

---

# Gestión de datos maestros

---

PID\_00246837

Josep Curto Díaz

---

Tiempo mínimo de dedicación recomendado: 3 horas

---



---

Universitat  
Oberta  
de Catalunya

---

*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

<b>Introducción</b>	5
<b>Objetivos</b>	6
<b>1. Gestión de datos maestros</b>	7
1.1. La necesidad de la gestión de datos maestros	7
1.2. ¿Qué es <i>master data management</i> ?	8
1.3. Tipos de datos maestros	10
1.4. ¿Qué beneficios aporta la gestión de datos maestros?	11
<b>2. Programa de gestión de datos maestros</b>	14
2.1. Fases de un programa de gestión de datos maestros	14
2.2. MDM en contexto de gobierno del dato	17
2.3. Evaluando la madurez de la gestión de datos maestros	18
2.3.1. Nivel 1: inicial	18
2.3.2. Nivel 2: reactivo	19
2.3.3. Nivel 3: gestionado	20
2.3.4. Nivel 4: proactivo	21
2.3.5. Nivel 5: estratégico	21
<b>3. Desarrollando un programa de gestión de datos maestros</b>	23
3.1. Enfoques del programa MDM	23
3.2. Cómo decidir qué es necesario mantener y administrar	25
3.3. Creando y manteniendo una lista maestra	27
3.3.1. Creación de una lista maestra	27
3.3.2. Mantenimiento de una lista maestra	28
3.4. Información por dominio	29
3.4.1. Cliente B2B	30
3.4.2. Cliente B2C	30
3.4.3. Producto	31
3.4.4. Datos de referencia	31
3.4.5. Localización	32
3.4.6. Organización	32
3.4.7. Proveedor	33
3.4.8. Empleado	33
<b>4. Técnicas y tecnologías para la gestión de datos maestros</b>	34
4.1. Técnicas para la gestión de datos maestros	34
4.2. Tecnologías para la gestión de datos maestros	35
4.3. Ejemplo de caso de uso	37

4.4. MDM y <i>big data</i> .....	38
<b>Resumen</b> .....	40
<b>Glosario</b> .....	41
<b>Bibliografía</b> .....	42

## Introducción

El modelo de datos de las organizaciones se está haciendo más y más complejo, al tener que dar soporte a lo que conocemos como la transformación digital. Esta transformación se traduce en que los activos digitales empresariales están distribuidos en múltiples sistemas de información, dispositivos y en la nube, y desacoplados en diferentes capas de abstracción y servicios, y al mismo tiempo deben ser analizados mediante técnicas analíticas y cumplir los marcos de regulación pertinentes tanto generales como vinculados al sector de actividad.

Sin una estrategia adecuada, frecuentemente se generan silos de información que conforman una visión dispersa, fragmentada e incoherente de clientes, productos, empleados, proveedores y procesos.

Para dar respuesta a la necesidad de gestionar esta complejidad, dentro del gobierno del dato existe lo que se conoce como la **gestión de datos maestros** (*master data management*), que permite limpiar, racionalizar e integrar los datos en un sistema de registro empresarial para dar soporte a todas las funciones básicas de la organización.

En una primera instancia, podemos entender que la gestión de datos maestros pone el foco en el principio de estandarización del gobierno del dato.

En este módulo estudiaremos qué es y en qué consiste la gestión de datos maestros; qué es un programa de gestión de datos maestros; cómo se desarrolla dicho programa; qué aspectos es necesario tener en cuenta para llevar a buen puerto esta iniciativa; y qué técnicas y tecnologías son necesarias para la gestión de datos maestros.

## Objetivos

Este material didáctico está dirigido a:

- 1) Desarrolladores y consultores que quieren conocer qué significa gestión de datos maestros o *master data management*.
- 2) Desarrolladores y consultores que quieren ayudar al desarrollo de estrategias de negocio que incluyan gestión de datos maestros.
- 3) Gestores que están interesados en la transformación digital de su organización y en la inclusión de gestión de datos maestros 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 *master data management*, las situaciones en las que es necesario desplegar una solución de este tipo y las ventajas que proporciona.
- 2) Conocer en qué consiste un programa de gestión de datos maestros.
- 3) Enumerar y dar a conocer mejores prácticas de gestión de datos maestros.
- 4) Conocer técnicas y tecnologías para la gestión de datos maestros.

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. Gestión de datos maestros

### 1.1. La necesidad de la gestión de datos maestros

La gestión de datos maestros está recibiendo cada vez más atención en el contexto empresarial. Encontramos distintos motivos para este interés. Por un lado, el dato juega un papel fundamental en la generación de ingresos, en el control de costes y en la mitigación de riesgos. Ser capaces de compartir, almacenar, proteger y recuperar los datos relevantes cuando es necesario puede derivar en importantes ventajas competitivas. En definitiva, las organizaciones deben cumplir normativas (propias del sector y genéricas) y verificar requerimientos estratégicos de negocio como los siguientes:

- Con el objetivo de capturar, retener y generar valor de los clientes, toda organización necesita una visión de 360 grados de los datos de clientes, desde sus datos de perfil y contrato hasta las transacciones e interacciones con cada uno de los departamentos y canales de venta y soporte.
- Con el objetivo de establecer un análisis completo de gastos, el departamento financiero debe acceder a toda la información relacionada, con independencia del centro de coste propietario del dato (por ejemplo, marketing).
- Desde junio del 2007, las organizaciones que realizan procesos industriales deben cumplir la Ley de la Unión Europea (UE) sobre las sustancias químicas y su uso y seguridad, conocida como REACH (registro, evaluación, autorización y restricción de sustancias y preparados químicos). Los objetivos de esta ley son: proteger mejor la salud y el medio ambiente de los posibles riesgos derivados de las sustancias químicas, promover el uso de métodos alternativos de comprobación de la seguridad, y mejorar la manipulación y el uso de estas sustancias con total seguridad en todos los sectores de la industria.
- En el sector financiero, las organizaciones que participan en el sector deben tener en cuenta directivas como MiFID II (*markets in financial instruments directive*), que armoniza la regulación sobre los mercados de valores, los instrumentos financieros que en ellos se negocian y la organización y relación con los clientes de las entidades financieras que prestan servicios de inversión y protección al inversor. En esencia, esta directiva obliga a una mayor transparencia de los mercados, en aras de la mayor proyección de los inversores, de modo que todo tipo de dato de la entidad debe ser consistente, actualizado y completo.

Todas estas responsabilidades empujan a las organizaciones a dedicar tiempo a gestionar el dato. Sin embargo, muchas organizaciones o bien no han desarrollado estas estrategias de forma integrada y transversal, o bien solo son iniciativas puntuales o manuales. De este modo, frecuentemente se dedican a arreglar y mejorar el dato en lugar de usarlo y tomar mejores decisiones. Cuando se llega a esta situación, se ha alcanzado el límite de las capacidades de mantenimiento manual y el dato se convierte en un activo tóxico para la organización, tal y como apunta Nassim Nicholas Taleb. El principal problema de tener que arreglar el dato de forma recurrente es la pérdida de confianza de los usuarios de negocio, puesto que no pueden usar el dato para la toma de decisiones cuando realmente se necesita y tiene su máximo valor.

### Lectura complementaria

Taleb, N. N. (2016). *The Bed of Procrustes: Philosophical and Practical Aphorisms*. Penguin

## 1.2. ¿Qué es *master data management*?

Existen muchas formas de clasificar los datos. En el contexto de estos materiales, vamos a introducir una nueva. Según Reichert, Otto y Österle, los datos de una organización pueden dividirse en datos maestros, datos transaccionales y datos de inventario. Veámoslos:

### Lectura complementaria

Otto, B.; Reichert, A. (2010). *Organizing master data management: findings from an expert survey*. *Proceedings of the 2010 ACM Symposium on Applied Computing* (págs. 106-110). ACM.

**1) Datos maestros:** hacen referencia a las características de los objetos básicos de negocio dentro de una organización. Este tipo incluye datos de materiales, productos, proveedores, clientes, empleados y activos. Por ejemplo, respecto al cliente, su nombre y apellidos forman parte de los datos maestros.

**2) Datos transaccionales:** describen las transacciones comerciales y representan la entrada y salida de los procesos empresariales. Se crean constantemente en el curso de las actividades empresariales, hacen referencia a datos maestros e indican cambios en los datos de inventario. Por ejemplo, imaginemos que tenemos una cafetería: la compra de un café y una galleta para desayunar y los datos vinculados a dicha compra (importe, hora, etc.) son datos transaccionales.

**3) Datos de inventario:** hacen referencia a los niveles de existencias y cuentas. Por ejemplo, si continuamos con el tema de la cafetería, la cantidad en existencia de café, en qué fecha se compró o la fecha de caducidad son datos de inventario.

Cada uno de estos tipos de datos tiene características diferentes que podemos comparar y que nos permiten comprender mejor las diferencias entre ellos:

**1) Referencia temporal:** es la dependencia del dato respecto a un determinado punto en el tiempo.

**2) Frecuencia de actualización:** hace referencia a la frecuencia con la que un dato se actualiza o cambia durante un determinado periodo de tiempo.



**3) Volatilidad de volumen:** hace referencia al crecimiento del dato en volumen respecto al tiempo.

**4) Independencia existencial:** es el grado en el que el dato hace referencia a otros tipos de datos.

Y que se resumen en la tabla 1.

Tabla 1. Tipos de datos y características

Tipo	Referencia temporal	Frecuencia actualización	Volatilidad volumen	Independencia existencial
Datos maestros	Baja	Baja	Baja	Alta
Datos transaccionales	Alta	Baja	Alta	Baja
Datos de inventario	Alta	Alta	Baja	Baja

### Ejemplo

Vamos a ejemplificar estas características:

- 1) El nombre y los apellidos de un cliente, que sabemos que son datos maestros, no suelen cambiar en el tiempo, por lo que está claro que las tres primeras características son baja y la cuarta, alta.
- 2) En el caso de la compra del café y la galleta, este dato está vinculado a un cierto momento en el tiempo, no se va a actualizar, depende de los datos maestros (cliente y productos).
- 3) En el caso del nivel de existencia de café, depende del tiempo, se actualiza tras cada transacción en la que se vende un café, depende de los datos maestros y las transacciones, pero tan solo cambia un registro.

Teniendo en cuenta estos tipos de datos, así como sus características, queda patente que necesitamos una disciplina para poder gestionar los datos maestros.

Se entiende como **gestión de datos maestros**, o *master data management* (MDM), el conjunto de procesos, gobierno, políticas, estándares y herramientas que definen y gestionan constantemente los datos críticos de una organización para proporcionar un único punto de referencia.

De este modo:

- La gestión de datos maestros es una pieza fundamental para las operaciones y la toma de decisiones analíticas.
- Los datos maestros se pueden encontrar dispersos por toda la organización, e incluso fuera de ella.
- El dato maestro establece una definición estándar para datos críticos desde la perspectiva de negocio que se comparte a lo largo de la organización, y representa la única fuente de verdad.

En el contexto de MDM, frecuentemente se habla del registro de oro (*golden record*), o versión única de la verdad.

Se entiende por *golden record* la versión única y bien definida de todas las entidades de datos en un ecosistema organizacional.

La gestión de datos maestros busca generar este tipo de registros dentro de la organización, respondiendo a las siguientes preguntas:

- ¿Cuáles son las entidades importantes de negocio que se usan como referencia repetidamente?
- ¿Qué datos describen la misma entidad de negocio?
- ¿Dónde están almacenados los datos?
- ¿Qué datos son los más exactos?
- ¿Qué datos son más relevantes para tareas específicas?
- ¿Qué datos de diferentes y múltiples fuentes de origen pueden ser integrados para crear una vista más completa y comprensible de la entidad de negocio?
- ¿Qué reglas de negocio se pueden establecer de forma automática para combinar, emparejar, identificar, etc.?
- ¿Cómo identificamos y restauramos datos que han sido combinados y emparejados de forma incorrecta?

### 1.3. Tipos de datos maestros

Aunque en general hablamos de gestión de datos maestros, la realidad es que existen diferentes tipos, más allá del ámbito de negocio al que hacen referencia, que, como ya sabemos, pueden ser materiales, productos, proveedores, clientes, empleados y activos. Podemos diferenciar los siguientes:

- **Datos de referencia:** es el conjunto de valores admisibles que deben ser utilizados por otros campos de datos (maestros o transacciones). Los datos de referencia normalmente cambian lentamente, reflejando cambios en los modos de operación de la empresa, en lugar de cambiar en el curso normal de los negocios.

- **Datos de maestros:** son una fuente única de datos empresariales comunes utilizados en múltiples sistemas, aplicaciones y/o procesos.
- **Datos maestros corporativos:** hacen referencia a la fuente única de datos empresariales comunes utilizados en todos los sistemas, aplicaciones y procesos de toda una empresa (todos los departamentos, divisiones, empresas y países).
- **Datos maestros de mercado:** son la fuente única de datos comerciales comunes para todo un mercado. Los datos maestros de mercado se utilizan entre empresas dentro de la cadena de valor. Un ejemplo de datos maestros de mercado es el UPC (*universal product code*), que se encuentra en los productos de consumo.

Estos tipos de datos maestros tienen asociados metadatos que son relevantes como, por ejemplo, significado y propósito del dato; en qué fuentes de origen aparece el dato; la versión del dato en cada fuente de origen; origen del dato; versión disponible en el sistema MDM; fecha de actualización del dato maestro; cómo se efectúa el mantenimiento del dato; y quién es responsable de la gestión del dato y del metadato.

#### 1.4. ¿Qué beneficios aporta la gestión de datos maestros?

El *master data management* aporta múltiples beneficios a una organización, que van más allá de tener una visión única del dato. La creación y gestión de datos maestros proporciona:

- Una comprensión consistente de los activos de datos más críticos y comunes de la organización.
- Mecanismos para el uso consistente de los datos maestros a lo largo de la organización.
- Está diseñado para la gestión de cambio, lo que ayuda a una organización a adaptarse al mercado.

Vamos a entrar en detalles en tres beneficios que aportan valor en un ámbito operativo.

El primero beneficio hace referencia a **optimizar y evitar errores** en diferentes procesos de negocio, al trabajar con datos maestros consistentes, confiables y compartidos entre todos ellos. Entre estos procesos, destacamos la mejora de:

- Informes financieros.
- Gestión y atención de clientes.

- Gestión de reclamos.
- Gestión del cumplimiento regulatorio de procesos y productos.
- Gestión de la privacidad y la seguridad.
- Gestión de la cadena de suministro.
- Gestión de la seguridad.
- Ventas cruzadas y aumento del ticket promedio.

El segundo hace referencia al hecho de que permite **ganar eficiencia y efectividad**, dado que, por ejemplo:

- Se evita destinar recursos para gestionar problemas de datos de forma reactiva.
- Se evita dañar la credibilidad con clientes, proveedores y socios, al poder confiar en los datos compartidos con ellos.
- Se reducen efectos negativos en los ingresos y beneficios, ya que los datos pueden reflejar la realidad del negocio, y es posible tomar decisiones más eficientes.

En tercer lugar, la gestión de datos maestros es una pieza angular para que la organización **trabaje en contextos VUCA**. En particular, centrémonos en algunos escenarios específicos de aplicación:

- **Gestión del riesgo:** la volatilidad del mercado, las brechas en la seguridad cibernética e incluso la inestabilidad de la política internacional están contribuyendo al aumento del riesgo empresarial. Para mitigar ese riesgo, es necesario comprender el dato desde su origen y sus modificaciones. Por ejemplo, teniendo trazabilidad del dato se puede conocer el punto donde ha sucedido la brecha y, por lo tanto, cuál es el punto más débil de la cadena.
- **Cumplimiento normativo:** cada año, aparecen nuevos escenarios que requieren la actuación de las agencias reguladoras, desarrollando o extendiendo medidas como HIPAA (*health insurance portability and accountability*, 1996), SOX (Sarbanes-Oxley, 2002) o GDPR (*general data protection regulation*, 2016). Ser capaz de producir informes precisos es más importante que nunca, sea cual sea el ámbito que el organismo regulador ocupe: medio ambiente, impuestos, empleo, protección del consumidor, importación/exportación o empleados.
- **Gestión de inventarios:** los estándares para productos, que podemos encontrar en la tabla 2, requieren datos maestros limpios para garantizar que las empresas puedan mantenerse al día con sus inventarios y ventas. Por ejemplo, imaginemos que somos una empresa con tiendas físicas en España, y que cada una de las tiendas compra de forma independiente al

#### VUCA

Cuando hablamos de VUCA, hacemos referencia a un acrónimo de las condiciones de volatilidad (*volatility*), incerteza (*uncertainty*), complejidad (*complexity*) y ambigüedad (*ambiguity*) que describen el mercado actual.

mismo proveedor. El hecho de usar la gestión de datos maestros daría conocimiento de esta situación y, por lo tanto, se podría haber negociado el precio con el proveedor.

Tabla 2. Estándares en la gestión de inventarios

Acrónimo	Identificador	Se usa para identificar.	Ejemplos
GTIN	<i>Global trade item number</i>	Productos y servicios	Lata de sopa, barra de chocolate, álbum de música
EAN	<i>European article number</i>	Productos y servicios, en un ámbito europeo	Cereales, abrigos, aerosol
JAN	<i>Japanese article number</i>	Productos y servicios, en el ámbito de Japón	Arroz, soja, sake
ASIN	<i>Amazon standard identification number</i>	Productos y servicios, en el contexto de Amazon	Libro, videojuego, camiseta
ISBN	<i>International standard book number</i>	Publicaciones	Libros, revistas
GLN	<i>Global location number</i>	Localización	Empresas, almacenes, fábricas, tiendas
SSCC	<i>Serial shipping container code</i>	Unidades de logística	Cargas unitarias sobre palés, jaulas de rodillos, paquetes
GRAI	<i>Global returnable asset identifier</i>	Activo devolutivo	Palés, cajas
GIAI	<i>Global individual asset identifier</i>	Activo	Equipos médicos, de fabricación, de transporte y de TI
GSRN	<i>Global service relation number</i>	Relaciones entre proveedor de servicios y receptor	Médicos en un hospital, miembros de la biblioteca
GDTI	<i>Global document type identifier</i>	Documentos	Formularios de envío, licencias de conducción
GINC	<i>Global identification number for consignment</i>	Consignaciones	Unidades logísticas transportadas juntas en un contenedor oceánico
GSIN	<i>Global shipment identification number</i>	Envíos	Unidades logísticas entregadas a un cliente de forma conjunta
GCN	<i>Global coupon number</i>	Cupones	Cupones digitales
CPID	<i>Component/part identifier</i>	Componentes y partes	Componentes de un automóvil

## 2. Programa de gestión de datos maestros

Desarrollar una iniciativa de gestión de datos maestros no solo consiste en implementar una solución tecnológica. Es necesario desarrollar un programa para que llegue a buen puerto, y se debe fundamentar en los siguientes principios rectores:

- Los datos compartidos pertenecen a la empresa y no a un departamento.
- MDM es un esfuerzo progresivo y continuo.
- Los administradores de datos son la autoridad responsable para controlar los valores de los datos de referencia.
- Los valores de los *golden records* representan el mejor esfuerzo de la organización para tener la información más **exacta**, **actual** y **relevante** de las entidades de negocio clave.

En definitiva, se entiende como **programa de gestión de datos maestros** la metodología estratégica y sistemática para identificación, captura, almacenamiento, gestión y propagación de datos maestros en una organización.

Como cualquier otro proyecto, el programa debe tener en cuenta los requisitos, las prioridades, la disponibilidad de recursos, el tiempo y el tamaño del problema que se está abordando.

Cabe comentar que un programa de *master data management* tiene puntos en común con el gobierno del dato, con la calidad del dato, así como con la inteligencia de negocio. Esto significa que avances en estas iniciativas repercuten positivamente en el desarrollo del programa.

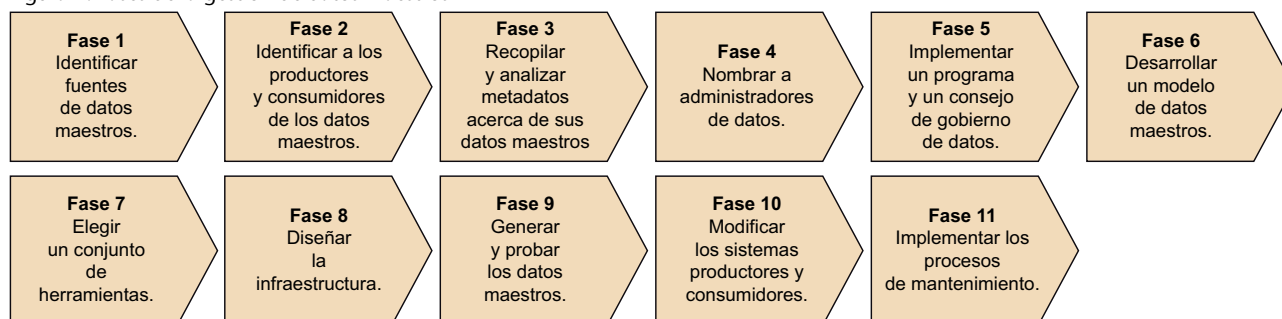
### 2.1. Fases de un programa de gestión de datos maestros

Aunque un programa MDM de una empresa a otra puede tener diferentes etapas, la mayoría de los proyectos MDM incluyen, al menos, estas fases:

- Identificar fuentes de datos maestros.
- Identificar a los productores y consumidores de los datos maestros.
- Recopilar y analizar metadatos acerca de sus datos maestros.
- Nombrar a administradores de datos.
- Implementar un programa y un consejo de gobierno del dato.
- Desarrollar un modelo de datos maestros.
- Elegir un conjunto de herramientas.
- Diseñar la infraestructura.
- Generar y probar los datos maestros.
- Modificar los sistemas productores y consumidores.
- Implementar los procesos de mantenimiento.

Estas fases mínimas son secuenciales, tal y como se ilustra en la figura 1.

Figura 1. Fases de la gestión de datos maestros



Fuente: Josep Curto

Vamos a explicar con detalle estas etapas:

- **Identificar fuentes de datos maestros:** en este paso, se crea un catálogo de todas las fuentes de datos (bases de datos, ficheros, etc.) que contienen los datos de materiales, productos, proveedores, clientes, empleados y activos. Este paso frecuentemente resulta muy revelador, dado que se descubre que los datos suelen estar dispersos en múltiples fuentes que el departamento de las tecnologías de información (TI) no sabía que existían.
- **Identificar a los productores y consumidores de los datos maestros:** tras tener el catálogo de datos, el siguiente paso es conocer qué actores en la organización generan y consumen datos. Por ej., qué aplicaciones/personas producen los datos maestros identificados en el primer paso y qué aplicaciones/personas los utilizan. En algunos casos, este punto puede ser más sencillo si todos los cambios se detectan y se manejan en el ámbito de base de datos.
- **Recopilar y analizar metadatos acerca de sus datos maestros:** este paso consiste en generar conocimiento sobre las fuentes de datos. Para todas las fuentes identificadas en el primer paso, ¿cuáles son las entidades y atributos de los datos, y qué significan? Responder a esta pregunta significa

conseguir nombre de atributo, tipo de datos, valores permitidos, restricciones, valores predeterminados, dependencias y quién posee la definición y el mantenimiento de los datos. Frecuentemente, determinar quién es el propietario del dato resulta sumamente difícil, y pone de manifiesto una necesidad.

- **Nombrar a administradores de datos:** en la etapa anterior, habrá quedado patente qué datos tienen propietarios y administradores. Un administrador de datos, también conocido como *data steward*, es una persona con el conocimiento de los datos fuente actuales y la capacidad de determinar cómo transformar la fuente en el formato de datos maestros. En general, los administradores deben ser nombrados por los propietarios de cada fuente maestra de datos, por los arquitectos responsables de los sistemas MDM y por representantes de los usuarios empresariales de los datos maestros.
- **Implementar un programa y un consejo de gobierno del dato:** la implementación de MDM necesita del gobierno del dato. Es sumamente aconsejable que exista un proceso de toma de decisiones respecto al dato bien definido, puesto que será necesario tomar decisiones sobre cómo se mantienen los datos maestros, qué contienen, cuánto tiempo se mantienen y cómo se autorizan y auditan los cambios.
- **Desarrollar un modelo de datos maestros:** normalmente, esta es una de las fases más importantes y difíciles. Aquí es necesario determinar cómo son los datos maestros. Es decir, qué atributos se incluyen, qué tamaño y tipo de datos son, qué valores se permiten, etc. También aquí se incluye la asignación entre el modelo de datos maestros y las fuentes de datos actuales. La dificultad, en este punto, reside en el hecho de que si se incluyen todos los atributos de origen en la entidad maestra, a menudo se termina con un modelo de datos maestros demasiado complejo y engorroso como para ser útil.
- **Elegir un conjunto de herramientas:** la implementación de la gestión de datos maestros requiere herramientas de soporte para el almacenamiento del dato, el tratamiento del dato, su limpieza y su gestión. Existen herramientas especializadas en MDM (que integran diferentes componentes necesarios), o se aplican tan solo en uno de los conjuntos de datos, como puede ser para cliente, conocidas como *customer data integration* (CDI) o para producto, conocidas como *product information management* (PIM); o bien es posible aprovechar herramientas existentes en la organización de ámbitos como la integración de datos o la calidad del dato. Como veremos en el apartado 4, el conjunto de herramientas debe soportar múltiples funciones, entre las que destacamos también mantener versiones y jerarquías de datos maestros.
- **Diseñar la infraestructura:** los datos maestros deben almacenarse en una infraestructura que permita exponerlos a las aplicaciones y proporcionar



al mismo tiempo procesos para administrarlos y mantenerlos. Existen diferentes opciones para el diseño, como veremos en el apartado 3. El foco de esta infraestructura es la fiabilidad y la escalabilidad.

- **Generar y probar los datos maestros:** en este paso, se utilizan las herramientas que han sido seleccionadas y se combinan los datos de origen para confirmar las listas de datos maestros. Durante el proceso, frecuentemente se requiere el ajuste de reglas de negocio al descubrir, por ejemplo, excepciones. Aunque las herramientas han avanzado considerablemente, la inspección manual puede ser necesaria para asegurar que los resultados son correctos y cumplir con los requisitos establecidos para el proyecto. Este proceso se conoce como **curación de datos**.
- **Modificar los sistemas productores y consumidores:** la implementación de MDM en una organización implica frecuentemente cambiar los sistemas que producen, mantienen o consumen datos maestros para trabajar con la nueva fuente de datos maestros. Si los datos maestros se utilizan en un sistema separado de los sistemas de origen (un almacén de datos, por ejemplo), es posible que los sistemas fuente no tengan que cambiar. Sin embargo, si los sistemas fuente van a utilizar los datos maestros, probablemente habrá cambios requeridos. Los sistemas fuente tendrán que acceder a los nuevos datos maestros o los datos maestros tendrán que sincronizarse con los sistemas fuente, de modo que los sistemas fuente tengan una copia de los datos maestros limpios que se utilizarán. En este punto, las técnicas de integración de datos jugarán un papel fundamental.
- **Implementar los procesos de mantenimiento:** MDM es un proceso continuo en el que se debe asegurar y mantener la calidad de los datos maestros. Esta fase consiste en establecer mecanismos para reconocer los problemas (por ejemplo, señalar coincidencias cuestionables), simplificar las correcciones (por ejemplo, la herramienta debe proponer soluciones y ayudar a la rápida identificación de los errores) y revisar el historial de cambios (por ejemplo, para aislar la fuente de errores y deshacer cambios incorrectos o revisar los elementos que se agregaron como nuevos, porque los criterios de coincidencia estaban cerca pero por debajo del umbral).

## 2.2. MDM en contexto de gobierno del dato

En el contexto de gobierno del dato, la gestión de datos maestros es una función más. Como ya sabemos, cada función tiene distintas actividades (planificación, control, de desarrollo y operativas), cada una de las mismas efectuada por el rol correspondiente.

Para la gestión de datos maestros, estas actividades son:

- Gestión de los datos maestros y de referencia (actividad de planificación).
- Comprensión de las necesidades de integración de datos maestros y de referencia (actividad de planificación).
- Definición de la arquitectura de MDM (actividad de desarrollo).
- Implementación de la solución de MDM (actividad de desarrollo).
- Validación de datos de control y de referencia (actividad de control).
- Integración de datos maestros (actividad operativa).
- Replicación de datos maestros y de referencia (actividad operativa).
- Mantenimiento de las jerarquías dimensionales (actividad operativa).

El programa de gestión de datos maestros cubre estas funciones, que forman parte del marco más general.

### 2.3. Evaluando la madurez de la gestión de datos maestros

Cuando hemos hablado de las diferentes fases que componen un programa de gestión de datos maestros, hemos comentado que no todas las organizaciones pasarán por las mismas etapas. Esto es así puesto que frecuentemente ya han empezado a desarrollar algunas de las capacidades necesarias para llevar a cabo el programa.

Aplicar un modelo de madurez en este ámbito permite conocer qué capacidades ya han sido desarrolladas y en qué medida, y sugerir cuáles no están presentes o se deben sofisticar para continuar el desarrollo del programa.

Vamos a presentar un modelo de madurez para la gestión de datos maestros fundamentado en el tradicional modelo de madurez de capacidades (CMM), desarrollado por el Instituto de Ingeniería de Software de la Carnegie Mellon University y creado por David Loshin.

El modelo de madurez tiene cinco niveles de madurez: inicial, reactivo, gestionado, proactivo y estratégico; y cada nivel se evalúa respecto a seis capacidades: arquitectura, gobierno, gestión, identificación, integración, gestión de procesos de negocio. Una organización puede encontrarse respecto a cada una de las capacidades en un nivel diferente, y no necesariamente todas en el mismo nivel.

#### 2.3.1. Nivel 1: inicial

En este nivel, las capacidades se caracterizan por:

- **Arquitectura:** la arquitectura de aplicaciones se define para cada aplicación empresarial. Existe una consolidación empresarial limitada de modelos representativos y colecciones de diccionarios de datos en varias formas, pero sin modelos de datos maestros.

- **Gobierno:** se lleva a cabo una limpieza limitada de datos por aplicación/línea de negocio, con fines específicos (por ejemplo, normalización de direcciones), pero hay ausencia de modelos definidos de propiedad de datos. Emerge el reconocimiento de la necesidad de supervisión.
- **Gestión:** se lleva a cabo la gestión de identidades por aplicación cuando es necesario (por ejemplo, clientes). Se hace la configuración de la aplicación, pero no coordinada a través de la gestión centralizada.
- **Identificación:** uso limitado de la gestión de identidad por línea de negocio. Se intenta consolidar los datos del cliente según lo requieran las aplicaciones (por ejemplo, actualizaciones de software o transición de aplicaciones contables).
- **Integración:** existen copias replicadas de datos de referencia, pero con reuso limitado de datos, y no hay reutilización de servicios de aplicaciones.
- **Gestión de procesos de negocio:** participación de negocio limitada o no existente, excepto en el nivel más alto de definición de requisitos.

### 2.3.2. Nivel 2: reactivo

En este nivel, las capacidades se caracterizan por:

- **Arquitectura:** se intenta reunir los diccionarios de datos en un único repositorio. Empieza la exploración inicial en servicios de aplicaciones de bajo nivel y la revisión de opciones para compartir información (por ejemplo, integración de información empresarial o integración de aplicaciones empresariales).
- **Gobierno:** se usan aplicaciones externas para administrar metadatos, y se introduce la gestión de la calidad de los datos para análisis, normalización y consolidación.
- **Gestión:** se asignan recursos para administrar el uso de conjuntos de herramientas introducidos para la gestión de datos maestros. La capacitación para el despliegue de herramientas y tecnología de la empresa hace que las capacidades estén disponibles en una base más amplia. Hay administración centralizada de metadatos e índices maestros.
- **Identificación:** se habilita la búsqueda de identidades y coincidencias utilizadas para reducir la duplicación, así como la búsqueda de identidad y concordancia utilizadas para la vinculación de registros rudimentarios con fines domésticos.

- **Integración:** exploración inicial de la consolidación de datos para aplicaciones analíticas recientemente desarrolladas (por ejemplo, CRM). El almacén de datos es utilizado como un repositorio central para datos maestros. No hay integración de nuevo en las aplicaciones contribuyentes.
- **Gestión de procesos de negocio:** se describen modelos de procesos empresariales conceptuales y emerge el uso inicial de reglas empresariales integradas en aplicaciones. También aparece la integración analítica de aplicaciones de datos consolidados.

**CRM**

Cuando hablamos de CRM, *customer relationship management*, hacemos referencia a las estrategias, prácticas y tecnologías que habilitan la gestión de las interacciones con un cliente a lo largo de su ciclo de vida.

### 2.3.3. Nivel 3: gestionado

En este nivel, las capacidades se caracterizan por:

- **Arquitectura:** existe un modelo de datos de maestros definido para la persistencia, y una arquitectura fundamental para el repositorio maestro compartido. Así como un marco operacional identificado para actividades de ciclo de vida de los datos maestros de bajo nivel, y servicios definidos para la integración con el repositorio maestro.
- **Gobierno:** herramientas de calidad de datos en su lugar. Políticas y procedimientos para la gestión de la calidad de los datos. Seguimiento de problemas de calidad de datos. Procesos de estándares de datos en el lugar. Línea de administración de datos empresariales.
- **Gestión:** gestión de identidades centralizada en el índice maestro y utilizada en numerosas aplicaciones. Jerarquías identificadas (dentro de una clase de datos) y utilizadas por las aplicaciones analíticas. Configuración avanzada y administración del uso de los datos maestros. Un plan de migración está disponible para aplicaciones seleccionadas.
- **Identificación:** servicios de búsqueda de identidad disponibles para las aplicaciones. Enlace de registro integrado en la capa de servicio MDM. Normas de fusión y consolidación estandarizadas y gestionadas bajo control centralizado. Fusión y consolidación de procesos establecidos y repetibles.
- **Integración:** servicios de componentes disponibles para la integración de aplicaciones. Los servicios sincronizan las aplicaciones con el repositorio.
- **Gestión de procesos de negocio:** integración de reglas de negocio con operaciones de datos maestros. Conectividad fundamental entre aplicaciones empresariales y objetos de datos básicos. Los analistas de procesos de negocio participan en los requisitos de ingeniería de datos maestros.

#### 2.3.4. Nivel 4: proactivo

En este nivel, las capacidades se caracterizan por lo siguiente:

- **Arquitectura:** se establecen modelos maestros. Existe la capacidad para pasar del marco de datos maestros a la transacción. Asimismo, existe una arquitectura orientada a servicios (SOA), establecida para la arquitectura de aplicaciones, que se combina con la gestión centralizada de metadatos empresariales.
- **Gobierno:** programa de gestión de datos empresariales en vigor con normas de datos empresariales y gestión de metadatos en lugar. La monitorización proactiva para el control de calidad de datos alimenta el programa de gobernabilidad.
- **Gestión:** gestión de identidades completamente integrada y establecida en toda la empresa, con identificación única de todas las instancias de objetos maestros. La gestión de la jerarquía de ciclo completo apoya tanto las actividades analíticas como las operacionales. La administración de la jerarquía permite el restablecimiento de errores de consolidación de falsos positivos.
- **Identificación:** los servicios para el ciclo de vida de datos integran búsqueda, coincidencia y resolución de identidades. Todas las operaciones del ciclo de vida de los datos son estructuradas sobre los servicios de fusión y consolidación. Esta consolidación se produce mayoritariamente de manera automática y, de forma puntual, cuidada por expertos de negocio.
- **Integración:** sincronización completamente integrada en los servicios de ciclo de vida. La capa de componentes admite la integración de aplicaciones en un ámbito de objeto maestro. SOA impulsa la integración de aplicaciones empresariales.
- **Gestión de procesos de negocio:** la lógica empresarial se reutiliza. Las reglas de negocio están integradas en un motor de reglas y disponibles en el nivel de procesos empresariales. Analistas de negocio integrales para el desarrollo de aplicaciones. Relaciones personalizadas con los clientes. Procesos de negocio automatizados.

#### 2.3.5. Nivel 5: estratégico

En este nivel, las capacidades se caracterizan por:

- **Arquitectura:** existe una completa integración de toda la arquitectura empresarial con el sistema MDM que propaga los cambios de forma automática entre las diferentes fuentes.

- **Gobierno:** el gobierno de datos maestros está integrado con las otras iniciativas de gobierno del dato como una función más, y asegura un intercambio de información de alta calidad.
- **Gestión:** existe una gestión transparente de la identidad de todos los objetos de datos sincronizados, con representaciones internas y externas.
- **Identificación:** hay mecanismos automáticos para resolución de conflictos entre datos maestros. Se ha vinculado el rendimiento empresarial con los datos maestros.
- **Integración:** el desarrollo de aplicaciones de negocio se vincula con los datos maestros, y se tienen en cuenta las necesidades de integración con el sistema MDM.
- **Gestión de procesos de negocio:** MDM completamente integrado en los procesos de negocio, tanto en un ámbito operacional como analítico. Destaca el uso para el perfilado de clientes, proveedores y productos.

### 3. Desarrollando un programa de gestión de datos maestros

Aunque es cierto que la aplicación de un programa de gestión de datos maestros será más efectiva si se aplica a todos los datos maestros, el alcance, el coste, el esfuerzo y los riesgos asociados a un proyecto en un ámbito corporativo son barreras infranqueables y difíciles de justificar en una fase inicial. Por ello, es mucho más sencillo empezar con solo algunas fuentes de datos maestros e ir expandiendo progresivamente el proyecto. Este enfoque permite demostrar el éxito de la iniciativa y, al mismo tiempo, consolidar el conocimiento y las lecciones aprendidas.

De este modo, desarrollar un programa de MDM supone:

- Identificar un enfoque para el programa.
- Identificar qué datos maestros se incluyen dentro de la iniciativa.
- Crear y mantener una lista maestra.

Tratamos con detalle estos puntos en los siguientes subapartados.

#### 3.1. Enfoques del programa MDM

Teniendo en cuenta la necesidad de iniciar el programa de forma acotada, es necesario identificar el enfoque que tendrá el programa de gestión de datos maestros.

Encontramos diferentes dimensiones para el enfoque, tal y como ilustra la figura 2.

Figura 2. Dimensiones enfoque MDM



Fuente: Josep Curto

#### Lectura complementaria

Haug, A.; Stentoft, J.; Zachariassen, F.; Schlichter, J. (2013). «Master data quality barriers: an empirical investigation». *Industrial Management & Data Systems* (núm. 113, vol. 2, págs. 234-249).

Definir el programa supone definir un enfoque combinado de estas tres dimensiones:

- **Dominio de datos (maestros):** esta dimensión se enfoca en el dominio de datos que hay que considerar en el proyecto, si debe ser acotado o completo.
  - **Acotado:** hace referencia a un conjunto limitado, pero cerrado de datos. Estos datos pueden ser de clientes (B2B o B2C), de producto, datos de referencia, localizaciones, de proveedores, etc.
  - **Completo:** hace referencia a todos los datos maestros de la organización.
- **Estilo del proyecto:** esta dimensión pone foco en el estilo del proyecto MDM. Tenemos, principalmente, seis estilos:
  - **Estilo de registro:** en este tipo de proyecto, los distintos sistemas de origen publican sus datos y el sistema MDM almacena solo los identificadores (ID) de sistema de origen, las claves foráneas (ID de registro en sistemas de origen) y los valores de datos clave necesarios para la coincidencia. El sistema ejecuta los algoritmos de limpieza y emparejamiento, y asigna identificadores globales únicos a los registros coincidentes, pero no envía ningún dato de nuevo a los sistemas de origen. En este caso, el sistema se apalanca en la federación de datos para crear una vista de los datos maestros.
  - **Estilo de consolidación:** los registros maestros se consolidan físicamente en un sistema central. La creación de los datos sigue estando distribuida a través de los sistemas operacionales, y los datos maestros pueden actualizarse a partir de eventos, pero no se garantiza que estén actualizados. Los datos maestros, en este caso, generalmente no se usan para transacciones, sino que apoyan la generación de informes.
  - **Estilo de coexistencia:** el sistema MDM incluye datos maestros escritos y almacenados en numerosos sistemas operacionales, así como un registro maestro físicamente almacenado en el sistema central y datos maestros armonizados en toda la cartera de aplicaciones de la organización.
  - **Estilo de transacción:** el sistema MDM almacena, mejora y mantiene todos los atributos de datos (maestros) relevantes. Se convierte en la fuente autorizada de la verdad, y publica esta información valiosa de nuevo en los respectivos sistemas de origen. El sistema MDM publica y escribe los distintos elementos de datos en los sistemas fuente, después de que los algoritmos de vinculación, limpieza, emparejamiento y enriquecimiento hayan hecho su trabajo. Los sistemas operacionales están suscritos a actualizaciones.
  - **Estilo de transacción adaptable:** es similar al estilo de transacción, pero además proporciona la capacidad de responder a diferentes solicitudes de

#### B2B y B2C

Cuando hablamos de un negocio y sus clientes, podemos encontrarnos empresas que solo comercializan con empresas (B2B, acrónimo de *business to business*) y/o que lo hacen con consumidores (B2C, acrónimo de *business to consumer*).



información y proceso en toda la empresa. El sistema consolida información interna y externa, y proporciona servicios a sistemas operacionales, analíticos o los dos. Proporciona información en tiempo real y linaje (o trazabilidad) de cambios.

- **Estilo de confederación:** encontramos diferentes sistemas MDM que se mantienen en un ámbito departamental/territorio/agencia, y que están conectados a un sistema central. Cada sistema, en un ámbito de dominio, puede diseñarse con uno de los estilos anteriores, pero el central suele ser de registro. Este enfoque acostumbra a aplicarse en el sector público, donde no es posible almacenar datos de diferentes agencias de forma conjunta.
- **Caso de uso:** existen dos casos de uso:
  - **Operacional:** que proporciona servicios de MDM para sistemas operacionales.
  - **Analítico:** que proporciona servicios de MDM para sistemas analíticos, como el almacén de datos.

El enfoque de la iniciativa MDM es la suma, por lo tanto, de los tres tipos de dimensiones anteriores, teniendo en cuenta siempre en qué sector se opera y si se proporciona el servicio a un área determinada o a toda la organización.

### 3.2. Cómo decidir qué es necesario mantener y administrar

Uno de los puntos relevantes es qué entidades de datos maestros deben formar parte del programa, puesto que no todos los datos que se ajustan a la definición de datos maestros deben gestionarse necesariamente como tales.

Existen varios criterios que hay que tener en cuenta para justificar si un dato debe ser tratado como un dato maestro:

- **Comportamiento:** los datos maestros se pueden identificar por la forma en que interactúan con otros datos. Los datos maestros normalmente están relacionados con los datos transaccionales, siguiendo una relación sustantivo/verbo. Los datos transaccionales captan verbos (por ejemplo, una venta) y los maestros, los sustantivos (por ejemplo, cliente y producto).
- **Ciclo de vida:** los datos maestros también se pueden identificar por la forma en que se crean, se leen, se actualizan, se eliminan y se buscan. Este ciclo de vida se conoce como CRUD, acrónimo de *create*, *read*, *update* y *delete*. Este proceso depende de las reglas empresariales de la empresa, el segmento de la industria y los sistemas de datos. La tabla 3 ilustra ciclos CRUD para cuatro dominios de datos maestros comunes.

Tabla 3. Datos maestros según CRUD

CRUD	Cliente	Producto	Activo	Empleado
Crear	Visita cliente (vía web o en la tienda); cuenta creada	Producto comprado o creado	Unidad adquirida; proceso de aprobación	Contratación, orientación, etc.
Leer	Vistas contextualizadas basadas en los credenciales del usuario	Inventario, catálogo de producto	Informes periódicos, verificación, depreciación	Acceso oficina, revisiones, etc.
Actualizar	Dirección, descuento, teléfono, preferencias, cuenta de crédito	Cambios en los materiales o embalaje	Transferencia, mantenimiento, accidentes	Matrimonio, promoción, transferencia
Borrar	Muerte, bancarota, liquidación	Cancelado, reemplazado, no disponible	Obsoleto, vendido, destruido, robado	Despido, muerte
Buscar	CRM, call center, gestión de contactos	ERP, sistema de pedidos	Gestión de activos	Sistema de RR. HH.

- **Cardinalidad:** a medida que la cardinalidad de un dato decrece, la probabilidad de que sea considerado un dato maestro desciende. Por ejemplo, si una empresa tiene una cantidad limitada y reducida de clientes, los beneficios de implementar una solución de gestión de datos maestros serán limitados.
- **Vida del dato:** los datos maestros tienden a ser menos volátiles que los datos transaccionales. La duración del dato depende del sector. Un tipo de dato en una industria particular puede considerarse un dato maestro (en términos de baja volatilidad), mientras que, en otras, no. Por ejemplo, un contrato para pintar una casa es una transacción, ya que es un acuerdo único y de corta duración, mientras que un contrato para representar a un artista puede ser considerado un dato maestro, puesto que se trata de un acuerdo a largo plazo, e incluir datos no volátiles.
- **Complejidad:** los datos maestros suelen ser complejos, por lo que requieren mayor atención y cuidado en su gestión. Aquellos activos que son simplemente guardados y almacenados (lo que puede entenderse como baja complejidad) no necesitan ser considerados como datos maestros.
- **Valor:** a mayor valor de un dato, mayor es la necesidad de protegerlo y mantener su calidad, y también la probabilidad de que sea considerado como un dato maestro.
- **Volatilidad:** aunque los datos maestros son menos volátiles, no significa que no lo sean. Un dato que no presenta esta característica no necesita una solución para gestionarlo.
- **Reusabilidad:** los datos maestros se caracterizan por su alta reusabilidad en la organización. Por ejemplo, consideremos los datos de cliente que se comparten en múltiples aplicaciones y departamentos. Sin una solución de MDM, existirán múltiples versiones del mismo concepto.

#### Cardinalidad

Cuando hablamos de cardinalidad, hacemos referencia al número de elementos en un conjunto.

Resumiendo, el principal reto de una organización es identificar qué datos de la misma deben considerarse como dato maestro. De este modo, crear una lista de datos maestros no solo se fundamenta en aplicar la definición del concepto, sino en categorizarlos en función de su comportamiento y atributos.

### 3.3. Creando y manteniendo una lista maestra

Crear y mantener los datos maestros es la función más importante del programa de gestión de datos maestros. Es necesario comprender de manera detallada estas dos tareas.

#### 3.3.1. Creación de una lista maestra

Para crear datos maestros, hay dos pasos básicos:

- Limpiar y estandarizar los datos.
- Combinar los datos de todas las fuentes para consolidar duplicados.

Sin embargo, antes de limpiar y normalizar los datos, es absolutamente necesario comprender el modelo de datos para los datos maestros. Es decir, es necesario identificar para cada dato maestro y de referencia el contenido del mismo y la asignación con el sistema de origen. Esta información se utiliza para definir las transformaciones necesarias para limpiar los datos de origen.

Para la limpieza de datos y su transformación en el modelo de datos maestros, necesitamos tecnologías y procesos de calidad de datos y de integración de datos. En particular, son similares a los procesos de extracción, transformación y carga de datos (ETL) utilizados para cargar los datos en la factoría de información y sus elementos (como el *data warehouse*).

De hecho, en algunas organizaciones, frecuentemente se utilizan las herramientas ETL disponibles, en lugar de implementar y/o aprender una nueva herramienta. Sea cual sea el escenario, algunas de las funciones típicas que hay que hacer serán:

- **Normalizar formatos de datos:** consiste en establecer y mantener un formato común para los datos maestros. Por ejemplo, que todos los números de teléfono sigan el mismo formato (código país y número).
- **Reemplazar los valores faltantes:** consiste en completar los datos maestros con valores predeterminados. Por ejemplo, completar los códigos postales de una dirección a partir de una base de datos especializada.
- **Estandarizar los valores:** consiste en unificar y estandarizar valores con el objetivo de poder compararlos. Por ejemplo, transformar los precios en una moneda común.
- **Mapa de atributos:** consiste en analizar un dato para identificar los diferentes atributos que contiene. Por ejemplo, en nombre de contacto, el nombre y apellido.

#### Lectura complementaria

Curto, J. (2017). *Introducción al Business Intelligence*. Barcelona: Editorial UOC.

Hay situaciones en las que no será posible automatizar completamente el proceso de limpieza de datos. En dicho momento, será necesario un procesamiento manual, pero este proceso debe fundamentarse en la curación por parte de los expertos de negocio.

Los datos de calidad formarán parte de la tabla (o las tablas) de datos maestros. Los errores formarán parte de otra tabla, y a medida que cada fuente se limpia, la salida deberá ser examinada para asegurar que el proceso esté funcionando correctamente.

### 3.3.2. Mantenimiento de una lista maestra

Existen diferentes aproximaciones para gestionar y usar datos maestros. Vamos a revisar tres de los escenarios más comunes:

- **Copia única:** en este enfoque, solo hay una copia maestra de los datos maestros. Todos los cambios en el modelo de datos maestros se llevan a cabo sobre esta copia, y a partir de esta, se propagan a las aplicaciones.
- **Múltiples copias, mantenimiento único:** en este enfoque hay una copia maestra, pero en los sistemas fuente se almacena una copia. Estas aplicaciones no pueden cambiar o agregar datos maestros, pero pueden actualizar las partes de los datos que no forman parte de los datos maestros.
- **Combinación continua:** en este enfoque, las aplicaciones pueden cambiar su copia de los datos maestros. Estos cambios se remiten a la copia maestra, donde se fusionan con el modelo maestro.

Cada uno de estos enfoques tiene beneficios e inconvenientes que se resumen en la tabla 4.

Tabla 4. Escenarios de mantenimiento

Enfoque	Beneficios	Inconvenientes
Copia única	Se garantiza la consistencia del dato.	El proceso de copia a las fuentes de origen puede derivar en la modificación de los esquemas que, en ciertos casos, es imposible.
Múltiples copias, mantenimiento único	Se garantiza la consistencia del dato, se evita la modificación de esquemas.	Será necesario modificar las aplicaciones para evitar la modificación de los datos maestros, lo que puede significar pérdida de funcionalidades.
Combinación continua	Se minimizan los cambios en aplicaciones, al mismo tiempo que se evita la modificación de esquemas y se garantiza la consistencia.	Pueden aumentar los conflictos de actualización, consistencia y reconciliación de datos, y se deben definir más reglas.

Sin importar cómo se lleve a cabo el mantenimiento, es importante entender de dónde vienen los cambios, qué proceso se ha seguido, cuál es el estado anterior y a qué jerarquía de datos pertenece. Esto significa que el manteni-

miento requiere capacidades de **gestión de versiones, auditoría y gestión de jerarquías**.

El hecho de que el sistema MDM gestione los puntos anteriores permite:

- Identificar errores y volver a versiones anteriores válidas del dato maestro.
- Tener trazabilidad del proceso de cambio de los datos.
- Usar el sistema MDM para propagar un cambio de jerarquía en un dato maestro.

### 3.4. Información por dominio

Cuando hablamos de los dominios de datos maestros, es necesario considerar que tenemos datos en diferentes niveles y que conforman el conocimiento sobre un cierto dominio. Esta estratificación va desde la información básica del dominio, que podemos entender como los datos maestros básicos, a las observaciones, transacciones e interacciones que hacen referencia al comportamiento. Tenemos:

- Información básica del dominio.
- Conocimiento extendido del dominio.
- Información financiera, legal y reguladora.
- Observaciones, transacciones e interacciones.

Estos datos normalmente se obtienen de las siguientes fuentes de origen:

- **Datos financieros:** este tipo de datos puede obtenerse del ERP (*enterprise resource planning*), que es un sistema informático destinado a la administración de recursos en una organización.
- **Datos de clientes:** este tipo de datos puede obtenerse del CRM (*customer relationship management*), que es un sistema informático para la administración y gestión de la relación con clientes. Sin embargo, con la explosión de las redes sociales e información de terceros, es posible extender esta información más allá de los sistemas internos de la organización.
- **Datos de empleados:** este tipo de datos puede obtenerse del HRM (*human resources management*), que es un sistema informático para la administración y gestión de la relación con empleados. También es posible encontrar información en sistemas para la gestión de proyectos. Sucede lo mismo que con la información de cliente que es posible extender con información de terceros.
- **Datos de productos:** este tipo de datos se puede obtener del PLM (*product lifecycle management*), que es un sistema para la gestión del ciclo de vida de un producto, desde su diseño hasta el fin de su vida útil y reciclaje.
- **Datos de localización:** que pueden incluir datos geopolíticos (como países, estados, provincias, etc.) y datos de negocio (como direcciones de oficinas, de almacenes, etc.).

### 3.4.1. Cliente B2B

En el caso de cliente B2B, la figura 3 ilustra los niveles de conocimiento.

Figura 3. Información de cliente B2B

<b>Observaciones, transacciones e interacciones</b>	<b>Oportunidades</b> Ventas actuales, previsión, ratio conversión, etc.	<b>Social</b> Eventos, noticias, blogs, alertas, canales, PR	<b>Servicio</b> Nivel de servicio, nivel de riesgo, etc.	<b>Ventas</b> Facturas, términos, línea de crédito, etc.	<b>Órdenes</b> Cantidad, productos, pedidos, etc.	<b>Call center</b> Llamadas, satisfacción, métricas	<b>Contrato</b> Tipo, fecha, riesgo, etc.
<b>Información financiera, legal, regulatoria</b>	<b>Legal</b> Propietarios, estructura del capital, etc.	<b>Financiera</b> Informes financieros, ratings, etc.	<b>Riesgo</b> Ratings, evaluación crediticia, etc.	<b>Mercados</b> Segmentos, geografía, etc.	<b>Regulación</b> Certificados, regulación, etc.		
<b>Conocimiento extendido</b>	<b>Canales</b> web, correo electrónico, teléfonos, LinkedIn, etc.	<b>Contactos</b> Decisores, administrador, etc.	<b>Jerarquías</b> Empresa madre, mercado, rating, etc.	<b>Localizaciones</b> Tipo, localización, certificaciones, etc.	<b>Direcciones</b> De envío, de facturación, etc.		
<b>Información básica</b>	<b>Cliente B2B</b> Nombre legal, año creación, IBEX 35, etc.						

Fuente: Josep Curto

### 3.4.2. Cliente B2C

En el caso del cliente B2C, la figura 4 ilustra los niveles de conocimiento.

Figura 4. Información de cliente B2C

<b>Observaciones, transacciones e interacciones</b>	<b>Oportunidades</b> Ventas actuales, previsión, ratio conversión, etc.	<b>Social</b> Ventas actuales, previsión, ratio conversión, etc.	<b>Servicio</b> Nivel de servicio, nivel de riesgo, etc.	<b>Ventas</b> Facturas, términos, línea de crédito, etc.	<b>Órdenes</b> Cantidad, productos, pedidos, etc.	<b>Call center</b> Llamadas, satisfacción, métricas	<b>Contrato</b> Tipo, fecha, riesgo, etc.
<b>Información financiera, legal, regulatoria</b>	<b>Conformidad</b> Registro criminal, fraude, seguridad, etc.	<b>Salud</b> Registros, enfermedades conocidas, etc.	<b>Riesgo</b> Ratings, evaluación crediticia, etc.	<b>Ocupación</b> Fechas, posiciones, empleadores, etc.	<b>Legal</b> Identificador, tipo, año, entidad certificadora, etc.		
<b>Conocimiento extendido</b>	<b>Canales</b> Canal de contacto preferido	<b>Social media</b> Facebook, Twitter, LinkedIn, Google, etc.	<b>Direcciones</b> De envío, de facturación, etc.	<b>Relaciones</b> Amigos, familia, hijos, etc.	<b>Puntos de contacto</b> Correo electrónico, teléfono, etc.		
<b>Información básica</b>	<b>Cliente B2C</b> Nombre, apellido, fecha de nacimiento, etc.						

Fuente: Josep Curto

### 3.4.3. Producto

En el caso de producto, la figura 5 ilustra los niveles de conocimiento.

Figura 5. Información de producto

Observaciones, transacciones e interacciones	<b>Ventas</b> Ventas actuales, previsión, ratio conversión, etc.	<b>Social</b> Eventos, noticias, blogs, alertas, canales, PR	<b>Inventario</b> Inventario, almacenes, disponibilidad, etc.	<b>Logística</b> Capacidad, trazabilidad, etc.	<b>Recomendación</b> Detección productos similares, etc.	<b>Call center</b> Llamadas, satisfacción, métricas	<b>Campañas</b> Actividad, de marketing, etc.
Información financiera, legal, regulatoria	<b>Mercado y precio</b> Precios, modelos, disponibilidad, etc.	<b>Ciclo de vida</b> Ciclo de fabricación, empaquetado	<b>Marca y PI</b> Marca, patentes, propiedad intelectual, etc.	<b>Contrato</b> Contratos, términos, resellers	<b>Cumplimiento regulatorio</b> Ambiente, salud, etc.		
Conocimiento extendido	<b>Taxonomía</b> Familias, categorías, clasificación, etc.	<b>Variantes</b> Colores, tamaños, dimensiones, variaciones, etc.	<b>Marketing</b> Imágenes, media, descripciones, documentos, etc.	<b>Atributos de negocio</b> Tipo, localización, certificaciones, etc.	<b>Materiales</b> Lista de materiales, jerarquía, etc.		
Información básica	<b>Producto</b> Identificación, nombre, descripción, componentes, etc.						

Fuente: Josep Curto

### 3.4.4. Datos de referencia

En el caso de datos de referencia, la figura 6 ilustra los niveles de conocimiento.

Figura 6. Información de datos de referencia

Observaciones, transacciones e interacciones	<b>Análisis de impacto</b> Uso en aplicaciones		
Información financiera, legal, regulatoria	<b>Estándares de la industria</b> NAICS, SWIFT, ISO, etc.	<b>Regulación</b> Certificados, regulación, etc.	<b>Mapa externo</b> Mapeo de datos internos frente a externos
Conocimiento extendido	<b>Taxonomías</b> Jerarquías, clasificaciones, diccionarios, etc.	<b>Glosario de negocio</b> Definiciones, roles, etc.	<b>Mapa interno</b> Mapeo de datos, traducciones, etc.
Información básica	<b>Información básica</b> Códigos internos geográficos, estados, etc.		

Fuente: Josep Curto

### 3.4.5. Localización

En el caso de localización, la figura 7 ilustra los niveles de conocimiento.

Figura 7. Información de localización

<b>Observaciones, transacciones e interacciones</b>	<b>Ventas</b> Facturas, términos, línea de crédito, etc.	<b>Social</b> Eventos, noticias, blogs, alertas, canales, PR	<b>Campañas</b> Actividad, de marketing, métricas, etc.	<b>Planificación</b> Capacidad, previsión, presupuesto, etc.	<b>Web</b> Tráfico, <i>page rank</i> , SEO, etc.	<b>Producción</b> Rendimiento, indicadores, etc.
<b>Información financiera, legal, regulatoria</b>	<b>Marca y PI</b> Marca, patentes, propiedad intelectual, etc.	<b>Contratos</b> Acuerdos, alquiler, socios, etc.	<b>Cumplimiento tributario</b> Local, nacional, exportaciones, etc.	<b>Regulación</b> Certificados, regulación, etc.	<b>Costes</b> Inversión, mantenimiento, amortizaciones, etc.	
<b>Conocimiento extendido</b>	<b>Geografía</b> Localización, dirección, etc.	<b>Horarios</b> Horarios, calendario, apertura, cierre, etc.	<b>Marketing</b> Imágenes, <i>media</i> , descripciones, documentos, etc.	<b>Contactos</b> Mánager, correo electrónico, teléfonos, etc.		
<b>Información básica</b>	<b>Localización</b> Identificador, nombre, descripción, etc.					

Fuente: Josep Curto

### 3.4.6. Organización

En el caso de organización, la figura 8 ilustra los niveles de conocimiento.

Figura 8. Información de la organización

<b>Observaciones, transacciones e interacciones</b>	<b>Talento y habilidades</b> Formación, certificaciones, etc.	<b>Actividades sociales</b> Eventos, noticias, artículos, etc.	<b>Rendimiento</b> Informes, evaluaciones, etc.	<b>Nómina</b> Cantidad, fecha pago, etc.	<b>Viajes y alojamiento</b> Cantidad, fecha, frecuencia, etc.	<b>Contrato</b> Tipo, fecha, riesgo, etc.
<b>Información financiera, legal, regulatoria</b>	<b>Seguro y beneficios</b> Seguros, beneficios, etc.	<b>Sueldo y bonos</b> Rango salarial, impuestos, bono, etc.	<b>Salud</b> Revisión anual, enfermedades conocidas, etc.	<b>Conformidad</b> Registro criminal, fraude, seguridad, etc.	<b>Legal</b> Identificador, tipo, año, entidad certificadora, etc.	
<b>Conocimiento extendido</b>	<b>Coste de centro</b> Coste de centro, jerarquía, etc.	<b>Unidad de negocio</b> Unidad, jerarquía, etc.	<b>Historial</b> Fechas, posiciones, empleadores, etc.	<b>Trabajos</b> Descripción, responsabilidades, etc.	<b>Jerarquía RR. HH.</b> Mánager, responsable, equipo, etc.	
<b>Información básica</b>	<b>Empleado</b> Nombre, apellidos, correo electrónico, DNI, pasaporte, etc.					

Fuente: Josep Curto

Un posible enfoque consiste en centrarse en las personas como eje central de la organización, por lo que estos niveles coinciden con los de empleado.



### 3.4.7. Proveedor

En el caso de proveedores, la figura 9 ilustra los niveles de conocimiento.

Figura 9. Información de proveedores

Observaciones, transacciones e interacciones	<b>Financiero</b>	<b>Social</b>	<b>Call center</b>	<b>Ventas</b>	<b>Órdenes</b>
	Precio acciones, eventos financieros	Eventos, noticias, blogs, alertas, canales, PR	Llamadas, satisfacción, métricas	Facturas, términos, línea de crédito, etc.	Cantidad, productos, pedidos, etc.
Información financiera, legal, regulatoria	<b>Legal</b>	<b>Financiera</b>	<b>Riesgo</b>	<b>Acuerdos</b>	<b>Regulación</b>
	Propietarios, estructura del capital, etc.	Informes financieros, ratings, etc.	Ratings, evaluación crediticia, etc.	Propiedad intelectual, SLA, NDA, etc.	Certificados, regulación, etc.
Conocimiento extendido	<b>Jerarquías</b>	<b>Emplazamiento</b>	<b>Contactos</b>	<b>Banco</b>	<b>Direcciones</b>
	Mercado, ratings, grupo empresarial	Tipo, localización, certificación, etc.	Decisores, administrador, etc.	Cuenta, autorizaciones, etc.	Web, correos electrónicos, dirección, etc.
Información básica	<b>Proveedor</b>				
	Nombre legal, año SIC, NAIC, etc.				

Fuente: Josep Curto

### 3.4.8. Empleado

En el caso de empleados, la figura 10 ilustra los niveles de conocimiento.

Figura 10. Información de empleados

Observaciones, transacciones e interacciones	<b>Talento y habilidades</b>	<b>Actividades sociales</b>	<b>Rendimiento</b>	<b>Nómina</b>	<b>Viajes y alojamiento</b>	<b>Contrato</b>
	Formación, certificaciones, etc.	Eventos, noticias, artículos, etc.	Informes, evaluaciones, etc.	Cantidad, fecha pago, etc.	Cantidad, fecha, frecuencia, etc.	Tipo, fecha, riesgo, etc.
Información financiera, legal, regulatoria	<b>Seguro y beneficios</b>	<b>Sueldo y bonos</b>	<b>Salud</b>	<b>Conformidad</b>	<b>Legal</b>	
	Seguros, beneficios, etc.	Rango salarial, impuestos, bono, etc.	Revisión anual, enfermedades conocidas, etc.	Registro criminal, fraude, seguridad, etc.	Identificador, tipo, año, entidad certificadora, etc.	
Conocimiento extendido	<b>Coste de centro</b>	<b>Unidad de negocio</b>	<b>Historial</b>	<b>Trabajos</b>	<b>Jerarquía RR. HH.</b>	
	Coste de centro, jerarquía, etc.	Unidad, jerarquía, etc.	Fechas, posiciones, empleadores, etc.	Descripción, responsabilidades, etc.	Mánager, responsable, equipo, etc.	
Información básica	<b>Empleado</b>					
	Nombre, apellidos, correo electrónico, DNI, pasaporte, etc.					

Fuente: Josep Curto

## 4. Técnicas y tecnologías para la gestión de datos maestros

Hemos revisado hasta ahora los aspectos vinculados a los procesos, gobierno, políticas y estándares en el contexto de la gestión de datos maestros.

En este apartado, vamos a centrarnos en las herramientas, técnicas y tecnologías que soportan la gestión de datos maestros.

Las herramientas para la gestión de datos maestros han evolucionado de forma considerable en los últimos años. El concepto emerge vinculado al CRM, el ERP y el *data warehouse*, que necesitan información fidedigna y compartida para la creación de informes y el análisis de las entidades más relevantes de negocio. No se considera una disciplina independiente de la inteligencia de negocio, y no existen herramientas dedicadas. Poco a poco, a medida que más aplicaciones de negocio e incluso sistemas operacionales necesitan el uso de datos maestros, empiezan a emerger herramientas integradas para MDM, vinculadas a la información de clientes o a la de producto.

Actualmente, existen ya plataformas dedicadas a la gestión de datos maestros. Por este motivo, al hablar de técnicas y tecnologías dentro de estas plataformas, muchas de ellas ya serán conocidas. En cierta medida, podemos considerar la gestión de datos maestros como un caso específico y combinado de todas estas, que ha alcanzado la suficiente entidad propia como para escindirse de la inteligencia de negocio.

### 4.1. Técnicas para la gestión de datos maestros

Teniendo en cuenta la evolución de la gestión de datos maestros, no es extraño descubrir que las técnicas que soportan MDM sean una combinación de técnicas que frecuentemente pueden encontrarse de forma independiente en una organización pero que, en este contexto, deben trabajar de forma orquestada e integrada.

Estas técnicas incluyen:

- **Técnicas de integración de datos:** permiten la captura, el procesamiento y la distribución de datos desde fuentes de origen hacia el repositorio central de datos maestros, y viceversa. Estas técnicas incluyen la propagación, consolidación, federación de datos y *change data capture*.

#### Change data capture

Cuando hablamos de *change data capture*, hacemos referencia a una técnica que propaga los cambios en una base de datos en el momento de la ocurrencia hacia las fuentes destino.

- **Técnicas de almacenamiento de datos:** permiten el almacenamiento de datos maestros y de referencia en un repositorio central (y aquellos que sean necesarios en función del enfoque). Estas técnicas incluyen la creación de modelos de datos y el establecimiento de reglas de negocio.
- **Técnicas de calidad de datos:** permiten mantener y mejorar la calidad de los datos maestros y de referencia. Estas técnicas incluyen el perfilado de datos, análisis semántico y estadístico de datos, emparejamiento por reglas de negocio, semántico y estadístico de datos, y normalización y estandarización de datos. Parte de estas técnicas se fundamentan en algoritmos de aprendizaje automático.
- **Técnicas de jerarquías:** posibilitan ordenar y categorizar los datos maestros de forma flexible. Estas técnicas tienen puntos en común con las técnicas OLAP (en el sentido de definición).
- **Técnicas de exploración de datos, aplicaciones y sistemas:** permiten identificar automáticamente los datos de una organización (presentes en múltiples sistemas y aplicaciones) y, además, proponer qué datos deben formar parte del repositorio MDM. Parte de estas técnicas también se fundamentan en algoritmos de aprendizaje automático.

**OLAP**

Cuando hablamos de OLAP, *online analytical processing*, hacemos referencia al método para organizar y consultar datos sobre una estructura multidimensional.

#### 4.2. Tecnologías para la gestión de datos maestros

En este subapartado, vamos a hacer referencia a la plataforma de gestión de datos maestros como herramienta integrada. Este tipo de plataformas debe tener los componentes siguientes:

- **Modelización de datos:** componente que se encarga de la creación y el mantenimiento del modelo de datos maestros. Incluye el diseño fundamentado en reglas de negocio, glosario de negocio, soporte para reglas complejas de negocio, etc.
- **Semántica y calidad de datos:** componente que se encarga de la calidad de datos, incluyendo normalización, estandarización, uso de capacidades semánticas y estadísticas para el emparejamiento y la calidad de datos, enriquecimiento de datos, etc.
- **Flujos y procesos de negocio:** componente para la gestión del flujo de trabajos vinculados con datos maestros, incluyendo asignaciones basadas en roles y usuarios, API para la integración con sistemas externos, métricas para la comprensión de los flujos y actividades fundamentadas en reglas de negocio.

- **Gobierno y definición de políticas de datos:** componente que habilita la creación y el mantenimiento de políticas de datos, incluyendo perfilado y descubrimiento de datos, creación colaborativa de reglas de datos, glosario de definiciones, versión de políticas y cuadro de mando de control.
- **Gestión de jerarquías de datos:** componente para la gestión de jerarquías de datos maestros que incluye autodescubrimiento de jerarquías, jerarquías basadas en relaciones predefinidas, búsqueda por jerarquías, soporte para jerarquías múltiples por rol y dato maestro, así como jerarquías balanceadas, desequilibradas y recursivas.
- **Administración y aplicación de políticas de datos:** componente que es el único punto para la aplicación de políticas de datos maestros, incluyendo autoría y combinación de datos, aplicación de políticas en un ámbito de registro o atributo, gestión de colas de aplicación de políticas, creación y gestión de eventos y notificaciones y métricas y cuadro de mando de control.
- **Integración de datos:** componente que se encarga de integración de datos maestros, incluyendo conectores a todo tipo de fuentes de datos (bases de datos relacionales, NoSQL, ficheros, aplicaciones, API, etc.), ETL, CDC, etc.
- **Arquitectura de datos:** componente que se encarga del almacenamiento de datos maestros que debe soportar despliegue en la nube o en el centro de datos propio de la organización, así como escalabilidad.
- **Seguridad y confianza de datos:** componente que se encarga de la seguridad del dato, incluyendo: seguridad basada en roles, en un ámbito de registro y campo; integración con sistemas corporativos de la gestión de protocolos de acceso como LDAP (*lightweight directory access protocol*); auditoría de datos, trazabilidad del dato y análisis de impacto.

En mercado, podemos encontrar plataformas genéricas que incluyen la gran mayoría de los componentes (como Semarchy\*), plataformas que a partir de la integración de datos han evolucionado hacia MDM (como Informatica\*\*), soluciones comerciales *open source* (como la de Talend\*\*\*), así como soluciones específicas para datos de producto (como la de inRiver\*\*\*\*) o de cliente (como la de Pitney Bowes\*\*\*\*\*).

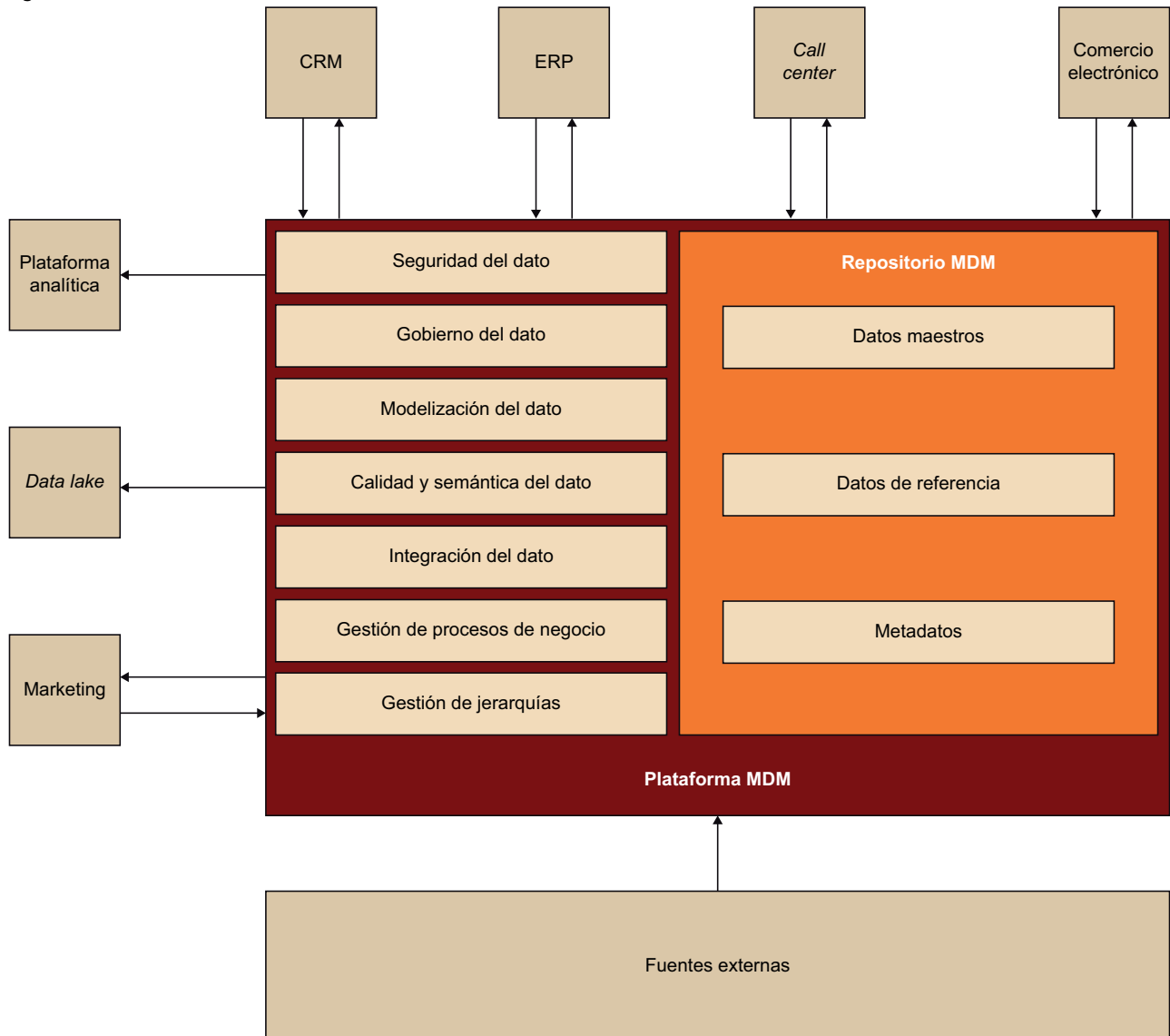
La figura 11 ilustra la plataforma MDM que se complementa con el repositorio para los datos maestros, de referencia y metadatos. La plataforma se conecta con aquellos sistemas de información (internos y externos) que tienen datos relevantes para el sistema MDM. Esta comunicación, como ya sabemos, es bidireccional para poder propagar los datos desde el sistema de almacenamiento de datos maestros (frecuentemente denominado *hub*).

#### LDAP

Cuando hablamos de LDAP, hacemos referencia a un protocolo en un ámbito de aplicación que permite el acceso a un servicio de directorio ordenado y distribuido para buscar distinta información en un entorno de red.

\* [www.semarchy.com](http://www.semarchy.com)  
 \*\* [www.informatica.com](http://www.informatica.com)  
 \*\*\* [www.talend.com](http://www.talend.com)  
 \*\*\*\* [www.inriver.com](http://www.inriver.com)  
 \*\*\*\*\* [www.pitneybowes.com](http://www.pitneybowes.com)

Figura 11. Plataforma MDM



Fuente: Josep Curto

### 4.3. Ejemplo de caso de uso

Imaginemos una empresa de comercio electrónico que se ha dado cuenta de la necesidad de mejorar la información de sus productos para:

- Mejorar el seguimiento de productos, con el objetivo de reducir incidencias vinculadas con la regulación asociada a su sector.
- Evitar la frustración de clientes y empleados, fundamentada en la existencia de datos de productos erróneos.

#### ¿Cómo puede empezar su iniciativa de MDM?

Imaginemos que la organización dispone de una herramienta de MDM. La organización debe seguir estos pasos:

- **Descubrir** qué aplicaciones contienen datos relevantes de producto y susceptibles de formar parte de los datos maestros. Asimismo, es necesario evaluar su nivel de calidad. En este punto, se usan los componentes de integración y calidad de datos de la plataforma.
- **Crear** el modelo de datos maestros a través del consenso con las diferentes unidades de negocio implicadas. En este punto, se usa el componente de modelización de datos maestros.
- **Definir** la reglas de validación (de negocio), de calidad de datos y de convergencia entre fuentes. En este punto se usan diferentes componentes, como el de administración de las políticas.
- **Identificar** y definir las interfaces de acceso de usuarios a la plataforma MDM. En este punto, interviene el componente de integración de datos y la seguridad del dato y el gobierno.
- **Definir** las políticas de gobierno de datos maestros y sus flujos de trabajo. Aquí es relevante el componente de administración y definición de políticas.
- **Integrar** MDM con las aplicaciones de negocio necesarias. Aquí juega un papel fundamental el componente de integración de datos.
- **Poner** en producción y gestionar el gobierno del dato maestro, fundamentado en el componente de gobierno.

En esencia, la plataforma MDM se convierte en un mecanismo para articular cualquier iniciativa. Está claro que cuanta más integración haya entre los componentes de la plataforma, estos pasos serán más fluidos.

#### 4.4. MDM y *big data*

Como ya sabemos, *big data* es una de las iniciativas actuales más relevantes para generar valor a partir de los datos. En general, múltiples estudios, como los de Capgemini\*, sugieren que tan solo el 35% de los proyectos de *big data* pueden describirse como de éxito o de mucho éxito y, por otro lado, Gartner\*\* indica que tan solo el 15% de empresas son capaces de poner sus iniciativas de *big data* en producción. En definitiva, estos estudios confirman que estos proyectos son complejos y que pueden acabar en fracaso.

Dentro de las iniciativas que se enmarcan dentro de *big data*, nos encontramos con aquellas vinculadas con el *data lake*. El propósito de esta iniciativa es presentar una visión no refinada de los datos, que permita explorar libremente técnicas de análisis de datos e investigar casos de uso. Inicialmente, estas iniciativas se han fundamentado en bases de datos NoSQL del tipo clave-valor.

#### **Big data**

*Big data* hace referencia al conjunto de estrategias, tecnologías y sistemas para el almacenamiento, el procesamiento, el análisis y la visualización de conjuntos de datos complejos.

\* Colas, M.; y otros (2014). *Cracking the Data Conundrum: How Successful Companies Make Big Data Operational*. Capgemini.

\*\* Heudecker, N.; Hare, J. (2016). *Survey Analysis: Big Data Investments Begin Tapering in 2016*. Gartner.

### **Data lake**

*Data lake* hace referencia a la colección de instancias de almacenamiento de todos los activos de datos referentes a una organización almacenados en una copia casi exacta, o incluso exacta, del formato fuente.

Desafortunadamente, las iniciativas de *data lake* no están dando los frutos esperados\*. El foco inicial ha estado en la captura y el almacenamiento de datos, sin tener en cuenta capacidades de gestión de datos y su posterior explotación.

\* Heudecker, N.; White, A. (2014). *The Data Lake Fallacy: All Water and Little Substance*. Gartner.

A medida que ha quedado patente la dificultad de implantación del *data lake*, los fabricantes han empezado a crear plataformas integradas que van más allá del almacenamiento masivo de datos. Estamos hablando de la integración de las capacidades siguientes:

- **Ingerir datos:** capacidad para la ingesta de datos de autoservicio con limpieza de datos, validación y perfiles automáticos.
- **Preparar datos:** capacidad para manipular y preparar datos a través de SQL, y transformación interactiva a través de una interfaz de usuario simple.
- **Descubrir datos:** capacidad para explorar datos y metadatos, y comprender su linaje y su perfilado.
- **Monitorizar datos:** capacidad para comprender el rendimiento y funcionamiento del *data lake* (es decir, quién lo usa, cómo, etc.), así como el seguimiento de SLA y resolución de problemas de rendimiento.
- **Diseño de flujos de datos:** capacidad para diseño de flujo de datos para autoservicio.

#### **SQL**

Cuando hablamos de SQL, *structured query language*, hacemos referencia al lenguaje de consulta de bases de datos relacionales.

#### **SLA**

Cuando hablamos de SLA, *service level agreement*, hacemos referencia al contrato entre un proveedor de servicios (interno o externo) y el usuario final que define el nivel de servicio esperado por parte del proveedor.

Como se apreciar, estos componentes tienen puntos en común con los componentes de una gestión de datos maestros que hemos revisado anteriormente como, por ejemplo, la integración, calidad, validación y gobierno del dato.

Otra aproximación que está emergiendo es el uso de una plataforma MDM que extiende el *data lake* y lo complementa, es decir, la plataforma de *big data* se convierte en otra fuente de datos que puede recibir la propagación de datos maestros.

En esencia, MDM crea un contexto para *big data* proporcionando información confiable sobre cómo los datos no estructurados y entrantes encajan en el entorno empresarial. Por otro lado, *big data* crea un nuevo contexto de datos para expandir los datos maestros, lo que ayuda a la organización a incrementar el valor del dato.

## Resumen

En este módulo didáctico hemos presentado el concepto de gestión de datos maestros, que tiene como objetivo crear definiciones estándar de los datos críticos y estables de una organización. Los datos maestros se transforman en un único punto de referencia para múltiples sistemas, incluyendo la inteligencia de negocio.

Primero hemos introducido el concepto, los diferentes tipos de datos maestros existentes y los beneficios que aporta esta estrategia para una organización.

A continuación, hemos revisado en qué consiste un programa de gestión de datos maestros, cómo encaja dentro del gobierno del dato, las diferentes metodologías existentes para su desarrollo y las mejores prácticas. Asimismo, hemos revisado qué significa MDM para diferentes tipos de datos, como de cliente o producto.

Por último, hemos revisado las técnicas y la tecnología que componen la gestión de datos maestros y los componentes que forman parte de una plataforma.



## 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 y que permite tomar mejores decisiones a los usuarios de una organización.

**CRM** *m* Sistema informático para la administración y gestión de la relación con clientes.

**data quality** *f* Técnicas para la identificación, el control, el incremento y el mantenimiento de la calidad de datos en una organización.

**data warehouse** *m* Repositorio de datos que proporciona una visión global, común e integrada de los datos de la organización, de manera independiente de cómo vayan a ser utilizados posteriormente por los consumidores o usuarios, con las propiedades siguientes: estable, coherente, fiable y con información histórica.

**ERP** *m* Sistemas informáticos destinados a la administración de recursos en una organización.

**ETL** *m* Procesos que permiten la extracción, transformación y carga de datos desde fuentes de origen hasta el destino para su correcto consumo.

**gestión de datos maestros** *f* Conjunto de procesos, gobierno, políticas, estándares y herramientas que definen y gestionan constantemente los datos críticos de una organización para proporcionar un único punto de referencia.

**golden record** *f* Versión única y bien definida de todas las entidades de datos en un ecosistema organizacional.

**HRM** *m* Sistema informático para la administración y gestión de la relación con empleados.

**PLM** *m* Sistema para la gestión del ciclo de vida de un producto.

## Bibliografía

**Berson, A.; Dubov, L.** (2010). *Master Data Management and Data Governance, 2/E*. Nueva York: McGraw Hill Professional.

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

**Curto, J.** (2017). *Introducción al Business Intelligence (nueva edición ampliada y revisada)*. Barcelona: Editorial UOC.

**Loshin, D** (2010). *Master data management*. Nueva York: Morgan Kaufmann.

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