
4.1.3. Gestión de datos maestros	31
4.1.4. Gestión de la seguridad del dato	32
4.2. Madurez	32
4.3. Modelo de madurez del <i>data governance</i>	33
4.3.1. Modelos de <i>data governance</i>	35
4.4. ¿Qué modelo de madurez usar?	37
Resumen	39
Glosario	40
Bibliografía	41

Introducción

A medida que las organizaciones fundamentan sus decisiones e iniciativas en datos, surgen preguntas como: ¿de dónde provienen los datos?, ¿tienen la calidad necesaria?, ¿qué sabemos sobre nuestra información?, ¿están estos datos alineados con nuestra política de empresa?, ¿tiene toda la organización la misma visión sobre el dato?

Llevar a buen puerto iniciativas de inteligencia de negocio, *data warehousing*, analítica de datos o *big data*, requiere tener un enfoque holístico respecto al dato y poder responder a las preguntas anteriores.

Para dar respuesta a dichas necesidades, emerge el **gobierno del dato**, que permite a una organización administrar de forma más eficiente y eficaz el dato. De hecho, esta iniciativa corporativa aúna personas, procesos y tecnología para cambiar la forma en que los datos son adquiridos, gestionados, mantenidos, transformados en información, compartidos en el contexto de la organización como conocimiento común y sistemáticamente obtenidos por la empresa para mejorar la rentabilidad.

Aunque el gobierno del dato se apoya sobre soluciones de tecnología, a lo largo de este módulo comprenderemos que está fundamentado en el desarrollo de procedimientos y marcos de actuación dentro de una organización.

En este módulo estudiaremos la necesidad e importancia de conocer el ciclo de vida del dato y en qué consiste el gobierno del dato. Asimismo, comprenderemos qué aporta el gobierno del dato y dónde impacta, cuáles son sus componentes y, finalmente, cómo evaluar la madurez de nuestra propia organización.

Objetivos

Este material didáctico está dirigido a:

- 1) Desarrolladores y consultores que quieren conocer qué significa gobierno del dato o *data governance*.
- 2) Desarrolladores y consultores que quieren ayudar al diseño e implementación de estrategias de negocio que incluyan gobierno del dato.
- 3) Gestores que están interesados en la transformación digital de su organización y en la inclusión de gobierno del dato como uno de sus pilares fundamentales.

En los materiales didácticos de este módulo encontraremos las herramientas indispensables para asimilar los siguientes objetivos:

1. Entender el concepto de *data governance*, las situaciones en las que es necesario desplegar una solución de este tipo y las ventajas que proporciona.
2. Conocer el ciclo de vida del dato y las necesidades en cada una de las fases del ciclo.
3. Enumerar y dar a conocer los diferentes componentes del gobierno del dato.
4. Conocer qué significan las diferentes etapas de madurez en el gobierno del dato y cómo analizar la madurez de la propia organización.

Si bien la obra es autocontenida en la medida de lo posible, los conocimientos previos necesarios son:

- 1) Conocimientos básicos sobre *business intelligence* y *big data*.
- 2) Conocimientos sobre estrategia y gestión de las tecnologías de la información (TI).

Se introducirán los conceptos necesarios para el seguimiento de este material.

1. El dato como activo de valor para la organización

1.1. Estrategias competitivas basadas en activos

Durante las últimas décadas, mejorar el posicionamiento de una organización ha sido objeto de estudio tanto dentro como fuera de la academia. Cuando una organización cuenta con una mejor posición que sus rivales, hablamos de que tiene una **ventaja competitiva** que permite asegurar a los clientes y defenderse contra las fuerzas competitivas. Por ejemplo, Delta Airlines, aerolínea de Estados Unidos, ha sido considerada durante muchos años un ejemplo en el uso de la eficiencia operacional para destacar por encima de otras compañías en su mismo mercado. Otro ejemplo es Apple, que ha destacado durante varios años por ofrecer productos altamente integrados en un ámbito de software y hardware combinado (como iPod, iPhone o AirPods), con diseños orientados al cliente como mecanismo de diferenciación.

Tradicionalmente, las ventajas competitivas se han fundamentado en:

- **Liderazgo en coste:** la organización es capaz de producir un producto o proporcionar un servicio a menor coste que sus competidores.
- **Diferenciación:** la organización es capaz de ofrecer un servicio o producto superior a los clientes. La diferenciación es, frecuentemente, el resultado de optimizar la cadena de valor del mercado en el que se opera o de llevar a cabo cambios en el alcance competitivo.

A mediados de los años ochenta, emerge un modelo que considera que los recursos y las capacidades que tiene la organización son la clave para generar un rendimiento superior al de sus competidores. En este enfoque estratégico, es necesario poder determinar si los recursos y las capacidades son relevantes, cosa que proporciona el marco de referencia que se ilustra en la figura 1.

El marco anterior busca identificar y clasificar los activos de una organización (**financieros, físicos, humanos, tecnológicos, de reputación u organizacionales**) en función de su valor. El objetivo es identificar las fortalezas y debilidades en relación con sus competidores y, de este modo, encontrar nuevas oportunidades o un uso más eficiente de los mismos. De hecho, este mismo marco sirve para analizar las capacidades de una organización, puesto que una capacidad puede entenderse como un **recurso intangible**.

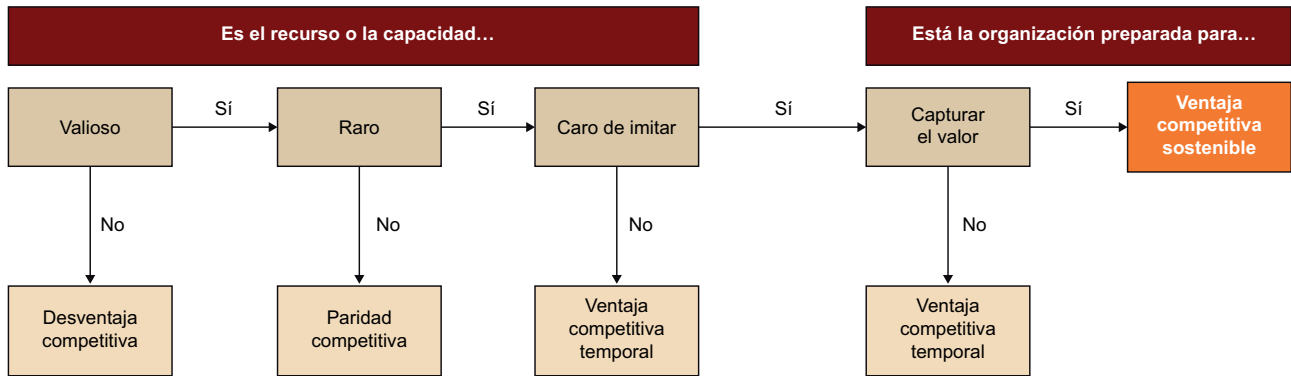
Lectura complementaria

Porter, M. (1998). *Competitive Advantage Creating and Sustaining Superior Performance*. Free Press.

Lectura complementaria

Wernerfelt, B. (1984). «A Resource-Based View of the Firm». *Strategic Management Journal* (núm. 2, vol. 5, págs. 171-180).

Figura 1. Marco de referencia



Fuente: Frank Rothaermel.

Está claro que el principal objetivo de esta estrategia es generar ventajas competitivas sostenibles, fundamentadas en:

- **Duración**, es decir, la velocidad a la que un recurso se deprecia o se convierte en obsoleto.
- **Transparencia**, es decir, a menor transparencia del recurso, menos posibilidades de que sea imitado por un competidor.
- **Transferencia**, es decir, a menor capacidad de transferencia de recursos —por ejemplo, no se pueden adquirir—, menor capacidad de que sea imitado por un competidor.
- **Replicación**, es decir, a mayor complejidad de procesos y capacidades organizaciones, mayor será la dificultad de replicación.

El hecho de que una organización tenga recursos y capacidades susceptibles de generar ventajas competitivas no significa que realmente las genere. En definitiva, es necesario diseñar una estrategia que identifique los recursos y las capacidades de la organización, mida su valor, maximice su explotación y, además, se lleve a cabo de manera continua para aumentar este tipo de recursos dentro de la organización.

1.2. El dato como activo de valor

Vivimos una época de profunda transformación de las organizaciones, fundamentada en tecnología (como *cloud computing*, impresión 3D, etc.), en la que la gran mayoría de sus procesos de negocio y su cadena de valor están siendo revisados e interpretados de nuevo.

Esta transformación tiene distintos efectos. Entre los mismos, destaca la generación de datos como subproducto o como intención principal de la trans-

formación. Cuando, de forma consciente, la organización gestiona el dato, es capaz de tomar decisiones más eficaces y eficientes, y de competir en el mercado de una forma distinta.

Es destacable, por ejemplo, el caso de Netflix, que usa los datos de cliente (su comportamiento y preferencias) combinados con algoritmos para evitar el abandono de clientes. De hecho, su conocimiento y la eficiencia de sus algoritmos se incrementan a medida que la compañía tiene más clientes, lo que genera un efecto de red.

Hasta tal punto es importante esta iniciativa que, según Netflix, su impacto en el negocio es de un billón de dólares.

Tal y como apuntan Higson y Waltho, estas organizaciones han cambiado su percepción respecto al dato. Han pasado de considerar el dato como un activo tóxico, cuyo uso es necesario controlar y limitar, a un activo de valor, cuyo uso hay que gestionar y maximizar.

Es decir, aquellas empresas que se están transformando en organizaciones orientadas al dato contemplan el dato como un recurso que genera ventajas competitivas, siguiendo las ideas de Wernerfelt.

El foco del dato no solo está en la eficiencia, como en el caso de Netflix. Ian Davis, analista de McKinsey, postula que las organizaciones deben ser ágiles para adaptarse a los cambios del mercado. Y defiende el rol del dato como fuente de la agilidad y el aprendizaje continuo. Para que una empresa sea flexible ante el mercado, ha de poder adaptar sus decisiones de forma continua y, por lo tanto, será tan flexible como lo sean su dato y sus capacidades de análisis.

Desde la perspectiva de un activo, el dato tiene sus peculiaridades si lo comparamos con otro tipo de activos:

- **Los datos frecuentemente tienen una naturaleza transitoria**, es decir, están vinculados a un tiempo y un momento. Por ejemplo, los valores bursátiles, como activo, solo tienen sentido si son capturados y analizados de forma continua. Esto es muy diferente, por ejemplo, en el caso de un edificio, cuyo valor presenta una duración más amplia.
- **El dato requiere un mantenimiento continuo**, es decir, es necesario manipularlo para mantener o incluso incrementar su valor.
- **El dato puede reusarse**, es decir, a diferencia de otros activos que con el uso se desgastan, como por ejemplo una pelota, el dato tiene la posibilidad de ser usado en diferentes escenarios, a un coste marginal.
- **El dato posee un gran potencial como activo**, puesto que tiene una baja o nula transparencia, transferencia y replicación, y una larga duración.

Lectura complementaria

Higson, C.; Waltho, D. (2010). *Valuing Information as an Asset*.

Sin embargo, la visión del dato como un activo aún no está completamente implantada en las organizaciones. El principal problema reside en el hecho de que no se tienen mecanismos establecidos para medir el coste y el valor del dato, ni tampoco, frecuentemente, procesos para gestionar de forma adecuada este activo.

Si se poseyeran estos valores, al igual que se hace con otros activos, el dato podría tener una entrada en los informes financieros, y sería factible dirimir su impacto en el negocio. Además, esto permitiría:

- Asegurar un cuidado y administración apropiados de los activos de datos.
- Incentivos para utilizar estos activos de forma eficiente.
- Potenciar la búsqueda de rentabilidad sobre estos activos, y clasificarlos en función de su rentabilidad.

Sin embargo, este escenario no es posible de manera generalizada. El valor del dato está relacionado con la capacidad de diferenciación y creación de ventajas competitivas, como en el caso de Netflix, lo que supone que cada organización podría tener mecanismos diferentes y propios para medir el valor.

Si se dispusiera de un proceso sistemático para gestionar los datos como un activo, sería factible poder seguir el ejemplo de Netflix y determinar qué significa valor y cómo evoluciona.

Por lo que nuestro objetivo es responder a las siguientes preguntas: ¿cómo es posible identificar qué datos de la organización son realmente activos de valor? ¿Cómo medimos el valor de dichos activos? ¿Cómo determinamos la evolución del valor a lo largo del tiempo?

En los siguientes apartados, discutiremos estas preguntas, que se articulan a través del **gobierno del dato**.

2. El ciclo de vida de los datos

2.1. ¿Qué es el ciclo de vida de un activo?

En el momento en el que una organización identifica el dato como un activo de valor, el siguiente paso es la gestión de dicho activo de forma precisa. Para ello, es necesario poder conocer y asociar a cada dato lo que definimos como el ciclo de vida de un activo:

Entendemos como **ciclo de vida de un activo** las diferentes etapas por las que pasa un activo desde su nacimiento hasta el fin.

La visión del ciclo de vida de un activo vincula directamente el activo con una actividad o proyecto. Por lo que, en general, el ciclo de vida de un activo puede estar vinculado a las siguientes etapas:

- Idea inicial y estudios preliminares (del proyecto).
- Evaluación del proyecto respecto a diferentes perspectivas, como viabilidad técnica o económica.
- Definición y planeamiento de las etapas del proyecto, y en cuáles de ellas el activo es relevante.
- Detalle del proyecto y diseño de los procesos.
- Ejecución del proyecto de acuerdo a la planificación.
- Puesta en marcha, prueba de todas las instalaciones y aceptación de las mismas.
- Operación de las instalaciones, uso o consumo de los bienes o servicios.
- Evaluación del proyecto para identificar siguientes mejoras o cambios.
- Descarte, reciclaje o venta de la instalación.

Es decir, se trata de conocer todo lo que ocurre con el dato: desde cómo se crea o incorpora a un proceso, hasta el descarte final, reciclaje o venta del mismo. Comprender el ciclo de vida del dato significa analizar las condiciones del dato en la organización en vinculación con el negocio y, por lo tanto, gestionar este ciclo. Esto nos lleva a definir el siguiente concepto.

Entendemos como **gestión del ciclo de vida del dato** el enfoque basado en políticas para gestionar el flujo del dato a través de uno o varios sistemas de información a lo largo de su ciclo de vida.

Algo que hay que recalcar es que el dato es un activo con condiciones muy definidas que caracterizan todo su ciclo. Centrémonos, por un momento, en la parte final del ciclo. La retención de datos dentro de la organización puede estar definida por obligaciones legales, por la legislación vigente o por el valor que el dato suministra a una organización. En países como España, es necesario tener en cuenta la ley de conservación de datos relativos a las comunicaciones electrónicas y a las redes públicas de comunicaciones. También es necesario recordar que aunque los costes de almacenamiento son cada vez menores, existe la necesidad de identificar cuándo los datos deben ser descartados, mediante su completa eliminación o desplazándolos a un almacenamiento externo para, así, mantener un tamaño manejable en el entorno de datos dentro de la organización.

Lectura complementaria

Ley 25/2007, de 18 de octubre, de conservación de datos relativos a las comunicaciones electrónicas y a las redes públicas de comunicaciones.
Fuente: BOE-A-2007-18243.

2.2. ¿Cómo medir el valor del dato?

Como hemos comentado, considerar el dato como un activo abre la puerta a preguntarse qué valor tiene este activo. Gestionar el dato no solo consiste en saber dónde está el dato, cómo ha evolucionado o qué significa. También es necesario conocer el valor que aporta un dato a la organización y cómo medir este valor.

Aunque hemos discutido que a veces puede resultar complicado determinar esta medición, vamos a introducir un punto de partida general. De acuerdo con la teoría de la contabilidad, el valor de un activo proviene de dos fuentes:

- Uso del activo, caso en el que hablamos de valor en uso.
- Venta del activo, caso en el que hablamos de valor de intercambio.

A partir de estas fuentes, se han usado tres métodos para calcular el valor del dato:

- **Fundamentado en el coste:** el valor del activo está basado en el precio pagado para adquirir el activo. Por ejemplo, cuando se compra una base de datos a Crunchbase*.
- **Fundamentado en el mercado:** el valor del activo se fundamenta en el precio que está dispuesto a pagar un tercero por adquirirlo.
- **Fundamentado en la utilidad:** el valor del activo está basado en el valor actual de los beneficios esperados futuros.

Cada uno de estos métodos tiene sus limitaciones.

- El valor fundamentado en coste puede ser inferior o superior al valor real para un conjunto de datos. Pensemos en el valor de la interacción de un

Lectura complementaria

Moody, D., Walsh, P. (1997). *Measuring The Value Of Information: An Asset Valuation Approach*.

*Más información en:
www.crunchbase.com

usuario en la página de Amazon, donde lo que está sucediendo tiene valor tan solo durante un corto periodo de tiempo.

- El valor fundamentado en el mercado requiere más tiempo para poder determinar qué precio es correcto para el dato.
- El valor fundamentado en la utilidad presenta la complejidad de que es necesario medir los flujos de caja en el futuro, lo que supone trabajar con hipótesis.

Cómo calcular este valor se ha convertido en un tema candente, al transformarse el dato en un activo de valor para crear ventajas competitivas o nuevas fuentes de ingresos, tal y como se ha comentado en el apartado 1. Según Gartner*, es necesario ampliar las formas en las que determinamos el valor del dato, y las hay de dos tipos: fundacionales y financieras. Veámoslas:

1) Medidas fundacionales, con foco en la mejora de la gestión de la información.

- **Valor intrínseco:** evalúa el grado de corrección del dato y si es completo. Se calcula mediante la fórmula:

$$IVI = \text{Validez} * \text{Compleitud} * (1 - \text{Escasez}) * \text{Ciclo de vida}$$

donde:

- Validez es el porcentaje de registros considerados correctos.
- Compleitud es el porcentaje del total de registros respecto al universo de potencial o supuesto de registros.
- Escasez es el porcentaje de su mercado o competidores que también, probablemente, tienen estos mismos datos.
- Ciclo de vida es la utilidad razonable y utilizable de cualquier unidad (registro) del activo de información (por ejemplo, en meses).
- **Valor de negocio:** evalúa si el dato es relevante para un propósito específico. Se calcula mediante la fórmula:

$$BVI = \sum_{p=1}^n \text{Relevancia}_p * \text{Validez} * \text{Compleitud} * \text{Oportunidad}$$

donde:

Lectura complementaria

Laney, D. (2015). *Why and How to measure the value of your information assets*. Gartner.

*Gartner es una de las principales empresas de prospección de mercado.

- Relevancia es lo útil que puede ser la información (o es) para uno o más procesos de negocio (0 a 1).
- Validez es el porcentaje de registros considerados correctos.
- Completitud es el porcentaje del total de registros respecto al universo de potencial o supuesto de registros.
- Oportunidad es con qué rapidez se capturan las instancias nuevas o actualizadas de los datos y es posible acceder a las mismas.
- **Valor de rendimiento:** evalúa cómo afecta el dato a los factores clave del negocio (en inglés, Key Performance Indicator, cuyo acrónimo es KPI). Se calcula mediante la fórmula:

$$PVI = \left[\frac{KPI_i}{KPI_c} - 1 \right] * T/t$$

donde:

- i es el número de instancias de proceso de negocio que utilizan el recurso de información (grupo informado).
- c son las instancias de proceso empresarial que no utilizan la información (grupo de control).
- T es el promedio de vida útil de cualquier instancia de datos.
- t es el tiempo durante el que se midió el KPI.

2) Medidas financieras, con foco en la mejora del rendimiento financiero:

- **Valor del coste:** evalúa el coste asociado a la pérdida del dato. Se calcula mediante la fórmula:

$$CVI = \frac{\text{Coste proceso} * \% \text{Atribuible} * T}{t} \left\{ + \sum_{p=0}^n \text{Ingresos perdidos}_p \right\}$$

donde:

- Coste proceso es el coste anual de los procesos que permiten capturar el dato.
- % Atribuible es el porcentaje atribuible a la captura del dato.
- T es el promedio de vida útil de cualquier instancia de datos.

- t es la duración del periodo sobre el que se mide el coste del proceso.
- n es el número de periodos de tiempo hasta que es necesario volver a adquirir o la continuidad de negocio ya no está afectada por la pérdida de información.
- **Valor del mercado:** evalúa el beneficio que se podría obtener al vender o intercambiar el dato. Se calcula mediante la fórmula:

$$MVI = \frac{\text{Precio exclusivo} * \text{Número de socios}}{\text{Prima}}$$

- **Valor económico:** evalúa cómo el dato contribuye al resultado final. Se calcula mediante la fórmula:

$$EVI = (\text{Ingresos}_i - \text{Ingresos}_c - \text{Costes}) * T/t$$

donde:

- Ingresos_i son los ingresos generados con el dato como activo.
- Ingresos_c son los ingresos generados sin el dato como activo.
- T es el promedio de vida útil de cualquier instancia de datos.
- t es la duración del experimento.

2.3. Fases del ciclo de vida del dato

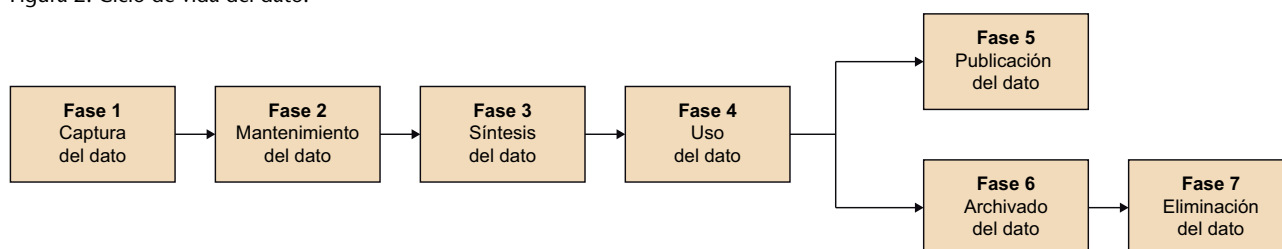
En el subapartado 2.1, hemos comentado que el dato es un activo diferente a otros activos de la organización, ya sea por su naturaleza, complejidad o incluso por el hecho de que sea un metadato.

Para comprender el dato y su valor, debemos hacerlo a lo largo de su ciclo de vida, que tiene diferentes fases, cada una de las cuales con sus propias características, como se ilustra en la figura 2.

Lectura complementaria

Curto, J. (2015).
Fundamentos de big data.
UOC.

Figura 2. Ciclo de vida del dato.



A continuación, entramos en detalle en cada una de estas fases:

- **Fase 1. Captura del dato.** Esta fase puede considerarse como el acto de crear datos que no existen aún en la organización y que nunca han existido en la misma. Encontramos distintos métodos para la captura de datos, entre los que destacan:

- 1) **Adquisición de datos:** la ingesta de datos que han sido creados y existen fuera de la organización.
- 2) **Introducción de datos:** la creación de nuevos datos en la organización por personal humano, o generados mediante dispositivos dentro de la propia organización.
- 3) **Recepción de señales:** la captura de datos creados por dispositivos, normalmente en sistemas de control, pero cada vez más importantes para los sistemas de información como el Internet de las cosas.

Los tres métodos identificados suponen importantes retos asociados. Por ejemplo, la adquisición de datos suele implicar la creación de contratos o normas que definen el modo en que la organización puede usar los datos que obtenga del exterior. La introducción de datos implica una detallada gestión de la seguridad del dato dentro de la organización, es decir, quién tiene permisos para crear o modificar datos en la organización. Por último, un reto asociado a la recepción de señales es cómo captar y procesar esos datos, los cuales pueden ser de cualquier naturaleza y tamaño, como por ejemplo los datos generados en tiempo real por los sensores de control en los yacimientos petrolíferos.

Internet de las cosas

Es un concepto que se refiere a la interconexión digital de objetos cotidianos con Internet. En inglés, *Internet of things*, abreviado IoT.

- **Fase 2. Mantenimiento del dato.** El objetivo de esta fase es procesar el dato, pero sin que se genere aún un valor claro para la organización. Este procesamiento incluye tareas de integración, limpieza, enriquecimiento, así como procesos de extracción, transformación y carga del dato (conocidos en inglés como ETL - *extract, transform, and load*). Debido a la diversidad de actividades en esta fase, existen numerosos retos asociados, como por ejemplo, cómo mejorar el proceso de envío del dato al destino final para su síntesis y uso, previniendo que se genere un elevado número de movimientos de datos durante todo el procesamiento, de inicio a fin.
- **Fase 3. Síntesis del dato.** Esta fase consiste en la creación de datos de valor aplicando un procesamiento o lógica inductiva determinada, usando otros datos como fuente. Esta es el área del procesamiento analítico donde se usa el modelado de datos, como, por ejemplo, el modelo de riesgos de una organización. La lógica inductiva requiere algún tipo de experiencia o conocimiento como parte de la lógica de negocio, como por ejemplo la forma en la que se crean los informes de créditos bancarios. Sin embargo, la lógica deductiva no formaría parte de esta fase, y quedaría relegada a la

fase anterior, como crear un nuevo valor aplicando una ecuación simple:

$$\text{Valor neto} = \text{Valor bruto} - \text{Impuestos asociados}$$

- **Fase 4. Uso del dato.** Una vez que el dato ha sido capturado y transformado dentro de la organización y se ha usado como fuente en la fase anterior, el dato se usa para beneficio de la propia organización, en tareas que la organización utiliza y gestiona. Aunque normalmente se trata de tareas que no pertenecen al ciclo de la vida del dato, el dato es cada vez más considerado como un activo fundamental en los modelos de negocios de muchas organizaciones. En esta fase, también existen retos importantes, como por ejemplo el uso permitido del dato, o lo que es lo mismo, si es legal o no el uso del dato de la manera en la que los usuarios de negocio pretenden usarlo.
- **Fase 5. Publicación del dato.** Esta fase puede ser definida como el envío del dato. Este envío puede ser interno (a una intranet) o externo (a un lugar fuera de la organización, gestionado por terceros). Un ejemplo es una agencia de inversión que envía informes mensuales a sus clientes. Una vez que el dato ha sido enviado fuera de la organización, es imposible recuperarlo para una posterior modificación. El gobierno del dato ayuda a decidir cómo actuar con los datos incorrectos o incompletos que han sido enviados fuera de la organización. Los accesos no autorizados a los datos también se incluirían en esta fase.
- **Fase 6. Archivado de datos.** Esta fase consiste en copiar los datos en un entorno donde son almacenados, en caso de que se vuelvan a necesitar en el futuro en un entorno activo de producción, y la completa eliminación de estos datos en todos los entornos activos. Un archivo de datos es simplemente un almacenamiento de datos, pero no de mantenimiento, uso o publicación de datos. En caso necesario, los datos pueden ser recuperados en un entorno donde es posible llevar a cabo cualquiera de estas actividades.
- **Fase 7. Eliminación del dato.** Esta es la fase final del ciclo de vida del dato, y consiste en la destrucción de cualquier copia del dato que exista dentro de la organización. Idealmente, este proceso será llevado a cabo mediante un archivado de datos. Un reto, en esta fase, sería verificar que la eliminación se ha llevado a cabo de manera satisfactoria.

2.4. ¿Cómo es posible detectar los cambios en un dato?

El dato no es un activo estático durante su ciclo de vida. Dentro de las organizaciones, se crean nuevas fuentes de datos continuamente, y es necesario mantener un registro del dato a la vez que se mueve a través de distintos sis-

temas dentro de la organización. Para ello, necesitamos establecer lo que se conoce como trazabilidad del dato o linaje del dato.

Se entiende como **linaje del dato** la capacidad de conocer todo el ciclo de vida de un dato, desde la fecha y hora exacta en la que fue extraído, el momento en que se produjo su transformación, y hasta el instante en que tuvo lugar su carga desde un entorno fuente (servidor, fichero, tabla, campo, etc.) a otro de destino. En inglés, *data lineage*.

Mediante el linaje del dato, las organizaciones pueden obtener una mejor comprensión de lo que sucede con el dato, lo que posibilita el rastreo e identificación de errores y, así, aplicar protocolos más rigurosos de gobierno del dato.

El linaje del dato provee a los profesionales del dato una representación visual que permite visualizar el flujo del dato. De este modo, se puede identificar cómo y cuándo el dato es modificado en la organización. Por ejemplo, identificar los cambios efectuados en el dato por los diferentes procesos de extracción, transformación o carga definidos en la arquitectura de datos de la organización. Cabe también destacar que, gracias a la capacidad de monitorizar el dato de una forma continuada, los errores relativos al dato pueden ser detectados antes de que aparezcan y, por tanto, corregidos.

Un escenario de uso común del linaje del dato se da en el área de la inteligencia de negocio (*business intelligence*). Por ejemplo, esta capacidad muestra cómo se ha obtenido cierta información relativa al negocio y qué papel puede desempeñar en los distintos métodos disponibles de integración de datos en la organización. Otro escenario habitual tiene que ver con la reducción de riesgos y la protección de datos. Los profesionales del dato pueden usar el linaje del dato para gestionar de una manera óptima el dato y, en todo momento, controlar dónde se encuentran los datos sensibles, y evitar así su exposición o minimizar los efectos de una posible violación en la seguridad en la organización.

Business intelligence

Es el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información, que permite tomar mejores decisiones a los usuarios de una organización.

Otros casos de uso incluyen:

- **Resolución de errores o conflictos:** en la creación de informes, la trazabilidad del dato permite conocer cómo se han construido las métricas que se incluyen, qué transformaciones se han hecho y de dónde provienen.
- **Análisis del impacto:** en el desarrollo y evolución de sistemas de información, el linaje del dato habilita comprender de antemano qué sucederá con el dato y qué medidas se deben tener en cuenta.

- **Informes de conformidad:** en ciertos sectores, se han impuesto regulaciones y normas para una mayor transparencia. *Data lineage* habilita reducir los errores humanos y las brechas en la conformidad respecto a la normativa vigente.

En esencia, el linaje del dato es un paso necesario, pero no suficiente, hacia la gestión eficiente del dato.

3. Gobierno del dato

Aunque pueda parecer lo contrario, la gestión del dato (y en su forma más moderna, el gobierno del dato) no es un concepto nuevo, existe desde hace más de veinte años. Para poder entender su forma más moderna, es necesario comprender sus orígenes.

3.1. Orígenes del *data governance*

Como concepto, el gobierno del dato ha existido desde hace décadas, y ha pasado por diferentes épocas.

Desde los inicios de las tecnologías de la información, hasta finales de los años ochenta, las organizaciones fueron adoptando progresivamente múltiples sistemas de información, de cara a soportar procesos de negocio y reducir las tareas manuales. Sin embargo, el dato era considerado como un subproducto de poco valor, con poca necesidad de ser compartido y gestionado. En las pocas organizaciones en las que se identificó la necesidad, trataron de gobernar el dato mediante el modelado de datos corporativos, pero con escaso éxito, debido principalmente a dos motivos: la falta de soporte en un ámbito corporativo, ya que solo el departamento de informática estaba implicado en estas tareas, y la rigidez de las aplicaciones disponibles en la época. Para resumir, en esta época el gobierno del dato era considerado principalmente un ejercicio meramente académico.

Durante las siguientes dos décadas, muchas organizaciones empezaron a reconocer que el valor del dato iba más allá de las transacciones. Se inició en muchas organizaciones el largo proceso de tomar decisiones fundamentadas en datos. Poco a poco, también se empezó a identificar la necesidad de cruzar los datos de diferentes unidades de negocio y departamentos.

En una primera instancia, las organizaciones abordaron estas necesidades construyendo repositorios de datos consolidados, ya fuera mediante sistemas ERP* o *data warehouse*. La implementación de estos sistemas, aunque proporcionaba claros beneficios, no estaba exenta de retos y problemas, como su coste, duración y los conflictos generados de tener que responder a múltiples necesidades de negocio al mismo tiempo. En esta época, también debemos destacar la caída de compañías como Enron, que hizo que el Gobierno de Estados Unidos estableciese nuevas leyes para mejorar la certeza y fiabilidad de la información dentro de las organizaciones. En particular, la Ley Sarbanes-Oxley, en inglés

*ERP es el acrónimo de *enterprise resource planning*.

Sarbanes-Oxley Act of 2002, Pub. L. núm. 107-204, 116 Stat. 745, cuya finalidad es evitar fraudes y riesgo de bancarrota, para proteger al inversor.

Por estas razones, se empezó a identificar el gobierno del dato como crítico para el negocio. De hecho, fue un componente clave en las medidas creadas por la Ley Sarbanes-Oxley y otros reglamentos, los cuales obligan a los ejecutivos de las organizaciones a conocer y ser personalmente responsables de los datos que utilizan en sus negocios. El gobierno del dato empezó a madurar rápidamente y, como resultado, la tecnología desarrollada comenzó a posibilitar la automatización a la hora de crear y gestionar leyes de negocio en un ámbito de dato. Sin embargo, el gobierno del dato estaba habitualmente centrado en distintos repositorios aislados; por ejemplo, el gobierno del dato para un *data warehouse* o un sistema ERP. Además, al principio era un gobierno del dato informal, que carecía de una estructura organizativa y procesos bien definidos.

En la actualidad, los ejecutivos están empezando a tener clara la necesidad de gestionar el dato como un activo para la organización. Hay una proliferación de más y más regulaciones vinculadas con el dato. Conjuntamente con la explosión de la complejidad del dato (*big data*), la transformación digital empuja a considerar formas más sofisticadas de explotar los procesos de negocio y los datos (por ejemplo, mediante *machine learning*).

Este escenario empuja a las organizaciones a replantearse cómo gestionar el dato, al necesitar nuevas estructuras de almacenamiento, de procesamiento y de análisis, y cómo afectan estas a los modelos de datos, a la calidad, a la seguridad y al ciclo de vida. Esto ha llevado a la aparición de plataformas integrales de gobierno del dato, en las que se definen políticas de datos en un ámbito empresarial.

El cambio actual también se fundamenta en que estas iniciativas ya no son solo responsabilidad del departamento de TI*, que aporta consistencia y control, sino también del negocio, que incrementa su responsabilidad sobre el dato y el reconocimiento del valor que puede aportar. En alguna de ellas (como es el caso de Telefónica o Barclays), incluso existe un rol especializado en el consejo de dirección, el denominado *chief data officer* (CDO).

3.2. Definición de *data governance*

La creciente necesidad de gestionar de forma eficiente el dato ha generado el hecho de que múltiples instituciones hayan dedicado recursos a formalizar la definición del gobierno del dato, así como a definir sus componentes. Esto ha provocado que existan numerosas definiciones del concepto. Algunas se exponen a continuación:

Ley Sarbanes-Oxley

Es una ley de Estados Unidos, también conocida como la Ley de reforma de la contabilidad pública de empresas y de protección al inversionista, y busca monitorizar a las empresas que cotizan en bolsa de valores, así como evitar que la valorización de las acciones de las mismas sea alterada de manera dudosa, mientras que su valor es menor.

Big data

Es el conjunto de estrategias, tecnologías y sistemas para el almacenamiento, procesamiento, análisis y visualización de conjuntos de datos complejos.

Machine Learning

Es un subcampo de la inteligencia artificial con foco en la creación de algoritmos con capacidades de aprendizaje a partir de datos y de los resultados buscados.

*TI es el acrónimo de tecnologías de información.

- Según *The MDM Institute**, es la coordinación formal de personas, procesos y tecnología que permite a una organización utilizar los datos como un activo de empresa.
- Según The Data Governance Institute**, es un sistema de decisiones y responsabilidades sobre procesos de información, ejecutados de acuerdo a modelos preestablecidos que describen quién puede tomar qué decisiones, usando qué información y cuándo, bajo qué circunstancias y mediante qué métodos.
- Según KiK Consulting***, es la ejecución y aplicación de autoridad sobre la gestión de activos de dato y el rendimiento de las funciones del dato.
- Según IBM Data Governance Council****, es una disciplina de control de calidad para asesorar, gestionar, utilizar, mejorar, monitorizar y proteger información de la organización.
- Según Sunil Soares, experto en *data governance*, es la elaboración de normas para optimizar, proteger y hacer uso de la información como un activo de empresa, mediante el alineamiento de objetivos de múltiples funciones.

*Más información en:
www.the-mdm-institute.com

**Más información en:
<http://www.datagovernance.com>

***Más información en:
<http://www.datagovernance.com>

****Más información en:
<http://www.ibm.com>

En el presente material, vamos a considerar la definición de DAMA***** publicada en el *Data Management Body of Knowledge* (DMbok):

*****Más información en:
<http://www.dama.org>

Se entiende como **gobierno del dato** (en inglés, *data governance*) el ejercicio de autoridad, control y toma de decisiones compartida (planificación, vigilancia y aplicación) sobre la gestión de los activos de datos.

Es importante destacar que todas las definiciones son similares, y establecen que el gobierno del dato trata sobre los procesos que controlan la creación, el acceso, el uso compartido, la utilización y la destrucción de la información, y qué sucede cuando existe un conflicto al respecto.

3.3. Objetivos del *data governance*

Así pues, el gobierno del dato es un conjunto de normas y prácticas bien definidas para asegurar que el dato sea:

1) **Accesible**, lo que significa que:

- a) El usuario puede acceder a los datos que necesita y estos están disponibles.
- b) El formato del dato coincide con el que el usuario demanda.

2) **Seguro**, lo que significa que:

- a) Solo pueden acceder al dato los usuarios autorizados.
- b) Se impide el acceso al dato a los usuarios no autorizados.

3) **Consistente**, lo que significa que:

- a) La información es la misma cuando distintos usuarios acceden al mismo dato.

4) **Alta calidad**, lo que significa que:

- a) El dato es riguroso y certero.
- b) Cumple con los estándares preestablecidos.

5) **Auditable**, lo que significa que:

- a) La trazabilidad del dato es clara.
- b) El departamento de TI sabe quién usa el dato y para qué propósitos.

El objetivo final del gobierno del dato no es solucionar todos los problemas de TI de la organización, sino centrarse en todos los aspectos vinculados con el dato. Por ello, los principales objetivos del gobierno del dato incluyen:

- Definir, validar y comunicar las estrategias del dato, normas, estándares, arquitectura, procedimientos y métricas.
- Monitorizar las políticas del dato, estándares, arquitectura y procedimientos.
- Patrocinar y supervisar la entrega de proyectos de gestión de dato y servicios.
- Gestionar y resolver conflictos relacionados.
- Entender y promocionar el valor del activo del dato.

En este sentido, implementar el gobierno del dato en una organización supone trabajar en los siguientes puntos:

- **Estrategia.** Desarrollar una estrategia, es decir, decidir qué datos se van a gestionar, identificar los datos maestros. Determinar el valor del dato basado en el coste de la captura, mantenimiento, valor de negocio, riesgo de pérdida o certeza del dato.
- **Comisión.** Establecer una comisión de las líneas de negocio que están en la posición de entender el dato, entender el negocio y ser capaces de tomar una decisión.

- **Normas.** Establecer un conjunto de normas para definir la integridad de datos, calidad, seguridad y utilización del dato.
- **Estándares.** Establecer un conjunto de estándares para controlar la implementación de las normas, modelado de datos, herramientas, tecnologías y metodologías.
- **Procedimientos.** Definir procedimientos para tratar conflictos de calidad del dato y de seguridad.
- **Conflictos.** Gestionar los conflictos y su resolución.
- **Supervisión.** Ofrecer una supervisión continuada de la gestión del proyecto.
- **Incumplimiento.** Establecer un conjunto de penalizaciones por incumplimiento de los estándares y procedimientos.
- **Documentación.** Proporcionar documentación relativa al dato y requisitos de metadatos.

El gobierno del dato no reemplaza la gestión del dato, sino que la complementa. Cuando diferentes usuarios no pueden o deben tomar decisiones independientemente en la organización, el gobierno del dato facilita que equipos que desempeñan diferentes funciones en el negocio se unan para crear normas comunes, solucionar conflictos o dar nuevos servicios a las personas involucradas. La tabla 1. presenta las principales diferencias entre los dos conceptos.

Tabla 1. Diferencias gobierno y gestión del dato

Dimensión	Gobierno del dato	Gestión del dato
Alcance	Toma de decisiones, monitorizado y ejecución que tiene autoridad en la gestión del dato de una organización.	Control de la arquitectura de datos, calidad, seguridad, normas, prácticas y procedimientos.
Función	Decidir qué hacer con el dato y asegurar que se haga lo decidido.	Implementar las arquitecturas, procesos, herramientas y normas para alcanzar los objetivos del gobierno del dato.

Es bastante común en las organizaciones el hecho de que diferentes grupos de usuarios creen distintos repositorios de datos no sincronizados entre sí. Esto suele dar lugar a una pésima eficiencia operacional y un alto riesgo, lo que produce la necesidad de que exista una autoridad que resuelva los conflictos generados. Esta es la función del gobierno del dato. Por ejemplo, si existen multitud de repositorios sobre datos de clientes, habrá que tomar decisiones, a menudo impopulares, para crear un repositorio de datos maestros.

Gestión del dato

Es el desarrollo y la implementación de arquitecturas, normas, prácticas y procedimientos para gestionar eficazmente las necesidades del ciclo de vida de los datos en una organización.

3.4. Fases del *data governance*

El diseño de un programa de gobierno del dato no es necesariamente una secuencia lineal. De hecho, es probable que se tomen decisiones sobre cada una de las fases del programa a partir de otras fases. Por tanto, a pesar de que a continuación se detallan las fases en una secuencia lineal, se deberá tener en cuenta un enfoque holístico a la hora de hacer el diseño.

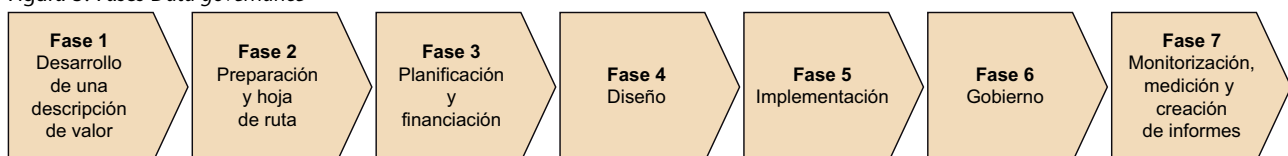
Según The Data Governance Institute, el ciclo de vida del gobierno del dato consta de siete fases, como se ilustra en la figura 3:

- 1) Desarrollo de una descripción del valor.
- 2) Preparación y hoja de ruta.
- 3) Planificación y financiación.
- 4) Diseño.
- 5) Implementación.
- 6) Gobierno.
- 7) Monitorización, medición, creación de informes.

Hoja de ruta

Hace referencia al plan detallado que refleja los objetivos a corto y largo plazo y guía el progreso para alcanzar esos objetivos, usando, a menudo, una solución tecnológica específica.

Figura 3. Fases *Data governance*



Fuente: The Data Governance Institute.

Como prerrequisitos, antes de crear el diseño del programa del gobierno del dato debemos tener en cuenta lo siguiente:

- El mensaje del valor del programa debe ser claro.
- Tiene que existir una hoja de ruta que será compartida con las personas involucradas, y que detalle los distintos aspectos; en particular, dar respuesta a las preguntas quién, qué, cuándo, dónde y por qué.

Es necesario explicar el detalle de las diferentes fases:

Fase 1. Desarrollo de una descripción del valor. La primera fase se centra en definir el valor que el gobierno del dato genera al negocio, así como los estándares para una medición continuada tras la puesta en marcha del programa. En esta fase, se revisa cuidadosamente la estrategia y objetivos del negocio y

se define y desarrolla una conexión entre el gobierno del dato y una mejora en la organización identificable desde un aspecto financiero. Para identificar el valor que el programa pueda generar al negocio, es preciso identificar qué sucede dentro de la organización en términos de la gestión del dato, ya que es posible que exista algún programa en ejecución, por ejemplo, relativo a la calidad del dato o a la gestión de datos maestros, y a cómo el gobierno del dato soportará el negocio, directamente o indirectamente a través de estos programas.

A continuación, es necesario definir cuál será el criterio de éxito para el programa del gobierno del dato, esto es, proveer unos estándares de métricas y medidas al respecto que permitan identificar el éxito o fracaso. Por ejemplo, se pueden identificar métricas financieras que permitan conocer cuál sería el coste de una carencia de gobierno del dato, o lo que es lo mismo, el uso de la información de una manera pobre e ineficaz dentro de la organización, en comparación con un programa gobernado.

Fase 2. Preparación y hoja de ruta. En esta fase, se planifican los detalles sobre los eventos en torno a la fecha de lanzamiento del gobierno del dato, momento en el cual los activos de datos de una organización pasan de no estar gobernados, a estarlo. A menudo, la hoja de ruta implica un plan que se desarrolla durante varios años, y que integra actividades con otros proyectos o iniciativas que se ejecutan dentro de la organización. Es necesario incluir frecuentes puntos de control y espacios para recoger críticas constructivas a lo largo del plan, y de este modo, modificarlo si fuera necesario y no llegar a un estado de bloqueo durante la ejecución del plan. Las actividades de esta fase implican:

- Integración del gobierno del dato con otras áreas. Dado que el gobierno del dato está relacionado con multitud de áreas dentro de la organización, será necesario contemplar las posibles relaciones en la hoja de ruta. Por ejemplo, integración del gobierno del dato con la gestión de datos maestros dentro de una organización.
- Diseño de métricas y requisitos de informes. Necesario para validar lo que se está gestionando y recibir información para asegurar la continuidad del plan.
- Definición de los requisitos de mantenimiento o gestión del cambio. Es necesario planificar los posibles cambios que puedan ocurrir para poder gestionarlos. Así, se identifican los posibles elementos que tienen lugar, durante una futura puesta en ejecución o lanzamiento del programa, y se determina cómo están coordinados entre sí.
- Definición del lanzamiento del gobierno del dato. Se especifican los detalles relativos a la puesta en marcha del programa, incluyendo información sobre el personal relacionado que lo gestionará.

Fase 3. Planificación y financiación. El programa del gobierno del dato a menudo supone una inversión significativa de tiempo y recursos, tanto humanos como financieros, por parte de la organización. Sin embargo, estos costes se compensan por el valor para el negocio que el programa proporciona. Mediante una descripción de valor bien definida, unos objetivos de negocio claros y un respaldo o patrocinio en un ámbito ejecutivo, el programa del gobierno del dato puede ser satisfactoriamente ejecutado y proveer un retorno de inversión significativo.

Fase 4. Diseño del programa. En esta fase, se especifican y desarrollan los detalles sobre cómo el programa va a ser usado y la funcionalidad que albergará. El resultado final de esta fase son los principios, normas y diseño de los procesos que definirán el programa. Las principales actividades que tienen lugar en esta fase son:

- Determinación de los procedimientos o principios centrales. Esta actividad resulta imprescindible para el éxito del programa, y consiste en identificar, documentar y aprobar los procedimientos clave que serán utilizados dentro de la organización para considerar el dato como un activo.
- Determinación de los procesos estándar para el negocio. Durante esta etapa, se determinan los procesos genéricos que definen los objetivos del programa del gobierno del dato. Por ejemplo, cómo reacciona el negocio ante un proceso de identificación o detección de incidencias y las acciones que se llevan a cabo en el proceso para su resolución.
- Determinación de las responsabilidades y el modelo de propiedad. Así pues, todos los procesos definidos previamente carecen de utilidad hasta que se ha definido quién hace qué, y los diferentes niveles de responsabilidad presentes.
- Exposición del modelo de gobierno del dato al equipo ejecutivo del negocio. De esta manera, se informa y educa al equipo de gestión de la organización sobre las diferentes responsabilidades y procesos.

Fase 5. Implementación del programa. Esta fase representa la ejecución de las actividades que se han definido en el paso previo. Es importante destacar que esta fase no implica que exista un definido comienzo y final de fase, ya que, en esencia, una vez que se ha comenzado a soportar el gobierno del dato, este proceso nunca acaba. Las actividades principales de esta fase son:

- Lanzamiento (en inglés, *rollout*) del gobierno del dato. Como su propio nombre indica, a partir de este momento la organización comienza a gobernar los datos. Al mismo tiempo, diferentes equipos dentro de la organización son educados y formados según lo acordado en la hoja de ruta, y se publican diferentes materiales creados durante el diseño del programa, tales como guías y normas.

- Ejecución de cambios en el plan. En esta etapa, se recoge toda la actividad relativa al soporte del gobierno del dato. Por ejemplo, se puede afirmar que, durante la ejecución del programa, distintos materiales, como las guías o materiales de formación, necesitarán una actualización. Asimismo, el nuevo personal necesitará una nueva orientación sobre el programa implementado.

Fase 6. Gobierno del dato. Una vez que el diseño funcional está definido, el siguiente paso es plasmar este diseño en un marco organizativo. En esta fase, se identifica al personal relativo al programa del gobierno del dato, tales como administradores o propietarios del dato, y algunas áreas o usuarios tienen diferentes actividades y rangos de responsabilidades relativos al gobierno del dato. En esta fase, las actividades se basan en:

- Diseño del marco organizativo. Esto es, determinar dónde y qué niveles del marco son los encargados de ejecutar, gestionar y ser responsables de manejar los activos de datos.
- Identificación de roles y responsabilidades. Esta tarea, aunque parezca trivial, puede suponer obstáculos dentro de la organización, como por ejemplo una percepción de dar poder sobre los datos a determinados usuarios que puede llevar a problemas internos dentro del marco organizativo, o el temor de perjudicar la productividad actual al añadir responsabilidades adicionales.

Fase 7. Monitorización, medición y creación de informes. El programa del gobierno del dato necesita ser examinado y revisado para saber si es eficaz para la organización tras su implementación. Normalmente, un grupo interno dentro de la organización llevará a cabo estas comprobaciones, las cuales implican la recogida de datos y generación de métricas que informan sobre distintos aspectos, como la efectividad de normas y estándares, así como de la actividad de usuarios definidos en la organización, tales como administradores del dato.

4. Componentes y madurez del *data governance*

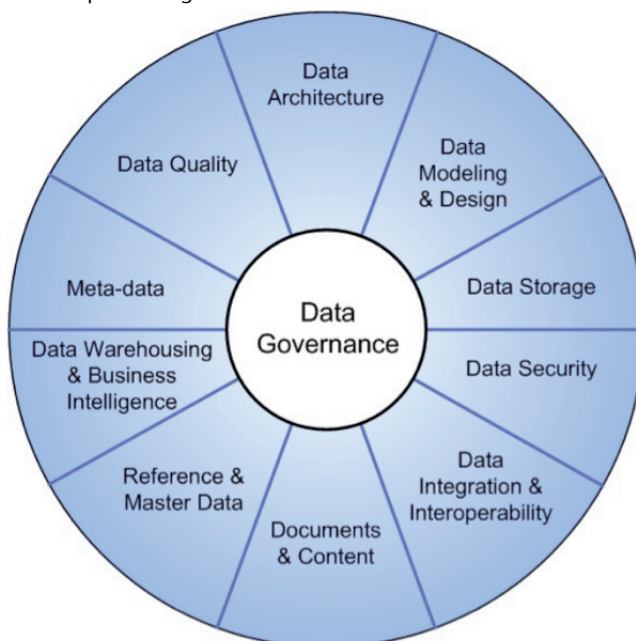
El gobierno del dato es un área muy grande. Incluye la gestión de aplicaciones de negocio (desde la perspectiva del dato), del gobierno TI (como la gestión de las bases de datos) y la gestión de la información. En estos materiales, nos centraremos en aquellos componentes de mayor relevancia para el análisis del dato. Asimismo, discutiremos cómo progresar el gobierno del dato a través de un modelo de madurez.

4.1. Componentes

Tal y como hemos comentado, el gobierno del dato es un ejercicio coral. Por un lado, las diferentes líneas de negocio son responsables de optimizar el uso y la comprensión de la información; por otro, el departamento de TI es responsable de optimizar el acceso y la protección de la información; y por último, la estructura organizativa que soporta el gobierno del dato es responsable de optimizar la disponibilidad y utilidad de la información.

Estas responsabilidades se entrelazan y se despliegan en las diferentes áreas de la gestión de la información, como se ilustra en la figura 4.

Figura 4. Componentes gobierno del dato.



Fuente: DAMA - DMBok.

Es decir, el enfoque actual del gobierno del dato incluye múltiples funciones. Según DAMA, estas funciones o componentes son:

- **Arquitectura de datos** (*data architecture*): consiste en las estructuras de almacenamiento y procesamiento del dato como parte de la arquitectura empresarial.
- **Diseño y modelos de datos** (*data modeling and design*): consiste en la gestión, el análisis, diseño, construcción, testeo y mantenimiento de los modelos de datos.
- **Almacenamiento de datos** (*database storage*): consiste en la gestión de las infraestructuras físicas de almacenamiento de datos.
- **Seguridad de datos** (*data security*): consiste en asegurar la privacidad, la confidencialidad y el acceso en los casos apropiados.
- **Datos maestros y de referencia** (*reference and master data*): consiste en la identificación, mantenimiento, gestión, acceso y propagación de datos maestros y de referencia.
- **Inteligencia de negocio y almacenes de datos** (*data warehousing and business intelligence*): consiste en la gestión de los procesos analíticos de la organización y del acceso de la forma correcta a la información para la toma de decisiones.
- **Integración e interoperabilidad de datos** (*data integration and interoperability*): consiste en la gestión de la adquisición, extracción, transformación, movimiento, propagación, replicación, federación y virtualización de datos.
- **Contenido y documentos** (*documents and content*): consiste en el almacenamiento, protección, indexación y habilitación de acceso de datos en documentos, permitiendo su integración e interoperabilidad con datos estructurados.
- **Metadatos** (*metadata*): consiste en la recopilación, categorización, mantenimiento, integración, control, gestión y distribución de metadatos.
- **Calidad de dato** (*data quality*): consiste en definir, monitorizar, mantener y mejorar la calidad y la integridad del dato.

Y por encima de todas ellas, el gobierno del dato, encargado de la planificación, vigilancia y aplicación de la gestión y del uso del dato.

De este modo, una plataforma tecnológica de gobierno del dato debe cubrir las funciones descritas en mayor o menor medida. Frecuentemente, es posible encontrar plataformas de gestión de datos maestros y de calidad de datos de manera independiente. Y con mayor frecuencia, la función de la seguridad de dato cae bajo la responsabilidad del departamento de seguridad tecnológica. De este modo, avanzar hacia un gobierno del dato requiere coordinar todas estas iniciativas independientes.

4.1.1. Normas, estándares y estrategia

Normalmente, constituyen el paso inicial cuando una organización inicia un proceso de gobierno del dato. Las principales actividades del gobierno del dato con foco en normas, estándares y estrategia incluyen:

- Identificación de personal involucrado, establecimiento de decisiones y clarificación de responsabilidades.
- Establecer, revisar, aprobar y monitorizar normas.
- Establecer, revisar, aprobar y monitorizar estándares.
- Establecer estrategias de datos corporativos.

Vamos a centrarnos de manera detallada en tres áreas: la gestión de la calidad del dato, la gestión de datos maestros y la gestión de la seguridad del dato.

4.1.2. Calidad del dato

Una parte relevante del gobierno del dato es la necesidad de mejorar la calidad del dato. Habitualmente, la calidad se aplica a los datos maestros e involucra algún software de gestión o limpieza de datos. Las principales actividades del gobierno del dato con foco en la calidad del dato incluyen:

- Identificación de personal involucrado, establecimiento de decisiones y clarificación de responsabilidades.
- Definir la estrategia de la calidad del dato.
- Monitorizar la calidad del dato.
- Informar sobre el estado de las iniciativas de calidad.

4.1.3. Gestión de datos maestros

Otra de las funciones relevantes del gobierno del dato es la explotación y gestión eficiente de la información. Esta eficiencia se fundamenta en el hecho de que no hay visiones divergentes respecto al dato, y se evitan los silos de datos. Las principales actividades del gobierno del dato con foco en la gestión de datos maestros incluyen:

- Identificar los datos maestros y los de referencia, tanto internos como externos.
- Consolidar y propagar el uso de los datos maestros y los de referencia.
- Asegurar la calidad de los datos maestros y de referencia.

4.1.4. Gestión de la seguridad del dato

En el contexto del gobierno del dato, hablar de la gestión de la seguridad incluye también la privacidad y el cumplimiento. No solo se trata de proteger el dato, sino de cumplir con los requisitos de los reguladores en el sector y país en los que se opera. Esta función incluye tecnologías para identificar información sensible, proteger el dato, gestionar la seguridad y el acceso y normas de auditoría y control. Una arquitectura de seguridad de datos corporativa es esencial para proteger todos los datos sensibles en uso y almacenados en la organización. Las principales actividades de esta función incluyen:

- Ayudar a proteger los datos sensibles mediante la gestión de acceso y requisitos de seguridad.
- Alineamiento entre la arquitectura de seguridad de datos e iniciativas de negocio.
- Ayudar a asesorar y gestionar el riesgo.
- Ayudar a hacer cumplir los requisitos reguladores y contractuales.
- Identificación del personal involucrado, establecimiento de decisiones y clarificación de responsabilidades.

4.2. Madurez

El desarrollo de un programa de gobierno del dato no está exento de problemas. Encontramos múltiples razones para que no llegue a buen puerto. Por ejemplo:

- Falta de apoyo ejecutivo, es decir, falta de un claro liderazgo en la organización.
- Falta de fundamentos de gestión del dato en la organización, es decir, falta de personas con conocimientos en la organización.
- Falta de métricas de seguimiento del programa, es decir, falta de indicadores y objetivos para comprender el avance del programa.
- Baja planificación en el programa y proyectos, así como no disponibilidad de presupuesto y/o tiempo adecuados, es decir, falta de organización respecto a estas iniciativas.
- Falta de perfiles adecuados, tanto técnicamente como en un ámbito de liderazgo y propiedad del dato.

Evaluar el estado de la organización respecto al gobierno del dato puede ayudar a comprender en qué situación se encuentra la organización, y cuáles son los siguientes pasos.

4.3. Modelo de madurez del *data governance*

Un buen punto de partida para desarrollar una estrategia efectiva del gobierno del dato es identificar cuál es el grado de madurez de la organización al respecto.

Como sucede con otros modelos de madurez, el grado de madurez está centrado en los procesos que ocurren dentro de la organización, su implementación y modo de operación. Estos procesos habilitan que la organización funcione eficazmente, y rigen su funcionamiento. Por tanto, resulta esencial que estos procesos estén implementados de una manera efectiva y que busquen una mejora de la optimización de forma continuada.

En la actualidad, encontramos distintos modelos de madurez relativos al gobierno del dato, los cuales no siempre coinciden en sus características. Entre ellos, destacamos:

- MIKE2.0, que es un marco de referencia de buenas prácticas relativas a la gestión de la información, y cuyo modelo es el *information maturity model*.*
- 2IBM Data Governance Council, formado por más de 50 instituciones que colaboran en el área y cuyo modelo es el *Data Governance Council maturity model*.**
- EDM Council, que es una asociación para el desarrollo y la implementación de estándares relativos al dato y publicación de buenas prácticas, y cuyo modelo es el *data management maturity model* (DCAM).***
- DAMA, una asociación internacional de profesionales focalizados en los conceptos y prácticas relativos a la gestión del dato y de la información, y cuyo modelo es el DMBOK (*data management body of knowledge*).****
- Gartner, consultora de prospección tecnológica, cuyo modelo es el *Enterprise Information Management maturity model*.*****

Aunque haya una falta de consenso entre los modelos existentes, en general se fundamentan todos ellos en el *capability maturity model* (CMM). Como norma general, es posible describir cinco niveles o fases diferentes. Cada uno de estos niveles consta de unos objetivos y prácticas que describen las características presentes, en un ámbito de procesos y actividades necesarias para gestionar y mejorar los procesos relativos.

Madurez

Hace referencia a una medida de la habilidad de una organización para la mejora continua respecto a una disciplina particular.

*Más información en:
<https://goo.gl/zCALoL>

**Más información en:
<https://goo.gl/rUQ12i>

***Más información en:
<https://goo.gl/12yEFS>

****Más información en:
<https://goo.gl/EKG9FE>

*****Más información en:
<https://goo.gl/wVvYUh>

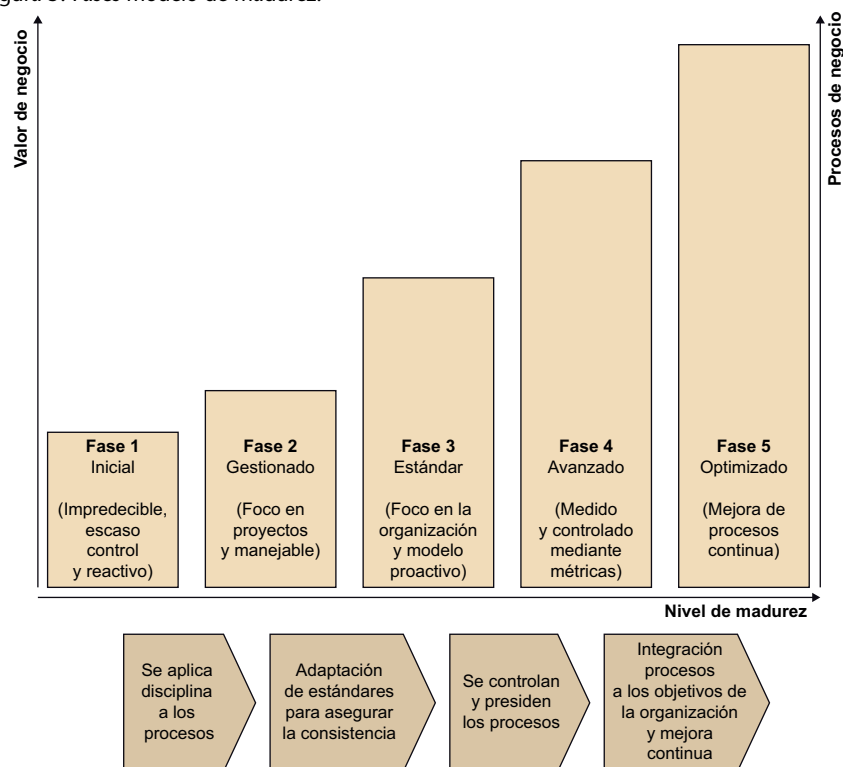
Esto da lugar a 5 + 1 fases (incluyendo el caso de que no hay absolutamente nada) respecto a la madurez en el gobierno del dato: ninguno, inicial, gestionado, estandarizado, avanzado y optimizado. Así, cuanto más alto sea el nivel de madurez de la organización, mayor será su eficiencia y sus costes y riesgos serán menores.

- **Fase 0. Ninguno.** No existe un proceso de gobierno del dato definido.
- **Fase 1. Inicial.** Existe una autoridad del dato en el departamento de TI, pero tiene poca influencia en los procesos de negocio. La colaboración entre áreas de TI y negocio no es consistente, y existe una total dependencia de determinados expertos del dato en cada área de negocio. Como resultado, los procesos no están integrados dentro de la organización.
- **Fase 2. Gestionado.** Dueño y administrador suelen estar presentes en líneas de negocio particulares. Existen procesos poco definidos en las aplicaciones clave en las líneas de negocio, y los problemas del dato son gestionados de una manera reactiva, sin identificar el origen del problema. Es una etapa temprana a la hora de estandarizar los procesos en las diferentes líneas de negocio.
- **Fase 3. Estandarizado.** El negocio está involucrado, existe un equipo de diferentes funciones, así como administradores de datos con responsabilidades claras. Hay establecidos procesos estandarizados y consistencia en las líneas de negocio. Existe un repositorio central de políticas de datos de fácil acceso, y la calidad del dato es regularmente monitorizada y medida.
- **Fase 4. Avanzado.** La estructura organizativa del gobierno del dato se considera como crítica para todas las funciones del negocio. El negocio es el dueño del contenido del dato y de la creación de las políticas del dato. Se crean y se documentan modelos de datos corporativos.
- **Fase 5. Optimizado.** El gobierno del dato es un proceso de negocio central, y las decisiones de negocio se toman evaluando el beneficio, coste y riesgo. Se crean objetivos de mejora de procesos para la organización, los cuales se revisan continuamente para reflejar posibles cambios en los objetivos del negocio. Los costes se reducen, debido a la automatización de procesos.

Así pues, las distintas fases muestran un gobierno del dato indisciplinado que va evolucionando con modelos reactivos en un principio, a proactivos, lo que al final del proceso da lugar a un gobierno que provee una visión única de la organización. En este nivel, las organizaciones usan datos de alta calidad para soportar la automatización de procesos que no necesitan, o no deberían necesitar, la intervención humana.

A medida que este nivel aumenta, el nivel de procesos involucrados será también mayor, como se ilustra en la figura 5.

Figura 5. Fases modelo de madurez.



Fuente: Marcos Pérez.

4.3.1. Modelos de *data governance*

Así pues, las distintas fases muestran un gobierno del dato indisciplinado que va evolucionando con modelos reactivos en un principio, a proactivos, lo que al final del proceso da lugar a un gobierno que provee una visión única de la organización. En este nivel, las organizaciones usan datos de alta calidad para soportar la automatización de procesos que no necesitan, o no deberían necesitar, de la intervención humana.

Vamos a detallar cuatro modelos en función de cuatro dimensiones —personal, normas, tecnología y riesgos y recompensas—:

1) Modelo indisciplinado

- Personal
 - El éxito depende de la competencia de unos pocos empleados.
 - Los analistas de negocio no están implicados en la definición de normas de calidad del dato.
- Normas
 - No existe calidad de dato.
 - Datos y procesos aislados.
 - Recursos no optimizados, debido a la redundancia de datos y su antigüedad.

- Tecnología
 - Estandarización y limpieza de datos en fuentes de datos aislados.
 - La mejora de datos se centra en determinadas aplicaciones.
- Riesgos y recompensas
 - El riesgo es muy alto. Los problemas del dato suponen la pérdida de clientes o procedimientos erróneos.
 - Las recompensas son bajas. La organización recibe pocos beneficios de la calidad de dato.

2) Modelo reactivo

- Personal
 - El éxito depende de un grupo específico de empleados de IT.
 - No existen procedimientos estándar que cubran distintas áreas funcionales, pero sí que hay iniciativas de calidad del dato.
- Normas
 - El foco es corregir errores cuando ocurren.
 - Los procesos de gestión se centran en problemas existentes.
 - Se estandarizan roles y tareas en grupos individuales y departamentos.
- Tecnología
 - Se dispone de herramientas de control de calidad.
 - No existe integración de datos en las unidades de negocio.
- Riesgos y recompensas
 - El riesgo es alto, debido a la falta de integración y consistencia de datos en la organización.
 - Las recompensas son limitadas y anecdóticas, a través de procesos individuales.

3) Modelo proactivo

- Personal
 - El equipo de gestión entiende el rol del gobierno del dato. Se asigna personal y recursos.
 - Se considera el dato como un activo estratégico para la toma de decisiones.
 - Surge la figura del administrador del dato para implementar la estrategia y trabajar con los distintos equipos funcionales.
- Normas
 - Surgen actividades en tiempo real y normas preventivas de calidad del dato.
 - Se comparan las métricas del dato con los estándares de la industria para obtener información para mejorar el modelo.
 - Los objetivos cambian de la corrección de errores a la prevención.
- Tecnología
 - Un grupo de administradores del dato mantiene las normas de negocio y las descripciones de los datos de la organización.
 - Un continuo monitorizado de los datos ayuda a mantener la integridad de la organización.

- Se dispone de procesos en tiempo real, y la calidad del dato es compartida por diferentes modos de operación.
- Riesgos y recompensas
- El riesgo es medio o bajo. Los riesgos se reducen al disponer de una mayor información que incrementa la fiabilidad en la toma de decisiones.
- Las recompensas son medias o altas. Cada vez más empleados adoptan el modelo.

4) Modelo gobernado

- Personal
- El gobierno del dato es soportado en un ámbito ejecutivo en la organización.
- Los usuarios de negocio tienen un rol activo en la estrategia del dato.
- Existe un grupo de gobierno o calidad del dato que trabaja directamente con los administradores del dato y otros usuarios.
- Normas
- Las nuevas iniciativas se aprueban después de considerar cómo van a afectar a la infraestructura del dato existente.
- Existen normas automatizadas para asegurar la consistencia y fiabilidad del dato en toda la organización.
- Tecnología
- Se estandarizan las herramientas de calidad e integración del dato en la organización.
- El dato es monitorizado continuamente, se soluciona inmediatamente cualquier desviación del estándar.
- Los modelos de datos capturan la lógica del negocio y los detalles técnicos de todos los datos organizativos.
- Riesgos y recompensas
- El riesgo es bajo. Los datos maestros están controlados en la organización, lo que permite disponer de una alta calidad de información.
- Las recompensas son altas. Existe una mejor comprensión sobre el negocio, que permite al equipo de gestión una plena confianza en las decisiones basadas en el dato.

4.4. ¿Qué modelo de madurez usar?

Al existir tantos modelos de madurez, las organizaciones no tienen claro cuál elegir y por dónde empezar. Y, a veces, esto se traduce en relegar esta elección al proveedor, que aplica el modelo de madurez asociado a su herramienta o servicio, en lugar de uno controlado por la organización.

Aplicar el modelo de madurez requiere hacerse múltiples preguntas respecto a todos los componentes presentados, así como las personas, las políticas y las capacidades.

La organización debe tomar una decisión: si apostar por modelos con propiedad y de pago (como el de DAMA), por modelos asociados a una empresa (como el de IBM) o, incluso, apostar por un modelo abierto y público como el de Stanford.

El modelo de Stanford es un poco diferente de los anteriores, en el sentido de que, aunque se fundamenta en CMM, limita las áreas que evalúa a personas, procesos y capacidades y se centra en aspectos como los datos maestros y la calidad del dato, dejando aparte la seguridad del dato. Dependiendo de la organización y de su madurez respecto a la seguridad del dato, puede ser también un modelo para tener en cuenta*. Por ejemplo, el Estado de Oklahoma ha aplicado este modelo en lugar de los anteriores**.

Los modelos de madurez son muy útiles para dominios de capacidades conocidos. A medida que en la disciplina del análisis del dato aparecen nuevas tecnologías, es necesario tener en cuenta que estos modelos no reflejan estos cambios y si los reflejan, probablemente, será en una medida de disminución de la madurez. Con ello queremos señalar que el modelo de madurez, si bien es útil como herramienta de guía, no debe convertirse en un freno a la innovación en el análisis y de nuevas técnicas y enfoques.

Lectura complementaria

VV. AA. (2013). *Stanford Data Governance Maturity Model*. Stanford University.

*Más información en:
<https://goo.gl/wo4WUh>

**Más información en:
<https://goo.gl/pw6jpg>

Resumen

En este módulo didáctico, hemos presentado el concepto de gobierno del dato para disponer de una visión holística en el resto de la asignatura.

Primero, hemos explicado el dato como activo de valor para la organización, y hemos visto que existe una necesidad cambiar su gestión.

Además, hemos introducido lo que se conoce como el ciclo de vida del dato, que describe en esencia todas las etapas por las que pasa como activo. En función de la etapa y el foco, el valor del dato será diferente y, por lo tanto, se han comparado distintas formas para su medición. La necesidad de conocer y controlar estas fases nos ha llevado a introducir el linaje del dato.

Esto nos ha llevado a introducir el gobierno del dato como punto central de la gestión de la información de una organización y a explicar en qué consiste, detallando los objetivos que persigue y las fases que incluye. En definitiva, el gobierno del dato busca establecer mecanismos de control suficientes para el control y la explotación eficiente del activo más importante: el dato.

También se han revisado con detalle los componentes que forman el gobierno del dato, para conocer qué es necesario tener en cuenta en la implementación de este tipo de programas. Estos componentes cubren desde la calidad del dato, hasta la inteligencia de negocio y el *data warehouse*.

Por último, se han revisado diferentes modelos de madurez para poder identificar la situación actual de una organización y determinar pasos futuros.

Glosario

big data *m* Conjunto de estrategias, tecnologías y sistemas para el almacenamiento, procesamiento, análisis y visualización de conjuntos de datos complejos.

business intelligence *f* Conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información, que permite tomar mejores decisiones a los usuarios de una organización.

ciclo de vida de un activo *m* Distintas etapas por las que pasa un activo, desde su nacimiento hasta el final.

gestión del ciclo de vida de un activo *m* Enfoque basado en políticas para gestionar el flujo del dato a través de uno o varios sistemas de información a lo largo de su ciclo de vida.

gobierno del dato *m* Ejercicio de autoridad, control y toma de decisiones compartida (planificación, vigilancia y aplicación) sobre la gestión de los activos de datos.

hoja de ruta *f* Plan detallado que refleja los objetivos a corto y largo plazo y guía el progreso para alcanzar estos objetivos, usando a menudo una solución tecnológica específica.

Internet de las cosas *f* Hace referencia a la interconexión digital de objetos cotidianos con Internet.

Ley Sabarnes Oxley *f* Ley que busca monitorizar a las empresas que cotizan en bolsa de valores, para evitar que la valorización de sus acciones sea alterada de manera dudosa, mientras que su valor es menor.

linaje del dato *m* Capacidad de conocer todo el ciclo de vida de un dato, desde la fecha y hora exacta en la que fue extraído y el momento en que se produjo su transformación, hasta el instante en el que tuvo lugar su carga desde un entorno fuente a otro de destino.

machine learning *f* subcampo de la inteligencia artificial con foco en la creación de algoritmos con capacidades de aprendizaje a partir de datos y de los resultados buscados.

Bibliografía

Berson, A.; Dubov, L. (2011). *Master Data Management and Data Governance, 2/E*. Nueva York: McGraw-Hill Education.

Brackett, M.; Earley, P. S. (2009). *The DAMA Guide to The Data Management Body of Knowledge (DAMA-DMbok Guide)*. Nueva York: DAMA.

Ladley, R. S. (2012). *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program*. San Francisco: Morgan Kaufmann.

Mosley, M (2009). *DAMA-DMbok functional framework*. Nueva York: DAMA.

Plotkin, D. (2013). *Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance*. San Francisco: Morgan Kaufmann.

Seiner, R. S. (2014). *Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success*. Basking Ridge: Technics Publications LLC.

VV. AA. (2007). *The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance*. Nueva York: IBM Data Governance Council.

