

Maestría en Análisis y Visualización de Datos Masivos

---

# Gobierno del Dato y Toma de Decisiones

Gobierno del Dato y Toma de Decisiones

---

# Tema 1. Dirección estratégica y gobierno de datos

# Índice

## Esquema

### Ideas clave

- 1.1. Introducción y objetivos
- 1.2. ¿Qué es la dirección estratégica (DE)?
- 1.3. El proceso de dirección estratégica
- 1.4. Análisis estratégico
- 1.5. Cuadro de mando integral (CMI)
- 1.6. Toma de decisiones
- 1.7. Gobierno de datos
- 1.8. Referencias bibliográficas

### A fondo

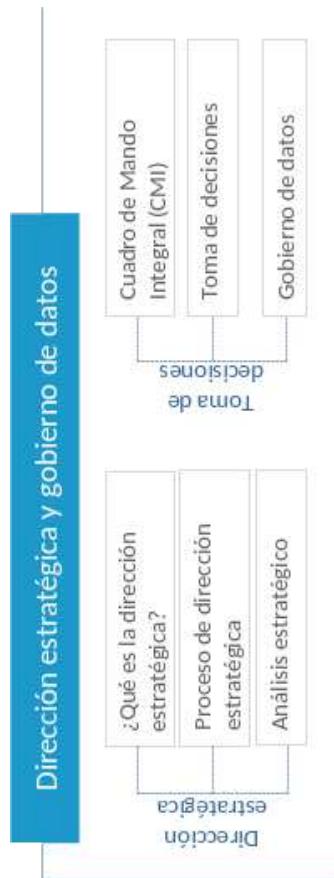
The Global Data Management Community

Plan estratégico e implantación del cuadro de mando integral. Guía de resultados y mejores prácticas

El arte de la guerra

Misión, visión y valores en una empresa

## Test



## 1.1. Introducción y objetivos

El direccionamiento estratégico busca alcanzar objetivos empresariales a mediano y largo plazo, es una disciplina que se ha ido adaptando con el tiempo y ha hecho uso de las tecnologías existentes, con el objetivo de tomar las mejores decisiones empresariales. En este tema veremos una introducción al direccionamiento estratégico y profundizaremos en detalle en el análisis estratégico para después estudiar el cuadro de mando integral como herramienta que permite visualizar y gestionar cada uno de los objetivos estratégicos, con sus respectivos indicadores o KPI.

En el cuadro de mando integral se pueden ver los KPI de diferentes áreas: financiera, cliente, procesos internos y aprendizaje. Una toma de decisiones efectiva depende de que existan herramientas y procesos que garanticen una información veraz. Para poder tomar decisiones con base en los datos, la empresa debe gestionarlo facilitando su disponibilidad, usabilidad, integridad y seguridad; y a esto le llamamos gobierno de datos. Veremos en qué consiste, cuáles son las tareas para realizar y por dónde podríamos empezar.

Los objetivos de este tema son:

- ▶ Comprender en qué consiste el análisis estratégico, su historia, qué es un objetivo estratégico y cómo se puede medir su cumplimiento.
- ▶ Esquematizar un cuadro de mando integral e identificar cada uno de sus componentes.
- ▶ Entender el significado del gobierno del dato y argumentar su importancia en un proceso de toma de decisiones.

## 1.2. ¿Qué es la dirección estratégica (DE)?

Es una **disciplina** que integra las distintas estrategias y tácticas empresariales, analiza decisiones tomadas y observa sus consecuencias o efectos durante un período de tiempo, aparte de su pretensión de alcanzar objetivos empresariales a largo plazo.

Es decir, precisa el camino que debería seguir la compañía para el logro de objetivos. Esto implica una revisión periódica de las metas trazadas versus los cambios de la empresa a nivel local, nacional o internacional, además de realizar los ajustes necesarios para cumplir con los objetivos trazados. Como ejemplo, aunque el ajedrez contiene un elevado contenido táctico y de cálculo, la verdadera maestría se alcanza con el dominio de los conceptos estratégicos. En el plano estratégico, el ajedrez provee dos tipos de enseñanzas: genéricas y específicas. Entre las genéricas destacamos la necesidad del estudio profundo del presente: el jugador debe recopilar toda la información acerca de la posición actual y evaluarla en profundidad antes de abordar el estudio de jugadas concretas. El trabajo con análisis y escenarios múltiples, así como el desarrollo del hábito estratégico son también algunas de las muchas aportaciones que nos ofrece este juego.

La definición de la dirección estratégica cambia en función de algunos autores. Por ejemplo, Drucker la define como: «la planificación estratégica no es una caja de trucos. Es pensar analíticamente y dedicar recursos a la acción: es el proceso continuo de tomar sistemáticamente, en el presente, decisiones empresariales con el mayor conocimiento posible de sus alcances futuros, organizar sistemáticamente los esfuerzos necesarios para llevar a cabo esas decisiones y medir los resultados de las mismas, comparándolas con las expectativas, mediante una sistemática retroalimentación» (Drucker, 2004).

Alfred Chandler, por su parte, define la planificación como la «determinación de las

metas y objetivos a largo plazo, la adopción de los cursos de acción y la consecución de recursos necesarios para lograr esas metas» (Chandler, 1969).

## ¿Por qué es necesario la dirección estratégica?

La dirección estratégica es necesaria para colmar las necesidades actuales de la alta dirección de orientar y mejorar las competencias gerenciales y llevar a las organizaciones al éxito. Con la gestión integral y la competitividad del negocio frente a las exigencias del mercado actual, es necesaria la aplicación de métodos eficaces, además de la obtención de beneficios tales como:

- ▶ La orientación a la alta gerencia.
- ▶ La gestión integral.
- ▶ El aumento de la competitividad.
- ▶ El aseguramiento una planificación estratégica.
- ▶ La aplicación y manejo del ciclo PHVA (planear, hacer, verificar y actuar).
- ▶ El establecimiento del cuadro de mando integral o CMI (*Balanced ScoreCard* o BSC).
- ▶ La implantación de indicadores de gestión y proceso.

## 1.3. El proceso de dirección estratégica

Se considera que la planificación estratégica consiste en **definir** y **organizar** sistemáticamente las tareas que se deben ejecutar para que la empresa cumpla su visión. En esencia, consiste en **establecer un enfoque competitivo y único**, basado en la creación de ventajas diferenciadoras. De esta forma, el plan estratégico es un documento que resume los principales elementos que determinan la competitividad de la empresa, las estrategias que se van a adoptar y los propósitos claves que se deben ejecutar para lograr la visión en un tiempo determinado. Es necesario aclarar que la planificación estratégica **no intenta predecir el futuro**, sino tomar decisiones que impacten en él.

Para la elaboración del plan estratégico es necesario tener en cuenta **tres elementos:**

- ▶ **Información:** no se pueden plantear estrategias competitivas sin tener una comprensión de nuestro mercado, los clientes (internos y externos), los competidores y las relaciones que se presentan en el entorno competitivo.
- ▶ **Metodología:** dada la gran cantidad de información que manejan las organizaciones, es necesario disponer de una metodología que ayude a seleccionar la información más importante y organizarla de tal manera que las decisiones tomadas sean las más recomendables para la situación actual de la empresa. La metodología debe facilitar al grupo directivo el diseño y la puesta en marcha del plan, con las actividades por realizar, así como el alcance de los objetivos propuestos.
- ▶ **Pensamiento estratégico:** no sirve de nada la información y el método si la organización en su conjunto, y en especial el grupo de planificación, carece de compromiso y formación, es decir, de pensamiento estratégico. Esto marcará la diferencia entre una estrategia y otra (Martínez y Milla, 2005).

En la Figura 1 podemos ver estos pasos de una forma esquemática, desde el presente al futuro de la organización. En los siguientes párrafos veremos los elementos más importantes de este tránsito.

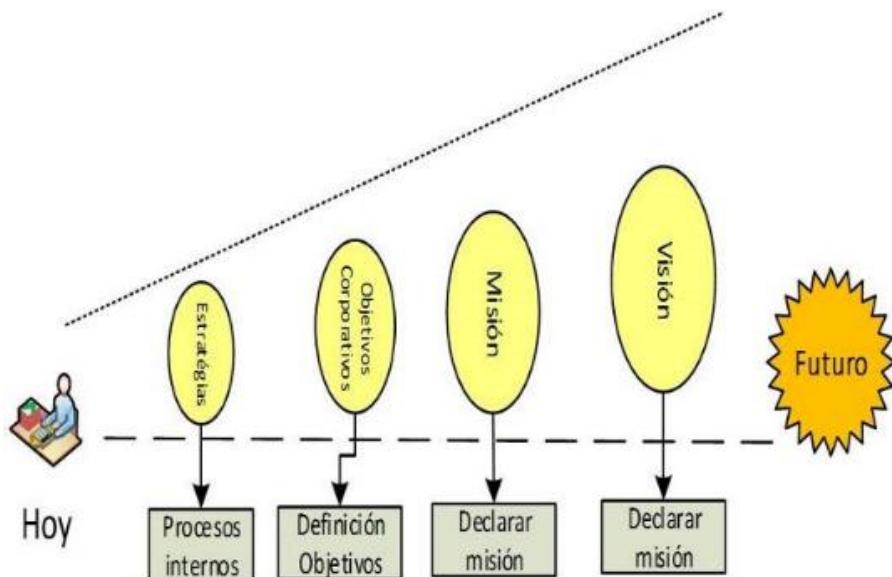


Figura 1. Modelo de gestión.

## Misión

Goodstein define la misión como «un enunciado breve y claro de las razones que justifican la existencia, el propósito o funciones que la organización desea satisfacer, su base de usuarios o consumidores y los métodos fundamentales para cumplir con un propósito» (Goodstein et al., 1999).

La misión es el **propósito** de la organización y marca la diferencia entre una empresa y otra con la misma función. La declaración de la misión debe responder a las siguientes preguntas: ¿quiénes somos?, ¿qué hacemos?, ¿cómo lo hacemos?, ¿dónde lo hacemos?, ¿para quién lo hacemos?, y nuestra ventaja competitiva.

- ▶ ▶ ¿Quiénes somos?: identifica la esencia de la empresa.
- ▶ ▶ ¿Qué hacemos?: comprende la definición del tipo de negocio y las capacidades

únicas que desarrolla.

- ▶ ¿Cómo lo hacemos?: define los recursos y las habilidades desarrolladas por la compañía.
- ▶ ¿Dónde lo hacemos?: supone la identificación de los productos y servicios, la localización geográfica.
- ▶ ¿Para quién lo hacemos?: define los clientes internos y externos a los cuales quiere satisfacer. Este es un punto muy importante para que la empresa desarrolle una ventaja competitiva.

Para desarrollar la misión, las empresas deben:

- ▶ Encontrar un público objetivo al que dirigirse, es decir, tanto los clientes internos como los externos, para satisfacer sus necesidades mediante productos y servicios.
- ▶ Llevar un seguimiento continuo del mercado, para conocer la situación de la competencia.
- ▶ Finalmente, diferenciarse de la competencia en el sector en el que se trabaja mediante características y habilidades especiales.

**Componentes** de la misión:

- ▶ Mercado del consumidor.
- ▶ Producto y/o servicio.
- ▶ Dominio geográfico.
- ▶ Tecnología.
- ▶ Valores.
- ▶ Autoconcepto.

- ▶ Ventaja competitiva.

Ejemplos de misión de algunas compañías:

## Ejemplo 1: estándar

En la empresa XXX, trabajamos para proveer \_\_\_\_\_, mediante \_\_\_\_\_.

## Ejemplo 2: empresa pública

La UA es una institución pública y socialmente responsable. Su MISIÓN, tal como la legislación (LEY ORGÁNICA 6/2001 y modificaciones posteriores) le confiere, supone:

- ▶ La formación integral de sus estudiantes. No sólo en conocimientos y disciplinas, sino también como fomento del sentido crítico, social, responsable, saludable y sensible a los principios de sostenibilidad, para contribuir de manera efectiva al bienestar de la sociedad donde se inserta. A esto, cabe añadir la garantía de la dignidad personal, el libre desarrollo de las personas sin ningún tipo de discriminación y por último, el derecho a la igualdad efectiva entre mujeres y hombres.
- ▶ La investigación como principio básico debe incrementar la mejora del conocimiento. Por un lado, por su transferencia a través de la docencia. Por otro, la contribución directa de la universidad a la sociedad a través de su ineludible compromiso con el desarrollo cultural, científico y tecnológico. De este modo, gracias a la colaboración con otros agentes sociales, dicha investigación pueda concretarse en innovación para el desarrollo sostenible y la mejora de la calidad de vida.

Figura 2. Modelos de misión empresarial.

## Visión

Definida por Jack Fleitman como «el camino al cual se dirige la empresa a largo plazo y sirve de rumbo y aliciente para orientar las decisiones estratégicas del crecimiento» (Fleitman, 2000).

La visión es el **estado ideal** de la empresa, lo que esta aspira a ser o estar en un futuro cercano o lejano, y está fundamentada en la realidad a través del desarrollo de metas y de indicadores de desempeño. Es más útil cuando se centra en las

capacidades que la compañía necesita para satisfacer tanto a sus clientes internos como externos.

Para desarrollar la visión, los dirigentes de las empresas deben:

- ▶ Saber muy bien la situación actual de la compañía.
- ▶ Conocer la competencia.
- ▶ Ser visionarios para desarrollar las habilidades y requerimientos que la empresa necesita para su desempeño futuro.
- ▶ Ser creativos.

Habitualmente la visión tiene los siguientes **elementos**:

- ▶ Horizonte de tiempo a cinco años.
- ▶ Posicionamiento en el mercado: líder.
- ▶ Ámbito de acción: nacional o internacional.
- ▶ Valores: ética y honestidad, entre otros.
- ▶ Negocio: prestación de servicios, fabricación de productos, etc.
- ▶ Principios organizacionales: calidad y competencia, entre otros.

Ejemplos de visión:

## Ejemplo 1: estándar

Ser la institución / organización líder y de excelencia en \_\_\_\_\_ acción (la generación) \_\_\_\_\_, de productos/servicios \_\_\_\_\_, fundamentado en \_\_\_\_\_ valores/filosofía \_\_\_\_\_, que sirve para \_\_\_\_\_ misión \_\_\_\_\_.

## Ejemplo 2: Telmex

Consolidar el liderazgo de TELMEX INTERNACIONAL, expandiendo su penetración en los mercados donde opera para ser una de las empresas de más rápido y mejor crecimiento a nivel mundial.

## Ejemplo 3: Nestlé España

Ser la empresa reconocida como líder en nutrición, salud y bienestar a nivel mundial por parte de sus consumidores, empleados, clientes, proveedores y todos los grupos de interés relacionados con la actividad de la compañía.

Figura 3. Modelos empresariales de visión.

## 1.4. Análisis estratégico

Las empresas se enfrentan día a día a la identificación del mercado en el que quieren o desean competir y deben definir una estrategia (a dónde quieren llegar y qué quieren ser) para estar presentes en el mercado actual.

Es necesario plantear preguntas como:

- ▶ ¿Cómo está constituido el sector en que se mueve la empresa?
- ▶ ¿Cómo son los clientes internos y externos? Este análisis es clave para los sucesos futuros y, por consiguiente, la atención a escenarios alternativos. El análisis interno hace una evaluación del desempeño de la organización.
- ▶ ¿Qué capacidad de crecimiento tiene la empresa?

La satisfacción del cliente es el tema más importante al que se enfrenta la dirección de la empresa, lo que envuelve un amplio análisis del cliente en cuanto a:

- ▶ Sus necesidades.
- ▶ Cumplimiento del producto.
- ▶ Evaluación posventa.

### Análisis externo o análisis PEST/PESTE

El análisis PEST es una herramienta de gran utilidad para entender el **crecimiento o bajada de un mercado** y, como resultado, la posición, el potencial y la dirección de nuestro negocio. Es útil para medirlo y evaluarlo. PEST es una sigla compuesta por los siguientes aspectos: políticos, económicos, sociales y tecnológicos.

En la Figura 4 se puede observar en detalle cada uno de los aspectos del análisis PEST.

Análisis Pest			
Político-Legal	Económico	Socio-Cultural	Tecnológico
Legislación y protección ambiental	Crecimiento económico	Distribución de ingresos	Gasto Gubernamental en investigación
Impuestos	Política monetaria y de tasas de interés	Demografía, tasas de crecimiento de la población, distribución de edad	Industria enfocada al esfuerzo tecnológico
Legislación y restricciones de tratados internacionales	Gasto Gubernamental	Movilidad social y laboral	Nuevos productos y desarrollos
Leyes de protección al consumidor	Política de desempleo	Cambios en el estilo de vida	Tasa de transferencia tecnológica
Leyes de protección al empleo	Reforma Fiscal	Espíritu empresarial, actitud hacia el trabajo, carrera y descanso.	Ciclo de vida y velocidad de obsolescencia tecnológica
Actitud y organización del Gobierno	Tasa de cambio	Educación	Energía, uso y costos
Leyes de competencia	Tasa de inflación	Modas	Cambios en información tecnológica
Estabilidad política	Estado del ciclo del negocio	Conciencia de seguridad social y de salud	Cambios en internet
Legislación sobre seguridad	Confianza del consumidor	Calidad de vida	Cambios en tecnología móvil

Figura 4. Análisis PEST. Fuente: <http://planeacion-estrategica.blogspot.com/2008/07/anlisis-pest.html>

Los objetivos del análisis PEST se centran en:

- ▶ Conocer los factores externos que afectan a la empresa.
- ▶ Identificar los factores externos que pueden cambiar en el futuro.
- ▶ Explotar los cambios actuales para convertirlos en oportunidades.
- ▶ Identificar y analizar las posibles amenazas para buscar una pronta solución.

El PEST funciona como un **marco para estudiar una situación en particular** y,

como el análisis DAFO, es de beneficio para revisar la estrategia, posición y dirección de la empresa. Los elementos analizados en PEST son básicamente externos; es recomendable efectuar dicho análisis antes del DAFO, el cual está enfocado a factores internos (fortalezas y debilidades) y externos (oportunidades y amenazas).

El PEST mide el mercado; el DAFO, una unidad de negocio (específica).

En ocasiones, el análisis PEST se extiende a otros factores, como el **ecológico**, el **legislativo** y el **industrial**, para convertirse en PESTELI. Algunos consideran que esta extensión no es necesaria, dado que, si se realiza un adecuado PEST, este debería cubrir todos estos aspectos de forma natural.

## Análisis interno

La matriz DAFO también se conoce como matriz FODA o análisis SWOT (siglas en inglés de *strengths, weaknesses, opportunities y threats*). Es empleada para la formulación y evaluación de estrategias. Por lo general, se puede usar para las empresas y las personas. Su nombre se debe a las iniciales de las siguientes palabras:

- ▶ Debilidades.
- ▶ Amenazas.
- ▶ Fortalezas.
- ▶ Oportunidades.

En la Figura 5 se muestran los factores que se incluyen en el análisis DAFO:

Factores internos	Factores externos
<p>Fortalezas y debilidades</p> <p>Crean o destruyen valor para la empresa.</p> <p>Aquí se pueden incluir los recursos, activos, habilidades, etc.</p>	<p>Amenazas y oportunidades</p> <p>Están fuera del control interno de la empresa, se encuentran la competencia, la demografía, la economía (local, nacional e internacional), la política, la legislación, etc.</p>

Figura 5. Factores del análisis DAFO.

El proceso de crear una matriz DAFO es muy sencillo: en cada uno de los cuadrantes se hace una lista de factores. Posteriormente, se les puede asignar un peso o valor, según las necesidades y prioridades de la empresa que se evalúa. No obstante, la matriz DAFO resultante es atractiva, simple y fácil de entender. Los especialistas consideran que lo más valioso y revelador de la metodología es el propio proceso de análisis que se realiza para llegar hasta el punto deseado. En la Figura 6 se muestra un ejemplo de una matriz DAFO.

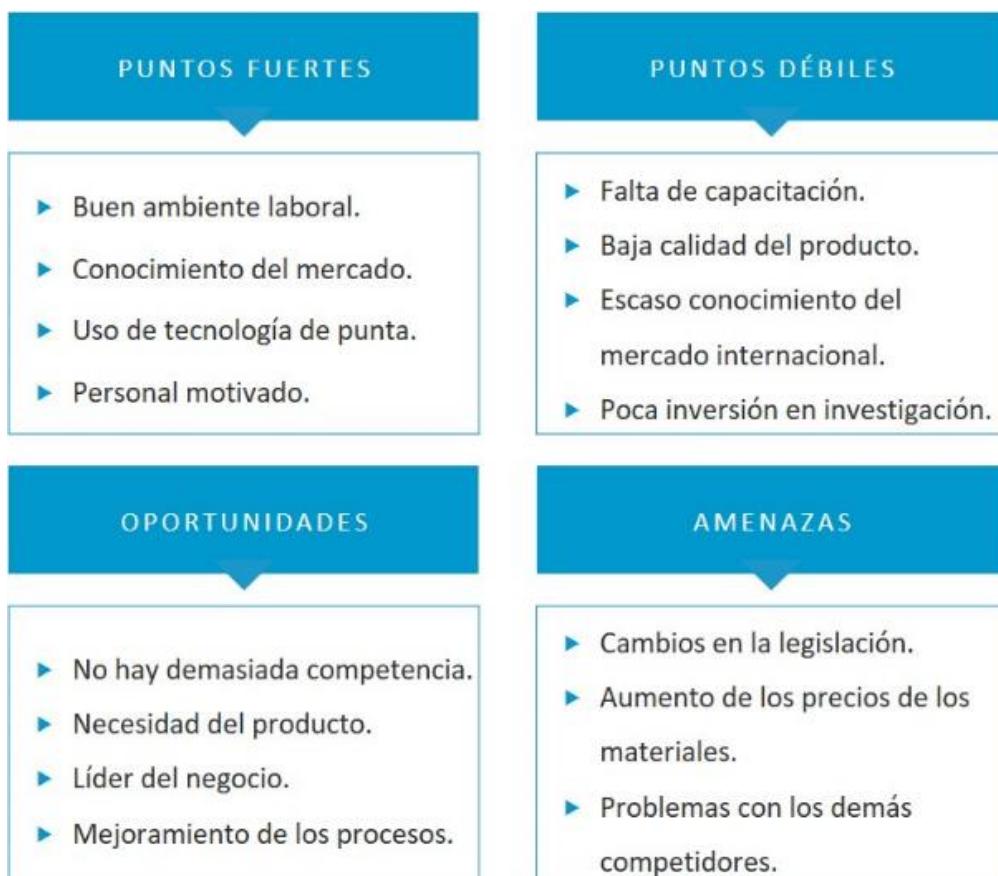


Figura 6. Ejemplo de matriz DAFO para cualquier empresa.

## 1.5. Cuadro de mando integral (CMI)

Es una **herramienta gerencial** que permite a los directivos y al personal de las empresas en general administrar y comunicar las estrategias, monitorear permanentemente la organización y utilizar los resultados de las medidas para la evaluación del desempeño, lo que facilita el logro de los objetivos (Wexler et al., 2017). Fue introducida por Robert Kaplan y David Norton, de la Universidad de Harvard, quienes desarrollaron un proyecto de investigación con doce compañías líderes en los procesos de medición. En 1992 inventaron la metodología llamada BSC (Kaplan y Norton, 1992).

Esta herramienta incluye los siguientes conceptos:

- ▶ Objetivos estratégicos.
- ▶ Perspectivas.
- ▶ Indicadores.
- ▶ Metas.
- ▶ Mapas estratégicos.

### Objetivos estratégicos

Son enunciados generales que definen las direcciones en las cuales pretende ir una empresa, sin comprometer niveles de desempeño específicos. Definen, cuantifican y cualifican el planeamiento estratégico mediante el logro de las metas que se ha propuesto la compañía en la visión:

- ▶ Diseñados para ver la empresa como un todo.
- ▶ Enfocados para alcanzar la visión de la empresa.

- Deben ser realistas y medibles.

## Perspectivas

Enmarcan los objetivos estratégicos, sus indicadores y sus metas, además de permitir tener una visión global de la empresa. En el siguiente gráfico se puede observar la integración de las cuatro perspectivas (Kaplan y Norton, 1992). Estas pueden adaptarse a las necesidades de las empresas.

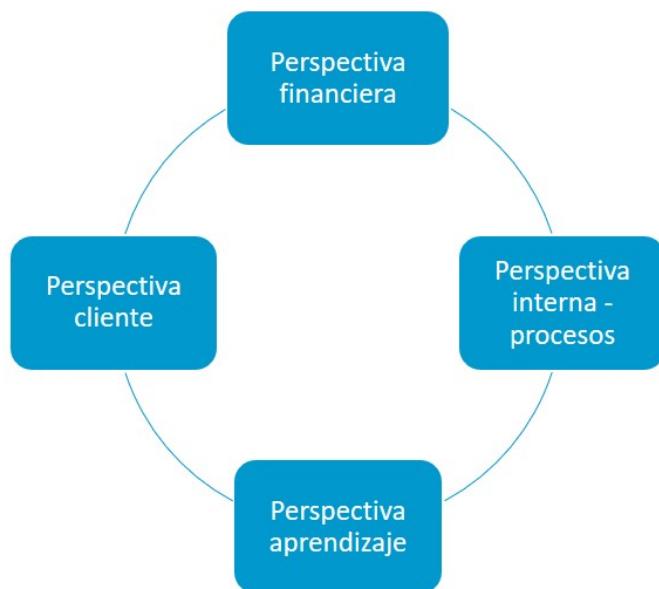


Figura 7. Perspectivas según Kaplan y Norton.

### Perspectiva financiera

Está centrada en la creación de valor para los dueños, con altos índices de rendimiento y garantía de crecimiento y mantenimiento del negocio.

Sus indicadores deben informar la habilidad y la capacidad que tiene la gerencia, para lograr convertir en ganancias los objetivos que se han acordado.

Los indicadores que se miden con mayor frecuencia en esta perspectiva son:

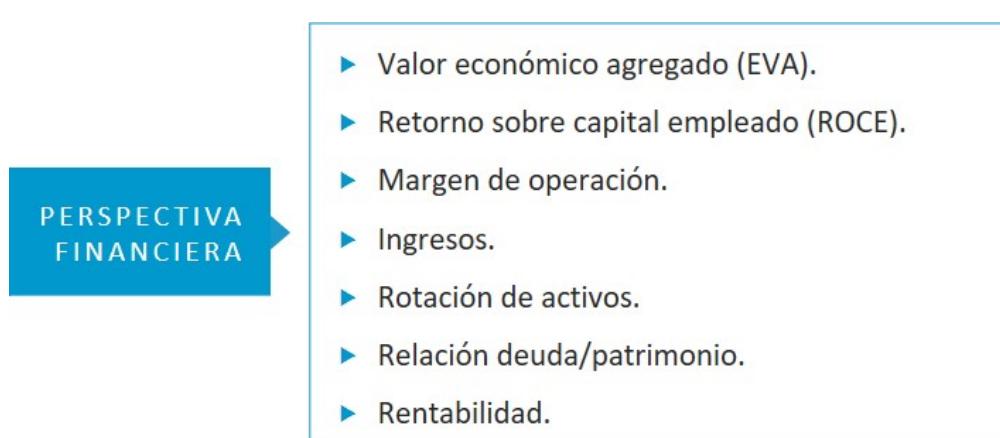


Figura 8. Indicadores de la perspectiva financiera.

## Perspectiva de cliente

Entender nuestro negocio a través de cómo nos ven nuestros clientes es importante para medir la capacidad de la empresa de satisfacer y retener a estos, además de atraer a aquellos potenciales.

Los indicadores que se miden con mayor frecuencia en esta perspectiva son:

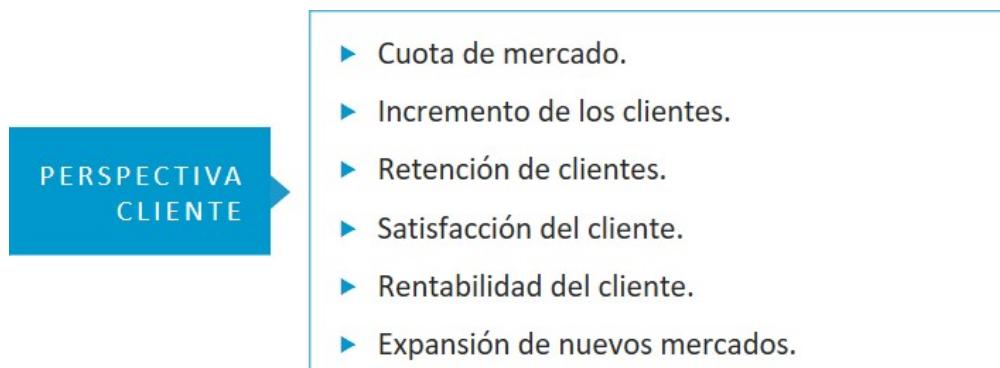


Figura 9. Indicadores de la perspectiva cliente.

## Perspectiva interna

Los indicadores deben mostrar la naturaleza misma de los procesos internos de la empresa enfocando las actividades más críticas de la misma.

Los indicadores que se miden con mayor frecuencia en esta perspectiva son:

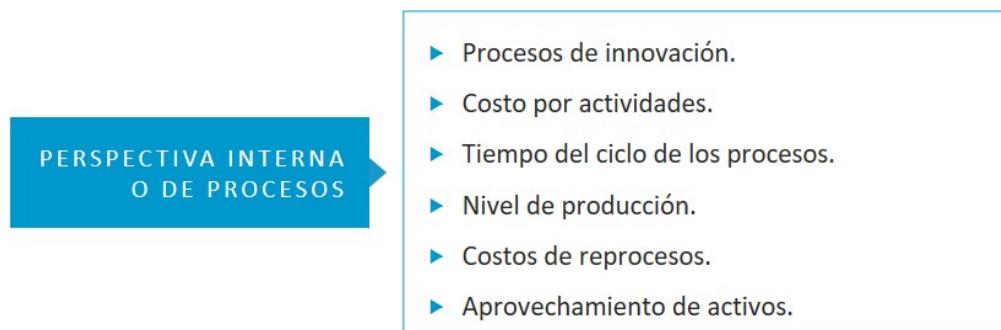


Figura 10. Indicadores de la perspectiva interna o de procesos.

Estos indicadores deben medir la eficiencia, la eficacia y la efectividad de los procesos.

CONCEPTOS	
Eficiencia	Proceso donde se minimizan los recursos y se elimina el desperdicio para lograr un objetivo en el menor tiempo posible.
Eficacia	Alcanzar el objetivo propuesto con éxito para satisfacer las necesidades de los clientes.
Efectividad	Eficiencia + eficacia.

Figura 11. Conceptos de eficacia, eficiencia y efectividad.

## Perspectiva aprendizaje

Informa sobre cómo crecen y se desarrollan los clientes internos de la empresa para poder estar en condición de anticiparse y satisfacer las necesidades de los clientes externos.

Los indicadores que se miden con mayor frecuencia en esta perspectiva son:

## PERSPECTIVA DE APRENDIZAJE

- ▶ Desarrollo de competencias básicas.
- ▶ Retención de los empleados.
- ▶ Satisfacción de los empleados.
- ▶ Productividad.
- ▶ Clima laboral.
- ▶ Tecnologías y sistemas de información.

Figura 12. Indicadores de la perspectiva de aprendizaje.

### Indicadores

Son cálculos que le permiten a la organización conocer si está cumpliendo o no con los objetivos estratégicos propuestos.

También se puede definir como una relación entre dos variables que se formula con el propósito de realizar una lectura distinta de un fenómeno o de algunos de sus componentes. Por ejemplo: A+B, A-B o A/B.

Existen dos tipos de indicadores:

- ▶ Estratégicos: miden las acciones que hace la compañía para, a su vez, medir sus resultados.
- ▶ De resultado: miden los logros finales del proceso.

Los indicadores deben tener las siguientes características, además de cumplir con la regla SMART (*specific* —específico—, *measurable* —mesurable—, *attainable*—alcanzable—, *relevant*—relevante— y *timely*—a tiempo—):

- ▶ Estar ligados a la estrategia empresarial.
- ▶ Tener una definición común.
- ▶ Ser accesibles.

- ▶ Ser relevantes.
- ▶ Ser cuantitativos.
- ▶ Ser de fácil comprensión.
- ▶ Tener frecuencia de medición.
- ▶ Tener un rango (mínimo, satisfactorio, sobresaliente). Los rangos se deben ajustar según las necesidades de cada empresa.
- ▶ Tener un tipo de unidad.
- ▶ Tener una fórmula.
- ▶ Tener una fuente de datos.

**Identificación visual de los indicadores:** el resultado del indicador en sí es un dato y para poder interpretarlo existen metas en las que se definen los valores mínimos, aceptables y deseables.

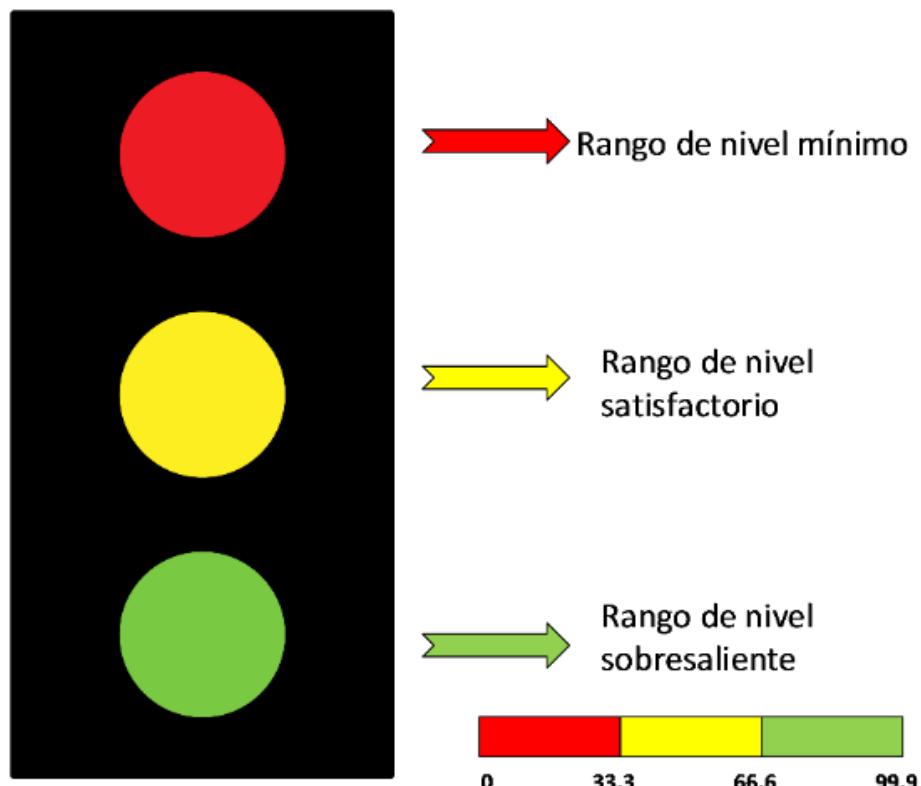


Figura 13. Representación visual de los indicadores.

Ejemplos de indicadores				
	Nivel de satisfacción de los clientes (perspectiva cliente)	Incremento de la rentabilidad económica (perspectiva financiera)	Reducción de tiempos-proceso (perspectiva interna)	Incremento de formación de profesores
<b>Rango mínimo</b>	75 %	5 %	10 días	60 %
<b>Rango satisfactorio</b>	76-90 %	6-10 %	5-9 días	61-80 %
<b>Rango sobresaliente</b>	91-100 %	> 10 %	> 5 días	81-100 %
<b>Meta</b>	95 %	10 %	5 días	95 %
<b>Valor real del indicador</b>	85 % (CMI: amarillo)	12 % (CMI: verde)	3 días (CMI: verde)	75 % (CMI: amarillo)

Figura 14. Ejemplos de indicadores.

Para la elaboración del Cuadro de Mando Integral o CMI se puede utilizar desde una hoja de cálculo sencilla a una solución tecnológica que nos permita automatizar el proceso, utilizando una herramienta de visualización de datos existentes en el mercado.

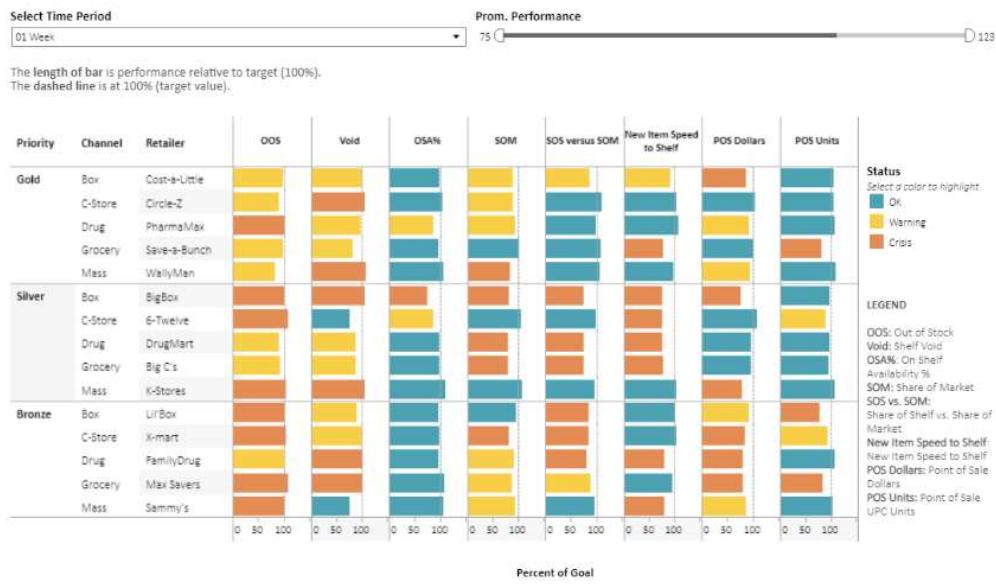


Figura 15. Ejemplo de CMI. Fuente: <https://www.tableau.com/es-es/solutions/workbook/achieve-your-goals-across-all-product-lines>

En la Figura 15 podemos observar un ejemplo de un cuadro de mando integral creado a través de la aplicación Tableau, una de las herramientas que nos ayuda a visualizar la información de la empresa.

## Metas

Las metas son declaraciones específicas concretadas en períodos de tiempo y deben ser medibles y cuantificables. Por lo general, también deben ser retadoras y realizables, teniendo en cuenta las fortalezas y las debilidades institucionales.

## 1.6. Toma de decisiones

Día a día, las personas se ven obligadas a escoger entre varias opciones, es decir, que muchas veces suponen una gran cantidad, algunas fáciles y otras difíciles de adoptar en función de las consecuencias o resultados derivados de cada una de ellas. De igual manera sucede en las empresas: deben escoger la mejor alternativa entre varias opciones.

Para llevar a cabo este proceso, es necesario tener información sobre cada una de las alternativas y sus consecuencias respecto a los objetivos definidos. La importancia de la información en la toma de decisiones ya la expresaba J. Forrester en su definición de decisión, entendida como «el proceso de transformación de la información en acción» (Forrester, 1968). La información es la esencial, es **la entrada al proceso**; luego, se trata de forma adecuada dentro del proceso de la toma de decisión y se obtiene como resultado una acción (Wexler et al., 2017).

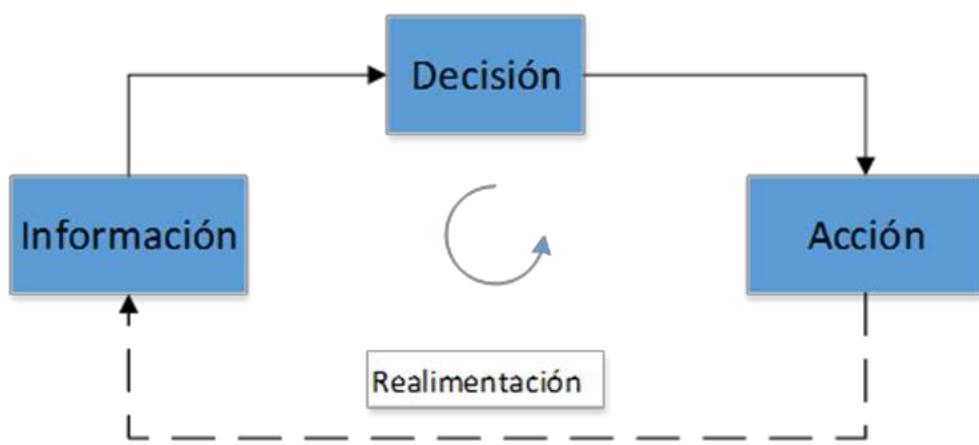


Figura 16. Proceso de toma de decisión. Fuente: adaptado de Menguzzato y Renau (1991).

La ejecución de la acción escogida genera nueva información y conocimiento que se integrará a la información ya existente para servir de base a una nueva decisión, que a su vez dará origen a una nueva acción y así repetidamente. Esto se conoce como

realimentación o **feedback**.

La Figura 17 muestra los pasos del proceso de toma de decisión: empieza por la identificación del problema, hasta seleccionar la alternativa que pueda resolverlo y, posteriormente, hacer una evaluación de la eficacia de la decisión. Este proceso se puede aplicar a las decisiones personales y, en especial, a las decisiones de una empresa.



Figura 17. Proceso para la toma de decisiones. Fuente: Robbins (2004).

## 1.7. Gobierno de datos

La toma de decisiones necesita una fuente de información veraz y confiable que generalmente no existe en las empresas. A principios del siglo XXI, los auditores en Estados Unidos de América prendieron las alarmas porque detectaron gran manipulación de datos financieros con el fin de mostrar una situación que no correspondía con la realidad (Trejo, 2020). Para evitar la manipulación de datos, nace la gobernanza del dato y de ella nace a su vez el gobierno de datos (GD), el cual se enfoca en la planificación y control de la administración de los datos a un alto nivel.

La meta del gobierno de datos es eliminar problemas presentes en las organizaciones: silos de datos, múltiples áreas generando informes manuales, dependencia entre áreas para acceder a datos, organización de informes manualmente, datos duplicados, KPI inconsistentes en cuanto a valores y criterios, y cálculos que utilizan diferentes fuentes (Cisterna y Arenas, 2020).

### **¿Por qué es importante implementar gobierno de datos?**

Facilita la disponibilidad, usabilidad, integridad y seguridad de los datos. Permite manejar un lenguaje común en la organización, mejorar la accesibilidad de los datos, asegurar la calidad e integridad de los datos y responder a las demandas actuales de ley de protección de datos (Starling, 2015).

The Global Data Management Community (DAMA International) es una comunidad global sin ánimo de lucro que promueve los conceptos y mejores prácticas de la gestión de información y datos. DAMA International define prácticas bien diferenciadas, como se puede ver en la Figura 18.

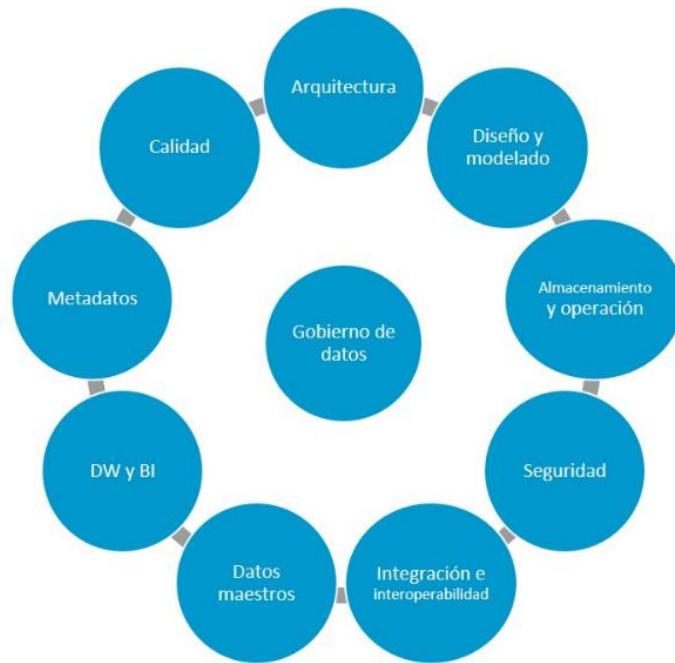


Figura 18. Mejores prácticas gobierno de datos. Fuente: adaptado de DAMA International (2015).

- ▶ Arquitectura de datos: estructuras de datos.
- ▶ Modelamiento y diseño de datos: diseño y construcción de aplicaciones para gestionar los datos.
- ▶ Almacenamiento, operación y persistencia de datos: dónde se va a guardar cada dato y qué operaciones se puede hacer con cada dato.
- ▶ Seguridad de los datos: estándar, clasificación, administración, autenticación y auditoría.
- ▶ Integración e interoperabilidad entre datos: dónde se producen los datos y cómo se integran.
- ▶ Gestión de datos maestros: de dónde salen los datos. Códigos externos, códigos internos, datos de clientes y datos de productos.
- ▶ *Data warehousing y business intelligence*: arquitectura, implementación, soporte y monitorización.

- ▶ Gestión de los metadatos: quién ha creado el dato y quién lo ha modificado.
- ▶ Calidad de los datos: test para comprobar integridad, datos únicos a pesar de diferentes fuentes. Análisis, medidas y mejoras.

En cada una de estas prácticas se puede medir el nivel de madurez con una puntuación que va del 0 al 5, como se explica en la Figura 19.

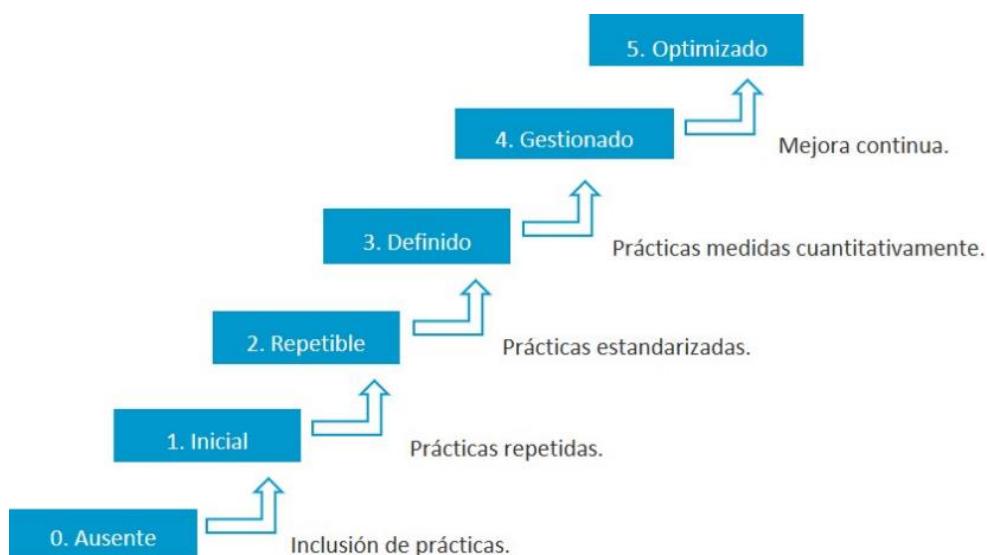


Figura 19. Niveles del modelo de madurez de gobierno de datos. Fuente: adaptado de Cisterna y Arenas (2020).

Cisterna y Arenas (2020) y U. S. Government Accountability Office (2020) presentan las siguientes recomendaciones para empezar a trabajar en el tema de gobierno de datos:

- ▶ Contar con un patrocinador importante dentro de la organización: es muy necesario el apoyo ejecutivo.
- ▶ Definir el lugar físico donde se instalará la oficina de gobierno con un equipo de gobierno conformado.
- ▶ Definir qué pilares estratégicos apoyarán el gobierno de datos, por lo general estos

pilares son las personas, las tecnologías y los procesos.

- ▶ Analizar en qué situación se encuentra la organización en cuanto a madurez o CMM (Capability Maturity Model).
- ▶ Definir el *roadmap* o ruta a seguir para la implementación del programa de gobierno.
- ▶ Definir roles internos de la oficina de gobierno de datos.
- ▶ Definir una política de datos para la compañía.
- ▶ Buscar aliados dentro de la organización.

La gobernanza de datos eficaz puede aumentar la disponibilidad y mejorar la calidad de los activos de datos federales. Esto es particularmente importante, ya que las agencias aprovechan los datos federales para monitorear y medir sus respuestas a la pandemia de la enfermedad del coronavirus (COVID-19).

**Figure 1: Data Governance Framework**

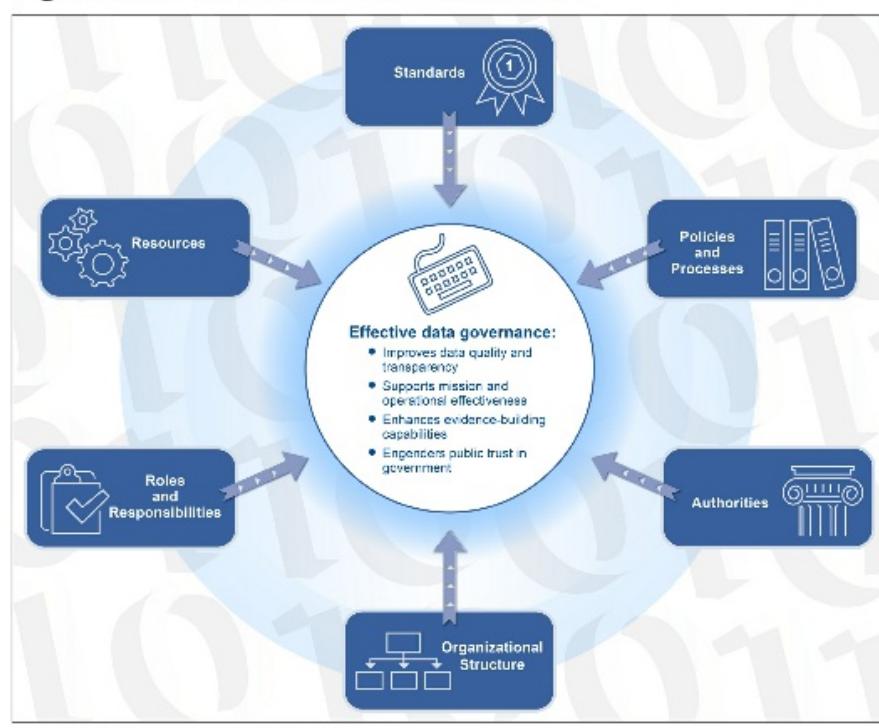
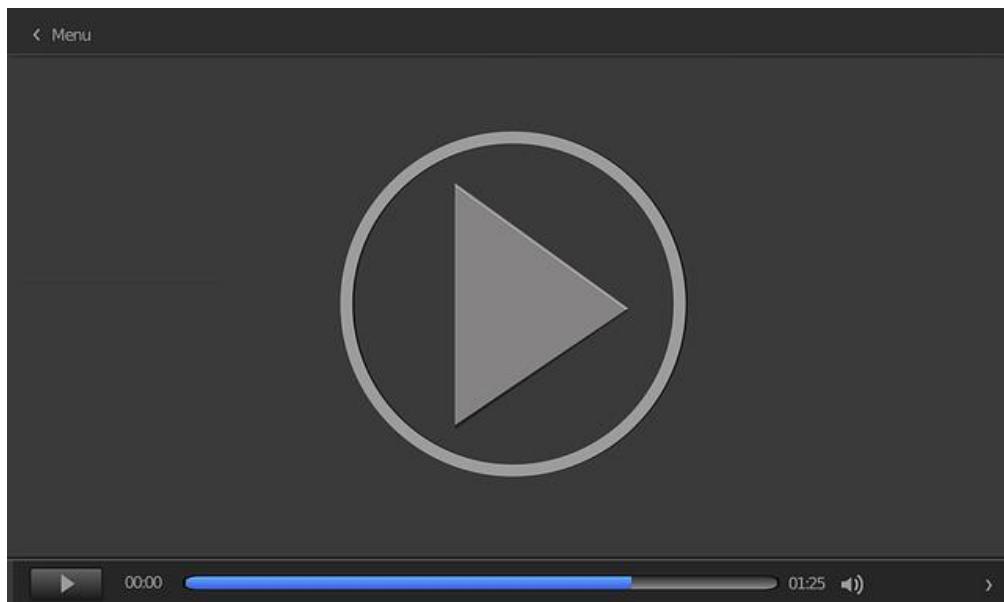


Figura 20. *Data governance framework*. Fuente: U. S. Government Accountability Office (2020).

Por último, accede al vídeo *Gobierno del dato en el framework DAMA*.



---

Accede al vídeo:

<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=d3bd48a5-de6d-4661-8ded-ad2b000b0e81>

---

## 1.8. Referencias bibliográficas

Chandler, A. D. (1969). *Strategy and structure: chapters in the history of the american industrial enterprise*. MIT Press.

Cisterna, D., y Arenas, K. [Club Data Governance Office Latam]. (6 de abril de 2020). *Resumen Webinar: Primeros pasos del Data Governance: Conformando una Oficina de Gobierno de Datos* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=DI21Z9yuCOY>

DAMA International. (2015). *DAMA: Guía de fundamentos para la gestión de datos*. Technics Publications.

Drucker, P. F. (junio de 2004). What makes an effective executive. *Harvard Business Review Press*. <https://hbr.org/2004/06/what-makes-an-effective-executive>

Fleitman, J. (2000). *Negocios exitosos: cómo empezar administrar y operar eficientemente un negocio*. McGraw-Hill.

Forrester, J. W. (1968). *Principles of systems*. Productivity Press.

Goodstein, L. D., Nolan, T. M., y Pfeiffer, J. W. (1999). *Planeación estratégica aplicada*. McGraw-Hill Interamericana.

Kaplan, R., y Norton, D. (1992). The Balanced Scorecard—Measures that Drive Performance. *Harvard Business Review*, 70(1), 47-54.

Martínez, D., y Milla, A. (2005). *La elaboración del plan estratégico y su implantación a través del cuadro de mando integral*. Ediciones Díaz de Santos.

Menguzzato, M., y Renau, J. J. (1991). *La dirección estratégica de la empresa*. Ariel.

Robbins, S. P. (2004). *Comportamiento organizacional*. Pearson Education.

Starling, H. (2015). *Data governance simplified: Creating and measuring trusted data for businesses.*

Trejo, D. (2020). *Gobierno de datos para directores: Realizando la transformación digital a partir de los datos.*

U. S. Government Accountability Office. (2020) *Data governance: agencies made progress in establishing governance, but need to address key milestones.*

Wexler, J. S., Shaffer, J., y Cotgreave, A. (2017). *The big book of dashboards: Visualizing your data using real-world business scenarios.* John Wiley & Sons.

## The Global Data Management Community

DAMA International. Página web oficial. <https://www.dama.org/cpages/home>

DAMA International es una comunidad sin ánimo de lucro dedicada a los avances y mejores prácticas del manejo de la información y los datos.

## Plan estratégico e implantación del cuadro de mando integral. Guía de resultados y mejores prácticas

Centro Europeo de Empresas e Innovación de Ciudad Real. (2009). *Plan estratégico e implantación del cuadro de mando integral. Guía de resultados y mejores prácticas*.  
[https://www.camaracr.org/uploads/tx\\_ictcontent/Manual\\_Experiencias\\_Plan\\_Estrategico\\_y\\_CMI\\_01.pdf](https://www.camaracr.org/uploads/tx_ictcontent/Manual_Experiencias_Plan_Estrategico_y_CMI_01.pdf)

La primera parte de este libro está dedicada al proceso de reflexión estratégica de la empresa, a través de la metodología para la realización de un plan estratégico que permita definir una estrategia adaptada al entorno, la historia empresarial y las posibilidades de cambio, entre otras. Sin embargo, también existen numerosas empresas que, teniendo una estrategia definida, no han alcanzado el éxito porque no han sido capaces de llevarla a la práctica. Por eso se ha dedicado la segunda parte del libro al proceso de implantación estratégica a través de la herramienta del cuadro de mando integral.

## El arte de la guerra

---

AMA Audiolibros. (14 de mayo de 2017). *Sun Tzu - El Arte de la Guerra (Audiolibro Completo en Español con Música) "Voz Real Humana"* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=DevNudhaGv4>

---

«Si las órdenes no son claras, la culpa es del general». *El arte de la guerra* es el mejor libro de estrategia de todos los tiempos, una inspiración para Napoleón y Mao Tse Tung, entre otros. Es un libro que, a pesar de haber sido escrito en el siglo IV a. C., no ha perdido vigencia, pues supone un tratado que enseña cómo aplicar la estrategia con sabiduría. Puedes conocerlo sin llegar a leerlo observando su influencia en este documental.

## Misión, visión y valores en una empresa

Enciclopedia de ejemplos. (2019). Misión, visión y valores de una empresa.

*Ejemplos.* <https://www.ejemplos.co/ejemplos-de-mision-vision-y-valores-de-una-empresa/>

En este artículo repasamos los conceptos de misión, visión y valores y vemos su aplicación en casos concretos variados.

- 1.** ¿Qué beneficio trae para la empresa la dirección estratégica?
  - A. La convierte en una empresa más competitiva.
  - B. Gestión integral en la evaluación de la decisión.
  - C. Ayuda en la implantación de indicadores.
  - D. Todas son correctas.
  
- 2.** Elementos del plan estratégico:
  - A. Información, metodología y alta gerencia.
  - B. Información y metodología.
  - C. Pensamiento estratégico.
  - D. B y C son correctas.
  
- 3.** Menciona los elementos del modelo de gestión estratégica:
  - A. Misión y visión.
  - B. Objetivos estratégicos.
  - C. A y B son correctas.
  - D. Pensamiento sistémico y procesos internos.
  
- 4.** ¿Qué preguntas se deben responder para crear la misión de una empresa?
  - A. ¿Qué?, ¿cómo?, ¿cuándo? y ¿por qué?
  - B. ¿Quiénes somos?, ¿qué hacemos? y ¿cómo lo hacemos?
  - C. ¿Para quién lo hacemos? y ¿dónde lo hacemos?
  - D. B y C son correctas.

5. ¿Qué elementos componen la visión?

- A. Ámbito, valores y misión.
- B. Horizonte de tiempo, negocio y datos.
- C. Ámbito de acción y estrategia.
- D. Ninguna es correcta.

6. ¿Por qué es importante implementar el gobierno de datos?

- A. Porque los datos van a ser de calidad.
- B. Porque se respeta la ley de protección de datos.
- C. Porque cada persona dentro de la empresa sabe qué debe hacer.
- D. Porque se logra gestionar el activo más importante de la organización (el dato) y se toman mejores decisiones.

7. Menciona las perspectivas del cuadro de mando integral:

- A. Financiera, cliente y valores.
- B. Financiera, cliente, procesos y aprendizaje.
- C. Aprendizaje, organizacional y procesos.
- D. Ninguna es correcta.

8. Menciona dos elementos que componen el cuadro de mando integral:

- A. Indicadores y metas.
- B. Mapa estratégico e indicadores estratégicos.
- C. Perspectivas y modelo de gestión.
- D. A y B son correctas.

**9.** Menciona cuatro etapas del proceso de toma de decisiones:

- A. Decisión, identificación de criterios, selección de opciones e información.
- B. Implementación de opciones, identificación del problema, desarrollo de opciones y evaluación de la decisión.
- C. Datos, identificación de criterios, identificación de problemas y análisis de alternativas.
- D. A y C son correctas.

**10.** ¿Qué elementos intervienen en la decisión?

- A. Decisión, información y acción.
- B. Realimentación, resultado e información.
- C. Acción, datos y decisión.
- D. B y C son correctas.

Gobierno del Dato y Toma de Decisiones

---

## Tema 2. Business intelligence y datos maestros

# Índice

[Esquema](#)

[Ideas clave](#)

[2.1. Introducción y objetivos](#)

[2.2. Datos, información y conocimiento](#)

[2.3. Datos maestros](#)

[2.4. Inteligencia de negocios](#)

[2.5. Business intelligence vs. business analytics](#)

[2.6. Referencias bibliográficas](#)

[A fondo](#)

[Introducción al business intelligence](#)

[Business intelligence y business analytics](#)

[Curso introductorio de business intelligence y business analytics](#)

[PowerData](#)

[Test](#)

# Esquema



## 2.1. Introducción y objetivos

¿Es lo mismo datos, información o conocimiento? La tendencia inicial es a responder afirmativamente a este interrogante, pero la respuesta correcta es «No». En este tema veremos en qué radica la diferencia principal y cómo los datos son la base de una buena generación de conocimiento. Veremos que existen diferentes tipos de datos, que el dato tiene un ciclo que debería cumplirse para poder garantizar seguridad y calidad. Además, veremos la importancia de crear repositorios de datos maestros, las ventajas de esta práctica y algunos pasos a seguir para su implementación.

Luego, seguiremos con las generalidades de los datos y entraremos en el mundo del *business intelligence* y el *business analytics* viendo sus principales diferencias y algunos pasos que nos faciliten su implementación.

Los objetivos de este tema son:

- ▶ Diferenciar los conceptos de dato, información y conocimiento.
- ▶ Analizar los pasos a seguir para implementar un repositorio de datos maestros.
- ▶ Comprender la importancia del ciclo que debe cumplir el dato en una estrategia de gobierno de datos.
- ▶ Conocer las semejanzas y diferencias entre *business intelligence* y *business analytics*.

## 2.2. Datos, información y conocimiento



Figura 1. Preguntas sobre datos e información.

¿Es lo mismo datos, información o conocimiento? La tendencia inicial es a responder afirmativamente a este interrogante, pero la respuesta correcta es «No».

Si empezamos a indagar un poco sobre ellas, las tres tienen significado diferente:

- ▶ **Los datos** son elementos sin procesar, sacados de la realidad que a su vez generan nuevos elementos y que por sí solos no generan nuevo conocimiento. Ejemplos de datos: el precio de un producto, la edad, el nombre de una persona, etc.
- ▶ **La información** es el principio del conocimiento. Son datos con un significado o función especial o el resultado de combinar diferentes datos, es decir, son datos con contexto.
- ▶ **El conocimiento** es la información analizada que hace nuevos aportes a un área específica.



Figura 2. Cómo se integran los datos y la información. Fuente: Weller, 2010.

Los datos son la principal fuente de información para el análisis de grandes volúmenes, por lo que es fundamental que las empresas realicen una selección lo más adecuada y cuidadosamente posible. La categorización de los datos es importante para cualquier proyecto, en especial cuando se trabaja con grandes volúmenes (*big data*). La principal categorización de los datos se basa en su estructura, en la que nos encontramos principalmente dos posibilidades:

- ▶ **Estructurados:** estos datos son aquellos que tienen una estructura definida y que no cambian independientemente de cuál sea su origen. Es decir, son aquellos datos que poseen un modelo (o estructura) definido. Entre estos tipos de datos podemos encontrar registros de bases de datos, que son el ejemplo más típico, datos de sensores o los que se obtienen a partir del API de Twitter.
- ▶ **No estructurados:** aquellos datos que no disponen de una estructura bien definida. Es decir, no poseen un modelo (o estructura) definido o no están ordenados de ninguna forma. Entre estos tipos de datos podemos encontrar fotografías, vídeos o documentos de texto (Word, PDF, etc.).



Figura 3. Introducción a la categorización de datos. Fuente: elaboración propia.

En cuanto a cómo se han generado los datos, nos podemos encontrar con los siguientes tipos de datos:

- ▶ **Creados:** son aquellos generados por la propia empresa a través de los sistemas de información.
- ▶ **Compilados:** son aquellos que se utilizan de otras grandes bases de datos, como censos electorales, información obtenida de las administraciones públicas en salud, vivienda, impuestos, etc.
- ▶ **Experimentales:** son los generados por simulaciones o pruebas para determinar la validez de los sistemas.

En la Figura 4 se muestran algunas fuentes de información que hoy en día se han ampliado gracias a las nuevas tecnologías de la información.

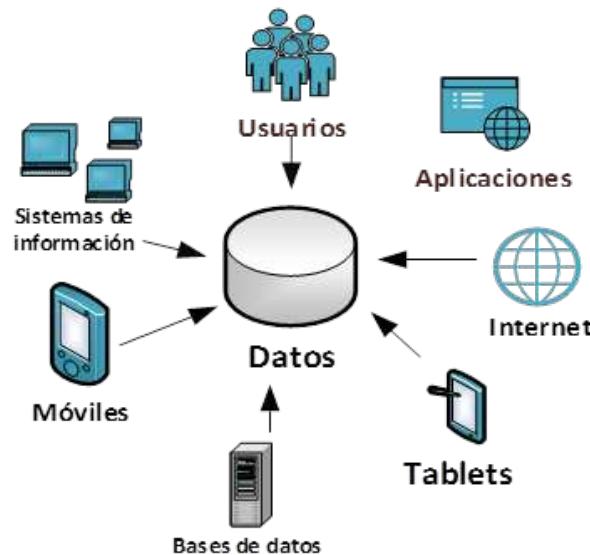


Figura 4. Fuente de datos.

- ▶ **Web (Internet) y medios sociales:** son aquellos que tienen origen en la red. Según los expertos, es la fuente más grande del *big data* y una de las más utilizadas en la actualidad. Se genera información en los clics de los vínculos y elementos, así como en las búsquedas que se hacen, las publicaciones en las redes sociales (Twitter, Facebook, LinkedIn...) y el contenido web como páginas, enlaces o imágenes.
- ▶ **Tabletas y móviles (smartphone):** son dispositivos móviles que permiten acceder a información desde cualquier parte gracias a su tamaño, las aplicaciones y su acceso a Internet. Entre todas las fuentes de datos utilizables, el teléfono inteligente o tableta es el que más potencial tiene.

Gracias a estos dispositivos, compañías como Google y Apple pueden saber dónde se encuentra una persona en cada momento y, de esta manera, poder conocer los gustos de comida, vestuario y diversión, entre otros.

- ▶ **Bases de datos y sistemas de información:** una base de datos es un almacén en el que se pueden organizar datos, para evitar la redundancia y mejorar el acceso a estos. Diferentes programas y usuarios deben poder utilizar los datos guardados. De ahí el término base «sistema de información».

## Tipos de datos

Para los diferentes procedimientos que se realizan en la inteligencia de negocios, es importante definir los tipos de datos que se van a usar para obtener conocimiento.

En la empresa se pueden encontrar los siguientes:

- ▶ **Numéricos:** aquellos valores enteros o reales. Por ejemplo: la edad, el salario y las horas trabajadas, entre otros.
- ▶ **Fecha/hora:** se identifican como campos de fecha/hora aquellos valores cuyo contenido encaja con formatos de fecha/horas más usuales. Sus componentes pueden ser: año, mes, día, hora, minutos o segundos.
- ▶ **Texto:** sus valores tienen texto libre, es decir, que no está limitado a un subconjunto de etiquetas. Por ejemplo, el nombre de los empleados y el nombre del departamento, entre otros.
- ▶ **Alfanuméricos:** son una combinación de los datos numéricos y datos de tipo texto como, por ejemplo, el número de DNI, la dirección, etc.
- ▶ **Booleano:** es un tipo con solo dos valores: verdadero y falso (Gagliardi et al., 1999).

La mayoría de las veces los datos no pueden ser utilizados de la manera en que se encuentran almacenados, pues pueden encontrarse en las ya mencionadas bases de datos o en archivos planos, entre otros. Debido a la poca rigidez de los sistemas de calidad y control al introducirlos a los diferentes sistemas, existen diferentes tipos de errores: datos incompletos, datos con ruido, datos inconsistentes o duplicados.

Es necesario tener un proceso de **calidad** de los datos para minimizar los errores y poder obtener un mayor rendimiento en el análisis; de lo contrario, los algoritmos generalmente ignoran el dato y se pierde información valiosa.

El autor Dorian Pyle define la preparación de los datos como «la manipulación y transformación de los mismos sin refinar para que la información contenida en el

conjunto de datos pueda ser descubierta o estar accesible de forma más fácil» (Pyle, 1999).

Los componentes para la preparación de datos son: limpieza, integración, transformación y reducción.

- ▶ **Limpieza:** esta tarea puede involucrar: completar los datos faltantes, resolver los problemas de ruido y eliminar valores extremos o corregir los datos inconsistentes (Han et al., 2006).
- ▶ **Integración:** los datos pueden provenir de diferentes fuentes y, por esta razón, es necesario agruparlos en diferentes tablas para crear información homogénea. Los datos que provienen de diferentes fuentes pueden tener diferentes formatos y escalas. La recopilación tiene que ser coherente. Frecuentemente, esta integración de datos se realiza en una base de datos (Herrera et al., 2004).
- ▶ **Transformación:** en este paso se crean nuevos atributos a partir de los atributos originales. Esta transformación puede facilitar una mejor interpretación de la información (Lin, 2002). Por ejemplo, el índice de masa corporal en datos médicos se calcula con el peso y la altura de una persona. Otros ejemplos de transformación de datos son la discretización, la normalización y derivación.
- ▶ **Reducción:** la reducción de la dimensionalidad consiste en aplicar una transformación para conseguir una representación reducida de los datos originales sin perder información. En esta parte se deben escoger aquellas variables o atributos que influyan en la obtención de conocimiento.

La capacidad de la organización de gestionar los datos de forma eficiente depende de que los datos cumplan un ciclo de vida y se garanticen las mejores prácticas en cuanto a calidad y seguridad en cada una de ellas. El ciclo de vida se puede ver en la Figura 5.

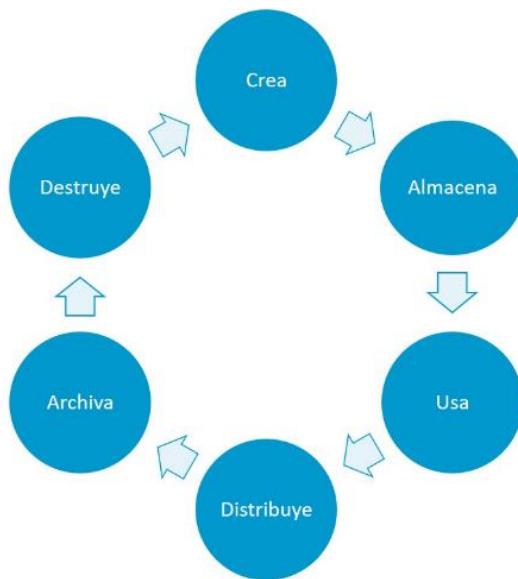


Figura 5. Ciclo de vida del dato en el gobierno de datos. Fuente: adaptado de Dataworks, s.f.

## 2.3. Datos maestros

Uno de los mayores problemas a la hora de consolidar los datos es que estos se encuentran en diferentes fuentes y pueden llegar a tener valores diferentes. Por ejemplo, un mismo cliente puede estar en dos bases de datos y tener una dirección o un teléfono diferente en cada una de ellas. La gestión de datos maestros (MDM, por sus siglas en inglés de *master data management*) nace como solución a este problema. El objetivo es que los datos críticos se encuentren en un solo sitio, simplificando de esta forma el intercambio de datos entre personas y departamentos de una misma organización (Trejo, 2020).

Los datos maestros se crean en general para unificar datos de clientes, empleados, proveedores y productos. Contienen la información central compartida dentro de la organización, la administración de los datos maestros es clave y necesaria, debe ejecutarse en coordinación con el grupo de gobierno de datos.

### ¿Es importante utilizar MDM?

Un error en un dato puede causar errores en todas las decisiones de una empresa. Si los datos maestros se crean pero no se administran, esto puede repercutir en todas las aplicaciones que utilizan dichos datos, así que es muy importante crear datos maestros y administrarlos. Un error muy típico se produce cuando un cliente cambia su dirección y esta es actualizada en una fuente de datos, pero no en todas, y nunca recibe notificaciones. Tiempo después, se dan cuenta de que la dirección no ha sido modificada en todas las fuentes de datos. Esto puede causar impagos o problemas serios al cliente. **Los datos maestros deben ser correctos y consistentes.**

Combinar datos maestros puede llegar a ser una tarea compleja. Por ejemplo, si se fusionan dos empresas, cada una de ellas tiene una estructura diferente en sus bases de datos. El mismo cliente puede tener varios nombres, números de

identificación diferentes, direcciones, teléfonos... Es necesario utilizar herramientas existentes que ayuden en esta tarea. Consolidar el maestro de datos es un desafío, pero a la vez tiene muchos beneficios: datos de factura consolidados, *marketing* más efectivo y codificación única de artículos. La Figura 6 explica los pasos más importantes a seguir para la creación de un repositorio de datos maestros.

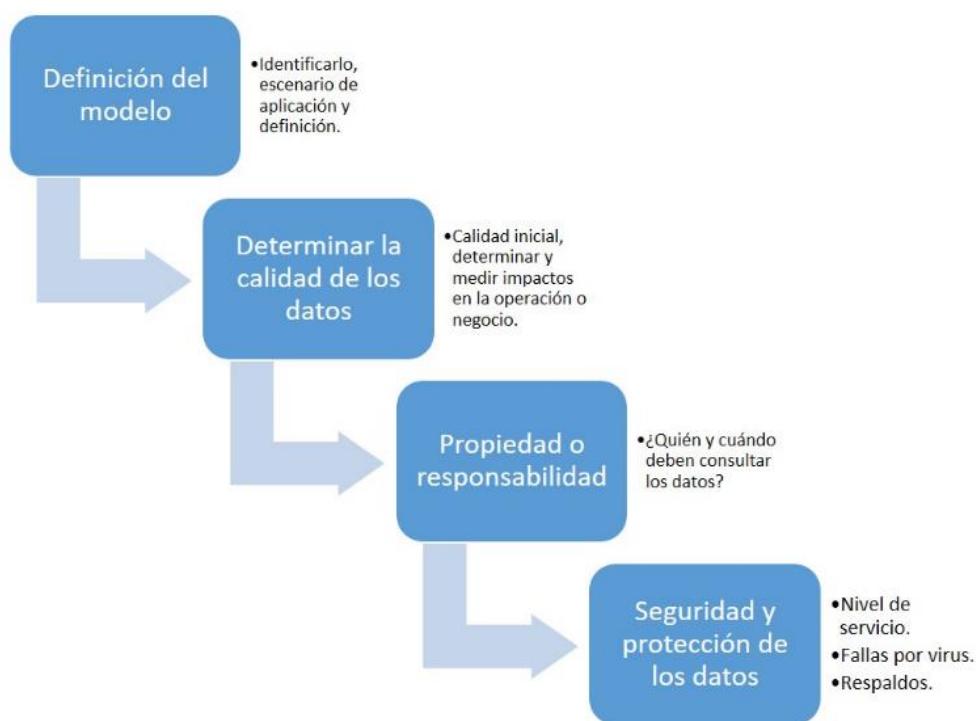


Figura 6. Vista *top down* de los pasos a seguir para un maestro de datos. Fuente: adaptado de Trejo, 2020.

La eficacia de la gestión de datos maestros se puede medir mediante varios criterios:

- ▶ **Disponibilidad de datos:** medido del 0 al 1, de acuerdo con lo interesantes que pueden ser los datos generados para cierta consulta.
- ▶ **Calidad de los datos:** rapidez, completitud y veracidad. También con una medición de 0 a 1.
- ▶ **Desempeño de la consulta de datos:** datos en tiempo real, incluyendo datos agregados. Medición entre 0 y 1.



## 2.4. Inteligencia de negocios

Si se cuenta con un repositorio de datos maestros unificados, es más fácil implementar una estrategia de inteligencia de negocios. Hasta hace algunos años estos datos maestros se construían en el proceso BI, ahora con el gobierno de datos la idea es crearlos antes de cualquier proceso BI, BA u otro que implique generación de conocimiento.

El *business intelligence* (BI), que se traduce como «inteligencia de negocios», es un **proceso de intercambio** para explorar y analizar información estructurada de la empresa o sobre una determinada área (con frecuencia almacenada en un *data warehouse*), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conocimiento para el mejoramiento de la empresa (Teixeira et al., 2019).

El proceso del *business intelligence* incluye la comunicación de los descubrimientos y la ejecución de los cambios. Las áreas que abarcan, por lo general, son clientes, proveedores, productos, servicios y competidores (véase en el subapartado de Webgrafía el ejemplo de Gather).

Implementar soluciones de inteligencia de negocios dentro de la empresa ayuda en las decisiones que se toman. También contribuyen a **nivel interno**, para apoyar la gestión del personal (Sharma et al., 2009), y a nivel externo, producir ventajas sobre sus competidores (Valenzuela, 2007).

En algunas ocasiones no se pueden lograr todos los beneficios que tiene el *business intelligence* debido al proceso que se lleva a cabo al implementar un proyecto de estas características. Se pueden cometer errores en la definición del planteamiento de las necesidades de conocimiento de la empresa y, al no determinar bien los problemas de información que necesitan solución, generalmente causan el fracaso

del proyecto.

## ¿Quién necesita el *business intelligence*?

Todas aquellas personas de la empresa que tienen que tomar decisiones. Dependiendo del tipo de negocio, se deben hacer las preguntas necesarias para responder y establecer el modelo de *business intelligence* que mejor se adapte.

## Beneficios del *business intelligence*

Uno de los objetivos básicos de los sistemas de información es contribuir a la toma de decisiones. Cuando se requiere tomar una decisión, es necesario pedir o buscar información, que servirá para minimizar la incertidumbre. En todo caso, no todos los responsables recogen la misma información: depende de factores como la experiencia, la formación y la disponibilidad, entre otros.

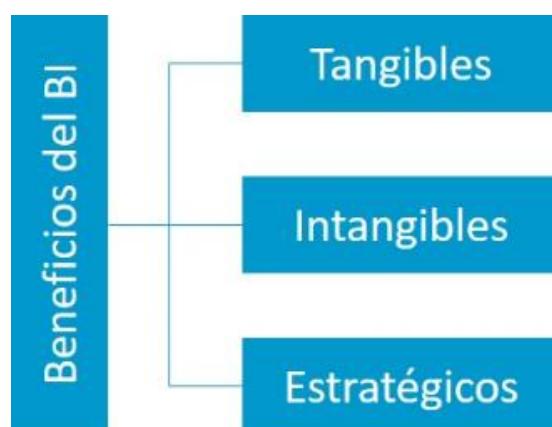


Figura 7. Beneficios del *business intelligence*. Fuente: elaboración propia.

Del mismo modo, los responsables pueden necesitar recoger más o menos información dependiendo del tipo de problema por resolver. A partir de los datos que proporciona el sistema de *business intelligence*, se puede descubrir nuevos aportes (conocimiento). Los beneficios pueden ser de distintos tipos (Figura 7) (Puklavec et al., 2018):

- ▶ **Beneficios tangibles:** son aquellos que la empresa puede cuantificar y que le

aportan beneficios económicos. Ejemplo: reducción de costes de producción, generación de nuevos ingresos, reducción en tiempo de producción, evitar pérdidas de clientes o materia prima, aumentar la rentabilidad.

- ▶ **Beneficios intangibles:** son aquellos que no se pueden cuantificar, pero que aportan valor agregado a los servicios o productos y mejoran la posición competitiva. Ejemplo: mejorar la atención al cliente, aumentar la satisfacción del cliente interno y externo, tener información más actualizada.
- ▶ **Beneficios estratégicos:** son aquellos que facilitan la creación de nuevas estrategias, respecto a qué clientes, mercados o con qué productos encaminar los esfuerzos de la empresa. Ejemplo: mejorar la toma de decisiones, identificar clientes potenciales, etc.

## Arquitectura del *business intelligence*

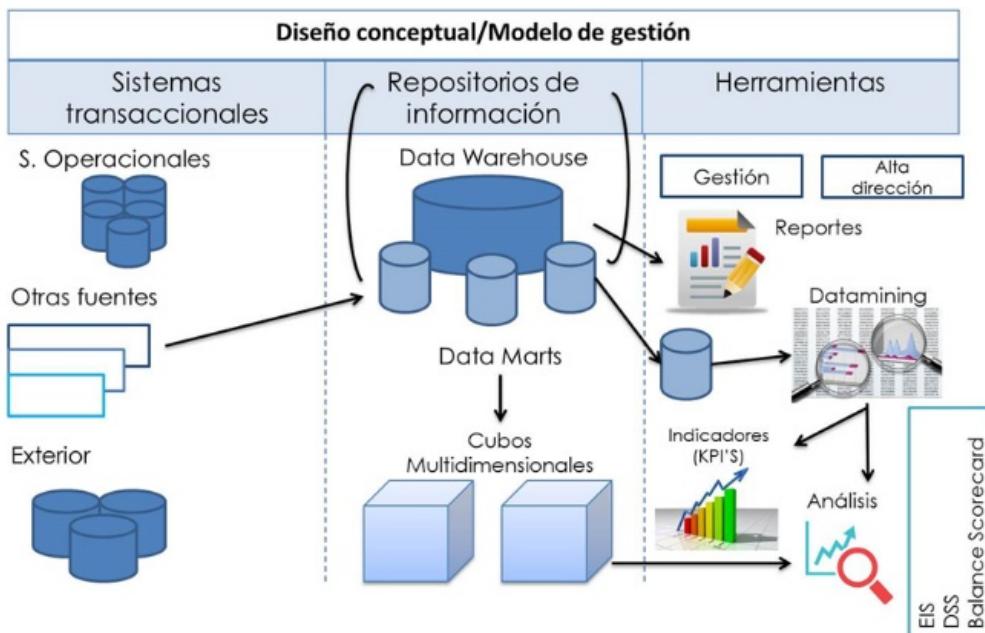


Figura 8. Arquitectura del *business intelligence*. Fuente: IDS2015, 2015.

Una solución de este sistema parte de varias fuentes de datos que suelen ser transformadas estructuralmente para optimizar el análisis, proceso al que se le denomina ETL. Una vez está unificada, la información se almacena en un *data*

*warehouse* que puede servir como base a distintos *data marts*. Los datos almacenados en el *data warehouse* o *data mart* se explotan utilizando herramientas de visualización o *reporting*. Esto lo podemos observar en la Figura 8.

- ▶ **Procesos ETL:** consisten en la extracción, transformación y carga de los datos en el *data warehouse*. Antes de guardarlos ahí, deben ser transformados, limpiados, filtrados y redefinidos. Como se mencionó anteriormente, la información que tienen las empresas en los sistemas transaccionales no está preparada para la toma de decisiones.
- ▶ **Data warehouse:** también llamado almacén de datos, con el *metadata* o diccionario de datos. Se busca almacenar los datos de una forma que facilite y maximice su flexibilidad, facilidad de acceso y administración. Surge como respuesta a las necesidades de los usuarios que necesitan información consistente, integrada, histórica y preparada para ser analizada y apoyar la toma de decisiones.
- ▶ **Herramientas OLAP:** para proveer la capacidad de cálculo, consultas, funciones de planeamiento, pronóstico y análisis de escenarios en grandes volúmenes de datos. En la actualidad, existen otras alternativas tecnológicas al OLAP. Siguiendo el modelo se deben analizar las tecnologías que permitirán tratar y visualizar la información que reside en un *data warehouse*. En este apartado también se tratarán las herramientas de visualización, ya que en muchas ocasiones van ligadas.

## Relación entre **business intelligence** y **big data** (BD)

El *business intelligence* se refiere a las habilidades, tecnologías, aplicaciones y prácticas para la exploración iterativa continua del pasado empresarial para proporcionar información útil en el presente. El *business intelligence* se centra en el desarrollo de nuevos conocimientos y la comprensión del rendimiento empresarial basados en métodos estadísticos.

El término *big data* se usa para caracterizar conjuntos de datos grandes, diversos y que cambian rápidamente, lo que es cada vez más frecuente en todas las

organizaciones. *Big data* requiere de sistemas de administración de bases de datos con capacidades más allá de las que se ven en los sistemas estándar basados en SQL.

La mayoría de las definiciones intuitivas de *big data* se centran en el volumen de datos que se producen, a menudo medidos en términos de **tera** (1012), **peta** (1015) o **exa** (1018) bytes. Algunos afirman que se está ingresando en la llamada «Era Petabyte» (Anderson, 2008) mientras que otros prefieren hablar de cuántos *exabytes* de datos se producen cada día (McAfee y Brynjolfsson, 2012).

Sin embargo, aunque el volumen es, sin duda, un aspecto del *big data* (probablemente el menos problemático), a medida que la tecnología se desarrolla, lo que fue grande en el pasado será normal mañana y probablemente se piense que es bastante pequeño en el futuro. Por consiguiente, para comprender qué hace que el *big data* sea diferente, también se debe considerar las dimensiones de la **velocidad**, la **veracidad** y la **variedad**.

- ▶ **Velocidad:** mientras que el volumen se refiere a lo que podría considerarse una reserva de datos, la velocidad se refiere a la que esa acción cambia; por ejemplo, la velocidad a la que se generan los datos, la frecuencia a la que se actualiza o la velocidad a la que son entregados.

Entre los ejemplos de datos de alta velocidad se incluyen datos financieros de mercados bursátiles, datos en tiempo real de sensores y cámaras de vídeo y datos de *stream* generados por visitantes a tiendas en línea.

- ▶ **Veracidad:** la veracidad hace referencia a la incertidumbre de los datos, es decir, al grado de fiabilidad de la información recibida. Para solventar el problema de la veracidad, es necesario conseguir datos de calidad, aplicando soluciones y métodos que puedan eliminar datos imprevisibles que puedan surgir como datos económicos, comportamientos de los consumidores que puedan influir en las decisiones de compra.

- ▶ **Variedad:** aunque tal vez no sea tan obvio como el volumen o la velocidad, en muchos sentidos la variedad plantea el mayor problema para el análisis de *big data*. La variedad se refiere a la cantidad de diferentes fuentes de las que pueden proceder los datos y los formatos, estructuras y semántica que están asociadas a ellas (la estructura se refiere tanto al formato en el que se almacenan los datos como el número y la longitud de los campos, y, más crucial, la semántica que debe asociarse con esos campos). Para que una computadora pueda procesar datos de manera que sean válidos y significativos para los seres humanos, primero los datos deben codificarse, es decir, un valor semántico —efectivamente un significado— tiene que asignarse a cada elemento de datos (Kimble, 2013).

Si bien estas cuatro variables (las 4 uves) son las más prominentes, revistas especializadas hablan de 5, 8 y hasta 10 uves. En un artículo de febrero de 2017, el portal *Transforming Data with Intelligence* proponía la siguiente lista de uves: volumen, velocidad, variedad, variabilidad, veracidad, validez, vulnerabilidad, volatilidad, visualización y valor (Firincan, 2017).

Las diferencias entre ***business intelligence*** y ***big data*** tienen consecuencias sobre cómo están organizadas. Tradicionalmente, los equipos de *business intelligence* están ubicados en organizaciones de consultoría internas, centros de excelencia o departamentos de TI, donde proporcionan a los gerentes y ejecutivos reportes con información bien definida, estable y útil (Burton et al., 2006; Davenport et al., 2012).

Sin embargo, dado que la mayoría de las iniciativas de *big data* carecen de preguntas predefinidas y son de naturaleza mucho más experimental (Casey et al., 2013), los especialistas de *big data* deben organizarse para que estén cerca de los productos y procesos en las organizaciones, es decir, que comparten negocios y unidades (Davenport et al., 2012). A continuación, se muestran algunas semejanzas y diferencias en las áreas de competencia de *business intelligence* y *big data*.

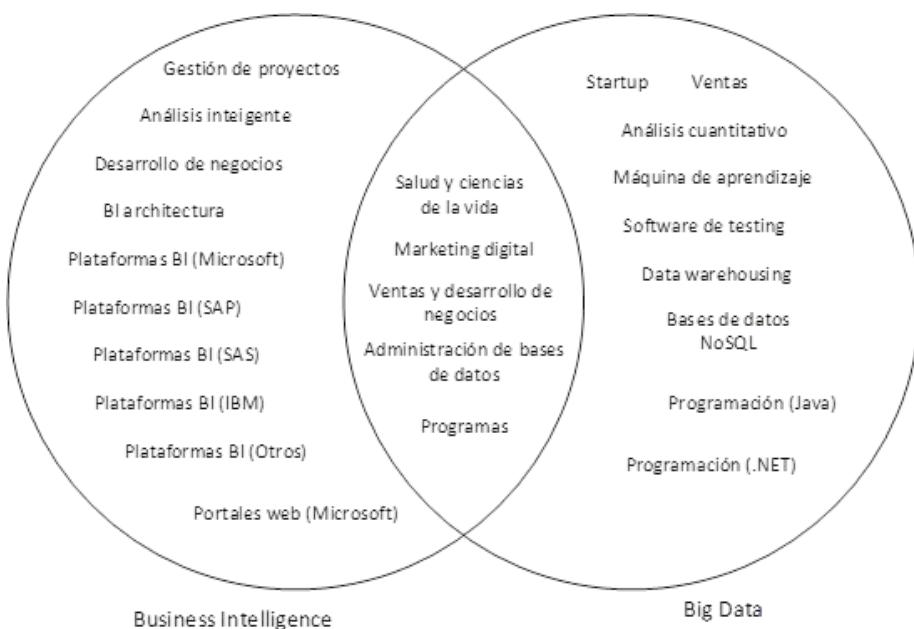


Figura 9. Semejanzas y diferencias en las áreas de competencia de *business intelligence* y *big data*. Fuente: Debortoli et al., 2014.

## ¿Cómo construir un ecosistema BI?

En la Figura 10 podemos ver las diferentes etapas que se deberían seguir para poder implementar un proyecto BI en una empresa.



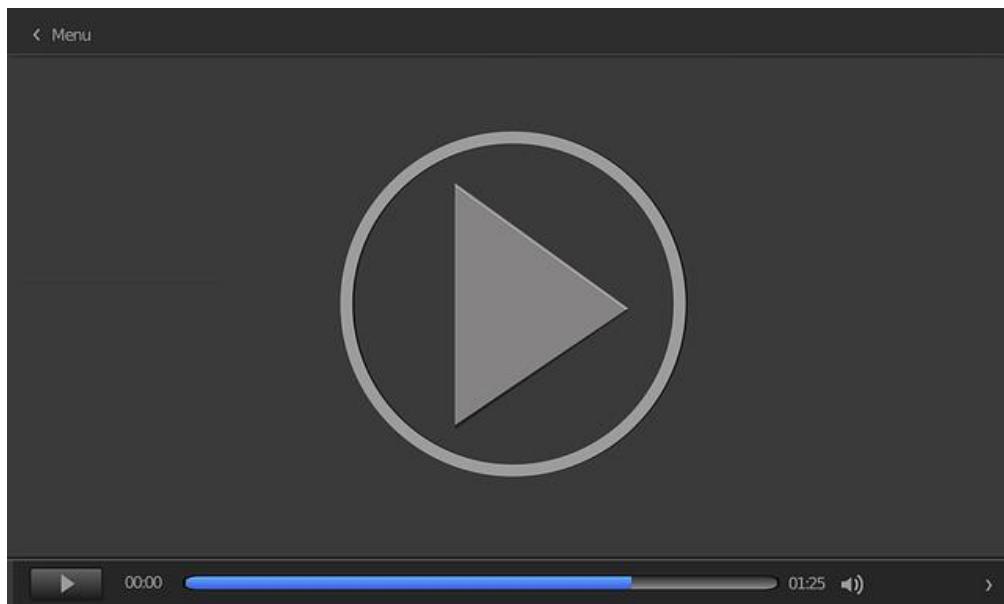
Figura 10. Fases para la construcción de un proyecto BI. Fuente: elaboración propia.

- ▶ Partimos de los objetivos estratégicos de la empresa.
- ▶ El estado actual hace referencia a si se está o no preparado para implementar una estrategia de BI. Puede que culturalmente estemos preparados; puede que haya empresas que estén muy evolucionadas en infraestructura, pero no en organización.
- ▶ En el estado deseado debemos hacernos la siguiente pregunta: ¿a dónde quiero llegar, según los objetivos estratégicos? Por ejemplo, yo quiero que todas las decisiones de la compañía se rijan por datos.
- ▶ En la estrategia BI debemos preguntarnos: ¿qué habilidades debo tener al interior de la organización?, ¿cómo promover el uso de las herramientas?, ¿cómo implementar el gobierno del dato? Si no definimos procesos de BI al interior de la compañía, no llegaremos a buen término.
- ▶ En el *roadmap* de soluciones, tenemos que buscar iniciativas para poder llegar al estado deseado. Priorizar iniciativas, de acuerdo con lo que la empresa tiene y

puede.

- ▶ Diseño, construcción y evaluación son tareas asignadas al departamento de tecnología. Teniendo en cuenta los anteriores pasos, ya se podría pensar en qué diseño y qué herramientas son las más adecuadas.

A continuación, accede al vídeo *Roadmap del ingeniero de datos*.



---

Accede al vídeo:

<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=7743072b-9c9a-4890-bd16-ad2b000b0e22>

---

## 2.5. Business intelligence vs. business analytics

La gestión de la empresa está fundamentada en la toma de decisiones más apropiada para cumplir con los objetivos del negocio, satisfacer las necesidades de los clientes y empleados y mantener o mejorar la calidad de los productos. Con el avance en las tecnologías de información y las comunicaciones, el aumento en la capacidad de almacenamiento de datos ha dado paso a nuevas metodologías tales como: el *business intelligence* (BI) y el *business analytics* (BA), que ayudan y facilitan el proceso de toma de decisiones.

- ▶ **Business intelligence:** la inteligencia de negocios (BI) es un instrumento mediante el cual diferentes organizaciones pueden apoyar la toma de decisiones basadas en información precisa y oportuna para garantizar la generación del conocimiento necesario que permita seleccionar la alternativa que sea más conveniente para el éxito de la empresa (Rosado y Rico, 2010).
- ▶ **Business analytics:** es un conjunto de técnicas (entre las que se encuentran algoritmos predictivos y modelos estadísticos) que le permiten a la organización predecir posibles eventos o resultados. Esto es, se enfoca en el análisis futuro en función de la información de la empresa y modelos predictivos para apoyar la toma de decisiones y mejorar los procesos y, por ende, la competitividad del negocio (Thorlund y Laursen, 2017).

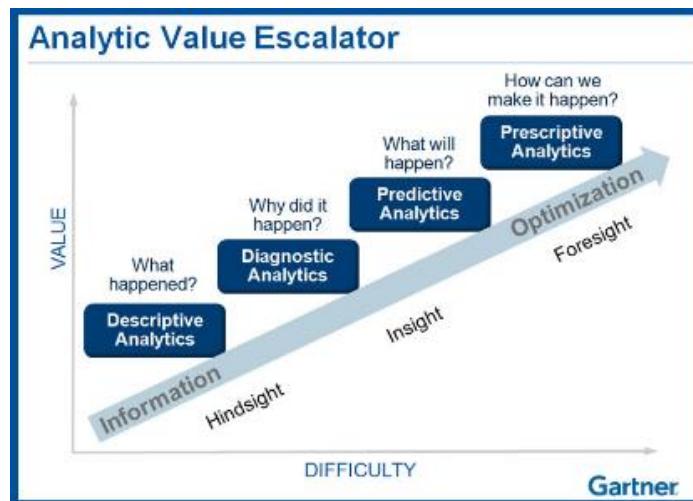


Figura 11. *Analytic value escalator*. Fuente: Pérez, s.f.

En resumen, se puede entender el *business intelligence* como las técnicas de **recoger** y **entender** datos del pasado, mientras que el *business analytics* permite alcanzar una visión más clara del futuro. Ambas metodologías se pueden complementar para construir un análisis minucioso de la actividad y futuro de la empresa, con el propósito de mejorar la toma de decisiones.

	<i>Big data</i>	<i>Business intelligence</i>	<i>Business analytics</i>
Herramientas		Consultas, alertas, reportes, OLAP	Predicción, clasificación, regresión, agrupación
Centro		Qué, cómo paso y qué está pasando	Qué puede pasar
Uso		Reactivo	Proactivo, predictivo
Tipo de datos		Estructurado	Estructurados, no estructurados
Alcance		Dirección	Proceso

Figura 12. Diferencias entre *business intelligence*, *business analytics* y *big data*. Fuente: Curto, 2012.

Por otro lado, el *business analytics* es el análisis de las respuestas proporcionadas por el *business intelligence*. Mientras que el *business intelligence* responde a la pregunta «¿qué sucedió?», el *business analytics* responde a «¿por qué sucedió, volverá a pasar?» El *business intelligence* incluye informes, monitoreo automatizado y alertas, tableros y cuadros de mando integral; el *business analytics*, por el contrario, incluye análisis estadísticos cualitativos y cuantitativos, minería de datos, modelado predictivo y pruebas multivariadas. Cuando escuchas el término «inteligencia empresarial», normalmente engloba todo *business intelligence* y *business analytics*.

### **¿Qué beneficios tiene para la toma de decisiones usar el *business analytics*?**

El conocimiento adquirido en *business analytics* permite a las organizaciones automatizar y optimizar sus procesos.

De hecho, las organizaciones impulsadas por datos que utilizan *business analytics* obtienen una ventaja competitiva porque pueden usar los conocimientos para (Thorlund y Laursen, 2017):

- ▶ Realizar minería de datos (explorar datos para encontrar nuevos patrones y relaciones).
- ▶ Realizar un análisis estadístico cualitativo y cuantitativo para explicar por qué ocurren ciertos resultados.
- ▶ Evaluar decisiones anteriores utilizando diferentes formas de análisis de datos.
- ▶ Utilizar el modelado predictivo y el análisis predictivo para pronosticar resultados futuros.

*Business analytics* también ofrece soporte a las organizaciones en el proceso de

tomar decisiones tácticas proactivas y hace posible que esas organizaciones automaticen la toma de decisiones para respaldar las respuestas en tiempo real.

La construcción básica de *business analytics* se fundamenta en los cuatro pilares de Gartner (Popkin y Hayward, 2004), que se constituyen con cuatro componentes básicos:

- ▶ Los datos.
- ▶ Las personas.
- ▶ Los procesos.
- ▶ La tecnología.

Ayudan a recordar las cuatro áreas clave que se deben tener en cuenta al considerar la implementación del *business analytics*.

La intención detrás de estos pilares, como podemos observar en la Figura 13, es simplemente ayudar a hacer mejores preguntas y obtener mejores respuestas al desarrollar aplicaciones analíticas comerciales.

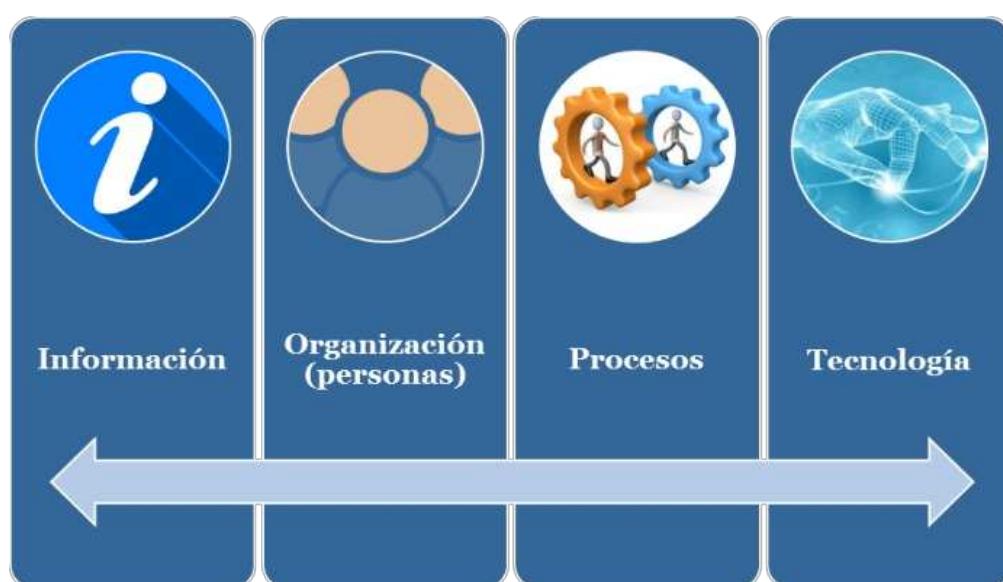


Figura 13. Pilares del *business analytics* según Gartner. Fuente: adaptado de Marrow, 2018.

- ▶ **Información:** el pilar de datos equilibra el manejo de la información. Requiere conectarse a fuentes de datos dispares (bases de datos, sistemas de información de otras empresas y datos en la web, entre otros), independientemente de su tipo y ubicación.

Ser capaz de aprovechar todos los datos disponibles y servirlo al usuario es un componente básico crítico para una estrategia sólida. La información como base fundamental se ha agregado a la estructura para reflejar la conectividad y la coexistencia con todas las fuentes de datos que utiliza el análisis empresarial, no simplemente el almacén de datos. Esto se ha expandido para incorporar datos estructurados y no estructurados (contenido), datos locales y basados en la nube, y hemos visto surgir nuevos términos como *big data* para representar nuevos desafíos de información extrema, no solo de volumen, sino también de velocidad, variedad y complejidad de información.

- ▶ **Personas:** a medida que el *business analytics* se separa de un modelo centralizado hacia un modelo descentralizado, las personas deben estar capacitadas en saber cómo usar los datos. En este marco se ajustan las actividades de las personas para representar tareas en lugar de roles. Anteriormente, los roles eran los que mejor representaban la relación tradicional entre el negocio y las TI, las nuevas formas de análisis superan estas distinciones. Por lo tanto, un usuario de análisis empresarial puede participar fácilmente en producir, consumir y habilitar nuevas actividades que se conviertan en conocimiento.
- ▶ **Procesos:** el pilar del proceso requiere tener la información correcta en el momento adecuado para tomar mejores decisiones y más rápidas. Debido a que diferentes roles toman decisiones diferentes, es importante aprovechar los mismos datos para respaldar una variedad de procesos. Por ejemplo, los usuarios operativos y ejecutivos requieren cuadros de mando; los clientes o ciudadanos requieren declaraciones, propuestas e informes. Esto es para reforzar el punto de que el

marco está compuesto por personas, procesos, plataformas y aspectos de desempeño.

- ▶ **Tecnología:** el pilar tecnológico abarca el desarrollo y la implementación de sistemas que permitan desplegar todo el conocimiento. Las organizaciones deben construir una arquitectura flexible que se adapta a las necesidades del negocio.

### Proceso de *business analytics*

El proceso completo de *business analytics* implica tres pasos principales aplicados secuencialmente a una fuente de datos (véase Figura 14). El resultado del proceso debe estar relacionado con las empresas, intentando mejorar el rendimiento constantemente.

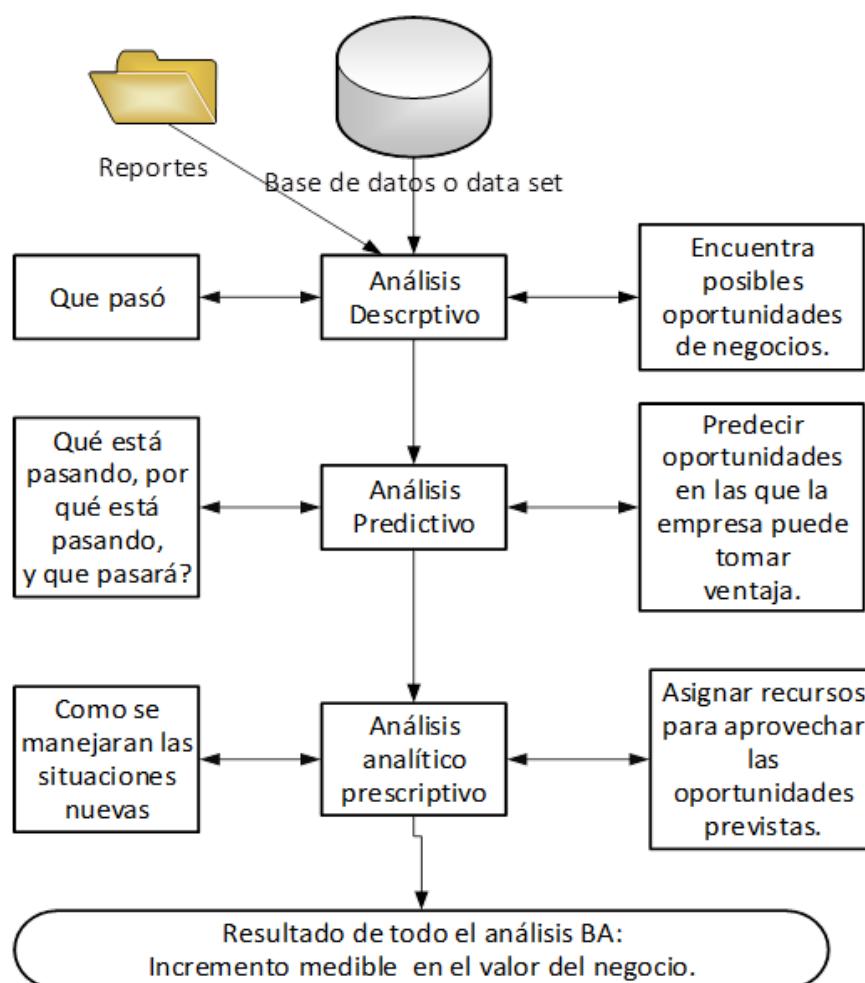


Figura 14. Proceso de *business analytics*. Fuente: Schniederjans et al., 2014.

La lógica del proceso en la Figura 14 se basa inicialmente en una pregunta: ¿qué valiosa información está encerrada en las fuentes de datos que la organización tiene disponibles? En cada uno de los tres pasos que componen el proceso se deben responder las preguntas adicionales. Responder a todas requiere extraer la información de los datos a través de los tres pasos de análisis que comprenden el proceso.

El tamaño de algunas fuentes de datos puede ser inmanejable, demasiado complejo y generalmente confuso. La organización de los datos y el intento de dar sentido a su valor informativo requieren la aplicación de análisis descriptivos como primer paso en el proceso de *business analytics*. Uno puede comenzar simplemente clasificando los datos en grupos usando las cuatro clasificaciones posibles presentadas en la Figura 15.

Tipo de datos	Descripción
Datos categóricos	<p>Datos que están agrupados por una o más características. Los datos categóricos usualmente involucran números cardinales contados o expresados como porcentajes.</p> <p><b>Ejemplo 1:</b> mercados de productos que pueden caracterizarse por categorías de productos «con altos ingresos» o productos «de bajos ingresos», basados en ventas.</p> <p><b>Ejemplo 2:</b> una encuesta donde se recoge información sobre variables como el género, estado civil o afiliación política. Es común usar este término para aplicar a conjuntos de datos que contienen elementos identificados por categorías, así como a observaciones resumidas en tabulaciones cruzadas o tablas de contingencia.</p>
Datos ordinales	<p>Datos clasificados u ordenados para mostrar preferencia relacional.</p> <p><b>Ejemplo 1:</b> clasificaciones de equipos de fútbol no basadas en puntos anotados sino en victorias.</p> <p><b>Ejemplo 2:</b> <i>ranking</i> de empresas comerciales basadas en la calidad del producto.</p> <p><b>Ejemplo 3:</b> <i>ranking</i> de universidades teniendo en cuenta la calidad de sus investigaciones.</p>
Datos de intervalo	<p>Los datos que se organizan a lo largo de una escala donde cada valor es igualmente distante de los demás. Son datos ordinales.</p> <p><b>Ejemplo 1:</b> un indicador de temperatura.</p> <p><b>Ejemplo 2:</b> instrumento de encuesta que usa una escala Likert para medir la satisfacción del cliente (es decir, 1, 2, 3, 4, 5, 6, 7). De 1 a 2 se percibe como equidistante al intervalo de 2 a 3, y así sucesivamente. Nota: En los datos ordinales, la clasificación de las empresas puede variar mucho del primer lugar al segundo, pero en los datos de intervalo, deberían ser relativamente proporcionales.</p>
Ratios de datos	<p>Datos expresados como una relación en una escala continua.</p> <p><b>Ejemplo 1:</b> la proporción de empresas con programas de fabricación ecológica es el doble que la de empresas sin dicho programa.</p>

Figura 15. Tipos de escalas de clasificación de medición de datos. Fuente: Schniederjans et al., 2014.

Además, también se pueden incorporar algunos de los datos en hojas de cálculo como las de Excel y preparar tabulaciones cruzadas y tablas de contingencia para

restringir los datos a una estructura de datos más manejable. Se pueden calcular medidas simples de tendencia central y dispersión para intentar capturar posibles oportunidades de mejora de los procesos. Otros métodos descriptivos de resumen analítico, que incluyen trazado y gráficos, pueden ayudar a los responsables de la toma de decisiones a visualizar los datos para comprender mejor las oportunidades de negocio.

Desde el **paso 1**, el **análisis descriptivo analítico** (véase Figura 14), algunos patrones o variables del comportamiento de la empresa se deben identificar para que representen los objetivos y las oportunidades de negocio, sumado al posible comportamiento futuro de las tendencias. Es probable que se requiera un esfuerzo adicional: la generación de informes estadísticos detallados estrechamente enfocados en los datos y relacionados con los objetivos del negocio para explicar lo que está ocurriendo (lo que sucedió en el pasado).

Esto es como una búsqueda estadística de variables predictivas en los datos que pueden conducir a encontrar patrones de comportamiento que una empresa podría aprovechar, si los patrones de comportamiento ocurren en el futuro. Por ejemplo, una empresa puede hallar en su información general de ventas que, durante los tiempos de inactividad económica, ciertos productos se venden a clientes de un nivel de ingresos específico y con una determinada publicidad. Las variables de ventas, clientes y publicidad pueden tener la forma de cualquiera de las escalas de datos descritas en la Figura 15 (Schniederjans et al., 2014).

Teniendo en cuenta los resultados del paso 1, se pueden determinar tendencias observadas y usarlas para pronosticar el futuro en el paso 2.

**Paso 2: análisis predictivo del proceso de *business analytics*.** Hay muchos métodos que se pueden emplear aquí. Una metodología comúnmente utilizada es la regresión múltiple. Esta es ideal para establecer si existe una relación estadística entre las variables predictivas encontradas en el análisis descriptivo. La relación

podría mostrar que una variable dependiente se asocia de manera predictiva con el valor comercial o el rendimiento de algún producto.

Explorar las bases de datos de la empresa utilizando procedimientos estadísticos avanzados para verificar y confirmar las mejores variables predictivas es una parte importante de este paso en el proceso. Esto responde a las preguntas sobre qué está sucediendo actualmente y por qué sucedió. Un modelo de regresión único o múltiple con frecuencia puede ayudar a pronosticar una línea de tendencia en el futuro. Cuando la regresión no aporta información, se pueden aplicar otros métodos de pronóstico (*exponential smoothing, smoothing averages*) como análisis predictivo para desarrollar los pronósticos necesarios de las tendencias del negocio (Schniederjans et al., 2014).

La **identificación de las tendencias futuras** es el resultado principal del paso 2. Esto ayuda a responder la pregunta «¿qué pasará?».

En el **paso 3, el análisis prescriptivo**, las metodologías de investigación de operaciones se pueden utilizar para asignar de manera óptima los recursos limitados de una empresa y aprovechar al máximo las oportunidades que se encontraron en las tendencias futuras previstas. Los límites en recursos humanos, tecnológicos y financieros impiden que una empresa busque todas las oportunidades que puede tener disponibles en el momento. El uso del análisis prescriptivo le permite a la empresa asignar recursos limitados para alcanzar los objetivos de la mejor manera posible en el menor tiempo posible (Schniederjans et al., 2014).

En resumen, los tres componentes principales del proceso de BA (descriptivo, predictivo y prescriptivo) pueden ayudar a una empresa a encontrar oportunidades en sus datos, predecir tendencias que pronostican oportunidades futuras y ayudar a seleccionar una línea de acción que optimice la distribución de recursos de la empresa para maximizar el valor, el rendimiento y el desempeño.



## 2.6. Referencias bibliográficas

Anderson, C. (23 de junio de 2008). The end of theory: the data deluge makes the scientific method obsolete [Página web].

*Wired.* [http://archive.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory)

Burton, B., Geishecker, L., Hostmann B, Friedman, T., y Newman, D. (2006). *Organizational Structure: Business Intelligence and Information Management*. Gartner.

Casey, T., Krishnamurthy, K., y Abezgauz, B. (12 de agosto de 2013). Who should own big data? [Página web]. *Strategy+Business*. <https://www.strategy-business.com/article/00211?gko=44b8e>

Curto, J. (2012). *Introducción al business intelligence*. UOC.

Dataworks. (s.f.). Integrity in the data lifecycle [Página web]. *Dataworks*. <https://www.dataworks.ie/5-stages-in-the-data-management-lifecycle-process/>

Davenport, T. H., Barth, P., y Bean, R. (2012). How big data is different. *MIT Sloan Management Review*, 54, 22-24.

Debortoli, S., Müller, O., y Vom-Brocke, J. (2014). Comparing Business Intelligence and big data skills. *Business and Information Systems Engineering*, 6(5), 289-300.

Firincan, G. (8 de febrero de 2017). The 10 Vs of Big Data [Página web]. *TWDI*. <https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx>

Gagliardi, E. O., Grosso, A., Turull, J. M., Piffaretti, P., y Pereyra, S. R. (1999). *Computación de queries a bases de datos relacionales utilizando circuitos booleanos* [Conferencia]. V Congreso Argentino de Ciencias de la Computación. San Luis,

Argentina.

Han, J., Kamber, M., y Pei, J. (2006). *Data mining, concepts and techniques*. Morgan Kaufmann.

Herrera, F., Riquelme, J., y Ruiz, R. (2004). Preprocesamiento de datos [Diapositivas]. Reunión Red Nacional DM & ML. <http://www.lsi.us.es/redmidas/Ilreunion/trans/prepro.pdf>

IDS2015. (5 de abril de 2015). Inteligencia de Negocios (BI) [Mensaje en un blog]. *Ingeniería del software UAH*. <https://ingenieriadelssoftwareuah2015.wordpress.com/2015/04/05/inteligencia-de-negocios/>

Kimble, C. (2013). Knowledge management, codification and tacit knowledge. *IR Information Research*, 18(2). <http://informationr.net/ir/18-2/paper577.html>

Lin, T. Y. (2002). Attribute transformation for data mining I: theoretical explorations. *International Journal of Intelligent Systems*, 17, 213-222.

Marrow, G. (2018). *Business analytics & intelligence. An introduction and considerations for getting started*. Durham County Government.

McAfee, A., y Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 61-67.

Pérez, M. F. (s.f.). Transformación digital en la empresa: análisis predictivo [Página web]. *Sistel*. <https://www.sistel.es/transformacion-digital-empresa-analisis-predictivo>

Popkin, J., y Hayward, B. (2004). *Top 10 strategic technologies for 2005*. Gartner Symposium/ITxpo.

Puklavec, B., Oliveira, T., y Popović, A. (2018). Understanding the determinants of business intelligence system adoption stages. *Industrial Management and Data*

*Systems*, 118(1), 236-261.

Pyle, D. (1999). *Data Preparation for data mining (the Morgan Kaufmann series in data management systems)*. Morgan Kaufmann.

Rosado, A. A., y Rico, D. W. (2010). Inteligencia de negocios: estado del arte.

*Scientia et technica*, 16(44), 321-326. <http://revistas.utp.edu.co/index.php/revistaciencia/article/view/1803/1209>

Schniederjans, M. J., Schniederjans, D. G., y Starkey, C. M. (2014). *Business analytics principles, concepts and applications: What, why, and how*. Pearson Education.

Sharma, S., Sharma, J., y Devi, A. (2009). Corporate social responsibility: the key role of human resource management. *Business Intelligence Journal*, 2(1), 205-213. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.514.7758&rep=rep1&type=pdf>

Teixeira, A., Oliveira, T., y Varajão, J. (2019). Evaluation of business intelligence projects success - a case study. *Business Systems Research*, 10(1) 1-12.

Thorlund, J., y Laursen, G. H. N. (2017). *Business analytics for managers*. Wiley.

Trejo, D. (2020). *Gobierno de datos para directores: Realizando la transformación digital a partir de los datos*.

Valenzuela, L. M. (2007). *La gestión del valor de la cartera de clientes y su efecto en el valor global de la empresa: diseño de un modelo explicativo como una herramienta para la toma de decisiones estratégicas de marketing* (Tesis doctoral). Universidad Complutense de Madrid. <http://eprints.ucm.es/8064/1/T29976.pdf>

Weller, K. (2010). *Knowledge representation in the social semantic web*. De Gruyter.

## Introducción al business intelligence

Curto, J. (2012). *Introducción al business intelligence*. UOC.

<https://docplayer.es/709686-Introduccion-al-business-intelligence.html>

El *business intelligence* (o inteligencia de negocio) es un concepto complejo. No por su definición, que es sencilla de enunciar y comprender, sino principalmente por el hecho de que en él confluyen una gran cantidad de tecnologías, metodologías, procesos y estrategias que complican sobremanera la iniciación al neófito. Además, a lo largo del ciclo de vida de estos sistemas de información, se incrementa la complejidad de la arquitectura, así como las necesidades de negocio y las tecnologías que las soportan.

Sin embargo, la gran mayoría de organizaciones necesita actualmente este tipo de sistemas de información para tomar mejores decisiones y ser más competitivas. El *business intelligence* se convierte en una de las principales necesidades. Por ello es necesario poder construir soluciones sólidas a partir de conocimientos profundamente asentados. Este libro introduce los principales conceptos de la inteligencia de negocio a través de las principales fases de diseño de un proyecto de este tipo para constituir una sólida base de adquisición de conocimientos más profundos.

## Business intelligence y business analytics

AddKw. (3 de abril de 2015). *Business Intelligence VS Business Analytics* [Vídeo].

Youtube. [https://www.youtube.com/watch?v=gGtb\\_FpJHTk](https://www.youtube.com/watch?v=gGtb_FpJHTk)

La organización sin ánimo de lucro LPI ADDKW explica de qué forma aplica tanto la inteligencia como la analítica de negocios y por qué estas pueden resultar útiles en cualquier otro organismo, ya sea gubernamental, empresarial, etc.

## Curso introductorio de business intelligence y business analytics

Consultora Re-Ingenia S. A. (23 de mayo de 2018). *Curso Introductorio Business Intelligence y Business Analytics* [Vídeo]. Youtube. [https://www.youtube.com/watch?v=\\_AOic35nLvM](https://www.youtube.com/watch?v=_AOic35nLvM)

La consultora Re-Ingenia S. A. tiene a disposición un curso completo para aplicar los conceptos vistos en este temario. Tal y como afirma: «*business intelligence* es la metodología técnica para transformar los datos en información y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios».

## PowerData

PowerData. Página web oficial. <https://www.powerdata.es/data-governance>

Presenta diferentes artículos en español donde presentan temáticas tales como *data governance*, calidad de los datos, MDM: ¿qué es y cómo debes implementarlo en tu empresa? y *data lake*, entre otros.

- 1.** Elige la respuesta correcta.
  - A. Los datos no tienen relevancia, pues no tienen ningún significado si no se contextualizan.
  - B. Si el dato está bien contextualizado, obtenemos información veraz y confiable.
  - C. El dato es una pieza fundamental en las empresas hoy en día, si los datos cumplen un ciclo de vida y se tienen maestros de datos, se garantiza el éxito en los procesos BI y BA.
  - D. La información sin ser procesada genera conocimiento.
- 2.** El ciclo de vida del dato en el gobierno de datos incluye:
  - A. Creación, uso y destrucción.
  - B. Almacenaje, distribución y archivo.
  - C. Creación, distribución y destrucción.
  - D. A y B son correctas.
- 3.** No forma parte de los pasos importantes para construir un ecosistema BI.
  - A. Entender y tener en cuenta los objetivos estratégicos es fundamental para poder dar respuesta a lo que la organización se traza como meta.
  - B. Evaluar el estado actual y el estado deseado en cuanto a madurez de la organización.
  - C. Definir una estrategia BI, definir una ruta y diseñar y construir las aplicaciones.
  - D. Crear modelos de predicción.

4. ¿Cuáles son los cuatro pilares de Gartner para la construcción básica de business analytics?

  - A. Proceso, clientes.
  - B. Información, personas.
  - C. Procesos, tecnologías.
  - D. B y C son correctas.
5. Menciona dos pasos del proceso de business analytics:

  - A. Predictivo, analítico.
  - B. Descriptivo, transformador.
  - C. Descriptivo, predictivo.
  - D. Ninguna es correcta.
6. ¿Por qué es importante crear datos maestros?

  - A. Evita duplicidad de datos.
  - B. Aumenta la confiabilidad de los datos.
  - C. A y B son correctas.
  - D. A y B son incorrectas.
7. Menciona una diferencia del business analytics frente al business intelligence.

  - A. En el BA se realiza análisis predictivo y prescriptivo, en el BI se analizan datos pasados.
  - B. Uso de bases de datos NoSQL.
  - C. Uso de administración de bases de datos.
  - D. Integración y divulgación.

- 8.** Identifica la afirmación falsa con respecto a los beneficios del business analytics:
- A. Las organizaciones obtienen ventaja competitiva.
  - B. Se puede realizar minería de datos para detectar patrones entre datos.
  - C. Se pueden simular diferentes escenarios y crear estrategias para mitigar riesgos.
  - D. Se puede realizar un análisis descriptivo y diagnóstico de los datos.
- 9.** El análisis que da respuesta a la pregunta de cómo se manejarán futuros escenarios es:
- A. Análisis descriptivo.
  - B. Análisis predictivo.
  - C. Análisis prescriptivo.
  - D. Ninguna es correcta.
- 10.** El business intelligence es:
- A. Reactivo.
  - B. Proactivo.
  - C. Predictivo.
  - D. Ninguno de los anteriores.

Gobierno del Dato y Toma de Decisiones

---

## Tema 3. Data warehouse y data lake

# Índice

[Esquema](#)

[Ideas clave](#)

[3.1. Introducción y objetivos](#)

[3.2. Procesos ETL](#)

[3.3. Almacén de datos \(data warehouse o DW\)](#)

[3.4. Lago de datos \(data lake\)](#)

[3.5. Referencias bibliográficas](#)

[A fondo](#)

[Creando una ETL con las herramientas de Pentaho 6](#)

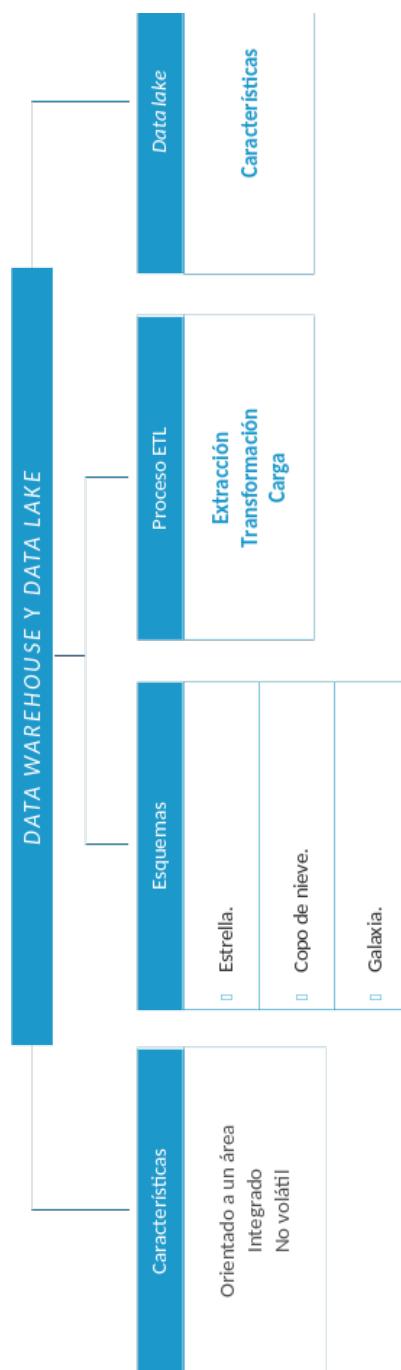
[Desarrollo de un cubo OLAP con Schema Workbench de Pentaho](#)

[Azure data lake storage tutorial](#)

[ETL vs. ELT](#)

[Test](#)

# Esquema



## 3.1. Introducción y objetivos

En el presente tema el estudiante podrá entender el proceso técnico que deben seguir los datos para transformarse de datos brutos a un *data warehouse* o un *data lake*, dependiendo de las necesidades empresariales.

Los objetivos de este tema son:

- ▶ Identificar cada uno de los pasos del proceso ETL: extracción, transformación y carga.
- ▶ Estudiar el concepto de *data warehouse* y diferenciar los tipos de esquemas.
- ▶ Comprender la diferencia entre un *data warehouse* y un *data lake*.

### 3.2. Procesos ETL

Como sus siglas indican, consiste en la extracción, transformación y carga de los datos, de modo que se puede afirmar que es una parte fundamental de este. Antes de guardar los datos, deben ser transformados, limpiados, filtrados y redefinidos. Como se mencionó anteriormente, la información que tienen las empresas en los sistemas no está preparada para la toma de decisiones (Ong et al., 2017).

El proceso de ETL consume entre el 60 y el 80 % del tiempo de un proyecto de *business intelligence*, por lo que es un proceso fundamental en el ciclo de vida del proyecto (Eckerson y White, 2003). Esta parte del proceso de construcción del *data warehouse* (DW) es costosa y consume una parte significativa de todo el proceso, razón por la que utilizan recursos, estrategias, habilidades especializadas y tecnologías. El proceso ETL va más allá del transporte de los datos de las fuentes a la carga dentro del DW, ya que añade un valor significativo a los datos. Una parte del proceso ETL se encarga de (Villanueva, 2011):

- ▶ Eliminar errores y corregir datos faltantes.
- ▶ Proporcionar medidas documentadas de la calidad de los datos.
- ▶ Supervisar el flujo de los datos transaccionales.
- ▶ Ajustar y transformar los datos de múltiples fuentes en uno solo.
- ▶ Organizar los datos para su fácil uso por los usuarios y las herramientas.

El proceso ETL es intuitivo y fácil de entender. La idea fundamental del proceso ETL es tomar los datos de las diferentes fuentes de información y depositarla sin errores en el *data warehouse*. Los procesos de limpieza y transformación de esa información son mucho más complejos de lo que se cree. Se pueden dividir en tareas específicas, dependiendo de las características de las fuentes de datos, los objetivos

de la empresa, las herramientas existentes y las características del DW final.

El desafío para un correcto desarrollo del proceso ETL es **planificar adecuadamente** la cantidad de tareas, para lo cual es preciso conservar la perspectiva sencilla e intuitiva del proceso.

El proceso ETL es obligatorio para acceder a los datos que formarán parte del *data warehouse*. El proceso ETL se divide en cuatro etapas:

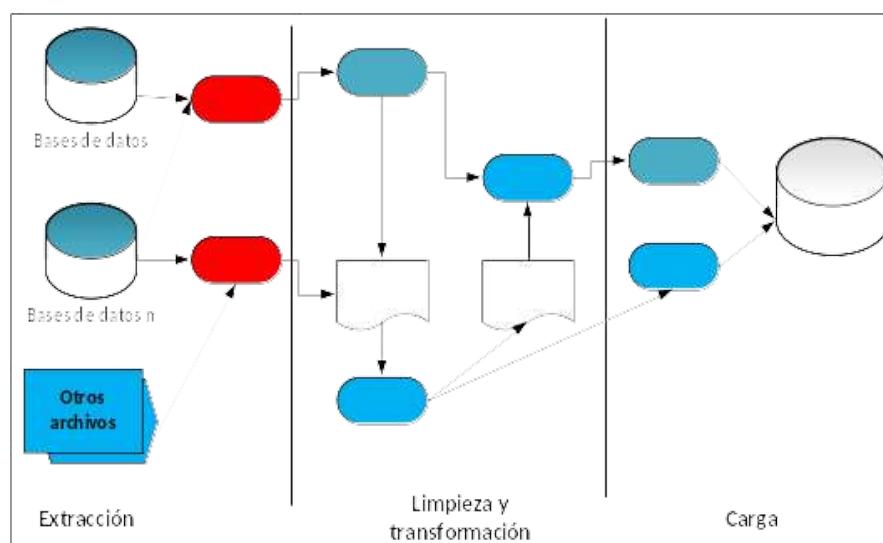


Figura 1. Etapas del proceso ETL.

## Etapas

### Extracción

Este proceso extrae los datos físicamente de las distintas fuentes de información. En este momento los datos están en la forma como se almacenan, en bruto. La extracción de los datos se puede realizar de forma manual o utilizando herramientas de ETL.

Durante el proceso de ETL, una de las primeras tareas que debe realizarse es la extracción de la información más relevante, es generalizar al *data warehouse* (Theodoratos et al., 2001). Para la extracción se pueden usar los siguientes métodos:

1. **La extracción estática**, que tiene lugar cuando el *data warehouse* necesita ser rellenado por primera vez. La detección de cambios se realiza físicamente mediante la comparación de dos imágenes (una correspondiente a la extracción anterior y la otra a la actual).
2. **La extracción incremental**, que es utilizada para actualizar los *data warehouse* de forma regular, aprovecha los cambios aplicados a los datos de origen desde la última extracción.

Finalmente, conviene recordar que el objetivo principal de esta etapa es extraer tan solo aquellos datos de los sistemas transaccionales que son necesarios y prepararlos para el resto de los subprocesos de ETL. Para ello, se deben determinar las mejores fuentes de información y de mejor calidad.

## Limpieza

Este proceso recupera los datos de la base de datos u otro tipo de fuente y comprueba la calidad, elimina los duplicados y, cuando es posible, corrige los valores erróneos y completa los valores incompletos, etc. Ejemplo de algunos errores más comunes:

- ▶ Datos duplicados: un cliente es registrado varias veces en la misma empresa.
- ▶ Inconsistencia en los datos: en la dirección de una persona, el código postal no corresponde a la ciudad donde vive.
- ▶ Inconsistencia de valores: aparece en primer lugar un valor y posteriormente aparece el mismo valor de otra forma. Por ejemplo: primero, escribir el país como USA y, luego, digitarlo completo (Estados Unidos de Norteamérica).

En particular, hay que tener en cuenta que estos tipos de errores son muy frecuentes cuando se manejan múltiples fuentes y se ingresan datos manualmente.

Las principales características de limpieza de datos que se encuentran en las

herramientas de ETL son la rectificación y la homogeneización. Utilizan diccionarios específicos para rectificar errores de digitalización y para reconocer sinónimos, además de la limpieza basada en reglas para imponer normas específicas de dominio y definir asociaciones apropiadas entre valores.

## Transformación

Este proceso recupera los datos limpios y de alta calidad, los organiza y resume en los distintos modelos de análisis. El resultado de este proceso es la obtención de datos limpios, consistentes, resumidos y útiles. La transformación incluye: cambios de formato, sustitución de códigos, valores derivados y agregados.

La transformación es el núcleo del proceso. Convierte los datos de su formato original a un formato de almacén de datos específico. Si se implementa una arquitectura de dos capas, esta fase genera su capa de datos conciliados.

Independientemente de la presencia de una capa de datos conciliados, establecer una correspondencia entre la capa de datos de origen y la de depósito de datos generalmente se dificulta, debido a la presencia de muchas fuentes diferentes y heterogéneas.

Los siguientes puntos deben rectificarse en esta fase:

- ▶ Los textos sueltos pueden ocultar información valiosa. Por ejemplo, Zapatos Zoe LTD no muestra explícitamente que se trata de una sociedad de sociedad limitada, ya que la sigla estándar en España es SL.
- ▶ Se pueden usar diferentes formatos para datos individuales. Por ejemplo, una fecha se puede guardar como una cadena de caracteres o como tres enteros.
- ▶ Seleccionar ciertas columnas para su carga (por ejemplo, que las columnas con valores vacíos no se carguen o se completen).
- ▶ Traducir códigos (por ejemplo, cuando se almacena una «H» para «Hombre» y «M»

para «Mujer», pero luego se cambia a formato numérico: «1» para Hombre y «2» para Mujer). Otro ejemplo: «V» para vivo y «M» para muerto se cambia a «1» para vivo y «0» para muerto.

- ▶ Codificar valores libres, como, por ejemplo: convertir «Hombre» en «1», «Mujer» en «2» o «Niños» en «3».
- ▶ Obtener nuevos valores calculados (por ejemplo, el índice de masa corporal = peso/altura).
- ▶ Calcular totales de múltiples filas de datos (por ejemplo, el total de una población, total de años, etc.).
- ▶ Dividir una columna en varias (por ejemplo, la columna de «Diagnóstico: pasar a tres columnas Diagnóstico\_1, Diagnóstico\_2, Diagnóstico\_3»).
- ▶ Datos erróneos: se pueden corregir o eliminar. Esto va a depender del valor que aporte las variables y los datos al *data warehouse*.

## La carga y actualización

Es la última etapa del proceso y valida que los datos cargados en el DW sean consistentes con las definiciones y formatos; los integra en los distintos modelos de las distintas áreas de negocio que se han definido. Estos procesos suelen ser complejos, por tanto, es necesario tener personal experto que ayude en el proceso. Aquí es esencial comprobar que se ha desarrollado correctamente, ya que, en caso contrario, puede llevar a decisiones erróneas a los usuarios.

Esta etapa es el momento en el que se cargan los datos y se comprueba si los elementos que se cargaron son equivalentes a la información que había en el sistema transaccional, así como los valores que tienen los registros cargados corresponden a los definidos en el *data warehouse*. Es importante comprobar que se ha desarrollado correctamente, ya que, de lo contrario, puede llevar a tomas de

decisiones equivocadas. La carga en un almacén de datos es el último paso para seguir.

La diferencia fundamental entre carga y actualización radica en el hecho de que la carga se realiza cuando el DW está vacío, mientras que la actualización se hace cuando ya existen datos en el mismo. En cualquier caso, tanto la carga como la actualización se pueden llevar a cabo de dos maneras:

- 1. Actualizar datos del almacén de datos completamente reescrito:** esto significa que los datos más antiguos se reemplazan. La actualización se usa normalmente en combinación con la extracción estática para poblar inicialmente un depósito de datos.
- 2. Actualización de datos solo con los cambios aplicados a los datos fuente:** la actualización generalmente se lleva a cabo sin eliminar o modificar datos preexistentes. Esta técnica se usa en combinación con la extracción incremental para actualizar los almacenes de datos regularmente.

### 3.3. Almacén de datos (data warehouse o DW)

A través del *data warehouse*, conocido también como el almacén de datos en el diccionario de datos, se busca almacenar los datos de forma que facilite y maximice su **flexibilidad, facilidad de acceso y administración**. Surge como respuesta a las necesidades de los usuarios que necesitan información consistente, integrada, histórica y preparada para ser analizada y poder tomar decisiones. Al recuperar la información de los distintos sistemas (transaccionales, departamentales o externos) y almacenarlos en un entorno integrado de información diseñado por los usuarios, el *data warehouse* permitirá analizar la información contextualmente y relacionarla dentro de la organización.

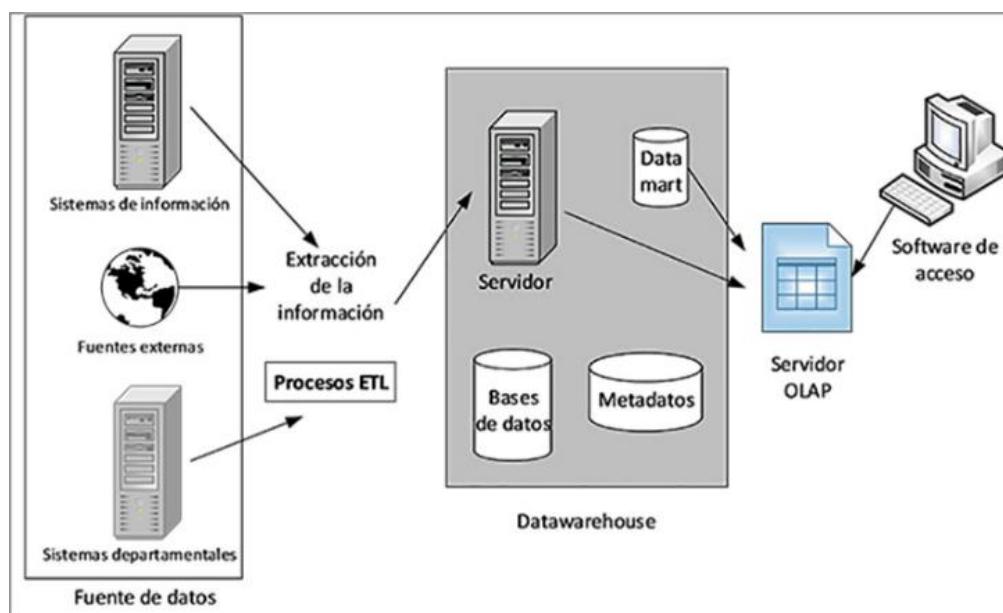


Figura 2. Componentes del *data warehouse*. Fuente: Cano (2007).

#### Fuentes de datos

Se parte de las fuentes para sostener la información del *data warehouse*. Las fuentes de información externas en algunos casos son compradas a otras empresas que gestionan información comercial, encuestas de satisfacción y estudios de mercado,

entre otros. Las fuentes de información externas son esenciales para enriquecer la información que se tiene de los clientes. En otras ocasiones es favorable para la empresa incorporar información como, por ejemplo: la población, el número de habitantes y los presupuestos públicos.

El autor Bill Inmon definió las **características** que debe cumplir un *data warehouse*: debe estar orientado sobre un área, integrado e indexado en el tiempo; es un conjunto no volátil de información que soporta la toma de decisiones (Inmon, 1992).

- ▶ **Orientado a un área:** significa que cada parte del DW está construida para resolver un problema de negocio, que ha sido definido por quienes toman las decisiones. Por ejemplo, entender los hábitos de compra de los adolescentes, analizar la calidad de los productos o analizar la productividad de una línea de producción. Para poder analizar un problema de negocio se necesita información que pueda venir de distintos sistemas: ventas, clientes y elementos de transporte, entre otros.
- ▶ **Integrado:** la información debe ser convertida en medidas comunes, códigos y formatos comunes para que pueda ser útil. La integración permite a las organizaciones implementar la estandarización de conceptos, por ejemplo: la moneda, las fechas, etc.
- ▶ **Indexado en el tiempo:** significa que la información histórica se mantiene y se almacena en determinadas unidades de tiempo, tales como horas, días, semanas, meses, trimestres o años. Ello nos permitirá analizar, por ejemplo, la evolución de las ventas, los inventarios en los períodos que se definan.
- ▶ **No volátil:** esta información no es mantenida por los usuarios, como se realizaría en los entornos transaccionales. La información se almacena para la toma de decisiones. La actualización no se realiza de forma continua, sino periódicamente, como lo define la empresa.

El *data warehouse* debe cumplir con algunos **objetivos**. Ralph Kimball (1996) define los siguientes:

- ▶ Acceder a la información de la empresa o del área funcional.
- ▶ Ser consistente.
- ▶ Separar la información para ser analizada a nivel individual o de manera conjunta.
- ▶ Utilizar herramientas de presentación de la información.
- ▶ Facilitar la publicación de la información.
- ▶ Tener alta calidad para soportar procesos de reutilización.

Los usuarios de negocio necesitan tomar decisiones basadas en la información del DW, por lo que se deben asegurar las siguientes características según Barrer (1998).

- ▶ Alta disponibilidad.
- ▶ Rendimiento.
- ▶ Copias de seguridad y recuperación.
- ▶ Recuperación física en caliente.

## Esquemas de un *data warehouse*

Existen varias estructuras bajo las cuales se construye un DW, las más utilizadas son los modelos estrella y copo de nieve, sus nombres se basan en el dibujo que forman al crearse.

### Esquema estrella

Este modelo es el más sencillo. Está formado por una tabla central de «hechos» y varias «dimensiones», incluida una dimensión de «tiempo». Lo más representativo de la arquitectura de estrella es que solo existe una tabla de dimensiones para cada dimensión. Esto quiere decir que la única tabla que tiene relación con otra es la de hechos, esto es, que toda la información relacionada con una dimensión debe estar

en una sola tabla. En la Figura 3 se observa un ejemplo de este modelo.

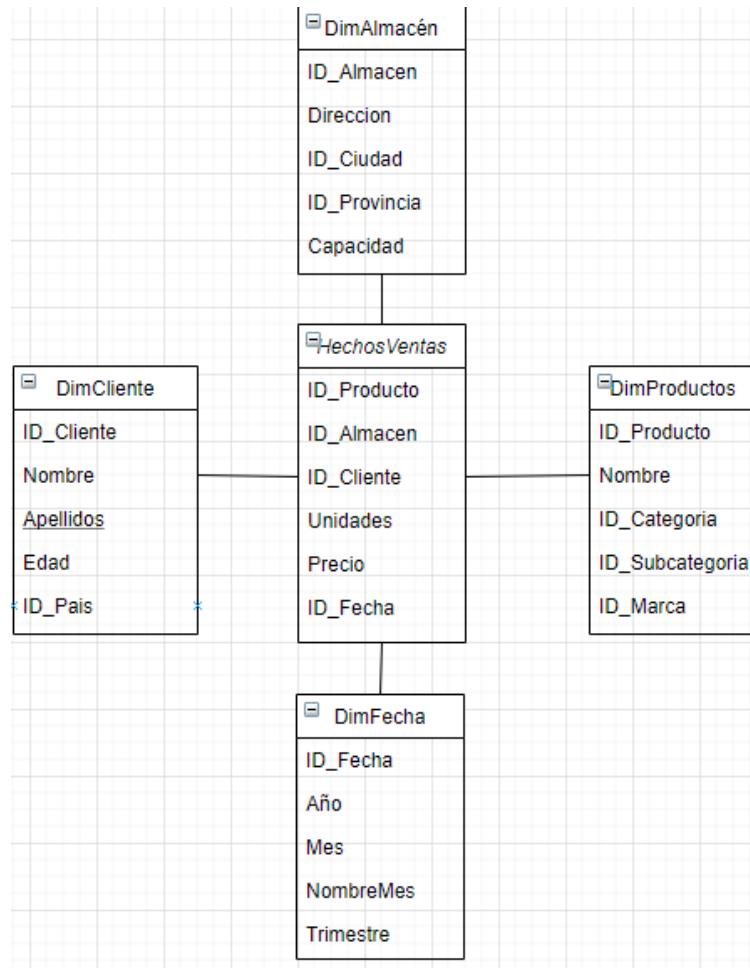


Figura 3. Ejemplo de modelo estrella. Fuente: adaptado de Esquema en estrella, 2021.

En un *data warehouse* de ventas, los hechos son las ventas. En uno financiero, los elementos del balance. En uno de análisis de la bolsa, los hechos serían los conceptos de apertura y precio de cierre. En la tabla de hechos, la clave está conformada por las claves foráneas que apuntan a las dimensiones: ID\_Producto, ID\_Almacen, ID\_Cliente, ID\_Fecha. Es decir, para un almacén, un día, un producto y un cliente, solo puede existir un registro de unidades y precio.

Un modelo estrella es un modelo desnormalizado, ya que lo que se busca es una mejora en el rendimiento de las consultas. Los *join* en las bases de datos

relacionales pueden ser muy pesados.

Ventajas y desventajas de este modelo:

- ▶ Simple y rápido para un análisis multidimensional. Permite consultar datos agregados y detalles.
- ▶ Permite implementar la funcionalidad de los datos multidimensionales y, a la vez, las ventajas de una base de datos relacional.
- ▶ En cuanto a rendimiento es la mejor opción, ya que permite indexar las dimensiones de forma individualizada sin que el rendimiento de la base de datos se vea afectado.

## **Esquema copo de nieve**

Es una variante del modelo anterior. En este modelo la tabla de hechos ya no es la única que se relaciona con otras tablas, existen otras tablas que se relacionan con las dimensiones y que no tienen relación directa con la tabla de hechos. El modelo fue concebido para facilitar el mantenimiento de las dimensiones, sin embargo, esto permite que se vinculen más tablas a las secuencias SQL. Este modelo es complejo de mantener, ya que permite la vinculación de muchas tablas.

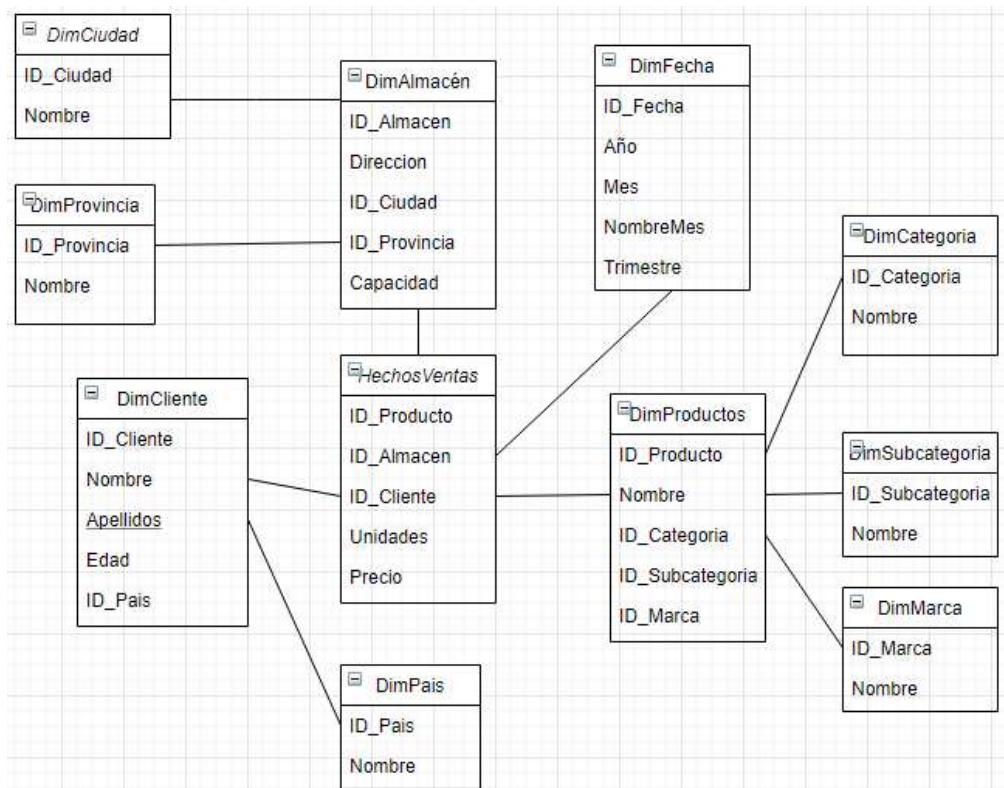


Figura 4. Ejemplo de modelo copo de nieve. Fuente: adaptado de Esquema en copo de nieve, 2020.

Ventajas y desventajas de este modelo:

- ▶ Al estar normalizado se evita la redundancia de datos.
- ▶ El tiempo de respuesta es muy elevado, por lo que, si es necesaria una respuesta rápida y es crítico para el sistema, puede no ser la mejor opción.

Los *data warehouse* se representan normalmente como una gran base de datos, que en algunas ocasiones pueden estar distribuidas en distintas bases de datos, es decir, centralizar toda la información que posee la empresa en un solo sitio, esto permite manejar la información fácilmente (ver Figura 4). El trabajo de construir un DW colectivo puede generar inflexibilidades, o ser costoso y requerir plazos de tiempo elevados.

## Esquema galaxia

Este esquema contiene varias tablas de hechos que comparten dimensiones. Es muy común encontrar este tipo de esquema, incluso es recomendable compartir dimensiones. El esquema se ve como una colección de estrellas, por eso su nombre.

Por ejemplo, pueden existir dos tablas de hechos: inventario y ventas que podrían compartir las dimensiones de producto y fecha.

Veamos algunas arquitecturas:

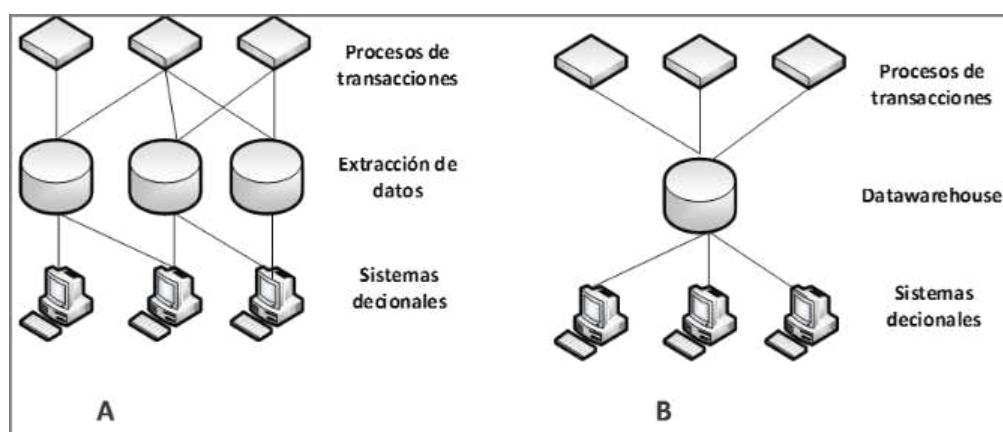


Figura 5. Almacenes de datos antes (A) y después de aplicar *data warehouse* (B).

Fuente: Abella et al. (2000).

## Arquitecturas

Para la realización del *data warehouse* se adoptan dos clasificaciones diferentes para su arquitectura:

- ▶ La primera clasificación está orientada a la estructura y depende del número de capas utilizadas por la arquitectura.
- ▶ La segunda clasificación depende de cómo se empleen las diferentes capas para crear vistas orientadas a los departamentos.

**Arquitectura de una sola capa:** no se utiliza con frecuencia en la práctica. Su objetivo es minimizar la cantidad de datos almacenados. Para alcanzar este objetivo,

se eliminan las redundancias de datos. Esto significa que un almacén de datos se implementa como una **vista multidimensional** de datos operacionales creados por un *middleware* específico o una capa de procesamiento intermedio (Devlin, 1997).

La debilidad de esta arquitectura radica en que no cumple con los requisitos de separación entre procesamiento analítico y transaccional. Las consultas de análisis se envían a los datos operativos después de que el *middleware* los interpreta. De esta manera, las consultas afectan a las cargas de trabajo transaccionales regulares. Además, aunque esta arquitectura puede cumplir los requisitos de integración y exactitud de los datos, no puede registrar más que las fuentes.

Por estas razones, un enfoque de este tipo para los almacenes de datos puede ser exitoso solo si las necesidades de análisis son particularmente restringidas y el volumen de datos a analizar es enorme (Rizzi y Golfarelli, 2009).

**Arquitectura de dos capas:** aunque normalmente se nombra «arquitectura de dos capas» para destacar la separación entre las fuentes físicamente disponibles y los almacenes de datos, en realidad consta de cuatro etapas de flujo de datos posteriores (Hüsemann et al., 2000).

- ▶ **Capa de origen:** es el sistema de almacén de datos que utiliza fuentes heterogéneas de datos. Los datos se guardan originalmente en bases de datos relacionales corporativas o pueden provenir de sistemas de información fuera de los muros corporativos. La prioridad en este tipo de sistema es la actualización y se mantienen pocos datos históricos.
- ▶ **Capa de almacenamiento de datos:** los datos almacenados en las diferentes fuentes deben extraerse, limpiarse para eliminar inconsistencias y llenar espacios, e integrarse para convertirlas en fuentes heterogéneas en un esquema común, proceso ETL. Pueden combinar esquemas heterogéneos, extraer, transformar, limpiar, validar, filtrar, quitar duplicados, archivar y cargar los datos fuente para ser utilizados en el *data warehouse* (Jarke et al., 2013).

- ▶ **Capa de depósito de datos:** la información se almacena en un solo depósito lógicamente centralizado. Se puede acceder directamente al almacén de datos, pero también se puede utilizar como fuente para crear nuevos productos de datos, que replican parcialmente los contenidos del almacén de datos y están diseñados para departamentos empresariales específicos. Los repositorios de metadatos almacenan información sobre fuentes, procedimientos de acceso, usuarios, esquemas de *data mart* (estos y los metadatos se amplían más adelante). Un DW está constituido por la integración de varios *data marts*.
- ▶ **Capa de análisis:** se accede de manera eficiente y flexible a los datos integrados para emitir informes, analizar la información y representar escenarios hipotéticos de negocios (adecuados para cada empresa). Tecnológicamente hablando, aquí se utilizan diferentes herramientas de visualización de datos, optimizadores de consultas para el apoyo para la toma de decisiones.

## Impacto del *data warehouse* (Mendez et al., 2003)

El éxito del *data warehouse* está enfocado en mejorar los procesos empresariales, operacionales y de toma de decisiones. Para que esto funcione se deben tener en cuenta los impactos producidos en los diferentes ámbitos de la empresa:

### Impacto en las personas

La construcción del *data warehouse* requiere de la participación de quienes lo utilizarán, depende de la realidad de la empresa y de las condiciones que existan en el momento de la creación, las cuales determinarán cuál será su contenido.

Como se ha visto, el *data warehouse* provee los datos que posibilitarán a los usuarios acceder a la propia información en el momento en que la necesiten. Para que se realice esta entrega hay que tener en cuenta:

- ▶ Los usuarios deberán adquirir nuevas destrezas; por lo tanto, van a necesitar programas de capacitación adecuados.

- ▶ Los largos tiempos de análisis y programación se reducen para usuarios pertenecientes a las áreas de tecnología, y se reduce también el tiempo de espera para los usuarios de negocio.
- ▶ Como la información estará lista para ser utilizada, es probable que aumenten las expectativas. Se reducirá considerablemente la gran cantidad de reportes en papel.

## **Impactos en los procesos empresariales y de toma de decisiones**

- ▶ Mejora del proceso para la toma de decisiones, ya que facilita la disponibilidad de la información. Las decisiones son tomadas más rápidamente y la gente entiende más del porqué de las decisiones.
- ▶ Los procesos empresariales se optimizan, se elimina el tiempo de espera de la información al encontrarse almacenada en un solo sitio.
- ▶ Se reducen los costos de los procesos, una vez desarrollado el *data warehouse* y en múltiples ocasiones se esclarecen sus conexiones y dependencias, lo que aumenta la eficiencia en dichos procesos.
- ▶ El *data warehouse* permite que los datos de los sistemas sean utilizados y examinados al estar organizados para tener un significado para la empresa.
- ▶ Aumenta la confianza en las decisiones tomadas con base en la información del DW , tanto los responsables de la toma de decisiones como los afectados conocen la información, que tendrá que ser de buena calidad, clara, precisa y concisa.
- ▶ La información que se comparte lleva a un lenguaje común, conocimiento común y mejora de la comunicación en la empresa.

### **Data mart**

El *data warehouse* es una gran estructura. En muchas ocasiones, para facilitar el manejo de los datos, es necesario utilizar estructuras de datos más pequeñas llamadas *data mart* (ver Figura 6). El propósito es ayudar a que un departamento

específico dentro de la empresa pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser resumidos, agrupados y explotados de múltiples formas para diversos grupos de usuarios.

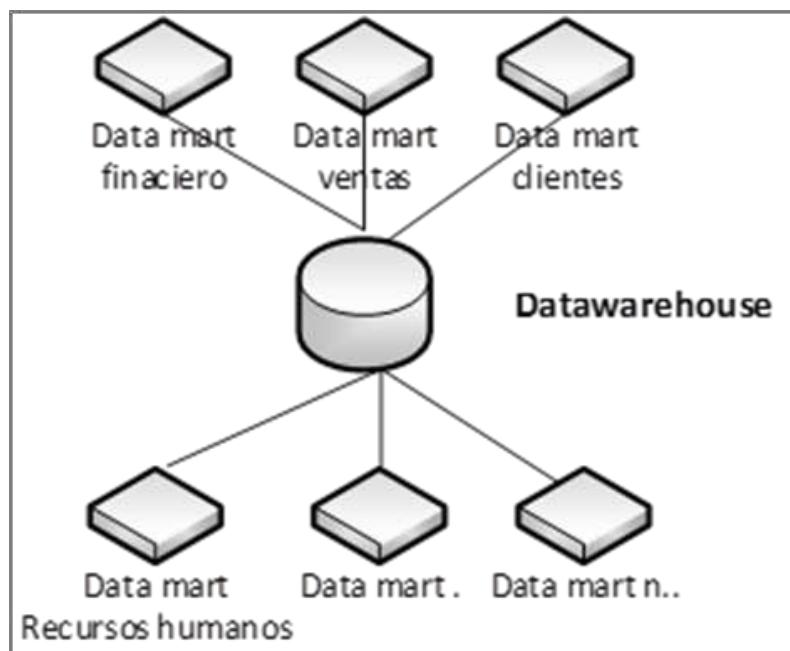


Figura 6. Ejemplo de *data mart*.

Los *data mart* están dirigidos a un conjunto de usuarios dentro de la empresa, que puede estar formado por los miembros de un departamento, por los usuarios de un determinado nivel administrativo o por un grupo de trabajo multidisciplinario con objetivos comunes.

Los *data mart* están compuestos por partes del DW primario, que en algunos casos pueden ser:

- ▶ **Dependientes:** utilizan los datos y metadatos del *data warehouse* directamente en lugar de obtenerlos de los sistemas de producción.
- ▶ **Independientes:** los datos son tomados de cada área de la empresa, siempre manteniendo los datos alineados con el DW, si este existe. Aunque los *data mart* no son estrictamente necesarios, son muy útiles para los sistemas de almacenamiento

de datos en medianas y grandes empresas debido a que:

- Se usan como bloques de construcción mientras se desarrollan depósitos de datos de forma incremental.
- Marcan la información requerida por un grupo específico de usuarios para resolver consultas más rápidas por el menor volumen de datos.
- Pueden ofrecer un mejor rendimiento porque son más pequeños que los *data warehouse* primarios. Por lo tanto, son más fáciles de implementar.
- Al ser pequeños los conjuntos de datos consumen menos recursos.

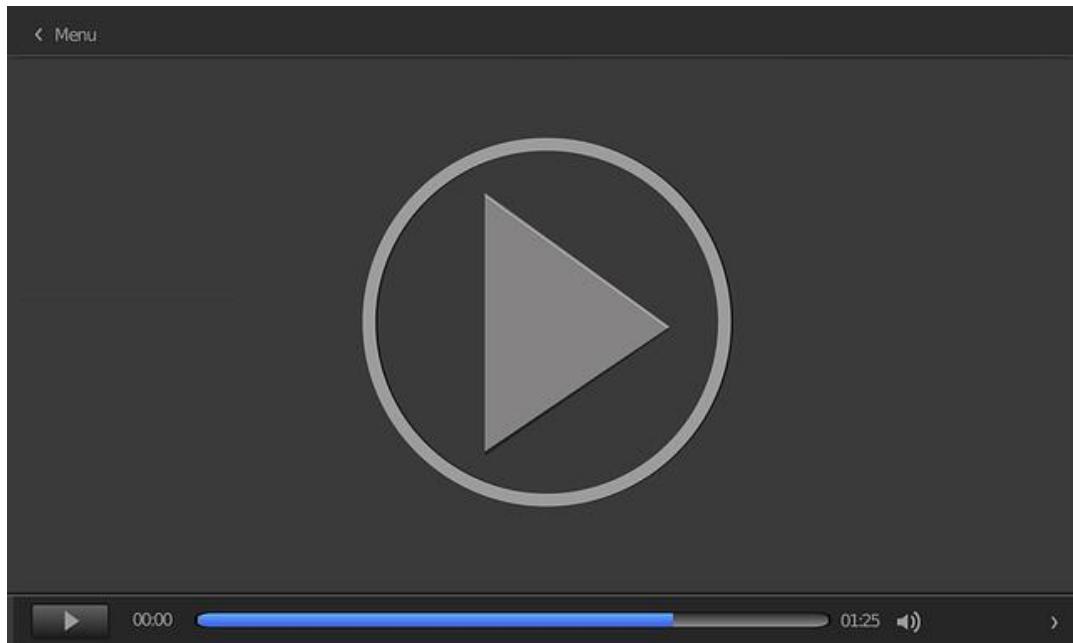
## Los metadatos

Un componente esencial de un *data warehouse* son los metadatos. Es el repositorio central de información que abarca todos los niveles. Da el significado de cada uno de los componentes, variables y atributos que residen en el DW o *data mart*. La información que contiene los metadatos es útil para los departamentos y los propios usuarios. Incluye localizaciones, estructura, definiciones de negocio, descripciones minuciosas de los tipos de datos, significado, formatos, la cantidad y otras características, como los valores máximos y mínimos de los datos. En otras palabras, mapean los datos.

La información más importante va dirigida hacia:

- ▶ **El usuario:** información sobre el significado de los datos utilizados y su localización en el *data warehouse*.
- ▶ **Equipo responsable de los procesos de transformación de los datos:** información sobre la ubicación del dato en los sistemas de producción y los procesos de transformación.
- ▶ **Equipo responsable de los procesos de creación de nuevos datos a partir de los datos detallados.**

A continuación, accede al vídeo *Metadatos*.



---

Accede al vídeo:<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=2fa9796d-6e13-4bb9-bb8a-ad2b000b0dad>

---

## 3.4. Lago de datos (data lake)

Puede definirse como un almacén de datos o un repositorio de grandes cantidades de datos que son útiles para realizar análisis. Los datos se almacenan en una arquitectura plana en lugar de una forma jerárquica, como se hace con los almacenes de datos o DW. Los datos almacenados pueden ser de cualquier tipo: datos estructurados (filas y columnas), semiestructurados (CSV, JSON, XML) y no estructurados (PDF, documentos, fotos, vídeos, correos). Es necesario crear metadatos para poder tener información adicional de cada dato almacenado. Si un lago de datos no proporciona valor para los usuarios o es inaccesible, se denomina pantano de datos.

Es necesario implementar un esquema de lectura para que los científicos y analistas de datos puedan realizar análisis predictivos, descubrir conocimiento y generar herramientas de visualización, entre otros procesos posibles. La transformación de datos se realiza en la etapa en la que se leen los datos.

Cuando se crea un *data lake*, el proceso ETL (extracción, transformación y carga) cambia a ELT (extracción, carga y transformación). Los datos se almacenan sin procesar (Nair, 2018).

En la siguiente tabla se encuentran las diferencias entre ETL y ELT.

	ETL	ELT
<b>Procesamiento</b>	Los datos se transforman en un servidor de almacenamiento intermedio antes de subirse al <i>data warehouse</i> .	Los datos se suben al <i>data lake</i> y allí permanecen. Las transformaciones se realizan en el sistema de destino.
<b>Tiempo de carga</b>	Los datos se cargan en un almacén intermedio y luego se mueven al sistema destino objetivo. Tiempo de carga intensivo.	Los datos se cargan una sola vez. Tiempo de carga muy rápido.
<b>Tiempo de mantenimiento</b>	Altos niveles de mantenimiento.	Bajo nivel de mantenimiento.
<b>Complejidad de implementación</b>	Lo complejo es tener la estructura y los procesos que llenarán esa estructura.	Se debe tener claro qué herramientas se van a utilizar y los <i>skills</i> necesarios.
<b>Madurez</b>	Este proceso se utiliza desde hace más de dos décadas, bien documentado y mejores prácticas disponibles fácilmente.	Es nuevo y complejo de implementar.

Tabla 1. Diferencias entre ETL y ELT. Fuente: adaptado de Ladrero, 2020.

## 3.5. Referencias bibliográficas

Barrer, R. (1998). *Managing a datawarehouse*.

Cano, J. L. (2007). *Business intelligence: competir con información*. ESADE Business School.

[http://itemsweb.esade.edu/biblioteca/archivo/Business\\_Intelligence\\_competir\\_con\\_informacion.pdf](http://itemsweb.esade.edu/biblioteca/archivo/Business_Intelligence_competir_con_informacion.pdf)

Devlin, B. (1997). *Data warehouse: From architecture to implementation*. Addison-Wesley.

Eckerson, W., y White, C. (2003). *Evaluating ETL and Data Integration Platforms*. TDWI Report Series.

Esquema en copo de nieve. (7 de junio de 2020). En Wikipedia. [https://es.wikipedia.org/wiki/Esquema\\_en\\_copo\\_de\\_nieve](https://es.wikipedia.org/wiki/Esquema_en_copo_de_nieve)

Esquema en estrella. (2 de mayo de 2021). En Wikipedia. [https://es.wikipedia.org/wiki/Esquema\\_en\\_estrella](https://es.wikipedia.org/wiki/Esquema_en_estrella)

Hüsemann, B., Lechtenbörger, J., y Vossen, G. (2000). Conceptual Data Warehouse Design. *Proc. of the International Workshop on Design and Management of Data Warehouses*.

Inmon, W. H. (1992). *Building the data warehouse*. Wiley. <https://epdf.pub/building-the-data-warehouse.html>

Jarke, M., Jeusfeld, M. A., Quix, C. J., Vassiliadis, P., y Vassiliou, Y. (2013). Data warehouse architecture and quality: impact and open challenges. En J. Bubenko, J. Krogstie, O. Pastor, B. Pernici, C. Rolland y A. Sølvberg (eds.), *Seminal contributions to information systems engineering* (pp. 183-189). Springer.

Kimball, R. (1996). *The data warehouse toolkit*. Wiley.

Ladrero, I. (12 de noviembre de 2020). ELT o ETL, ¿qué es mejor? [Página web].  
*Baoss*. <https://www.baoss.es/elt-o-ctl-que-es-mejor/>

Mendez, A., Mártire, A., Britos, P. y García-Martínez, R. (2003). Fundamentos de data warehouse. *Reportes técnicos en ingeniería del software*, 5(1), 19-26.

Nair, S., y Poornima, S. (2018). *Data lake: AWS & AZURE data lake, big data solutions & security*.

Ong, T. C., Kahn, M. G., Kwan, B. M., Yamashita, T., Brandt, E., Hosokawa, P., Uhrich, C., y Schilling, L. M. (2017). Dynamic-ETL: a hybrid approach for health data extraction, transformation and loading. *BMC Medical Informatics and Decision Making*, 17, 134.

Rizzi, S., y Golfarelli, M., (2009). *Data warehouse design: modern principles and methodologies*. McGraw-Hill Education.

Theodoratos, D., Ligoudistianos, S., y Sellis, T. (2001). View selection for designing the global data warehouse. *Data & Knowledge Engineering*, 39(3), 219-240.

Villanueva, J. (2011). *Marco de trabajo basado en ontologías para el proceso ETL* (Trabajo Fin de Máster). Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México.

## Creando una ETL con las herramientas de Pentaho 6

Joseph Reyes. (6 de mayo de 2016). *Creando una ETL con las herramientas de Pentaho 6* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=a6nMj6M7IUU&t>

Vídeo tutorial demostrativo para crear una ETL a partir de una base de datos transaccional, tomando como modelo un negocio de tipo tienda.

## Desarrollo de un cubo OLAP con Schema Workbench de Pentaho

Auribox Training. (17 de junio de 2017). *Desarrollando un CUBO OLAP con Schema Workbench de Pentaho / Tutorial* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=eYAgvsT5dd4>

En este vídeo podrás observar paso a paso la creación de un cubo con la herramienta Pentaho, de tipo *open source*, que integra todas las etapas de una estrategia BI.

## Azure data lake storage tutorial

---

Adam Marczak - Azure for Everyone. (12 de diciembre de 2019). *Azure Data Lake Storage (Gen 2) Tutorial / Best storage solution for big data analytics in Azure* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=2uSkjBEwwq0>

---

En este vídeo podrás ver una introducción a lo que sería construir un *data lake* en Azure, cómo trabaja y cómo aprovechar las ventajas de este tipo de almacenamiento en la nube.

## ETL vs. ELT

---

Astera Software. (28 de noviembre de 2019). [WEBINAR]: *ETL vs. ELT: A Data Integration Showdown* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=YOn9hGCwmrA>

---

En este webinar hablan sobre las capacidades de cada uno de estos enfoques, cómo pueden usarse individualmente y combinarlos para un mejor rendimiento.

- 1.** ¿Cuáles son etapas del proceso ETL?
  - A. Extracción.
  - B. Transformación.
  - C. Subida de datos brutos.
  - D. A y B son correctas.
  
- 2.** Los data mart:
  - A. Son los metadatos del data warehouse.
  - B. Son estructuras de datos específicas para un departamento, el conjunto de data marts compone un data warehouse.
  - C. Permiten acceder directamente al data warehouse.
  - D. Son una fuente de datos.
  
- 3.** Son arquitecturas para implementar un data warehouse:
  - A. Arquitectura mecánica.
  - B. Arquitectura de una sola capa.
  - C. Arquitectura de dos capas.
  - D. B y C son correctas.
  
- 4.** ¿Cuáles pueden ser dos posibles fuentes de datos para un data warehouse?
  - A. Bases de datos relacionales y archivos de texto plano.
  - B. Archivos XML y codificación de archivos HTML.
  - C. Archivos PDF y documentos en papel.
  - D. Ninguna de las anteriores.

5. ¿Cuáles pueden ser posibles fuentes de datos para un data lake?

  - A. Bases de datos relacionales y archivos de texto plano.
  - B. Archivos XML y codificación de archivos HTML.
  - C. Archivos PDF y fotos.
  - D. Todos las anteriores.
6. El autor Bill Inmon definió las características que debe cumplir un data warehouse. ¿Cuáles son?

  - A. Orientado a un área e integrado.
  - B. Portátil y fácil de manejar.
  - C. Indexado en el tiempo y no volátil.
  - D. Consistente y fácil.
7. ¿Cuál es la función del data warehouse y del data lake?

  - A. Aumentar el trabajo de los usuarios.
  - B. Ayudar en la toma de decisiones.
  - C. Centralizar los datos para facilitar el manejo.
  - D. Ninguna de las anteriores.
8. Es falso si hablamos de ETL:

  - A. Los datos se transforman en un servidor intermedio antes de subir al DW.
  - B. El tiempo de carga, sobre todo la primera vez, es muy rápido.
  - C. Altos niveles de mantenimiento.
  - D. Las estructuras pueden llegar a ser complejas.

**9.** Es falso si hablamos de ELT:

- A. Los datos se cargan y se transforman en un servidor intermedio antes de subir al DW.
- B. El tiempo de carga es muy rápido.
- C. Bajo nivel de mantenimiento.
- D. Es nuevo y complejo de implementar.

**10.** Es cierto si hablamos de metadatos:

- A. Son un repositorio central de información.
- B. Da significado a cada componente, variable y atributo que reside en el DW.
- C. Contiene información sobre la estructura del data lake.
- D. A y B son verdaderos.

Gobierno del Dato y Toma de Decisiones

---

## Tema 4. Metodologías y tendencias

# Índice

## Esquema

### Ideas clave

- 4.1. Introducción y objetivos
- 4.2. Metodología Kimball
- 4.3. Metodologías PMI
- 4.4. Metodología Inmon
- 4.5. Data-driven decision modelling
- 4.6. Metodología DevOps
- 4.7. Nuevos roles
- 4.8. Tendencias
- 4.9. Referencias bibliográficas

### A fondo

Metodología de Ralph Kimball para la implementación de DW/BI

El concepto de DevOps

Business intelligence governance best practices for 2021 and beyond

## Test

## METODOLOGÍAS Y TENDENCIAS

<b>Metodología Kimball</b>	Principios básicos del ciclo. Rutas de la metodología.
<b>Metodología PMI</b>	Ejecución de procesos. Áreas de conocimiento.
<b>Metodología Inmon</b>	Características del DW. Se debe tener presente.
<b>Data-driven</b>	Mejora de proceso de toma de decisiones.
<b>Tendencias</b>	Analisis de las últimas tendencias según los expertos en BI y gobierno del dato.

## 4.1. Introducción y objetivos

El tema de las metodologías es quizá el más importante, ya que, si no se tiene en cuenta los pasos que se deben seguir para implementar cualquier estrategia BI o BA, el proyecto tiende al fracaso. En este apartado presentamos las metodologías que han sido el cimiento de este tipo de estrategias y algunas que se han adaptado con el tiempo y se consideran más modernas.

Los objetivos de este tema son:

- ▶ Conocer las metodologías clásicas del BI, la de Inmon y la de Kimball, reconocer sus ventajas y desventajas.
- ▶ Utilizar metodologías aplicadas a otras temáticas y aplicarlas a la estrategia BI
- ▶ Profundizar en lo que es una empresa *data-driven*.
- ▶ Analizar las tendencias en el campo del BI y del gobierno del dato.
- ▶ Descubrir los nuevos roles que se están creando para poder afrontar los nuevos retos.

## 4.2. Metodología Kimball

Existen varias metodologías de diseño y construcción del *data warehouse* (DW). Cada productor de *software* de *business intelligence* intenta aplicar una metodología con sus productos. Las metodologías de Kimball y de Inmon son las más utilizadas actualmente. Para entender la diferencia entre estas dos metodologías, volvamos a los conceptos de **data warehouse** y de **data mart**. El segundo es un repositorio de información, similar a un DW, pero más pequeño y orientado a un área o departamento específico de la organización (por ejemplo: Compras, Ventas o RR. HH.), a diferencia del DW, que cubre toda la organización y su alcance (Kimball et al., 1998).

La metodología de Kimball se centra principalmente en el diseño de la base de datos que guardará la información para la toma de decisiones. El diseño se basa en la creación de tablas de hechos, que contienen la información numérica de los indicadores por analizar, es decir, la parte cuantitativa de la información.

La aplicación de cualquier metodología para el desarrollo del DW en las empresas siempre va a depender de las necesidades de las organizaciones y el compromiso institucional de quienes conforman dichas organizaciones.

La metodología se basa en lo que Kimball denomina ciclo de vida dimensional del negocio —*business dimensional lifecycle*— (Kimball et al., 1998; Mundy et al., 2006).

El ciclo de vida del proyecto de DW está fundamentado en cuatro principios básicos:



Figura 1. Principios básicos para el desarrollo del DW, según la metodología Kimball. Fuente: Kimball et al., 1998.

- 1. Centrarse en el negocio:** es necesario enfocarse hacia la tipificación de los requerimientos del negocio y el valor que le aportará, emplear estos esfuerzos en desarrollar relaciones sólidas con el negocio y con los encargados de implementar las herramientas.
- 2. Realizar una infraestructura de información:** se trata de diseñar una base de información única, integrada, fácil de usar y de alto rendimiento, en la cual deben verse reflejados todos los requerimientos de negocio identificados en la organización.
- 3. Cumplir las entregas en incrementos significativos:** define los tiempos de creación del almacén de datos en incrementos progresivos entregables. Los plazos están determinados por los desarrolladores del proyecto y estos pueden ser cuatro o seis meses cada entrega (ten en cuenta que estos son tan solo ejemplos, los tiempos de entrega se deben ajustar a las necesidades de la empresa).
- 4. Ofrecer la solución completa:** facilitar todos los elementos necesarios para entregar valor a todos los usuarios del negocio. Para empezar, se debe contar con un almacén de datos sólido, bien diseñado, con calidad probada y accesible. También se deberá entregar herramientas de consulta *ad hoc*, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación

(Rivadera, 2010).

La construcción de una solución de *data warehouse/business intelligence* no es una tarea fácil; sin embargo, Kimball propone una metodología que ayuda a facilitar esa complejidad. Las tareas de esta metodología (ciclo de vida) se muestran en la Figura 2.

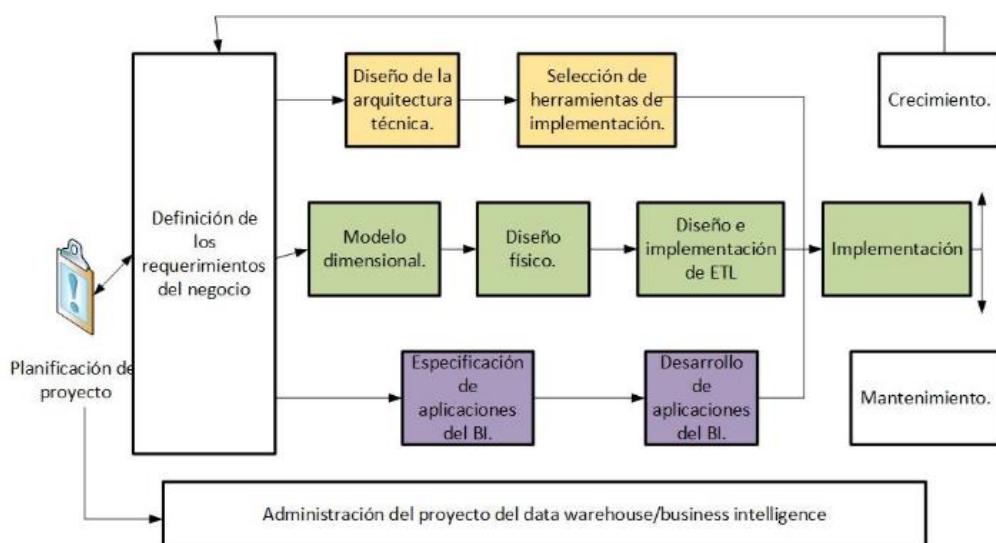


Figura 2. Tareas de la metodología de Kimball, denominada business dimensional lifecycle. Fuentes: Kimball et al., 1998; Mundy et al., 2006.

En la Figura 2, se observa que hay que destacar el papel central de la tarea de definición de requerimientos. Los requerimientos del negocio son la base inicial de las tareas siguientes. En segundo lugar, se observan tres rutas que se enfocan en tres áreas diferentes:

- ▶ **Tecnología (ruta superior):** envuelve tareas relacionadas con herramientas software específicas para cada empresa, por ejemplo: *data integration* —Kettle — Pentaho Open Source.
- ▶ **Datos (ruta interior):** se diseñará e implementará el modelo dimensional y desarrollo del subsistema de ETL (extracción, transformación y carga) para cargar el *data warehouse*.

- ▶ **Aplicaciones del *business intelligence* (ruta inferior):** en esta ruta se centran las tareas para el diseño, desarrollo e implementación de las aplicaciones de negocio para los usuarios finales.

Estas rutas se van mezclando a medida que se va realizando el proyecto hasta llegar a la instalación del sistema. En la base de la Figura 2, se observa la actividad general de administración del proyecto. A continuación, se amplía la información de cada una de las tareas.

## Planificación del proyecto

En esta parte se establece el propósito del proyecto de *data warehouse/business intelligence*, los objetivos específicos, el alcance del mismo, los riesgos y un acercamiento a las necesidades de información. Para Kimball et al. (1998), el proyecto hace referencia a una iteración simple del KLC (*Kimball Life Cycle*), desde el principio hasta el despliegue. Esta tarea incluye estas fases comunes para un plan de proyecto:

- ▶ Definir lo más certero posible el alcance (requerimientos del negocio).
- ▶ Identificar las tareas.
- ▶ Programar las tareas.
- ▶ Planear el uso de los recursos.
- ▶ Establecer la carga de trabajo a los diferentes recursos.
- ▶ Preparar una prueba piloto.
- ▶ Preparar y elaborar el documento final (informes).

En el apartado de la planeación de los recursos humanos, Kimball propone la conformación de un equipo de trabajo al que pertenece tanto personal del área de informática como personal directivo (expertos en el negocio). En la Tabla 1 se

muestran algunos ejemplos.

RECURSOS HUMANOS	
Negocios	Tecnologías de la información
Cliente final	Arquitecto técnico
Gerente	Modelador de datos
Líder	Coordinador de <i>data marts</i> y metadatos
Usuarios	Administrador de bases de datos
Expertos de áreas	Soporte del DW

Tabla 1. Ejemplo de recursos humanos según metodología Kimball. Fuente: Kimball et al., 1998.

## Análisis de requerimientos

Para poder conseguir los requerimientos del negocio, se debe planear cómo obtendremos dichos requerimientos, y para ello existen dos técnicas principales:

- ▶ Las entrevistas.
- ▶ Las sesiones facilitadoras.

Las entrevistas deben realizarse a los expertos en el negocio (**directivos**) y a los **técnicos**. Para ello es preciso tener una preparación previa con ayuda de un experto. De este proceso es imperativo aprender todo sobre el negocio, la competencia, la industria y los clientes (internos y externos), así como conocer las pautas y la terminología. A partir de este análisis se puede construir una herramienta, denominada **matriz de procesos/dimensiones**, *bus matrix* en inglés (Rivadera, 2010).

La dimensión es una **forma** o **criterio** por medio del cual se pueden resumir, cruzar o cortar datos numéricos para ser analizados. La matriz generada contiene en las filas los procesos de negocios detallados y, en las columnas, las dimensiones reconocidas. Un ejemplo de esta matriz de ventas se puede observar en la Tabla 2:

Dimensiones							
Hechos		Fechas	Artículo	Almacenes	Empleados	Clientes	Proveedores
	Ventas	X	X	X	X	X	
	Inventarios	X	X	X			
	Compras	X	X		X		X
	Facturación	X	X		X		

Tabla 2. Ejemplo de matriz hecho/dimensiones (*bus matrix*). Fuente: Zorrilla, 2011.

## Modelo dimensional

La creación de un modelo dimensional es un proceso eficiente y muy iterativo. En la Figura 3 se aprecia un esquema general.

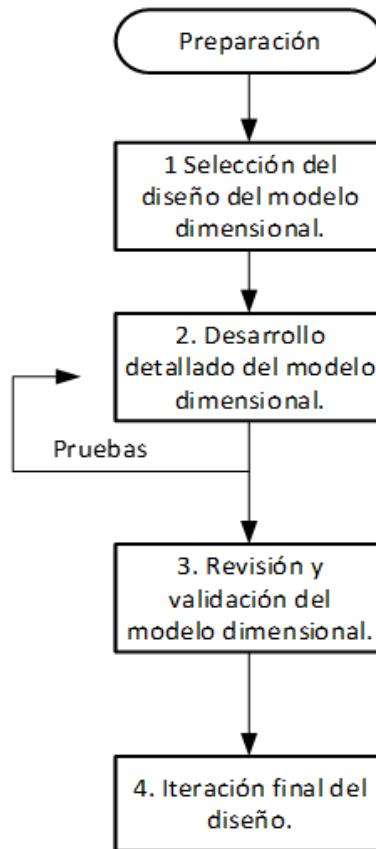


Figura 3. Diagrama de flujo del proceso dimensional de la metodología de Kimball, denominada *business dimensional lifecycle*. Fuentes: Kimball et al., 1998; Mundy et al., 2006.

El proceso de diseño empieza con un modelo dimensional de alto nivel que se realiza a partir de los procesos anticipados de la matriz descrita en el punto anterior. El proceso de pruebas consiste en cuatro pasos (Rivadera, 2010):

- ▶ Elegir el proceso de negocio: acción, tarea para ser modelada.
- ▶ Establecer el nivel de granularidad: especificar al máximo nivel de detalle.
- ▶ Elegir las dimensiones: estas nacen de las reuniones de los equipos de trabajo.
- ▶ Identificar medidas y las tablas de hechos: son los datos cuantificables del negocio que se desean analizar.

En la siguiente parte de la sesión inicial de diseño se completan las tablas con una lista de atributos bien definida. Esta se forma colocando en las filas los atributos de la tabla y en las columnas, la siguiente información: características relacionadas con la futura tabla dimensional del DW, origen de los datos y definición de los procesos ETL.

Una vez definidos los atributos y variables que van a conformar la **matriz**, se procede con la implantación del modelo dimensional. Se realizan pruebas de este y, si se presentan problemas, se hace una revisión y ajuste del modelo en caso de que sea necesario. En caso contrario, se procede a la validación del diseño.

### Diseño físico

El modelado dimensional es convertido en un modelo físico y se deben tener en cuenta aspectos técnicos de *hardware* (memorias, procesadores), *software* y redes (tarjetas de red, tipos de redes, etc.). Además, hay que considerar aspectos como el tamaño del sistema *data warehouse* que se va a implementar, la configuración del sistema y las definiciones de la complejidad del modelo y del tipo de usuario que va a acceder el *data warehouse*.

### Diseño del sistema de extracción, transformación y carga (ETL)

El sistema de ETL es la base sobre la cual se construye el *data warehouse*. Crear un sistema ETL que permita extraer los datos de los sistemas originales (bases de datos, tablas simples, etc.), aplicar las diferentes reglas para mejorar su calidad y consistencia, consolidar la información proveniente de distintos sistemas y, finalmente, cargar la información en el *data warehouse* en un formato adecuado (definido previamente) para la utilización por parte de las diferentes herramientas de análisis, adecuadas a las empresas (Rivadera, 2010).

Una vez finalizado el proceso ETL, se procede a la implantación del *data warehouse* en la organización. Se debe destacar que este puede ir sufriendo modificaciones

después de ser implantado; esto se debe a que, una vez interactúa con los usuarios, van surgiendo mejoras o ajustes.

Para tener una base sólida sobre cómo se ha llevado a cabo cada etapa del proceso del DW, es necesario ir documentando todo el proyecto en avances cortos para construir un informe final completo que sirva como medio de consulta y, además, como estándar para futuros *data warehouse*.

## 4.3. Metodologías PMI

Actualmente, existen varias metodologías para la dirección de los proyectos que también son aplicables a proyectos *business intelligence* y se adaptan a las necesidades de la empresa. Las más importantes son Scrum y PMI (siglas de Project Management Institute).

El PMI es una asociación profesional de las más grandes del mundo que cuenta con medio millón de miembros e individuos titulares de sus certificaciones en 180 países (Vahos et al., 2013).

La metodología del PMI de gestión de proyectos se compone de dos elementos esenciales: la ejecución de procesos y las áreas de conocimiento.

### Proceso

Para la Guía PMBOK (*A guide to the Project Management Body of Knowledge*), no puede hablarse de proyecto si este no se concibe como un proceso, es decir, una serie de actividades coordinadas e interrelacionadas entre sí que deben ejecutarse con un fin específico. No importa si son muchas o pocas las etapas que componen un proceso. Lo más importante es que este tenga tal entidad. Por supuesto, el número de etapas varía en función de las exigencias de cada caso: participantes, complejidad de las tareas y plazos de entrega, entre otros. Lo realmente decisivo en la ejecución de un proyecto es tener despejadas las etapas básicas que lo conforman, que para el PMI son, según la PMBOK:

- ▶ Inicialización.
- ▶ Planificación.
- ▶ Ejecución.
- ▶ Control.

- ▶ Cierre.

En la Figura 4 se observa la interacción entre las diferentes etapas de la metodología PMI:

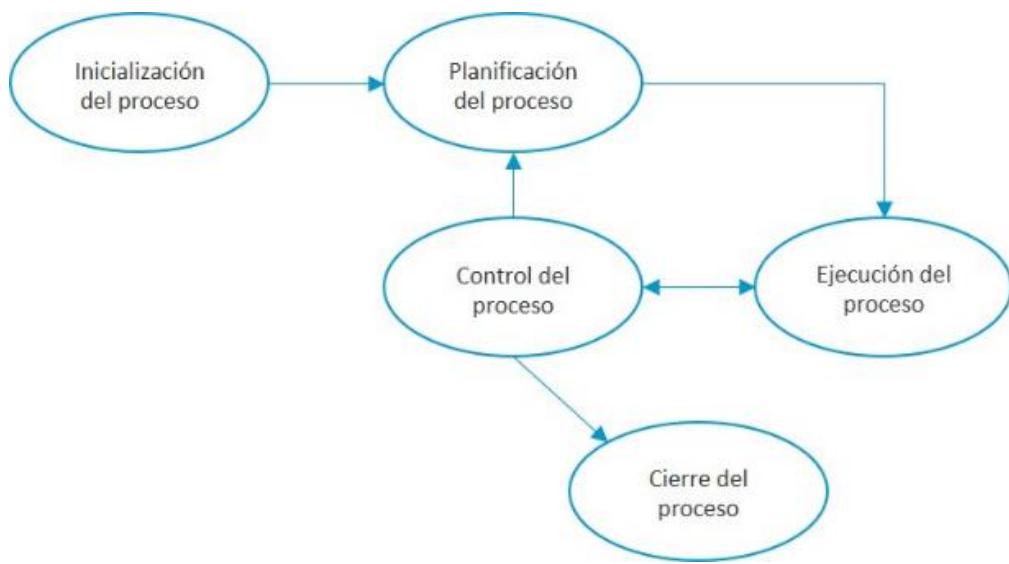


Figura 4. Etapas de la metodología PMI. Fuente: PMBOK Guide.

En el enfoque del PMI, el proceso del proyecto es realizado por personas y generalmente está en una de dos categorías:

- ▶ **Procesos de gestión de proyectos:** describe, organiza y completa el trabajo del proyecto.
- ▶ **Procesos orientados al producto:** especifica y crea el producto del proyecto, que generalmente está definido por el ciclo de vida del mismo (PMBOK).

## Áreas de conocimiento

Además de los conocimientos específicos de los diferentes sectores, los líderes de proyectos deben aplicar otro tipo de conocimientos adicionales, los cuales están relacionados con competencias específicas de la gestión. Ya no es suficiente con ser un especialista en el área en la que nos desenvolvemos; hace falta adquirir

competencias transversales para realizar un mejor desempeño del trabajo. Para el PMI, las áreas de conocimiento adicionales que no pueden faltar a la hora de gestionar un proyecto son:

- ▶ Integración.
- ▶ Recursos humanos.
- ▶ Costes.
- ▶ Alcance.
- ▶ Tiempo o plazos.
- ▶ Calidad de las tareas.
- ▶ Comunicación.
- ▶ Riesgos.
- ▶ Adquisiciones del proyecto.

## 4.4. Metodología Inmon

Bill Inmon es considerado el padre del concepto del *data warehouse* y menciona que debe cumplir con las características que se muestran en la Figura 5.

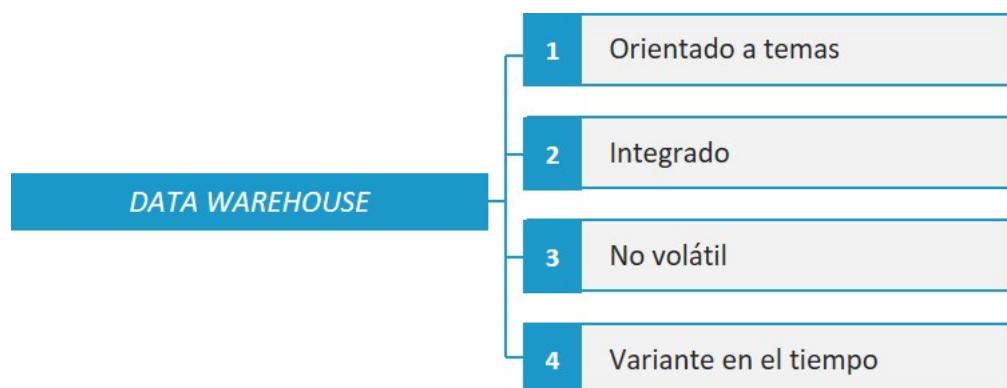


Figura 5. Características del DW según Inmon. Fuente: Inmon, 1992.

La metodología Inmon también es conocida como *top down*. Los datos son extraídos de las fuentes usando los procesos ETL y posteriormente son cargados en las áreas *stage*. Aquí son validados y agrupados en el *data warehouse* global, donde se encuentran los metadatos que contiene toda la información del *data warehouse*, es decir, los nombres de los atributos, los valores máximos y mínimos y los datos eliminados, entre otros. Una vez realizado este proceso, se actualizan los *data mart* con las respectivas transformaciones. Tener una orientación global suele ser más difícil de desarrollar e implementar.

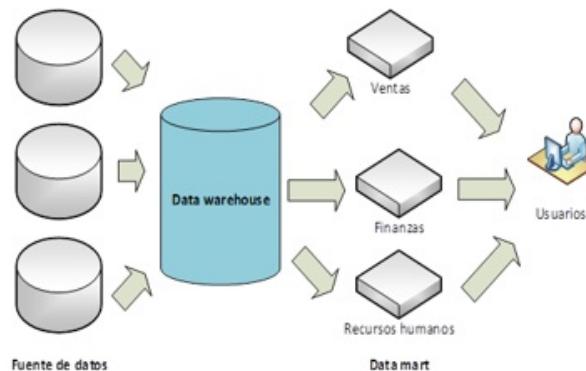


Figura 6. Metodología Inmon. Fuente: Inmon, 2002.

### Ciclo de vida Inmon

El ciclo de vida para el desarrollo del *data warehouse* comienza con los datos de las diferentes fuentes, los cuales están integrados y probados. Posteriormente, se desarrollan herramientas para leer dichos datos y, por último, se definen los requisitos para los sistemas de decisiones.

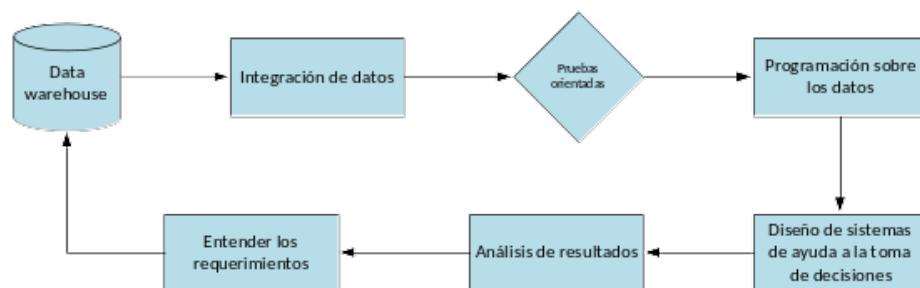


Figura 7. Ciclo de vida del *data warehouse*. Fuente: Inmon, 2002.

### Estructura de un *data warehouse*

La estructura propuesta por Inmon para este sistema está dividida en cuatro niveles que se muestran en la Figura 8:

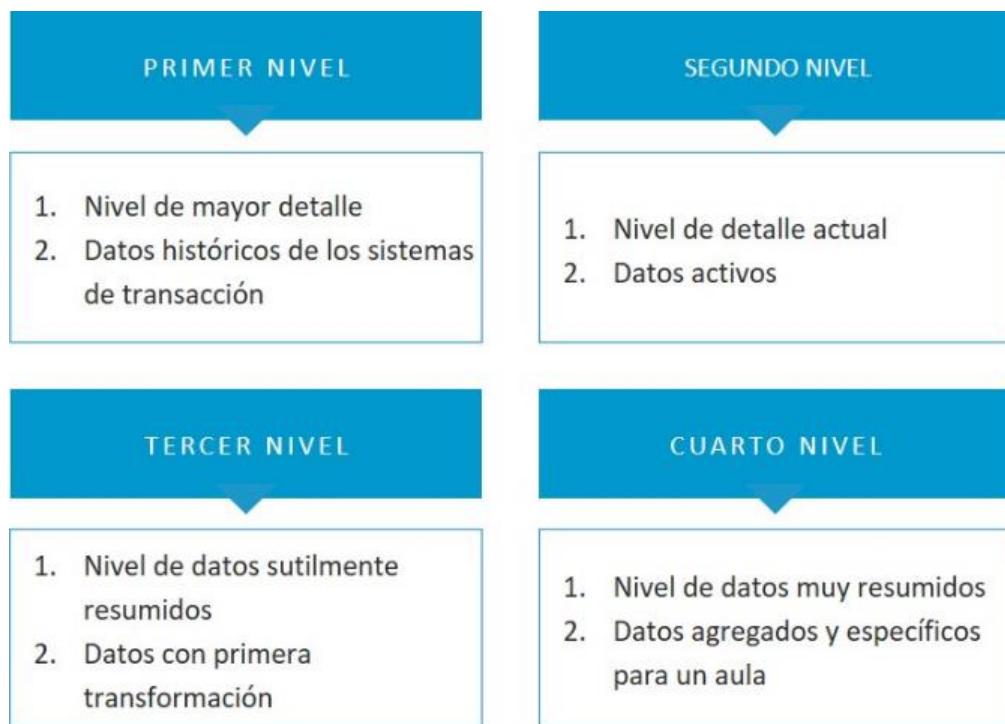


Figura 8. Estructura del *data warehouse*. Fuente: Inmon, 2002.

En la siguiente figura se muestran los cuatro niveles:

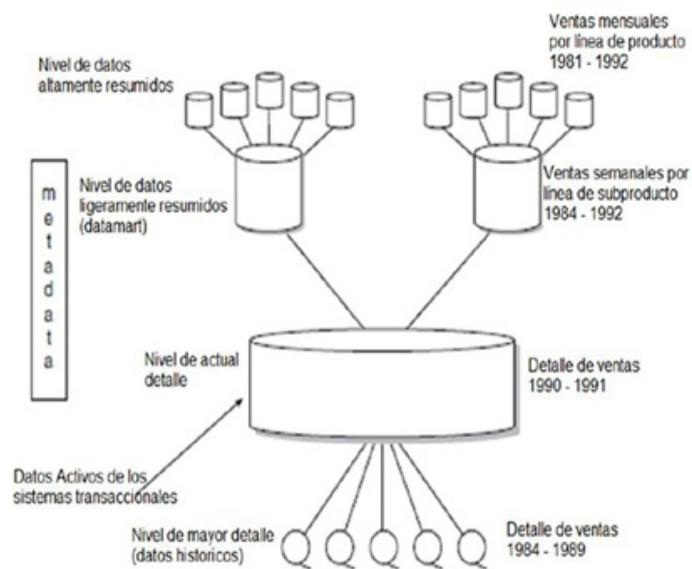


Figura 9. Estructura de los niveles del DW. Fuente: Inmon, 2002.

## Proceso para la construcción del DW

Inmon propone los siguientes pasos para la construcción del DW.

1. Reconocer los sistemas transaccionales de la empresa que van a servir como fuentes de datos del DW.
2. Se empiezan a llenar las primeras tablas en el DW de las respectivas unidades del negocio. Los usuarios empiezan a acceder a los datos integrados.
3. Se cargan más tablas al DW y aumentan también los usuarios que acceden al mismo.
4. El DW es cargado con los datos correctamente, esto trae como resultado la aparición de sistemas de apoyo a la toma de decisiones.
5. Se crean los *data mart* para cada unidad de negocio.
6. Si surgen nuevas necesidades, deben crearse nuevos *data mart*.
7. Después de todos estos pasos, finalmente la arquitectura está desarrollada. En algunas ocasiones los usuarios prefieren acceder a los *data mart*, ya que su acceso es más rápido.

## Granularidad

El aspecto más importante del diseño de un almacén de datos es la granularidad. De hecho, el problema de la granularidad afecta a la arquitectura que rodea el entorno del almacén de datos. La granularidad se refiere al **nivel de detalle o resumen de las unidades de datos en el DW**. Cuantos más detalles hay, menor es el nivel de granularidad. Cuantos menos, el nivel es mayor (Inmon, 2002).

La granularidad de los datos siempre ha sido un problema importante del diseño. Hace algunos años, la granularidad se daba por sentada. Cuando los datos detallados están siendo actualizados, es casi un hecho que los datos se almacenan

en el nivel más bajo de granularidad.

La granularidad es el principal problema de diseño en el entorno del almacén de datos porque afecta profundamente el volumen de datos que reside en el DW y el tipo de consulta que puede ser respondida. El volumen de datos en un almacén se intercambia con el nivel de detalle de una consulta. En casi todos los casos, los datos llegan al almacén de datos a un nivel de granularidad demasiado alto. Esto significa que el desarrollador debe gastar muchos recursos al separar los datos.

En ocasiones, los datos ingresan al almacén a un nivel de granularidad demasiado bajo. Un ejemplo de este nivel son los datos de registro web generados por el entorno del comercio electrónico. Los datos de la secuencia de clics del registro web se deben editar, filtrar y resumir antes de que su granularidad sea adecuada para el entorno del DW (Inmon, 2002).

Los beneficios de la granularidad son:

- ▶ Después de que el almacén de datos se haya construido correctamente, proporciona a la organización una base extremadamente flexible y reutilizable para diferentes procesos de sistemas de apoyo a la toma de decisiones.
- ▶ Los datos granulares que se encuentran en el almacén de datos son la clave para la reutilización, ya que pueden ser usados por muchas personas de diferentes maneras. Por ejemplo, dentro de una empresa, los mismos datos podrían ser utilizados para satisfacer las necesidades de *marketing*, ventas y contabilidad.
- ▶ Todos estos tipos de información están estrechamente relacionados, aunque ligeramente diferentes. Con un almacén de datos, las diferentes organizaciones pueden ver los datos como lo desean.
- ▶ Visualizar los datos de diferentes maneras es solo una ventaja de tener una base sólida. Un beneficio relacionado es la capacidad de conciliar datos, si es necesario.
- ▶ Otro beneficio relacionado es la flexibilidad. Tener una base de datos sólida permite

fácilmente dicha cualidad.

- ▶ Los datos granulares contienen un historial de actividades y eventos en toda la empresa. Y el nivel de granularidad debe ser lo suficientemente detallado como para que los datos puedan ser transformados.
- ▶ Tal vez el mayor beneficio de un almacenamiento de datos consiste en que los futuros requisitos desconocidos pueden ser integrados. Cuando surge un nuevo requerimiento y existe la necesidad de información, el almacén ya está disponible para el análisis y la organización está preparada para manejar los nuevos requisitos.

En la Figura 10 se muestra un ejemplo de granularidad:

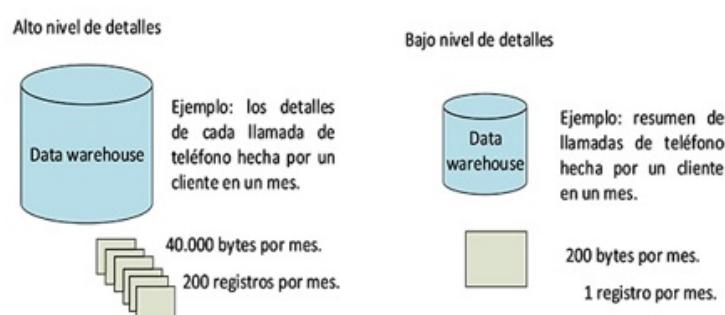


Figura 10. Ejemplo de granularidad. Fuente: Inmon, 2002.

## 4.5. Data-driven decision modelling

Las empresas están incluyendo el concepto de *data-driven* (literalmente, ‘impulsado por datos’) para mejorar el proceso de toma decisiones estratégicas basadas en análisis de datos y su interpretación.

Una perspectiva *data-driven* (DDM) permite a las empresas examinar y organizar sus datos con el objetivo de mejorar la atención a sus clientes y consumidores. Al usar datos para promover sus acciones, una empresa puede contextualizar y personalizar los mensajes a sus clientes o posibles clientes para un enfoque más centrado en el cliente.

Este método representa grandes avances en el modelo empírico convencional usado hasta el momento e incluye contribuciones de los siguientes campos (Solomatine et al., 2009):

- ▶ **Inteligencia artificial o AI (*artificial intelligence*):** estudio general de cómo la inteligencia humana puede ser incorporada a las computadoras.
- ▶ **Inteligencia computacional o CI (*computational intelligence*):** incluye redes neuronales, sistemas difusos y la informática evolutiva, así como otras áreas dentro de AI y máquina-aprendizaje.
- ▶ **Computación suave o SC (*soft computing*):** está cerca del CI, pero con especial énfasis en sistemas difusos basados en reglas inducidas a partir de datos.
- ▶ **Máquinas de aprendizaje automático o ML (*machine learning*):** una vez fue una subárea de AI que se concentraba en los fundamentos teóricos utilizados por CI y SC.
- ▶ **La minería de datos o DM (*data mining*) y el descubrimiento de**

**conocimiento en bases de datos** o KDD (*knowledge discovery in databases*): se enfocan a menudo en bases de datos muy grandes y se asocian con aplicaciones bancarias, servicios financieros y gestión de recursos de clientes. DM se ve como parte de un KDD más amplio. Los métodos utilizados provienen principalmente de estadísticas y ML.

- ▶ **Análisis inteligente de datos** o IDA (*intelligent data analysis*): tiende a centrarse en el análisis de datos en medicina e investigación e incorpora métodos de estadísticas y ML.

El DDM es un método que se centra en el uso de los métodos de CI, particularmente las máquinas de aprendizaje, para la construcción de modelos que complementan o reemplazan los modelos «basados en el conocimiento» que describen el comportamiento físico.

El proceso de construcción de un modelo DDM sigue los principios generales de: recolectar datos —seleccionar la estructura del modelo—, construir el modelo y finalmente probarlo y (posiblemente) iterar. Un enfoque general del modelo se muestra en la Figura 11.

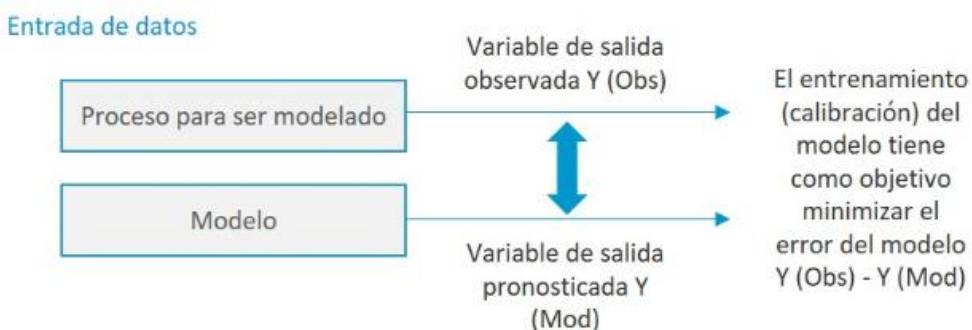


Figura 11. Enfoque general del modelo DDM. Fuente: Solomatine et al., 2009.

## 4.6. Metodología DevOps

Combina filosofía, prácticas y herramientas que incrementan la capacidad de la organización para desarrollar aplicaciones y servicios a gran velocidad.

Los equipos de desarrollo y operaciones ya no están aislados. Se utilizan prácticas y herramientas para automatizar los procesos que venían siendo manuales y lentos, ahora actúan de forma rápida y confiable (Microsoft, 2021).

Los beneficios de DevOps son (AWS, 2021):

- ▶ Velocidad: a través de microservicios y entrega continua.
- ▶ Entrega rápida: se automatiza el proceso de publicación de *software*, desde la creación hasta la implementación.
- ▶ Confiabilidad: monitoreo y registro ayudan a ver el desempeño en tiempo real.
- ▶ Escalado: la infraestructura del código ayuda a administrar los entornos de desarrollo, prueba y producción de forma repetible y eficaz.
- ▶ Seguridad: se puede definir y supervisar.

## 4.7. Nuevos roles

Con el avance de las tecnologías de la información para el análisis de los datos, se hace necesaria la incorporación de nuevos roles de usuarios que se deben integrar a los equipos de trabajo para su tratamiento.

- ▶ **Director de datos CDO (*chief data officer*):** lidera la gestión de datos y analítica asociada al negocio, es el responsable de los diferentes equipos especialistas en datos.
- ▶ **Científicos de datos (*data scientists*):** son los miembros principales del equipo de ciencia de datos, de los que extraen conocimiento e información valiosa. Tienen una perspectiva general de todo el proceso y pueden resolver problemas que se presenten en ciencias de datos, la construcción de modelos analíticos y algoritmos. Combinan diversas habilidades relacionadas con las matemáticas, la estadística, la programación y la visualización, pero también deben tener destrezas comunicativas para explicar los resultados obtenidos en la organización.
- ▶ **Ciudadano científico de datos (*citizen data scientist*):** el usuario dentro de la empresa que típicamente no está principalmente formado para este rol, pero que puede extraer valor y conocimiento a través de su experiencia, explorando los datos desde las diferentes unidades de negocio. Puede ejecutar una serie de sencillas tareas analíticas utilizando herramientas especializadas en el descubrimiento de datos. Por ejemplo, en una empresa de biotecnología hacen esta labor los mismos biólogos, expertos en el tema, conocedores de los datos e inquietos por saber sobre las últimas tecnologías de acceso y análisis de datos.
- ▶ **Ingeniero de datos (*data engineer*):** es el encargado de proporcionar los datos

de una manera accesible, clara y apropiada a los usuarios y al científico de datos. Es un usuario especializado en infraestructura *big data*. Desarrolla y explota técnicas, procesos, herramientas y métodos que deben servir para el desarrollo de aplicaciones *big data*. Tiene formación y conocimiento en gestión de bases de datos, arquitecturas de clúster (segmentación), lenguajes de programación y sistemas de procesamiento de datos.

- ▶ **Administrador de datos (*data steward*):** su responsabilidad es mantener la calidad, disponibilidad y seguridad de los datos. Busca siempre mejorar el almacenamiento y presentación de los datos en toda la empresa. Tiene conocimientos de todos los procesos de negocio y, por lo tanto, sabe cómo se utilizan dentro de estos procesos.
- ▶ **Analista de datos (*business data analyst*):** forma parte de las iniciativas y proyectos de análisis de datos. Es la persona que recopila las necesidades de los usuarios de negocio y se las pasa a los científicos de los datos de forma que sea inteligible técnicamente.
- ▶ **Artistas de datos (*data artist*):** son los responsables de crear los gráficos, infografías y otras herramientas visuales para ayudar a los diferentes usuarios de la organización a comprender datos complejos.

Además de estos roles propuestos, con la entrada en vigor del Reglamento General de Protección de Datos, RGPD (o GDPR por sus siglas en inglés), ha aparecido recientemente un nuevo rol: el ***data protection officer*** (DPO), quien se encarga de asegurar que el procesamiento de los datos personales de la plantilla, clientes, proveedores y cualquier otro individuo con los que se relacione cumple con el reglamento establecido en las reglas de protección de datos (EDPS, s.f.).

## 4.8. Tendencias

El *business intelligence* sigue asombrando por su capacidad de transformación, adaptación y fácil manejo, con lo que consigue que todos los usuarios puedan satisfacer sus necesidades. Las tendencias en BI ayudan a entender que este jugará un papel importante en la transformación digital de las empresas.

El autor Josep Curto, en su libro *Introducción al business intelligence*, opina que antes del año 2008 el mercado de BI había logrado una importante madurez con múltiples representantes ofreciendo soluciones que cubrían la mayoría de las funcionalidades del BI (ETL, *data warehouse*, *reporting*, cuadro de mandos y OLAP).

Esta madurez originó una fuerte consolidación del mercado durante el período 2005-2009 y lo agrupó de la siguiente forma:



Figura 12. Consolidación del mercado del BI según Josep Curto. Fuente: Curto, 2012.

- ▶ **Grandes agentes:** fabricantes externos que ampliaron su catálogo de soluciones empresariales con plataformas BI. En este entorno, las marcas más destacadas son: Oracle, que adquirió Hyperion; SAP, que adquirió Business Objects; IBM, que se hizo con el control de Cognos; y otras soluciones del mercado.

- ▶ **Empresas tradicionales del mercado:** aquellas que se mantienen con un producto especializado. Por ejemplo, Information Builders o Microstrategy.
- ▶ **Empresas de áreas especializadas:** aquellas que son expertas en un campo en concreto de la inteligencia de negocio como:
  - *Data warehouse:* Teradata, Netezza, Vertica.
  - Integración de datos: Informatica, Talend.
  - Análisis visual: Datawatch.
  - Análisis dinámico y flexible: QlikView, Tableau.
- ▶ **Empresas open source:** aquellas que cubren todo el proceso de la inteligencia de negocio y ofrecen soluciones con TCO (*total cost ownership*). Por ejemplo, Hitachi Vantara (quien adquirió Pentaho) y Jaspersoft (Curto, 2012).

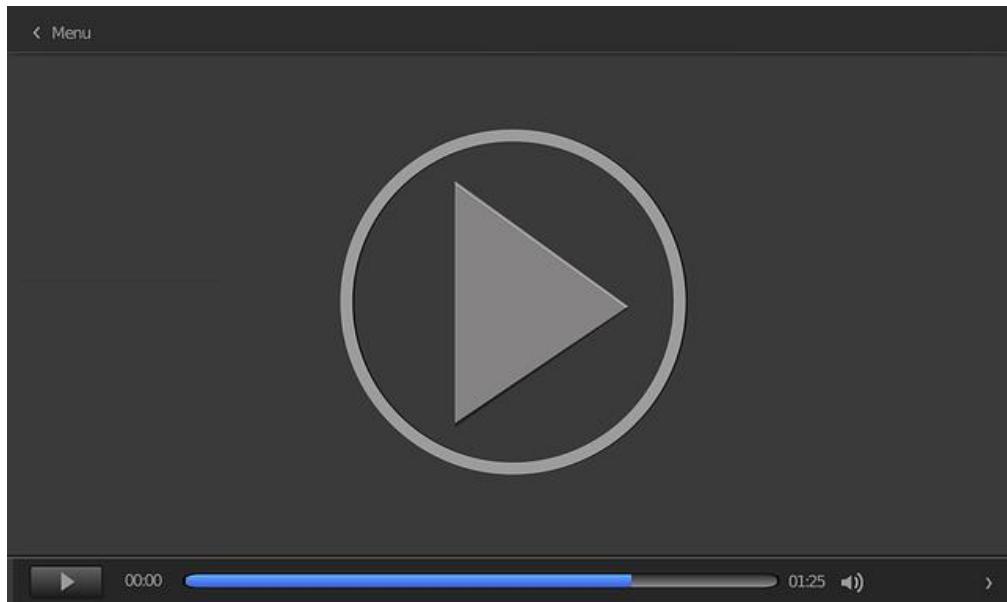
Cada año, BI & Analytics Trend Monitor de BARC brinda a los profesionales involucrados en procesos de *business intelligence* y gestión de datos una plataforma para opinar sobre las tendencias. Las opiniones son analizadas por expertos en BI de BARC, quienes complementan y redactan un informe anual. Las tendencias en los próximos años según los análisis de los reportes del 2020 están enfocadas hacia (BARC, 2021):

- ▶
  - **Avance en analítica de datos, *machine learning*:** la cantidad de casos de uso posibles es inmensa y comprende desde realizar pronósticos sobre ingresos, precios, ventas o valor para el cliente hasta prevenir cancelaciones de contratos, optimizar el tiempo de inactividad de la máquina no planificado y muchos más. Las consideraciones de sesgo en la toma de decisiones algorítmicas y los estándares éticos para tales soluciones están ganando importancia.

- **Desarrollo ágil de BI:** los principales beneficios del desarrollo ágil son la velocidad, la adaptabilidad y una alineación más estrecha entre el negocio e IT. Los nuevos paneles, informes y KPI se suministran utilizando canalizaciones de datos generadas por metadatos y controladas por modelos y otros conceptos de automatización del almacén de datos. El enfoque DevOps aporta una mentalidad y las mejores prácticas para una entrega continua automatizada, que permite un cambio rápido.
- **Generación de alertas:** *machine learning* se emplea en herramientas líderes para enfocar la conciencia de los usuarios sobre tendencias y valores atípicos que antes no estaban buscando. Aquí la monetización de los datos se vuelve obvia.
- **Datos en la nube y análisis:** el reto es implementar la gestión de los datos y el análisis a la nube. Existen factores que preocupan a las empresas: factores legales, seguridad y privacidad, escasez de consejos sobre mejores prácticas para construir arquitecturas híbridas o de múltiples nubes, falta de confianza en los proveedores y el deseo de mantener bajo control los datos.
- **Maestro de datos/administración de calidad de los datos:** las decisiones correctas solo pueden tomarse con base en datos confiables, consistentes. Solo los datos maestros son fundamentales para mantener una coherencia entre informes y operaciones basadas en datos. El desafío radica en definir roles y responsabilidades para asegurar la calidad y el monitoreo continuo de los datos de una empresa.
- **Descubrimiento y visualización de datos:** descubrir patrones y datos atípicos en los datos. El aprendizaje automático guía cada vez mejor a los usuarios comerciales y automatiza tareas desde la preparación hasta la visualización. Es importante que el sistema no solo dé una respuesta, sino que proporcione explicaciones.

- **Gobierno del dato:** a diferencia de BI o BA, la gobernanza se centra en los datos de todos los sistemas que tratan con datos. Debe ser el mecanismo de dirección para la administración de los datos. Es un esfuerzo a largo plazo.
- **Preparación de datos para usuarios del negocio:** lograr una preparación de datos ágil a escala es clave para aprovechar los datos empresariales y externos y poder tomar decisiones, automatizar procesos y monetizar los datos. Los usuarios comerciales deben contar con herramientas intuitivas y fáciles de usar que apliquen aprendizaje automático.
- **Modernización de los *data warehouse*:** IT debe estar preparado para cambio de requisitos analíticos rápidamente e implementar los DW en nuevas estructuras y más baratas a través de servicios externos.
- **Cultura *data-driven*:** cambio de las empresas que trabajan con datos aislados y orientados a proyectos a empresas completamente basadas en datos. Calidad y cantidad de datos pueden ser usados para soportar las decisiones de la empresa. No solo debe importar los beneficios de la empresa sino el nivel de cultura *data driven*.

Para finalizar, accede al vídeo Tendencias BI/BA.



---

Accede al vídeo:

<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=7b1721b5-a4dc-47e5-bb16-ad2b000b0d47>

---

Vídeo. Tendencias BI/BA

## 4.9. Referencias bibliográficas

AWS. (2021). ¿Qué es DevOps? [Página web].

Amazon. <https://aws.amazon.com/es/devops/what-is-devops/>

BARC. (2021). Data, BI & Analytics Trend Monitor 2021 [Página web]. BARC.

<http://barc-research.com/research/bi-trend-monitor/>

Curto, J. (2012). *Introducción al business intelligence*. Editorial UOC.

EDPS. (s.f.). Data Protection Officer (DPO) [Página web]. *European Data Protection Supervisor*. [https://edps.europa.eu/data-protection/data-protection/reference-library/data-protection-officer-dpo\\_en](https://edps.europa.eu/data-protection/data-protection/reference-library/data-protection-officer-dpo_en)

Inmon, W. H. (1992). *Building the data warehouse*. Wiley.

Inmon, W. H. (2002). *Building the data warehouse*. Wiley.

Kimball, R., Reeves, L., Ross, M., y Thornthwaite, W. (1998). *The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying data warehouses*. Wiley.

Microsoft. (2021). ¿Qué es DevOps? [Página web]. Microsoft Azure. <https://azure.microsoft.com/es-es/overview/what-is-devops/>

Mundy, J., Thornthwaite, W., y Kimball, R. (2006). *The Microsoft data warehouse toolkit: with SQL server 2005 and the Microsoft business intelligence toolset*. Wiley.

Rivadera, G. (2010). La metodología de Kimball para el diseño de almacenes de datos (data warehouses). *Universidad Católica de Salta*, 5, 56-71.

Solomatine, D., See, L. M., y Abrahart, R. J. (2009). Data-driven modelling: concepts, approaches and experiences. En *Practical Hydroinformatics* (pp. 17-30). Springer.

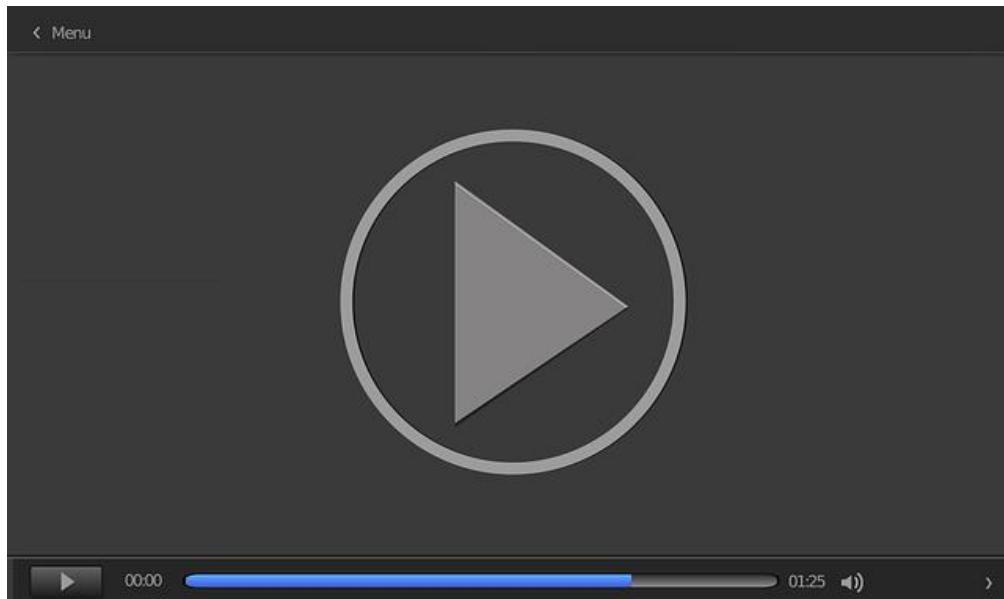
[https://www.springer.com/cda/content/document/cda\\_downloaddocument/9783540798804-c1.pdf?SGWID=0-0-45-620522-p173819207](https://www.springer.com/cda/content/document/cda_downloaddocument/9783540798804-c1.pdf?SGWID=0-0-45-620522-p173819207)

Vahos, L. E., Pastor, D. M., y Jiménez, J. A. (2013). Método para la formación de stakeholder en proyectos de ingeniería usando la metodología PMI y técnicas de inteligencia artificial. *Revista Ingenierías*, 12(23), 157-168.

## Metodología de Ralph Kimball para la implementación de DW/BI

Chávez, P. (10 de mayo de 2015). *Metodología de Ralph Kimball para la Implementación de DW/BI* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=f0SXElfOx6k>

Pedro Chávez nos da las claves del método Kimball, la aplicación de la inteligencia de negocio en diferentes sectores.



Accede al vídeo:

<https://www.youtube.com/watch?v=f0SXElfOx6k>

## El concepto de DevOps

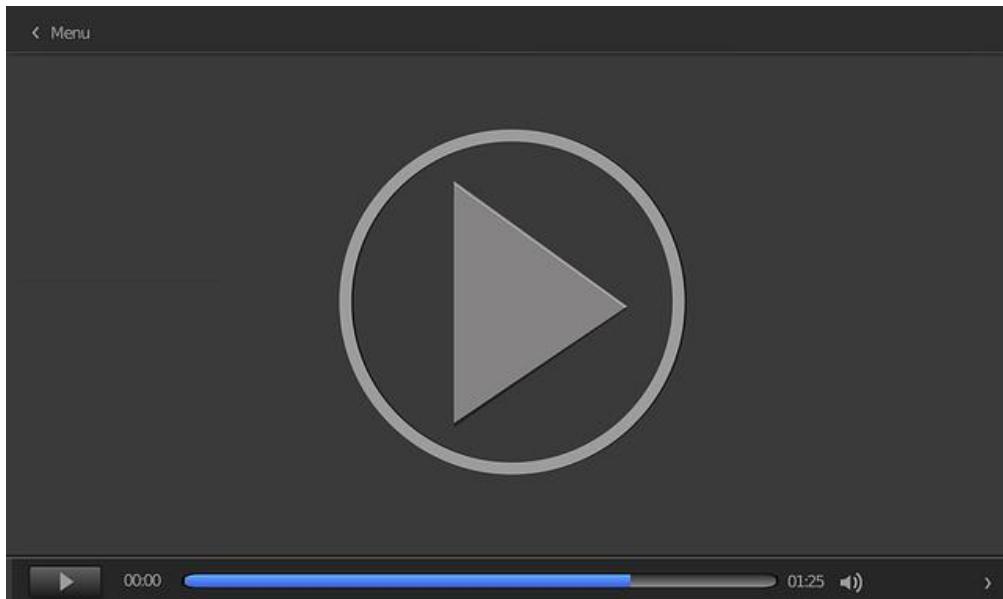
Red Hat. (2021). El concepto de DevOps [Página web]. Red Hat. <https://www.redhat.com/es/topics/devops>

Para usar DevOps se necesitan aplicaciones heredadas con las aplicaciones creadas en la nube y las nuevas infraestructuras.

## Business intelligence governance best practices for 2021 and beyond

Metric Insights. (19 de marzo de 2021). [Panel with PayPal & Veeco] Business Intelligence Governance Best Practices for 2021 and Beyond [Vídeo]. Youtube. <https://www.youtube.com/watch?v=mcZFF0x4pws>

Abhishek Rajpurohit, Gregory Zelo y Marius Moscovici discuten en un panel sobre gobernanza y BI.



Accede al vídeo:

<https://www.youtube.com/watch?v=mcZFF0x4pws>

1. Son principios básicos del desarrollo de DW según Kimball:

  - A. Centrarse en el negocio.
  - B. Incrementos significativos
  - C. Los datos como centro de análisis.
  - D. A y B son correctas.
2. No es una tarea dentro de la metodología Kimball:

  - A. Crear el maestro de datos.
  - B. Diseño e implementación de ETL.
  - C. Diseño de arquitectura técnica.
  - D. Selección de herramientas de implementación.
3. Son áreas de conocimiento dentro de la metodología PMI:

  - A. DevOps.
  - B. CRISP-DM.
  - C. Costes.
  - D. Alcance.
4. ¿Cuáles son las rutas para el desarrollo del *data warehouse* según la metodología Kimball?

  - A. Tecnologías.
  - B. Procesos.
  - C. Datos.
  - D. A y C son correctas.

5. ¿Cómo se pueden obtener los requerimientos del *data warehouse*?
  - A. A través de entrevistas.
  - B. Por medio de sesiones facilitadoras.
  - C. Por los clientes externos.
  - D. A y B son correctas.
6. ¿Cuántas características define Inmon para el DW?
  - A. 1.
  - B. 8.
  - C. 4.
  - D. 6.
7. Son nuevos roles incorporados en los últimos años:
  - A. Ingeniero de datos y analista de datos.
  - B. Consumidor de datos.
  - C. Usuarios finales.
  - D. Consumidores de información.
8. Según el estudio publicado por el BI & Analytics Trend Monitor de BARC, son tendencias del *business intelligence*:
  - A. Creación de grandes empresas.
  - B. Almacenamiento de gran cantidad de datos.
  - C. Desarrollo ágil de BI.
  - D. Generación de alertas.

**9.** Menciona dos etapas de la metodología PMI:

- A. Pretratamiento de los datos brutos.
- B. Inicialización y cierre de los datos maestros.
- C. Control y ejecución del proceso.
- D. Control y supervisión.

**10.** La metodología DevOps permite:

- A. Aplicar velocidad a los proyectos subidos a AWS.
- B. Aplicar velocidad a los proyectos subidos a Azure.
- C. Entrega rápida, confiabilidad y escalado de proyectos.
- D. Actuar lento pero seguro en proyectos BI.

Gobierno del Dato y Toma de Decisiones

---

## Tema 5. Introducción al marketing

# Índice

## Esquema

### Ideas clave

- 5.1. Introducción y objetivos
- 5.2. Fundamentos y concepto del marketing
- 5.3. Concepto de sistema de información
- 5.4. Captura de datos. Sistema de datos internos
- 5.5. La investigación del marketing. Las necesidades del cliente

### A fondo

- El tesoro de los datos masivos
- Los perfiles más demandados del marketing digital 2021
- La evolución del marketing
- Ojo con tus datos

### Test



## 5.1. Introducción y objetivos

Según el profesor Philip Kotler, los tres primeros pasos de cualquier proceso de *marketing* son:

- ▶ Entender el mercado y las necesidades de los clientes.
- ▶ Diseñar una estrategia adecuada con el foco en el cliente.
- ▶ Diseñar el programa de *marketing*.

## 5.2. Fundamentos y concepto del marketing

El *marketing* es una disciplina de reciente incorporación y su definición ha ido evolucionando y modificándose hasta llegar a los términos actuales.

### Evolución conceptual

El *marketing* surgió como una de las funciones de la empresa que se dedicaba a realizar acciones comerciales visuales necesarias para vender los productos de la empresa.

El *marketing* ha ido transformándose y progresando como concepto modificando su papel y finalidad tanto dentro de la empresa como en su aplicación en otras estructuras organizativas en función de los requerimientos e imposiciones hacia el mercado.

De esta evolución se obtiene una definición más amplia y generalizada, que es la que hoy se emplea. Esta concepción actual del *marketing* ya no se refiere solo a una actividad exclusivamente empresarial dedicada a intercambios básicamente comerciales. Se avanza y profundiza en las actividades objeto del *marketing* para abarcar todas las actividades que puede realizar cualquier tipo de organización, tengan o no estas instituciones ánimo de lucro.

**Lo que sí es imprescindible es que el marketing consiga llamar la atención del usuario final para que las empresas puedan vender sus productos.**



Figura 1. Impacto del *marketing*. Fuente: elaboración propia.

El impacto del *marketing* en la vida diaria de las personas ha convertido esta disciplina en una de las funciones más importantes en el actual entorno organizativo, adoptada por todo tipo de estructura empresarial.

La empresa como ruta de transmisión del flujo económico y, dentro de ella, el *marketing* como su elemento dinamizador se han convertido en protagonistas de las transacciones comerciales revitalizadores de la recuperación económica.

El *marketing* existe porque es clave para la supervivencia de la empresa al intentar generar una **actitud hacia el mercado para conseguir unos posibles comportamientos de compra**. De hecho, en función de cómo se va a concebir la **relación entre el intercambio y el consumidor** y los principios de *marketing* que aplica, la empresa adopta un método de negocio u otro.



Figura 2. Meta, misión y medio del *marketing*. Fuente: elaboración propia.

En este sentido, resulta importante reconocer que, aunque siempre se ha realizado una función comercial o de *marketing* en las empresas, conforme han ido variando las circunstancias socioeconómicas se han producido alteraciones en las estructuras organizativas que han modificado los procedimientos que se siguen para tomar decisiones.

Igual que ha evolucionado el concepto adaptándose a las necesidades del mercado, también se ha modificado la manera de concebir esta relación de intercambio entre la empresa y el consumidor.

Por esta causa, las orientaciones estratégicas de las empresas, en función de la evolución de la economía, del crecimiento de la competencia y del mercado, se han transformado hasta llegar a distinguir **cuatro orientaciones interrelaciones:**

1. Orientación a la producción.

2. Orientación al producto.
3. Orientación a las ventas.
4. Orientación al *marketing*.



Figura 3. Concepto de *marketing*. Fuente: elaboración propia.

### Concepto de *marketing*

Una vez introducidos los fundamentos del *marketing*, es importante definir cada uno de los conceptos que se utilizan sistemáticamente en esta disciplina científica, que constituyen las herramientas básicas de análisis y condicionan toda su metodología de trabajo.

El constante cambio de los escenarios actuales, la globalización y la incertidumbre que se refleja en los mercados hace que cada día resulten más importantes las acciones de *marketing* en las organizaciones. Por todo lo anterior es preciso razonar con claridad los términos fundamentales que se utilizan y precisar qué se entiende por cada uno de ellos, en qué sentido se aplican y cómo se trabajan desde el punto de vista empresarial.

## Productos, bienes, servicios e ideas

El producto es el **medio** del que dispone la empresa para satisfacer una necesidad en el consumidor, que posee un valor y es susceptible de satisfacer dicha carestía o insuficiencia.

El término *producto* se utiliza de forma genérica incluyendo también los servicios y las ideas.

## Necesidad, deseo y demanda

Conviene diferenciar estos tres conceptos, pues el *marketing* trata de satisfacer necesidades, formalizadas a través de un deseo y, para ello, actúa fundamentalmente sobre la demanda. Precisemos su contenido:

- ▶ Necesidad: sensación de carencia de algo.
- ▶ Deseo: expresión de la voluntad de satisfacer una necesidad.
- ▶ Demanda: formulación expresa del deseo condicionada por los recursos disponibles.

Distinguir claramente estos conceptos es imprescindible para asentar los principios de *marketing* y actuar en consecuencia.



Figura 4. Conceptos *marketing*: necesidad, deseo y demanda. Fuente: elaboración propia.

El mecanismo progresivo que se establece en el proceso de adquisición de cualquier producto o servicio se inicia cuando, registrada o despertada una **necesidad**, esta se convierte en un **deseo**, que correctamente encauzado a través de las diferentes fases del proceso de compra se convierte en una **demand**a específica materializada en la obtención de un **producto** concreto.

De esta condición escalonada, el deseo encaminado, que siempre es genérico, termina convirtiéndose en una determinada demanda satisfecha por medio de los planteamientos del *marketing* y la capacidad adquisitiva del individuo.

## Mercado

El *marketing* tiene como finalidad que las personas adquieran determinados productos (bienes tangibles o intangibles). Esta idea tan simple implica que el *marketing* va más allá del simple hecho de dar a conocer el producto.

En consecuencia, cada planteamiento empresarial exige tener un **conocimiento en profundidad del mercado**, real o potencial, al que se dirige y que estará constituido por personas que tienen necesidades específicas no cubiertas y que, por tal motivo, están dispuestas a adquirir bienes o servicios que los satisfagan.

Desde este enfoque en términos económicos, el mercado es el conjunto de personas u organizaciones que conforman un contexto (físico o virtual) y que tiene una necesidad que puede satisfacerse a través de la adquisición de un producto determinado por medio un conjunto regulado de transacciones e intercambios entre partes compradoras y partes vendedoras.

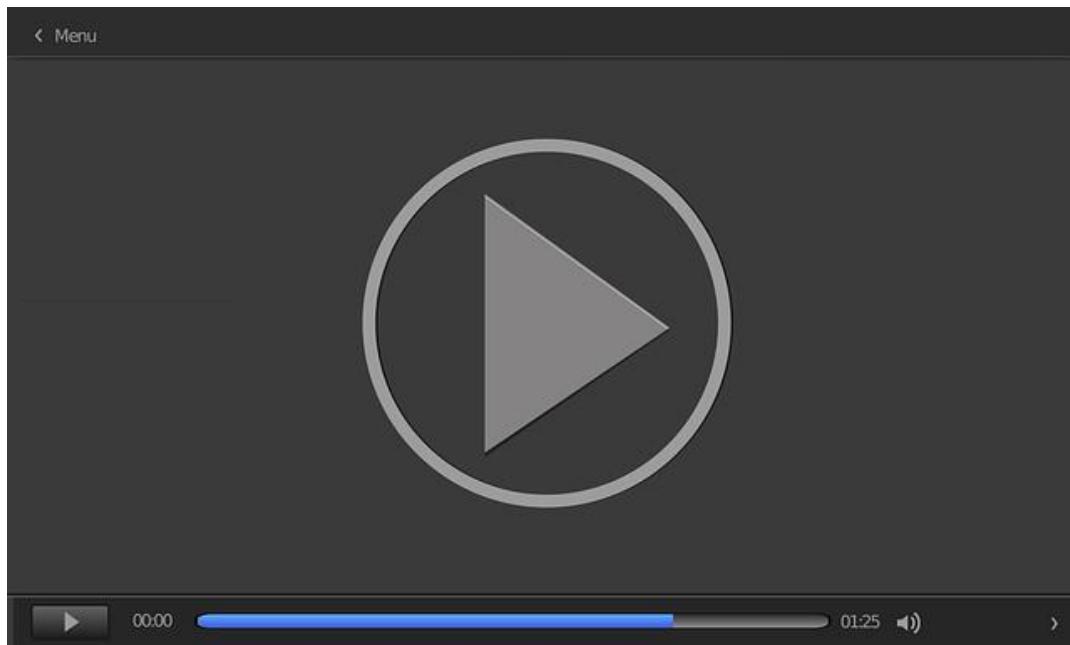
En esta contextualización del mercado resulta importante diferenciar tres conceptos en relación con las actividades de intercambio de *marketing* que son **evolutivas**, porque la superación de una significa situarnos en la siguiente.



Figura 5. Conceptos relacionados con las actividades. Fuente: elaboración propia.

Una vez vistos los fundamentos del *marketing* y la importancia de conceptos tales como necesidad, deseo, demanda, intercambio, transacciones y relaciones, vamos a estudiar los sistemas de información del *marketing* (SIM).

Pero antes, accede al vídeo *El marketing en todos los entornos empresariales actuales*.



Accede al vídeo:<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=2d9b8652-3a30-4778-846f-acd400c83f97>

---

Vídeo. *El marketing en todos los entornos empresariales actuales*

## 5.3. Concepto de sistema de información

Cada vez más, las organizaciones recurren a aplicar métodos racionales y versados para resolver sus dilemas y tomar decisiones, al mismo tiempo que reconocen los beneficios que se derivan de ello. De hecho, la mayoría de las empresas tienen automatizados y sistematizados los procesos de recogida de información.

Cualquiera de los métodos empleados, para que sean eficaces, conviene que sigan un sistema. **Sistematizar la información** es un recurso esencial para que la empresa alcance mayores niveles de competitividad. Un sistema de información fundamental es el que genera la propia organización en el desempeño de sus actividades cotidianas durante su labor diaria.

Pero el concepto de sistema de información que aludimos en esta materia va más allá de la mecanización de actividades rutinarias. Hace referencia a la gestión del conocimiento que encierran los datos disponibles y aprovechables en la información.

Hoy en día los datos que circulan en una empresa poseen un valor inmenso que debemos aprender a tramitar para poder ser utilizados con eficacia.

### La eficiencia de la información



Figura 6. Eficiencia de la información. Fuente: elaboración propia.

Cualquier empresa genera gran cantidad de información en su gestión diaria. En algunos casos es consciente de ella, pero en otros muchos, no. Resulta asombroso saber la inmensa cantidad de información de la que dispone una empresa, que en muchas ocasiones los propios directivos desconocen.

También asusta comprobar cuántos datos no son considerados como tales y se pierden, al no darles valor, por los entresijos de las estructuras organizativas. Sin embargo, es indudable que los directivos y gestores necesitan, en todo momento, información sobre el consumidor, el mercado y el entorno competitivo en el que se mueven para tener criterios de decisión y formarse opiniones contrastadas y fundamentadas.

En la actualidad, el entorno empresarial es tan dinámico y complejo que demanda **grandes cantidades de datos para tomar decisiones de forma eficaz** y ser resolutivos en la gestión empresarial. Una condición esencial del mundo digital en el que vivimos es una administración eficaz de la información procedente de fuentes internas y externas, que pasan por la empresa agrupadas en torno a un sistema de información, conocido como SIM, y que se agrega de los sistemas de datos internos, el sistema de inteligencia de *marketing* y la investigación de mercados que da soporte al sondeo de un problema u oportunidad delimitada detectado por alguno de los anteriores sistemas de *marketing*.

## Contenidos SIM

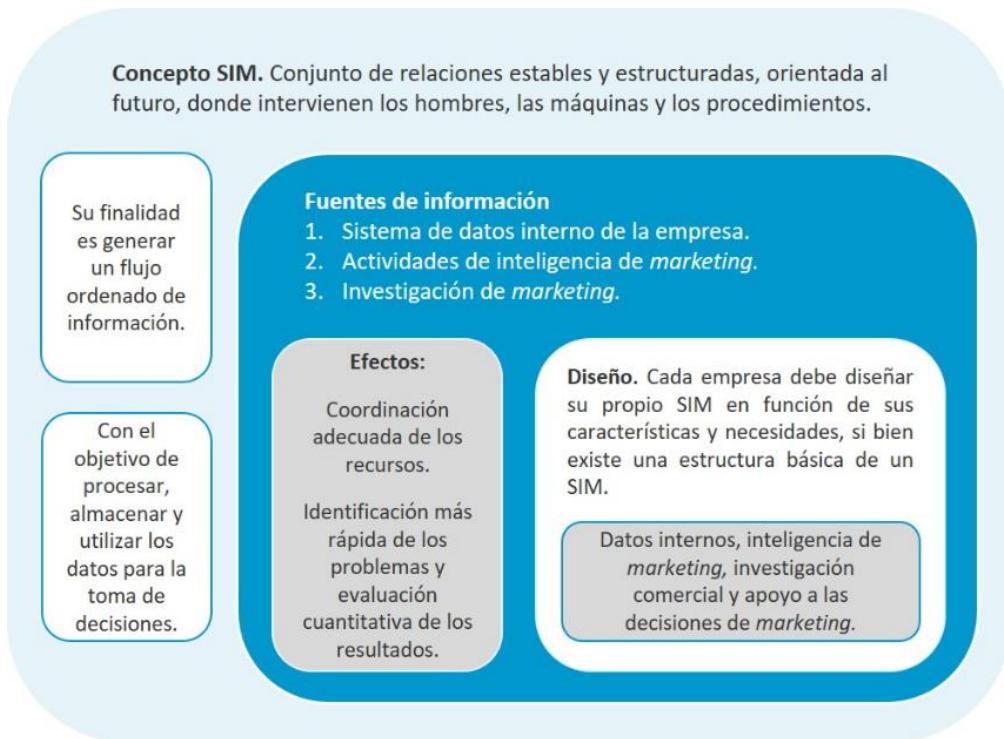


Figura 7. Concepto SIM. Fuente: elaboración propia.

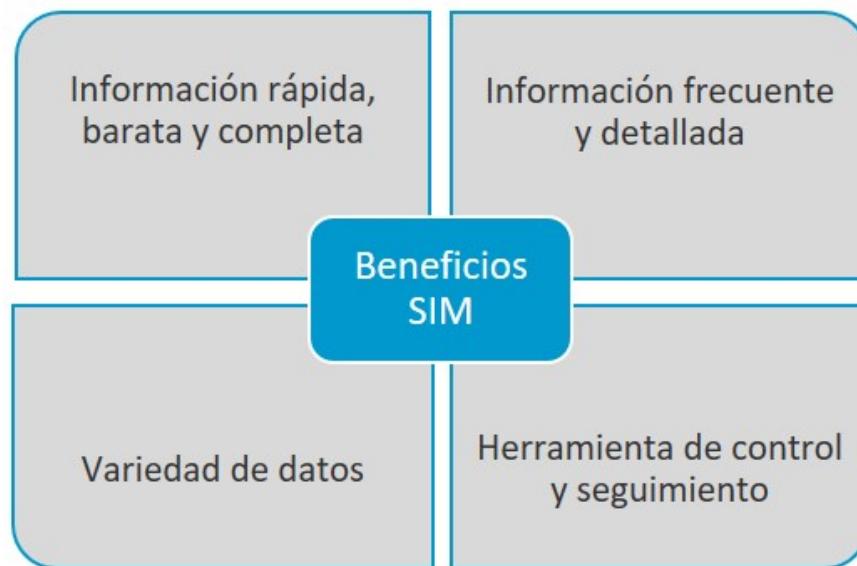


Figura 8. Beneficios SIM. Fuente: elaboración propia.

## 5.4. Captura de datos. Sistema de datos internos

Hoy en día, las empresas cuentan con **amplia información sobre sus clientes**, que obtienen a través de diferentes canales y se clasifican en función de su temporalidad en la obtención de los datos.

Clasificación de los sistemas de información en *marketing* SIM:

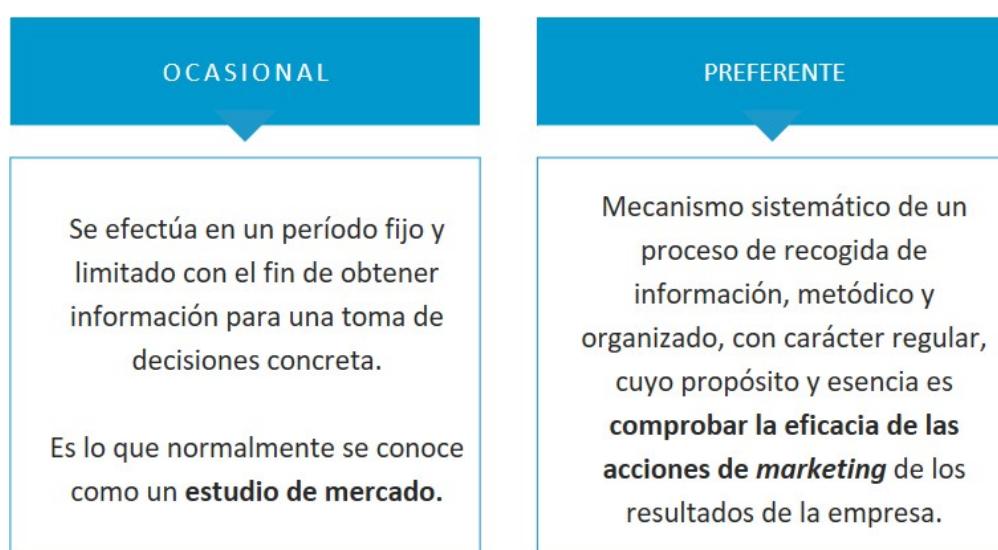


Figura 9. Clasificación sistemas SIM. Fuente: elaboración propia.

Esta clasificación responde a la imperiosa urgencia de invertir en la **obtención de datos en tiempo real**. Sea cual sea el SIM que se emplee, que dependerá del tiempo que se necesita en la respuesta, la captura empieza por la creación de un banco de datos que recoja toda la información que fluye por la empresa.

Determinar la calidad es fundamental para que resulten eficaces. Esta viene determinada por la utilidad que proporciona su empleo, para lo que deben estar sometidos a un diagnóstico previo que separará la información a través de un registro que normalizará los procedimientos de entrada y almacenamiento por medio de unos estándares establecidos.

Este banco de datos actualizado al instante es el que permite a la empresa contar con la información más actual de los clientes y así tomar medidas de la forma más rápida y eficiente para responder a sus necesidades.

Esta eficiencia de los datos internos se mide en función de:



Figura 10. Datos internos y externos. Fuente: elaboración propia.

Los **datos internos** son los primeros datos que hay que aprender a fiscalizar porque están dentro de la organización y no suponen costes adicionales. Pero no son los únicos.

Resulta importante **explotar todas las vías de información** de las que dispone la empresa, pudiendo sacar partido a todos los elementos que entran en juego desde el interior de la estructura hacia el mercado y con la competencia. Por eso, junto con esta actividad interna, debemos **completar la información con los datos externos**.

## 5.5. La investigación del marketing. Las necesidades del cliente

La supervivencia empresarial, en un principio, y el posterior éxito corporativo en el mercado y frente a la competencia de una empresa depende fundamentalmente de la demanda que realicen de sus productos los potenciales consumidores y sus clientes efectivos.

Esto significa que los departamentos de *marketing* no pueden quedarse solo en incidir en las necesidades que ya han exteriorizado los interesados en la oferta determinada. Deben ir más allá para enfocar sus esfuerzos, a la hora de analizar el mercado, en pensar cómo descubrir también las demandas todavía no manifestadas pero que están latentes. Adelantarse a las solicitudes todavía no formuladas no es un ejercicio de adivinación u oráculo sino un proceso de predicción y pronóstico en clave a datos del presente.

**El primer paso que es necesario emprender para lograrlo consiste en observar y estudiar detenidamente el comportamiento del consumidor con sus compras reales, sus inclinaciones, sus tendencias, predisposiciones y preferencias.**

La conducta del consumidor final es un indicativo de qué tipo de productos o servicios son los que desea adquirir, debiendo decidir así la empresa a qué precios venderlos, dónde y cómo hacerle publicidad al producto, qué canales de distribución emplear, etc. Precisamente para conseguirlo se emplea la investigación de *marketing*, cuya principal utilidad es conocer el nivel de satisfacción, el grado de bienestar y el estado de complacencia de sus clientes con lo adquirido.

Pero la empresa no puede quedarse en esta fase. Es solo el comienzo de una verdadera indagación del *marketing* estratégico que se ve obligado, por las circunstancias del entorno, a amplificar e incrementar los *inputs* que le llegan del

comportamiento del consumidor para traducirlos en conocimiento.

Esta comprensión acerca de los hábitos de compra proporciona información crítica sobre necesidades, preferencias y opiniones de los clientes, tanto habituales como futuros que tiene que traducirse en acciones empresariales.

Existen muchas formas de realizar una investigación de mercado, pero la mayoría de las empresas utilizan uno o varios de los cinco métodos básicos:

### Métodos de investigación de mercados



Figura 11. Herramientas investigación de mercados. Fuente: elaboración propia.

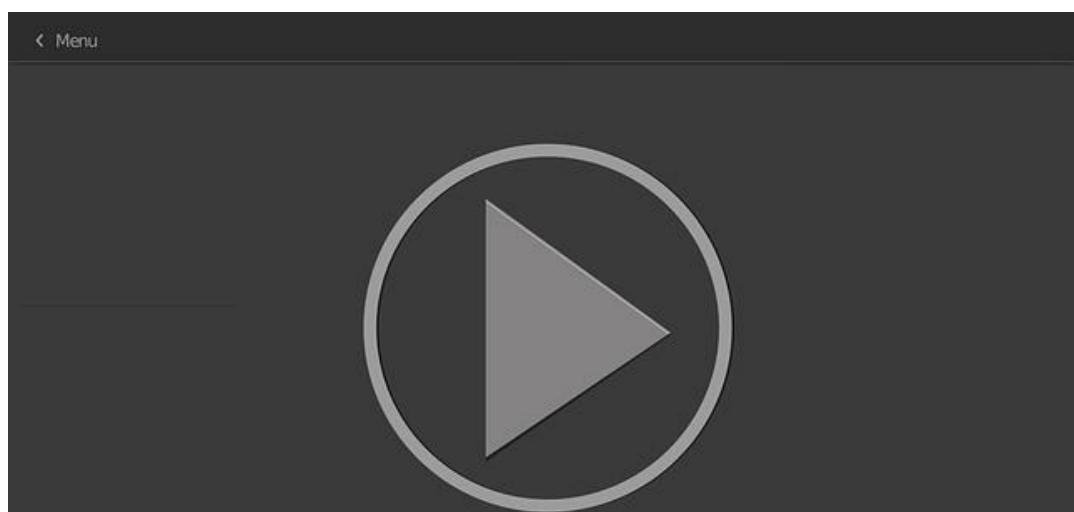
Estos métodos son las **herramientas más usadas** que permiten recabar información sobre distintos temas en investigaciones que pueden descubrir aspectos como:

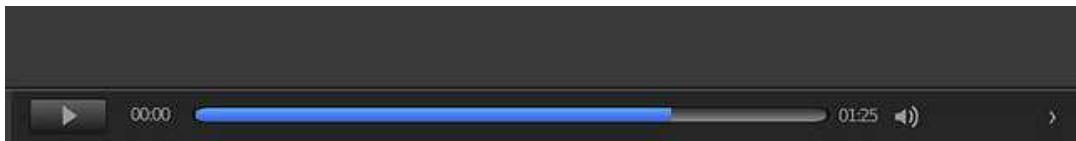


Figura 12. Utensilios del *marketing*. Fuente: elaboración propia.

La información combinada de todos estos utensilios proporciona una mejor idea de la acción de mercadotecnia, el impacto que se pretende conseguir y la influencia que se quiere conquistar, posibilitando saber cuál es la mejor combinación de los elementos 4P: producto, precio, distribución y comunicación, que permita a la empresa posicionarse de manera más ventajosa en el mercado.

Para finalizar, accede al vídeo Es imposible no influir.





---

Accede al vídeo:<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=a30354dc-15ca-4be5-8346-acd400c8404c>

---

Vídeo. Es imposible no influir.

## El tesoro de los datos masivos

---

García Campos, J. M. (13 de noviembre de 2013). El tesoro de los datos masivos. *La Vanguardia*. <http://www.lavanguardia.com/magazine/20131108/54392775355/big-data-datos-masivos-reportaje-en-portada-magazine-10-noviembre-2013.html>

---

En este interesantísimo artículo, el autor revela el cambio que estamos viviendo en el procesamiento de la información y en la generación de datos y nos señala cómo, en esta nueva era de la *datificación*, todo (incluso nuestro estado de ánimo, que revelamos a través de las redes sociales) se puede convertir en un formato cuantificado para su tabulación y análisis.

## Los perfiles más demandados del marketing digital 2021

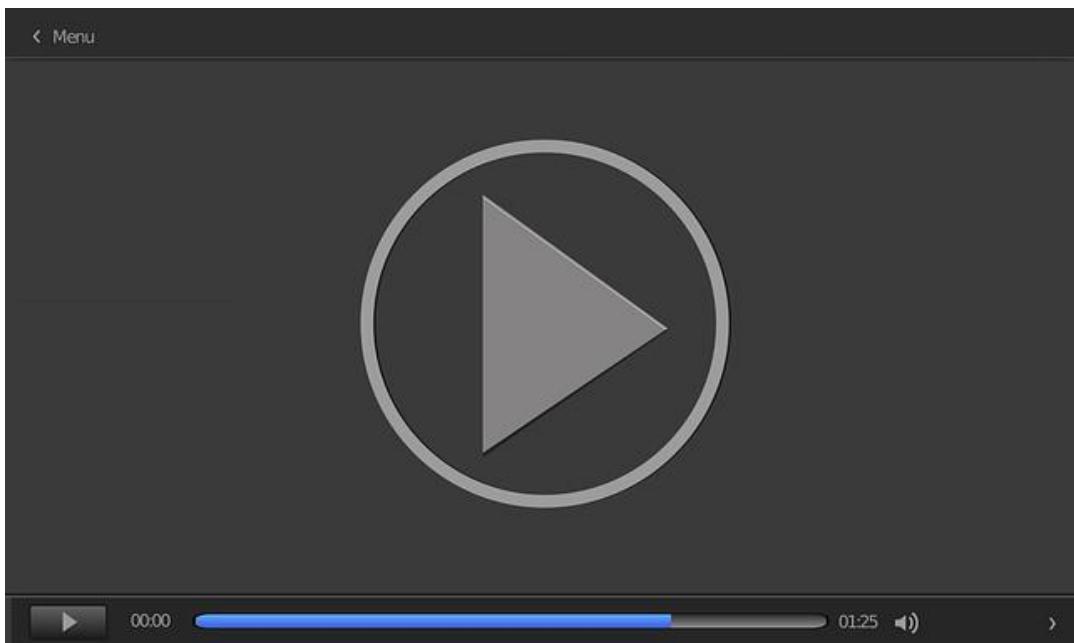
Redacción La Vanguardia. (12 de enero de 2021). Los perfiles más demandados del marketing digital. *La Vanguardia*. <https://www.lavanguardia.com/vida/formacion/20210111/6180827/perfiles-mas-demandados-marketing-digital.html/>

Actualmente, con el desarrollo de las nuevas tecnologías, las profesiones que más se demandan están relacionadas con este ámbito. Las empresas no centran sus estrategias en publicidad a través de medios como la televisión o grandes carteles; ahora se reúnen esfuerzos para crear oportunidades de crecimiento a través de la red.

## La evolución del marketing

TEDx Talks. (16 de enero de 2020). *La evolución del marketing* / Paul Soto / TEDxUANL [Vídeo]. Youtube. <https://www.youtube.com/watch?v=QyrL-K8AUuU>

Paul Soto nos habla en una charla TEDx sobre cómo la tecnología es un factor detonante para la evolución del marketing y los retos de la innovación.



Accede al vídeo:<https://www.youtube.com/watch?v=QyrL-K8AUuU>

## Ojo con tus datos

RTVE. (23 de diciembre de 2013). *Documentos TV - Ojo con tus datos* [Vídeo].

R T V E . <http://www.rtve.es/television/20131223/documentos-tv-ojo-tus-datos/826300.shtml>

En este reportaje del programa Documentos TV de RTVE se abordan los retos de la privacidad y la seguridad digital y es una ventana abierta a la reflexión sobre la información que se mueve y se genera en las redes sociales.

- 1.** El *marketing* como filosofía:
  - A. Defiende el establecimiento de unas relaciones permanentes con los clientes.
  - B. Defiende el establecimiento de unas relaciones no permanentes con los clientes
  
- 2.** La venta es un proceso:
  - A. En sentido único.
  - B. En doble sentido.
  
- 3.** El *marketing* como disciplina científica:
  - A. Es reconocido como una rama de la economía que se dedica a estudiar los canales de distribución.
  - B. Es reconocido como una ciencia aplicada del comportamiento que trata de comprender las relaciones entre compradores y vendedores.
  
- 4.** El concepto de *marketing* está estrechamente vinculado con la relación de intercambio.
  - A. Correcto.
  - B. Incorrecto.
  
- 5.** La satisfacción de necesidades es:
  - A. El elemento motivador que anula el proceso de intercambio.
  - B. El elemento motivador que facilita el proceso de intercambio.
  
- 6.** Cuando se habla de producto, desde el punto de vista del *marketing*, nos estamos refiriendo a:
  - A. Los bienes tangibles.
  - B. Los objetos tangibles e intangibles.

7. Los bienes son:

- A. Objetos físicos y, como tales, tangibles.
- B. Objetos materiales e inmateriales.

8. El enfoque actual del *marketing* hacia el mercado y en relación con el cliente:

- A. Solo es posible con un mejor conocimiento de las necesidades individuales del cliente.
- B. Implica un profundo conocimiento del mercado absoluto.

9. El sistema de inteligencia de *marketing* aporta información:

- A. Sobre aspectos relacionados con la situación y los resultados conseguidos por la empresa.
- B. Sobre los aspectos que tienen lugar en el entorno y que son más significativos para la empresa.

10. La investigación de *marketing* consiste en:

- A. El diseño sistemático, el análisis de datos y la recogida de información relevante para resolver un problema concreto al que se enfrenta la empresa.
- B. El diseño sistemático, el análisis de datos y la recogida de información relevante para resolver una complicación hipotética a la que se enfrenta la empresa.

Gobierno del Dato y Toma de Decisiones

---

## Tema 6. Métricas y métodos de análisis

# Índice

[Esquema](#)

[Ideas clave](#)

[6.1. Introducción y objetivos](#)

[6.2. Principios y fundamentos](#)

[6.3. Métricas básicas](#)

[6.4. Tipos de analíticas web](#)

[6.5. Herramientas de medición](#)

[6.6. Referencias bibliográficas](#)

[A fondo](#)

[Las sesiones en Google Analytics](#)

[Mapas de calor](#)

[Tutorial de Google Analytics](#)

[Eye tracking](#)

[Test](#)

MÉTRICAS Y MÉTODOS DE ANÁLISIS		
Principios y fundamentos	Métricas	Tipos de analíticas
Recopilar información (medición y recolección).	Visitantes únicos: número de visitantes no duplicados que han accedido al sitio web.	Logs vs. tags En la medición por logs, al navegar se envía información al servidor, siempre que se haya habilitado para este fin.
Realizar las preguntas correctas (análisis).	Visitas: conjunto de interacciones que tienen lugar en un sitio web durante un período determinado.	Censales vs. panel Los sistemas basados en una «huella», censales, analizan a todos los usuarios directamente.
Engagement es el compromiso que adquieren los usuarios con la marca y lo expresan a través de los medios sociales.	Páginas vistas: páginas a las que se accede dentro del sitio web.	Las cookies son archivos de tecnología capaz de almacenar y recuperar datos de un equipo terminal de una persona física o jurídica que utiliza para cualquier fin un servicio de la sociedad de la información.
Realizar los aportes adecuados (reporte).	Duración media de la visita: es la duración media de una sesión.	
Mejorar y optimizar un sitio web.	Porcentaje de rebote: porcentaje de visita a una sola página y en la que el usuario ha abandonado el sitio en su página de entrada.	

## 6.1. Introducción y objetivos

Este tema es una introducción básica a los conceptos de la analítica web y todas aquellas partes importantes para poder analizar las webs con éxito.

Para ello se van a estudiar los principios y fundamentos de la analítica web, así como los conceptos más importantes y las métricas más utilizadas.

Los objetivos que se pretenden alcanzar en este tema son:

- ▶ Conocer el alcance que se tiene con la analítica web.
- ▶ Entender cuáles serían las métricas claves o KPI que deben medirse para poder monitorizar la estrategia de *marketing* web.
- ▶ Revisar algunos indicadores de eficacia utilizados en el *marketing*.

## 6.2. Principios y fundamentos

La analítica web es una de las armas más potentes que existen para gestionar, solucionar, mejorar y enfocar la estrategia que se había propuesto con el sitio web.

La analítica web es la recolección, medición, análisis y reporte de los datos que se extraen de la navegación de los usuarios por un sitio web para poder comprender y optimizar su uso.

La analítica web sirve para:

- ▶ Conocer el comportamiento de los usuarios.
- ▶ Comparar el rendimiento de los diferentes medios de captación de tráfico (SEO, SEM...).
- ▶ Evaluar el rendimiento de nuestras páginas web y proponer mejoras.
- ▶ Analizar tendencias y comportamientos recurrentes en el tiempo.
- ▶ Tomar decisiones estratégicas a nivel de negocio, producto o precio.

**Objetivos:** para poder analizar los resultados que hemos obtenido, es imprescindible tener claro previamente los objetivos que queremos alcanzar como organización a nivel general y con la estrategia de *marketing* digital en particular. Estos objetivos estarán basados en una serie de KPI.

**Acciones en el sitio web:** es importante analizar, pero lo verdaderamente importante es plantear una serie acciones en función de todos los datos obtenidos. Tener documentos y presentaciones llenos de cifras y datos por sí solo no sirve para nada, es necesario convertir esa información en conocimiento para mejorar la gestión y optimizar el sitio web.

En concreto, si se desgrana su definición de forma natural podemos descifrar mejor en qué consiste el proceso de la analítica web:

**1. Recopilar información (medición y recolección):** accediendo a las herramientas analíticas, ya sea Google Analytics o Search Console, como cualquier otra, que nos proporcionan datos que configurar y a los que sacar partido. La estrategia es la clave para su interpretación. Enfrentarse a los datos de forma objetiva no es sencillo; además, es aconsejable no partir con prejuicios porque podemos llevarnos a equívocos. Lo primero que hay que hacer es asegurarse de que todas las páginas del *site* están bien etiquetadas y son accesibles por los usuarios. Como se aprecia en la siguiente ilustración.

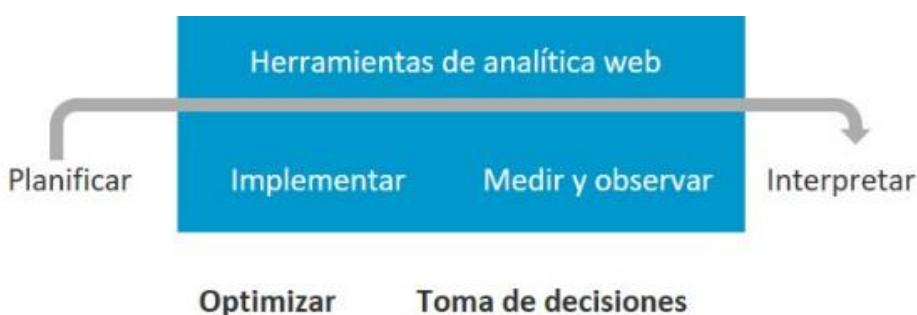


Figura 1. Herramientas de analítica web. Fuente: elaboración propia.

Un *site* es un sitio web o cibersitio. Es una colección de páginas web relacionadas y comunes a un dominio de Internet o subdominio en la World Wide Web en Internet.

**2. Realizar las preguntas correctas (análisis):** una vez se ha asegurado de que todas las páginas del sitio están disponibles para ofrecer datos y métricas, se deben interpretar estos datos correctamente. Lo primero es alinearnos con la estrategia y realizar las preguntas oportunas, y si las métricas básicas de visitantes únicos, visitas, páginas vistas, porcentaje de rebote y promedio de tiempo en la web son las esperadas.

Si la respuesta es negativa, lo mejor es hacer un regreso a los principios y preguntarse las cinco claves periodísticas: **qué, quién, cuándo, dónde y por qué**; y una sexta: **cómo**.

Según la Asociación Española de Analítica Web, la analítica web es la «recopilación, medición, evaluación y explicación racional de los datos obtenidos de Internet, con el propósito de entender y optimizar el uso de la página web de la organización» (SeedRocket, 2021).

## Qué están haciendo los usuarios, quién son estos usuarios y cuántos

Seleccionando un rango temporal para ir fijando la fecha raíz del problema que queremos solucionar, y gracias a los parámetros de *engagement* y profundidad de la visita, llegaremos con mayor claridad al por qué: en qué páginas no pueden acceder al sitio, por ejemplo, ya sea porque hay un error técnico o porque hay páginas que no se sirvan y devuelvan un error 404.

**Engagement** es el compromiso que adquieren los usuarios con la marca y lo expresan a través de los medios sociales.

Por tanto, analizar los datos conseguidos mediante la analítica web no es sencillo. La interpretación se debe basar en los KPI (*key performance indicators*) definidos y en los objetivos que la organización ha establecido en su estrategia general de *marketing*.

**3. Realizar los reportes adecuados (reporte):** tan importante como saber leer los datos para interpretarlos es generar unos cuadros de reporte sencillos, útiles y eficaces. Muy posiblemente estos reportes los realices para el CEO de tu empresa o para un comité directivo. Muchos de ellos no comprenderán lo que les mandas si es muy técnico y con muchos datos. Lo mejor es definir conforme al objetivo señalado las métricas relevantes para ello, ya sean visitas únicas, páginas vistas, visitas u otro tipo de métrica más sofisticada y, además, hacer que sean comprensibles.

**4. Mejorar y optimizar un sitio web:** la analítica web te dirá lo que están haciendo los usuarios en tu sitio web, pero no lo que no están haciendo, esto es algo que debes interpretar. Dicho esto, nos sirve diferenciar un reporte web con datos de un análisis web que conlleva un esfuerzo adicional de interpretación en orden al objetivo de la compañía. El analista web deberá obtener sus propias conclusiones y hacer recomendaciones, buscando siempre el conocimiento y la reflexión, no simplemente el dato.

Este proceso nos permitirá, entre otros:

- ▶ Diferenciar las características morfológicas y cuantitativas. Dependiendo de si es una web de contenidos o una de comercio electrónico, los objetivos y las métricas serán diferentes.
- ▶ Fijar objetivos (KPI) de crecimiento, ventas para tu página. Solo así se puede comprobar si la estrategia va encaminada al cumplimiento de los objetivos.
- ▶ Describir cuáles serán las métricas clave de nuestro *site* para medir su evolución y la respuesta de los usuarios.

**5. Detección de errores en el sitio web:** analizar los resultados de las métricas y extraer las conclusiones en relación con los objetivos para poder utilizar esas conclusiones en la mejora de tu sitio web.

Testear los cambios en comparación con la versión antigua para optimizar el rendimiento de la web.

## 6.3. Métricas básicas

Llegamos a uno de los aspectos fundamentales y con los que vamos a trabajar todos los días en la analítica web: las métricas. Podemos seleccionar de entre las docenas de métricas que se usan, aunque algunas de ellas son las relevantes y debemos tener claras:

- ▶ **Visitantes únicos:** número de visitantes no duplicados que han accedido al sitio web. La primera vez que un visitante accede a nuestra web se genera una *cookie* que es única y que se utilizará para identificarlo. Cuando el usuario tiene deshabilitadas las *cookies* se utilizan las direcciones IP, pero esto genera un dato poco preciso. Dependiendo del rango temporal que seleccionemos se puede presentar una problemática adicional. Un usuario puede visitar nuestra web en tres días distintos, generando una visita única cada día y, si analizamos el intervalo, nuestra herramienta nos dirá que hemos tenido tres visitantes únicos, cuando en realidad solo hemos tenido uno.
- ▶ **Visitas:** una visita es un conjunto de interacciones que tienen lugar en un sitio web durante un período determinado. Las visitas representan el número de sesiones individuales iniciadas por todos los usuarios para llegar al sitio web. Si un usuario permanece inactivo en su sitio durante al menos 30 minutos, toda actividad posterior se atribuirá a una nueva sesión. Los usuarios que abandonen su sitio y vuelvan en menos de 30 minutos se considerarán como parte de la sesión original.
- ▶ **Páginas vistas:** son las páginas a las que se accede dentro del sitio web. Las páginas están compuestas de *hits*, impactos o cantidad de respuestas de un servidor a la petición de un navegador o robot. Así que, si accedemos a una página que tiene cinco fotos y un CSS, el número de *hits* será alto, pero el número de páginas vistas aumenta solo en uno. La de las páginas vistas es una de las métricas más utilizadas por los equipos comerciales a la hora de hacer estimaciones para campañas y controlar su inventario, tanto el vendido como el no vendido.

- ▶ **Páginas/visita:** promedio de páginas vistas durante una visita al sitio. Este indicador tiene que ir creciendo con el tiempo dentro de la página web. En caso de no suceder así, puede indicar anomalías técnicas en el sitio o, sencillamente, que los usuarios no aceptan la estrategia de navegación que les disponemos, de forma que haya que revisarlo mirando más parámetros. En cierta forma, es uno de los parámetros que indican la sanidad de nuestro portal.
- ▶ **Duración media de la visita:** es la duración media de una sesión. Este indicador es muy relevante, ya que refleja el interés por nuestro sitio de los usuarios que han accedido a él. Junto con la de visitas, es una métrica de las más importantes en analítica web.
- ▶ **Porcentaje de rebote:** porcentaje de visitas de una sola página y en las que el usuario ha abandonado el sitio en su página de entrada. Es importante medir el porcentaje de rebote de las principales páginas de entrada para aplicar soluciones en caso de que no vayan según lo esperado. Suele ser sinónimo de la calidad del contenido, pero hay un caso en que esta métrica, aun siendo muy elevada, no debe preocuparnos: si nuestro sitio web solo consta de una página, los usuarios no tienen más alternativa que abandonar el sitio web habiendo consumido tan solo una página.

Una visita es un conjunto de interacciones que tienen lugar en un sitio web en un período determinado. Por ejemplo, una única visita puede contener páginas vistas, eventos, interacciones sociales, variables personalizadas y transacciones de comercio electrónico. Se puede decir que una visita es el elemento que engloba las acciones del visitante en su sitio.

**Por defecto una visita dura 30 minutos en Google Analytics**, momento en el que se elimina la cookie \_\_utmb. Pero esta se ve modificada con cada evento y, por tanto, cada una de las solicitudes adicionales restablecerá la caducidad de la cookie en 30 minutos. ¿Qué acciones implican que este tiempo se vea restablecido? Por ejemplo,

con una página vista, un evento, una interacción social o una transacción.

*Las cookies son archivos con tecnología capaz de almacenar y recuperar datos de un equipo terminal de una persona física o jurídica que utiliza para cualquier fin un servicio de la sociedad de la información.*

Para el caso de inactividad en la página, la primera visita que se inicia cuando el usuario llega al sitio acaba a los 30 minutos y la cookie\_\_utmb se elimina de su ordenador. Si después el usuario sigue navegando por el sitio web (viendo otra página, activando otro evento, etc.), Google Analytics instala una nueva cookie\_\_utmb con caducidad de 30 minutos y se inicia una nueva visita. Todas las cookies\_\_utmb que estén dentro de los 30 minutos comentados seguirán activas, aunque por ejemplo haya estado inactivo durante 29 minutos, y se considerará una única visita.

La otra forma de finalización forzosa de la cookie\_\_utmb se produce por la finalización del día. A las 00 horas del nuevo día esta *cookie* por lo general caduca, de forma que si la primera visita finaliza a las 11:59:59 p. m. del 28 de julio y la segunda empieza a las 12:00 a. m. del 29 de julio, aunque está dentro de los 30 minutos que anteriormente comentamos, se considerarán dos visitas.

Otro aspecto determinante es el de la caducidad por cambiar la fuente de entrada. Imaginemos que hacemos una búsqueda en un buscador y uno de los resultados nos llama la atención, hacemos clic y navegamos por esa web. Una vez hemos visto lo que nos interesa, lo dejamos sin más. A continuación, nos llega un boletín por correo electrónico sobre el que hacemos clic y aterrizamos en la misma web que antes visitamos. Esto se refleja en el panel como dos visitas.

Otras métricas interesantes para tener en cuenta son:

- ▶ **Porcentaje de visitas nuevas:** porcentaje estimado de visitas realizadas por primera vez.
- ▶ **Sesiones:** interacciones de un usuario en un sitio web. Termina una vez que transcurren 29 minutos de inactividad o cuando un mismo usuario vuelve al sitio web desde una fuente de tráfico diferente.
- ▶ **Tasa de conversión:** es el porcentaje resultante de dividir los resultados entre los visitantes únicos, teniendo en cuenta que los resultados pueden ser el número de compras, el número de registros, etc. (objetivo).
- ▶ **Métricas para obtener el ROI:**
  - Pago por tiempo: se paga por un determinado tiempo de exposición en un determinado contenido.
  - CPM (coste por mil): coste por cada mil impresiones. Se paga por la impresión.
  - CPC (coste por clic): se paga por la visita.
  - CPL (coste por *lead*): se paga por el *lead* (registro o *e-mail* normalmente).
  - CPA (coste de adquisición): se paga por venta conseguida una cantidad fija.
  - *Revenue sharing*: se paga un porcentaje de lo que se vende.
- ▶ **Medición de resultados:**
  - eCPM: CPM efectivo (el coste real final).
  - eCPC: CPC efectivo.
  - eCPL: CPL efectivo.
  - eCPA: CPA efectivo.

- ROI: *return on investment*, rentabilidad.
- eCPM: coste campaña / (impresiones/1000).
- eCPC: coste campaña / visitas reales.
- eCPL: coste campaña / *leads* reales.
- eCPA: coste campaña / ventas reales.
- ROI: margen (o ingresos) / coste campaña.

► **Métricas de rendimiento:**

- Impacto-visita. *Click through* (CTR).
- Visita-*lead*. *Lead through* (LTR).
- Visita/*lead*-venta. Tasa de conversión (CR).
- Venta-ingreso. Pedido medio.
- Venta-beneficio. Margen de contribución.
- Beneficio-inversión. ROI.
- *Click through* (CTR): porcentaje de clics (visitas) sobre impresiones.
- *Lead through* (LTR): porcentaje de *leads* sobre visitas.
- Tasa conversión (CR): porcentaje de ventas sobre visitas o *leads*.
- *Revenue sharing* (RS): porcentaje del importe de venta conseguido.

► **Indicadores de eficacia:**

- Tiempo: volumen de tráfico. Se miden las impresiones por día y se transforma a CPM.

- CPM: *click through CTR*. Se miden las visitas (clics) y se calcula el CTR y CPC.
- CPC: *lead through LTR*. Se miden los *leads* y se calcula el LTR y CPL.
- CPL: conversión a la venta TC. Se miden las ventas (pedidos) y se calcula la TC y el CPA.
- CPA: *revenue sharing*. Se mide el importe de venta y se calcula el RS.

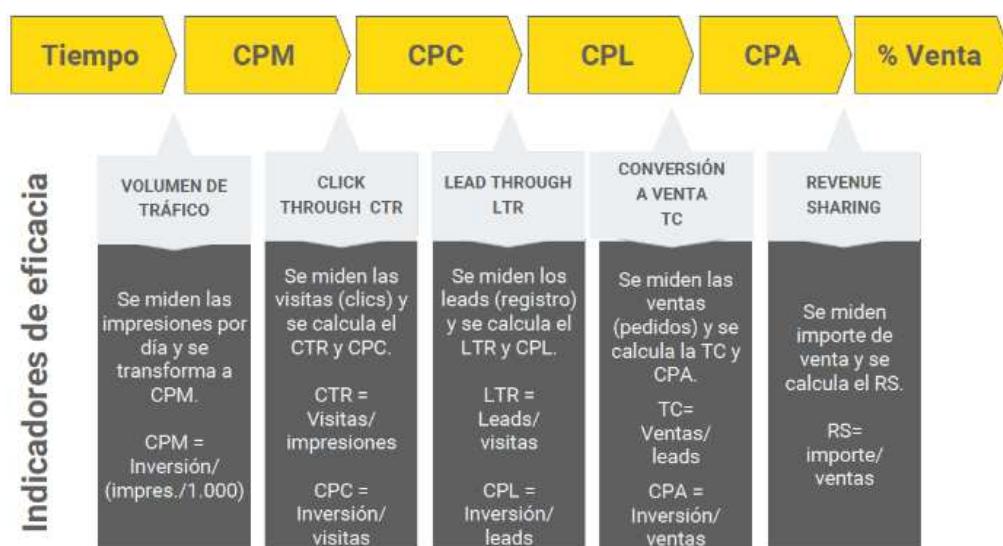
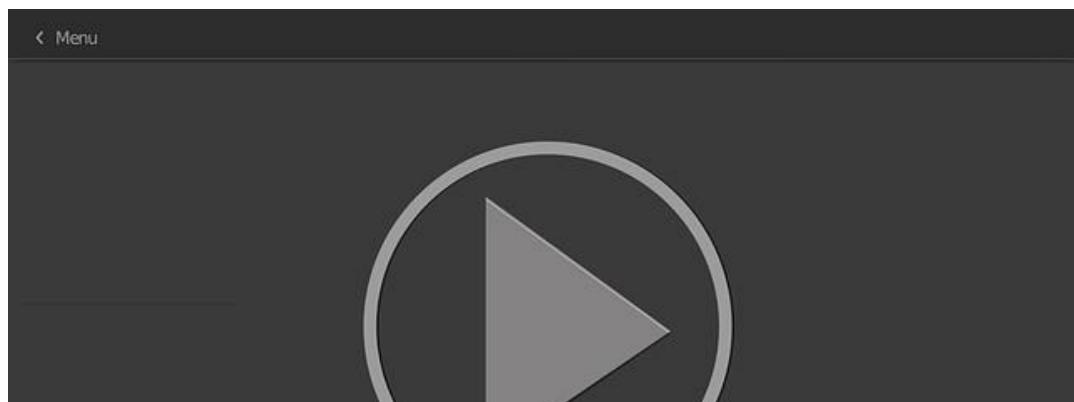
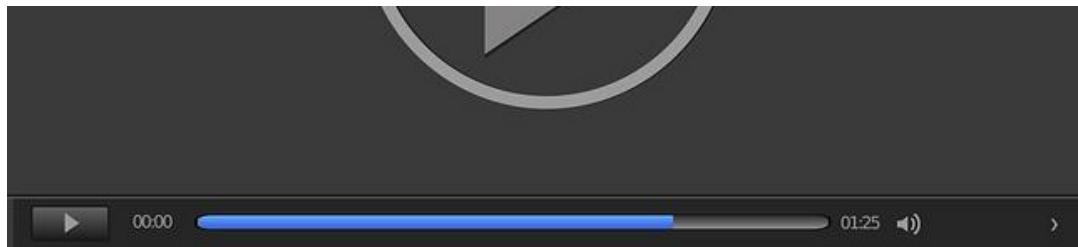


Figura 2. Indicadores de eficacia. Fuente: Google Activate, 2016.

Antes de continuar con el siguiente apartado, accede al vídeo *Segmentación y posicionamiento*.





---

Accede al vídeo:

<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=13534bf5-e3d0-407b-9e7d-acd400c84000>

---

Vídeo. *Segmentación y posicionamiento.*

## 6.4. Tipos de analíticas web

La analítica web se clasifica en dos tipos: la **cuantitativa** y la **cualitativa**:



Figura 3. Tipos de analítica web. Fuente: Taringa, 2007.

La **analítica cuantitativa** es lo que denominamos las métricas y se han visto en apartados anteriores, como pueden ser:

- ▶ Un simple número: visitas: 2114.
  - ▶ Pueden darnos referencias interesantes acerca de nuestros usuarios:
    - 1123 usuarios de España.
    - 35 % provienen de buscadores.
    - 2 % de [www.enlace.com](http://www.enlace.com)
    - 35 % hablan inglés.

- 70 % usan Mozilla Firefox.
- ▶ Pueden darnos referencias interesantes acerca de la evolución del negocio:
  - Hemos crecido un 50 % en número de usuarios en los últimos 3 meses.
  - Hemos duplicado la cifra de ventas.
  - Antes de la expansión internacional, un 5 % del tráfico mensual era extranjero y ahora es un 30 %.

Medio social	KPI	Enero	Febrero	Marzo	Total
Google Analytics	Visitas totales				
	Tiempo de permanencia				
	Reservas				
	Porcentaje de conversión				
	Porcentaje de rebote				
	Visitas 1ª vez				
	Porcentaje del total				
	Reservas 1ª vez				
	Porcentaje de conversión clientes nuevos				
	Páginas webs de referencia				
	Procedencia: países				
	Productos vendidos				

Tabla 1. Informe de métricas y KPI. Fuente: elaboración propia.

La **analítica cualitativa** nos aporta datos más visuales, relacionados con el comportamiento del usuario (*scroll*, clics...). En Google Analytics lo llamamos **dimensiones**.



Figura 4. Analítica cualitativa. Fuente: Ciotti, s.f.

En este caso, lo más difícil no es obtener el dato, sino ser capaz de interpretarlo adecuadamente y tomar las decisiones oportunas.

- ▶ Los datos nos ocultan cosas.
- ▶ Los datos no nos explican el porqué de los hechos.
- ▶ No tiene por qué haber solo una interpretación.

Hasta ahora hemos estado hablando de las métricas puras y duras, que son herramientas o indicadores del comportamiento de los usuarios en nuestro sitio web de forma objetiva. Entramos ahora en uno de los conceptos más usados y que es una métrica cualitativa: el *engagement*.

## El *engagement*

*Engagement* es una palabra con significado abierto y con una difícil y unívoca definición. Si lo centramos en el *marketing*, la idea de *engagement* se relaciona no solo con los consumidores de un sitio web, sino como define Avinash, es la acción de los usuarios en relación con una marca, compañía o producto. Este *engagement* puede ser positivo o negativo.

En general, decimos que una persona está vinculada, está *engaged* con una marca, compañía o producto cuando se siente preocupada y comprometida con ella.

Otra forma de verlo, muy actual y relacionada con redes sociales es la siguiente: no es lo mismo decir «me gusta» que **mantener una conversación continua con la marca**.

Cuando un usuario interactúa con la marca se toma el tiempo no solo de decir si le gusta o no, sino de **hablar de la experiencia que le ofrece**.

¿Qué métricas son las que determinan el *engagement*?

- ▶ Usuarios únicos.
- ▶ La última vez que nos visitó el usuario.
- ▶ Frecuencia de la visita.
- ▶ La profundidad de la visita.
- ▶ Tiempo empleado en el sitio.

Gracias a todo ello podemos llegar a obtener el nivel de *engagement* que tiene un usuario con nuestra compañía, marca o producto, pero difícilmente vamos a poder percibir si este es positivo o negativo. Pongamos un ejemplo:

Si somos un portal de Internet que quiere ofrecer un buen servicio de correo electrónico y los usuarios acceden desde la página de inicio y por más de 20 minutos, navegan en él, envían correos, etc., podríamos decir que los usuarios están satisfechos probablemente con nuestro portal.

Si, en cambio, los usuarios que quieren acceder a ese producto tan importante como es el correo electrónico para nuestro portal, están 20 minutos navegando por nuestro portal para encontrar el enlace que le dé acceso a ese servicio de correo electrónico, está claro que algunos usuarios no van a considerar que la web tiene un buen *engagement*.

Pues bien, encontramos a un usuario que ha visitado quince veces nuestro sitio web en el último mes y podríamos pensar que es una buena señal del *engagement* de este individuo hacia nuestra empresa. En cambio, lo que sucede es que nuestra decisión de dejar sin soporte a la versión de 2009 de nuestro sistema operativo ha provocado que este usuario haya accedido tantas veces precisamente para intentar localizar unos controladores para su ordenador, lo que le ha obligado a localizarlos por algún lado de nuestra web y con resultados dispares.

Con este ejemplo se puede ver que, cuando alguien nos diga que tiene una métrica que mide su *engagement*, en realidad nos estará diciendo que lo que mide es su nivel de *engagement* —alto o bajo—, pero no el tipo de *engagement* —positivo o negativo—. Posiblemente, si las métricas no terminan de declararnos el problema, aunque sí orientarnos, **deberíamos buscar el feedback o respuesta del usuario complementando todos aquellos datos con un elemento emocional como podría ser una encuesta.**

Lo importante es que cada empresa defina su propio concepto de *engagement*.

## La usabilidad

La usabilidad (del inglés *usability*) es la medida en la cual un sitio web puede ser usado por usuarios específicos para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un contexto de

uso concreto, mediante test de usuarios o directamente con la analítica web.

¿Qué métricas son decisivas en la analítica aplicada a la usabilidad?

- ▶ El tiempo en página o el porcentaje de rebote. Relacionado con el gusto o no de los usuarios hacia nuestro sitio web.
- ▶ *Funnels* o embudos de conversión. Son aquellos que hacen seguimiento de los procesos de nuestro sitio. El ejemplo habitual es: seguir y analizar un proceso de compra para ver en qué punto del proceso abandona el usuario o el ratio de conversión.
- ▶ Test A/B o test multivariante para testar distintas versiones de una misma página y comprobar cuál funciona mejor.
- ▶ Cuántos visitantes son nuevos y cuántos recurrentes, la frecuencia de la visita o la duración de la misma. Estos datos nos permiten conocer si nuestro sitio agrada a los usuarios o no. Tener usuarios recurrentes o no, que nos visitan frecuentemente y durante mucho tiempo es el mejor síntoma de que nuestro sitio no presenta problemas graves.
- ▶ Páginas de destino (aquellas de nuestro sitio a las que llegan los usuarios) así como las páginas de salida (la última página visitada antes de que el usuario abandone nuestro sitio). Observar con detenimiento las páginas de salida porque son las principales candidatas a albergar un problema de usabilidad que esté condicionando al usuario a abandonar el sitio.
- ▶ Páginas de salida que tengan un alto porcentaje de salidas serán las primeras que debemos comprobar.

## El comportamiento del usuario

Hoy en día la velocidad de descarga del sitio es uno de los factores fundamentales

que Google toma para el posicionamiento SERP de los sitios web. Pero no solo para los buscadores, para nuestros usuarios es también fundamental y es uno de los factores de la experiencia de usuario que debemos cuidar. La descarga lenta de un sitio es uno de los principales factores de abandono por los usuarios. Hoy en día esto ya no solo sucede en el caso de una web, sino también del móvil, las tabletas y todo tipo de plataformas donde nuestros contenidos estén disponibles para los usuarios.

¿Con qué herramientas podemos medir todo ello? Las herramientas para *webmasters* de Google (**Google Webmaster Tools**) proporciona información de la velocidad de descarga del sitio web que, unido a Google Analytics, si lo hemos vinculado con la cuenta de Google Webmaster Tools, nos da información de la velocidad del sitio, así como sugerencias para mejorar la velocidad de descarga.

También podemos ver estos datos accediendo a los informes de velocidad del sitio en la sección Comportamiento de nuestro Google Analytics.

¿Qué puede influir en la velocidad de descarga? Los *banners* de publicidad y la ubicación de los usuarios. Ambos son factores de corrección que deben tenerse en cuenta para evaluar los informes de velocidad. La tecnología con la que estén hechos los *banners* y la forma de mostrarlos puede hacer que se ralentice notablemente la descarga de la página y, por tanto, la experiencia del usuario. Por otro lado, si el sitio web tiene muchas visitas de usuarios residentes en países lejanos a la ubicación de los servidores donde esté alojado, dicho sitio web puede experimentar lentitud en la descarga.

## 6.5. Herramientas de medición

Las principales herramientas de medición analítica vienen dadas por aquellas que miden las *cookies* o los textos de JavaScript.

Por eso a la hora de decidir qué herramienta se va a utilizar, se deben plantear algunas cuestiones como:

- ▶ Las necesidades de nuestro negocio.
- ▶ La precisión y el volumen de datos.
- ▶ El precio: puede variar desde 0 € hasta los 100 000 € anuales.
- ▶ La herramienta más divulgada es Google Analytics, por su potencia y gratuidad.

En la siguiente figura se pueden ver algunas de las mejores herramientas web que existen en el mercado.



Figura 5. Mejores herramientas web para analítica. Fuente: Google Activate, 2016.

Estas son las más conocidas, pero hay más:



Figura 6. Herramientas de analítica web cualitativas. Fuente: Google Activate, 2016.

Evidentemente, como la más importante tenemos que citar la consola de Google, denominada Google Marketing Platform, ya que, además de ser gratuita, es muy fácil de instalar. Consiste en introducir un código JavaScript en todas las páginas del sitio web para que estas envíen la información que procesar.



Figura 7. Plataforma *marketing* Google. Fuente: Bunker DB, 2018.

Las plataformas que integra Google Marketing Platform son Google Analytics 360 Suite y los productos de DoubleClick. También incluye las siguientes herramientas:

- ▶ Analytics: para conocer mejor a los usuarios que visitan la web y mejorar sus experiencias.
- ▶ Data Studio: herramienta que nos permite crear informes personalizables.
- ▶ Optimaze: probar diferentes versiones de la web o de las campañas.
- ▶ Surveys: realizar encuestas y obtener opiniones reales fiables.
- ▶ Tag Manager: gestionar y crear etiquetas para poder medir aquellos eventos u objetivos que por defecto no mide ni Google Analytics ni Google AdWords.
- ▶ Display y Video 360: llegar a las audiencias conectadas a través del vídeo.
- ▶ Search Ads 360: obtener datos en tiempo real de las campañas.

## 6.6. Referencias bibliográficas

Bunker DB. (28 de junio de 2018). Google Marketing Platform, el radical reboot al que apuesta Google [Mensaje en un blog]. *Bunker DB Blog*. <https://bunkerdb.com/blog/tendencias/google-marketing-platform/>

Ciotti, G. (s.f.). 7 Marketing Lessons from Eye-Tracking Studies [Página web]. Neil Patel. <https://blog.kissmetrics.com/eye-tracking/>

SeedRocket. (2021). Taller de analítica web [Página web]. *SeedRocket*. <https://www.seedrocket.com/eventos/eventos-talleres/taller-de-analitica-web/>

Taringa. (2007). Dónde miran nuestros ojos cuando navegamos por internet [Página web]. Taringa. <http://www.taringa.net/posts/info/912167/Donde-Miran-Nuestros-Ojos-Cuando-Navegamos-Por-Internet.html>

## Las sesiones en Google Analytics

Google. (2021). Cómo se define una sesión web en Universal Analytics [Página web].

*Ayuda de Analytics.* <https://support.google.com/analytics/answer/2731565?hl=es>

En este artículo se especifica cómo configurar una sesión en Analytics, algo imprescindible para trabajar en la analítica web.

Se especifica cuánto dura una sesión y cómo etiquetar las diferentes posibilidades que Google nos ofrece por defecto.

## Mapas de calor

Mena, D. (8 de septiembre de 2015). Mapas de calor (heatmaps): Qué son, para qué sirven y por qué usarlos [Página web]. *Wanaleads*. <http://wanaleads.com/mapas-de-calor-heatmaps-para-optimizar-web/>

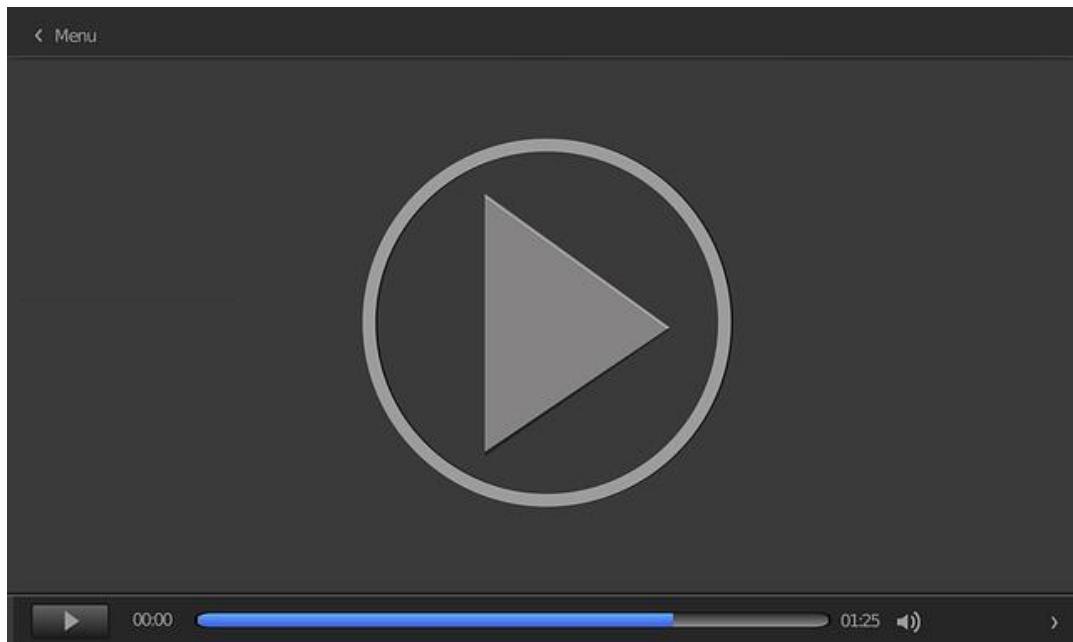
En este artículo se presenta cómo funcionan los mapas de calor y qué medidas básicas se deben utilizar para recoger toda la información.

Si aún no sabes lo que son los *heatmaps* o mapas de calor, David Mena nos muestra en este artículo qué zonas y elementos de nuestra web generan una mayor interacción con los usuarios

## Tutorial de Google Analytics

tímosfera. (21 de enero de 2015). *Tutorial Google Analytics Español 2015* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=i2rbGMHiR1Y>

En este vídeo se explican las claves principales para entender cómo funcionan las estadísticas que proporciona Google Analytics.



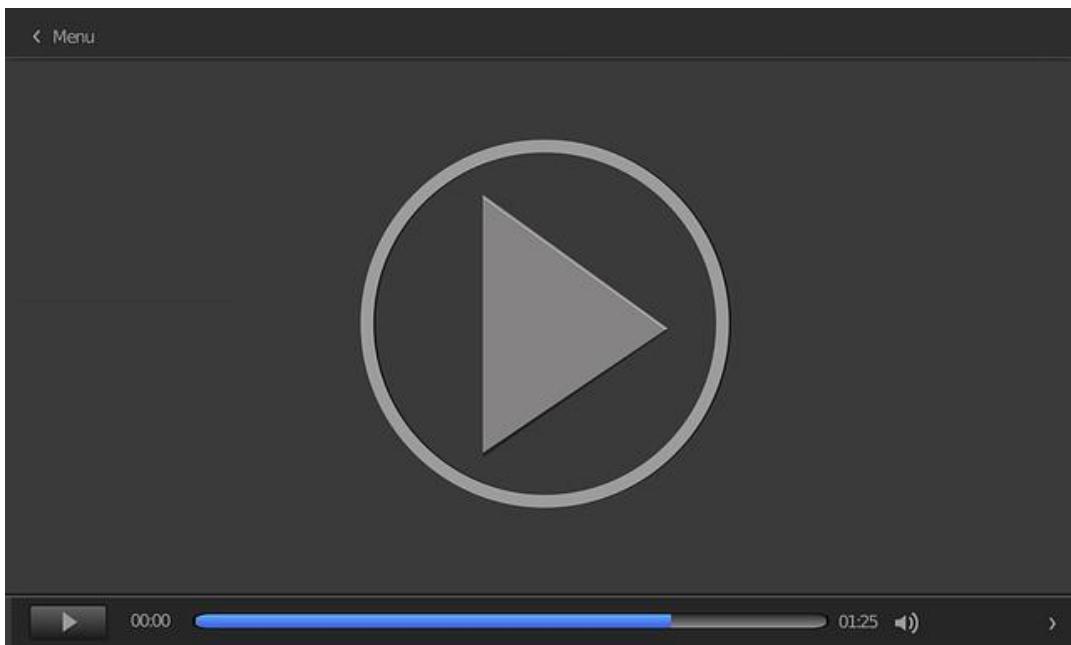
Accede al vídeo:

<https://www.youtube.com/embed/i2rbGMHiR1Y>

## Eye tracking

MarketingAmarillasCL. (25 de enero de 2013). *Seminario de Neuromarketing y Eyetracking - Publiguiás 2011* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=VvY8PyeKOJs>

El martes 27 de septiembre del 2011 se presentó a clientes Publiguiás, medios y agencias el seminario de *neuromarketing y eye tracking*, a cargo de Jurgen Klaric y Juan Pablo Rodríguez. En este vídeo podéis encontrar un resumen.



Accede al vídeo:

<https://www.youtube.com/embed/VvY8PyeKOJs>

1. ¿Qué métrica no es básica para el análisis web?

  - A. Visitas.
  - B. Porcentaje de rebote.
  - C. Tiempo real.
  - D. Todas son correctas.
  
2. ¿Puede un sitio web tener más páginas vistas que visitas?

  - A. No.
  - B. Sí.
  - C. Solo a veces.
  - D. Todas son correctas.
  
3. ¿Cuándo finaliza, en general, una sesión en Google Analytics?

  - A. A los 30 minutos de iniciarse la sesión.
  - B. A los 20 minutos de iniciarse la sesión.
  - C. A los 15 minutos de iniciarse la sesión.
  - D. Cuando apagas el ordenador.
  
4. ¿En un test A/B hay que insertar alguna etiqueta de Google Analytics en nuestra web?

  - A. Nunca.
  - B. Solo si no funciona.
  - C. Un código que se inserta en la página original y en la de objetivo.
  - D. Todas son correctas.

**5.** La velocidad de un sitio es importante para:

- A. La experiencia de navegación del usuario.
- B. La experiencia de usuario y el posicionamiento.
- C. Solo el posicionamiento.
- D. A y B son correctas.

**6.** Elige la opción que corresponda a una métrica:

- A. Visitas.
- B. Usuarios en España.
- C. Provienen de Yahoo.
- D. Todas son correctas.

**7.** ¿Qué herramientas sirven para analizar la velocidad del sitio web?

- A. Google Webmaster Tools.
- B. Photoshop.
- C. Yslow.
- D. A y C son correctas

**8.** Relaciona el indicador con lo que mide:

Transacciones.	1
Visitas totales.	2
Porcentaje de rebote.	3
Webs de referencia.	4

A	Visitas a la web.
B	Número de ventas.
C	Visitas muy cortas.
D	Páginas de procedencia.

**9.** Las herramientas web: ¿en qué indicador basan sus datos?

- A. *Cookies*.
- B. Código de Java.
- C. A y B son correctas.
- D. Ninguna es correcta.

**10.** ¿A qué corresponde la siguiente definición? Es la medida en la cual un sitio web puede ser usado por usuarios específicos para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un contexto de uso concreto, mediante test de usuarios o directamente con la analítica web.

- A. Analítica web.
- B. Usabilidad.
- C. *Engagement*.
- D. Ninguna es correcta.

Gobierno del Dato y Toma de Decisiones

---

## Tema 7. Marketing relacional y CRM

# Índice

[Esquema](#)

[Ideas clave](#)

[7.1. Introducción y objetivos](#)

[7.2. Introducción al marketing relacional](#)

[7.3. Características y beneficios del marketing relacional](#)

[7.4. CRM: definición y características](#)

[7.5. Factores clave y bases para un buen CRM](#)

[7.6. Referencias bibliográficas](#)

[A fondo](#)

[¿Por qué una plataforma de marketing relacional?](#)

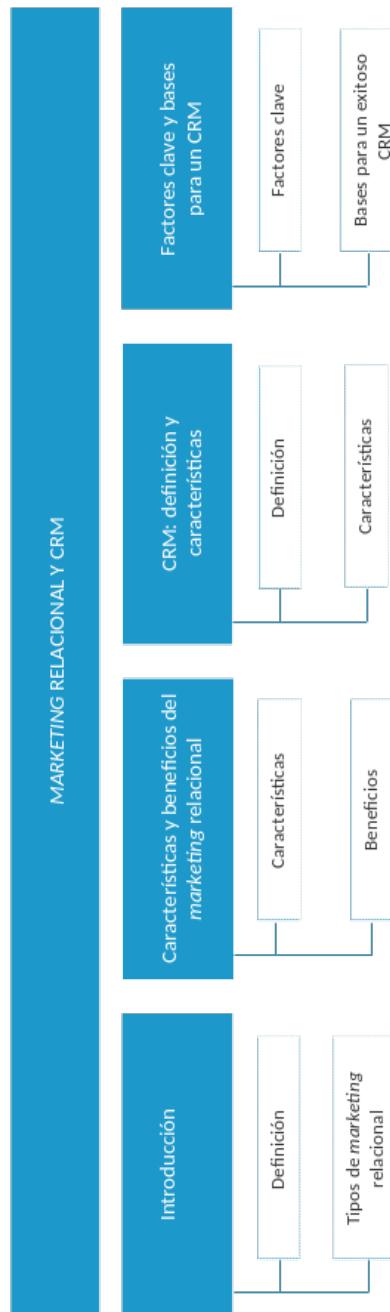
[Brain centric marketing: la evolución del customer centric para un nuevo marketing](#)

[Marketing digital vs. marketing tradicional: cinco diferencias](#)

[¿Por qué necesitamos un CRM?](#)

[Test](#)

# Esquema



## 7.1. Introducción y objetivos

Para estudiar este tema, debes leer el capítulo 17, «Marketing directo y marketing en línea: establecimiento de relaciones directas con los clientes», páginas 494-525 del libro:

Kotler, P., y Armstrong, G. (2012). *Marketing*. Pearson. Disponible en el aula virtual.

Los objetivos de este tema son:

- ▶ Entender la definición de *marketing* relacional, sus características y sus diferentes clasificaciones.
- ▶ Conocer la importancia de construir y mantener relaciones beneficiosas con los clientes, a través de los CRM y del modelo RATER.

## 7.2. Introducción al marketing relacional

Según el profesor Philip Kotler (Kotler y Armstrong, 2013), los tres primeros pasos de cualquier proceso de *marketing* son:

- ▶ Entender el mercado y las necesidades de los clientes.
- ▶ Diseñar una estrategia adecuada con el foco en el cliente.
- ▶ Diseño de programa de *marketing*.

Todos quedan pendientes del éxito o fracaso del cuarto y definitivo paso:

- ▶ La consecución de relaciones beneficiosas con los clientes.

Reflexionemos un poco y veamos que todo queda en nada si somos incapaces de alcanzar el «Santo Grial» al final de nuestro plan de convencer a nuestros clientes de que cualquier cosa en el mercado que necesiten, antes de mirar en otro lado, evaluar otras alternativas, realizar cálculos, etc., antes que todo ello, **estamos nosotros**.

Pero esto no se consigue con los mejores investigadores, los mejores psicólogos o los mejores financieros. Se consigue simplemente con **los mejores vendedores**.

Estamos hablando de **construir relaciones**. Y no solo de construirlas, sino también de saber **mantenerlas** luego. Piensa en una planta, la más bonita del mercado. ¿Qué pasa si no la riegas? ¿Qué pasa con aquellos amigos de la infancia que hace siglos que no hablas con ellos? Ya: que siguen siendo tus amigos..., pero ellos hacen su vida y tú la tuya. ¿Qué pasa en las familias cuando no se relacionan? ¿Te acuerdas de lo que estudiaste hace un año si no lo repasas?

Pues con los clientes sucede igual que con esa familia, esos amigos, ese club, esos apuntes: si no mantienes la relación, esta se marchita. No es tan importante conseguir clientes como **fidelizarlos**. Como la fama, el dinero o el éxito: lo importante no es conseguirlos, sino mantenerlos.

Figura 1. Obtener y mantener al cliente para maximizar la rentabilidad. Fuente: elaboración propia.

En el *marketing* actual todo pasa por mantener las relaciones duraderas  
con los clientes.

## Definición

El *marketing* relacional es, de hecho, un retorno a los orígenes del *marketing*: una vuelta, en el momento actual usando la tecnología, a la relación casi «familiar» con el cliente.

Es evidente que la gran ventaja actual en el mundo del *big data* es la gran cantidad de información que podemos tener de esa gran familia de clientes a los que **podemos conocer todos sus movimientos, hábitos y preferencias**, y así personalizar las comunicaciones con cada uno de ellos como si fuera el **único**.

Además, debido a las tendencias actuales del mercado, donde la competitividad es máxima, las compañías asumen que deben de acercar sus soluciones lo más posible «a la medida» del consumidor. El *marketing* relacional es una consecuencia de la situación actual, donde los consumidores (y más en Internet) tiene un sinfín de opciones. La respuesta de las compañías es desarrollar estrategias de *marketing* de relaciones que integren el cliente dentro de la compañía, **creando una relación de fidelidad permanente que, además, pueda ser referencia para nuevos clientes**.

## Tiempos de *marketing* relacional

«Un general victorioso, primero gana y después emprende la batalla. Un general perdedor, primero emprende la batalla y después espera ganar»

(Sun Tzu).

Este planteamiento es tan válido para el entorno bélico como para el deportivo, el académico y, por supuesto, para el mundo de la empresa y el *marketing*. **Si no sabemos qué queremos y a dónde nos queremos dirigir, difícilmente tendremos éxito en nuestro avance.** Es imposible ganar un partido sin haber sabido qué queríamos hacer, qué táctica queríamos emplear, qué *timing* aplicar y

qué recursos necesitábamos para alcanzar la victoria.

Este estudio de los planes al objeto de llevar a cabo nuestra estrategia forma parte del **marketing de análisis**, que obviamente tienen su continuación en el siguiente paso. Este consistiría en el despliegue de todas esas actuaciones (acciones) encaminadas a demostrar que nuestro **marketing analítico** ha «servido para algo». Este conjunto de recursos que la empresa dispone para la ejecución de sus objetivos predefinidos forma parte del *marketing relacional*. Ignasi Vidal, en su libro *Cómo conquistar el mercado con una estrategia CRM*, dice que es relacional «porque se trata de una serie de herramientas y recursos que persiguen la relación y la vinculación con un objetivo definido, que no es otro que la mente del consumidor y el medio de conquista, la relación» (Vidal, 2004, p. 153).



Figura 2. *Marketing* relacional. Fuente: Colina, 2015.

Por tanto, para conquistar la mente de nuestros consumidores tenemos que activar las «palancas» (las variables de nuestro *marketing mix*) del modo y manera que más favorezcan a nuestros intereses y que estén a nuestro alcance. Para ello **debemos ser conscientes de nuestras limitaciones y también**, lógicamente, de **nuestras fortalezas**. El «balanceo» en los recursos a la hora de utilizarlos dependerá mucho del momento, del estado en que vea en mi empresa, de las características del portfolio del producto/servicio por ofrecer y, cómo no, del poder económico que disponga la empresa para llevar a cabo los objetivos de mi plan.



Esto exige además una reflexión sincera sobre si realmente estoy en disposición de ganar la batalla por la conquista de la mente del consumidor o es mera utopía. Pero una vez que haya pasado este autotest previo y me salga la respuesta «Yes, we can», entonces debo decidir qué tipo de *marketing* relacional quiero en este momento, puesto que, como mencionaremos más adelante, **hay distintos enfoques de marketing relacional** que influirán decisivamente sobre la estrategia CRM por implantar y sus beneficios.

Según Vidal (2004), se definen **cinco tipos** de *marketing* relacional, que muy sucintamente serían:

- ▶ **Marketing relacional básico:** la empresa se dedica a vender sin contacto alguno con el cliente. Productos de gran consumo con bajo margen unitario. Pensemos en agua mineral o galletas.
- ▶ **Marketing relacional reactivo:** la empresa anima al cliente a intercambiar opiniones y experiencias. Es necesario, desde el punto de vista de estrategia CRM, recibir el *input* de las opiniones de los consumidores.
- ▶ **Marketing relacional estadístico:** la empresa toma un protagonismo más activo para estimular al consumidor. En esta etapa la respuesta del consumidor es necesaria y el CRM persigue las opiniones de los usuarios y fomenta la repetición del producto o servicio.
- ▶ **Marketing relacional proactivo:** entramos en esta fase de «verdadera relación» con el cliente. Para ello el CRM necesitará conocer cada momento del consumidor, sus hábitos, preferencias, etc. En definitiva, conocer para personalizar.
- ▶ **Marketing relacional de socio:** el máximo nivel de *marketing* relacional y a lo que idealmente tienden todos los sistemas. Relación de igual a igual en la que el cliente no se «despega» de su proveedor porque cree que es el único que le entiende. Esta etapa del *marketing* no es fácil y no siempre es factible para algunas empresas llegar a ella.

En cualquier caso, asumamos que cualquiera de estas modalidades del *marketing* de relaciones se verá ayudada por la estrategia CRM y en cada una de ellas podrá desarrollarse con mayor o menor grado.

El desarrollo de esta estrategia de *marketing* relacional ha dado pie a un **nuevo término en marketing: CRM**. Este concepto **es usado ahora como una estrategia global** que incluye *software*, tecnología en Internet y, en definitiva, un método sistemático que permite registrar, segmentar y seleccionar los clientes reales y potenciales de acuerdo con sus necesidades, motivaciones y deseos. Al mismo tiempo, el sistema CRM deberá proporcionar a los clientes toda la asistencia necesaria en cada momento que lo requieran. No nos olvidemos, por tanto, de la importancia de capitalizar toda la información (**dato**) en aras de tener la mayor y mejor información de nuestros clientes.

## 7.3. Características y beneficios del marketing relacional

### Características

La razón principal del *marketing* relacional, como hemos ido comprobando, es **mantener y fidelizar clientes.**

Veamos en el siguiente cuadro las diferencias más importantes.

Diferencias de enfoque	
Marketing Tradicional	Marketing “1to1®”
<ul style="list-style-type: none"><li>• Productos <i>estándar</i></li><li>• Clientes <i>masivos</i></li><li>• Mensajes <i>hacia los clientes</i></li><li>• Exito: <i>Adquisición</i> de un gran volumen de clientes (“<i>market share</i>”)</li></ul>	<ul style="list-style-type: none"><li>• Productos y servicios <i>personalizados</i></li><li>• Cliente <i>tratado como individuo</i></li><li>• Diálogo <i>permanente con los clientes</i></li><li>• Exito: <i>Adquisición y retención</i> de clientes <i>rentables de por vida</i> (“<i>client share</i>”)</li></ul>

Figura 3. *Marketing* tradicional vs. «1to1». Fuente: Brunetta, s.f.

En definitiva, sus características más importantes son:

- ▶ Perspectiva a largo plazo.
- ▶ Mantener clientes.
- ▶ Relaciones duraderas con clientes.
- ▶ Soluciones personalizadas.
- ▶ Cooperación mutua no solo con clientes, sino también con proveedores, *partners* y la propia organización de la compañía a todos los niveles.
- ▶ Foco en el servicio al cliente.

## Beneficios

Los beneficios evidentes de esta estrategia descansan sobre el máximo nivel de relaciones que vimos en el ejemplo anterior (*marketing* relacional de socio). Esto es la plena satisfacción de cliente y compañía basada en el conocimiento mutuo y duradero.

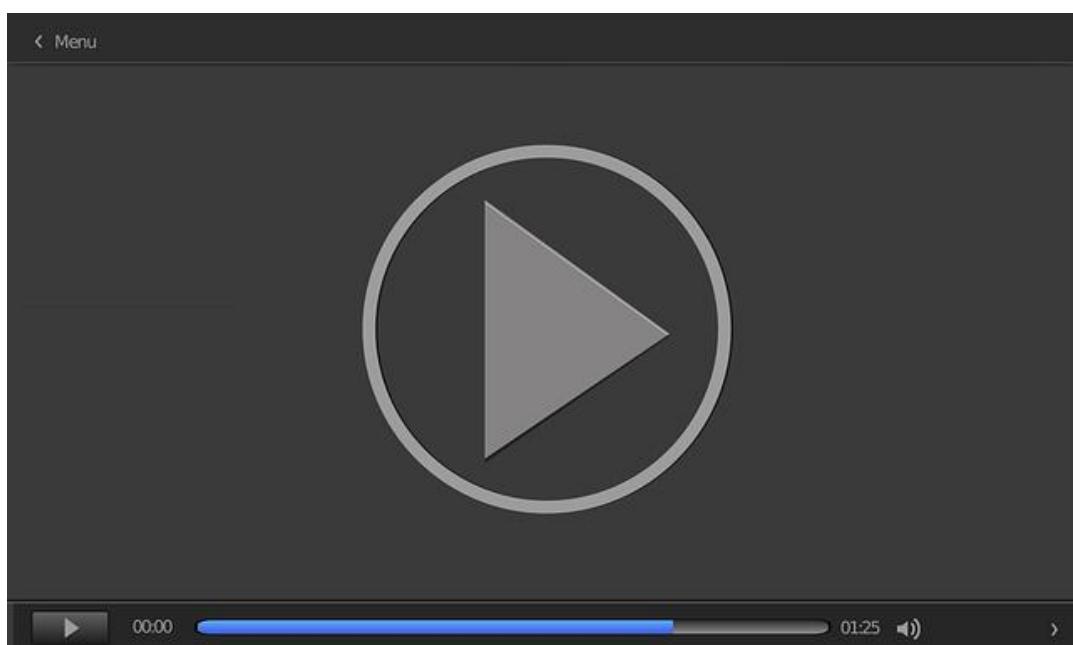
Don Peppers (Peppers y Rogers, 2004), uno de los mayores expertos en *marketing one-to-one*, recuerda que la estrategia debe ser diseñada en este caso en torno a la **dimensión cliente** y no a la dimensión (tradicional) de producto o servicio.

Las ventajas o beneficios obtenidos, entre otros, son los siguientes:

- ▶ Fidelidad, dado que la compañía conoce al cliente y le atiende como si fuera **único**.
- ▶ No es necesaria una política excesiva de descuentos, *low cost*, etc.; dado que lo que entendemos es una **relación prolongada** y no promocional.
- ▶ El coste de adquisición de clientes se optimiza y se centra ahora en el **mantenimiento de aquellos productivos**.

- ▶ **Costes de servicio mejorado**, dado que tenemos los datos de sus necesidades, deseos, demandas y hábitos.
- ▶ Este beneficio global obtenido a través de la gestión del conocimiento, basada en una estrategia de *marketing* relacional y de los datos conocidos de sus clientes, es estimado por el profesor Peppers (Peppers y Rogers, 2004) en un 42 %.

Llegado a este punto, accede al vídeo *El necesario arte de la persuasión*.



---

Accede al vídeo:

<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=54158c8c-1bd2-4a0f-b1e5-acd400c83fc7>

---

## 7.4. CRM: definición y características

### Definición

Muchos nos preguntamos cuál es la mejor definición de CRM. Por ejemplo, Pippanyk afirma que «el CRM es más que un *software* o solución tecnológica, el CRM involucra un cambio en las prácticas de negocio enfocadas al éxito».

Otros autores, como Philip Kotler (1999), lo definen como: «el concepto más importante del *marketing* moderno. Hasta hace poco, CRM ha sido definido simplemente como una actividad de gestión de datos de los clientes».

Dicha definición implica gestionar información detallada acerca de los clientes en particular, así como manejar cuidadosamente los puntos sensibles de nuestros clientes para maximizar la fidelidad de los mismos.

Más recientemente, CRM ha quedado definido por Kotler como «el proceso general de construir y mantener relaciones beneficiosas con los clientes a través del suministro de valor y satisfacción superior al cliente. Lógicamente ligado a aspectos como la adquisición, mantenimiento y crecimiento con los clientes» (Kotler y Armstrong, 2013).

También gusta esta definición de Thompson, que dice: «estrategia para crear y mantener relaciones de largo plazo con los clientes, alineando las actividades de la empresa con las necesidades del cliente». En particular, por lo afirmado al final: «alineando las actividades de la empresa con las necesidades del cliente».

¿Por qué? Principalmente, y lo veremos más adelante, si la empresa (grande, mediana o pequeña) no se alinea con la estrategia CRM de verdad, la estrategia CRM estará condenada al fracaso.

Recuerda que Piphanyk también expresaba en el primer párrafo de este apartado que el CRM es más que un *software* o solución tecnológica, el CRM involucra un cambio en las prácticas de negocio enfocadas al éxito. Por tanto, si no estamos dispuestos a replantearnos en serio un cambio en la manera de hacer negocio y tratar a nuestros clientes, es mejor que no empecemos, porque habremos malgastado tiempo, dinero y reputación.

No olvidemos la mala imagen que podemos tener grabada en nuestros cerebros de múltiples iniciativas de empresas que luego vieron que tenían que «batirse en retirada» por no haber adecuado su producto o servicio a las necesidades de sus clientes. ¿Qué quedó en nuestra mente de aquello? Pues como mínimo que fue un «sonoro patinazo». Por ello, aquellas organizaciones que quieran involucrarse en una estrategia CRM y no estén dispuestas a un cambio, mejor que lo dejen.

En definitiva, a modo de palabra clave que nos recuerde qué significa  
CRM: CRM = Cambio.

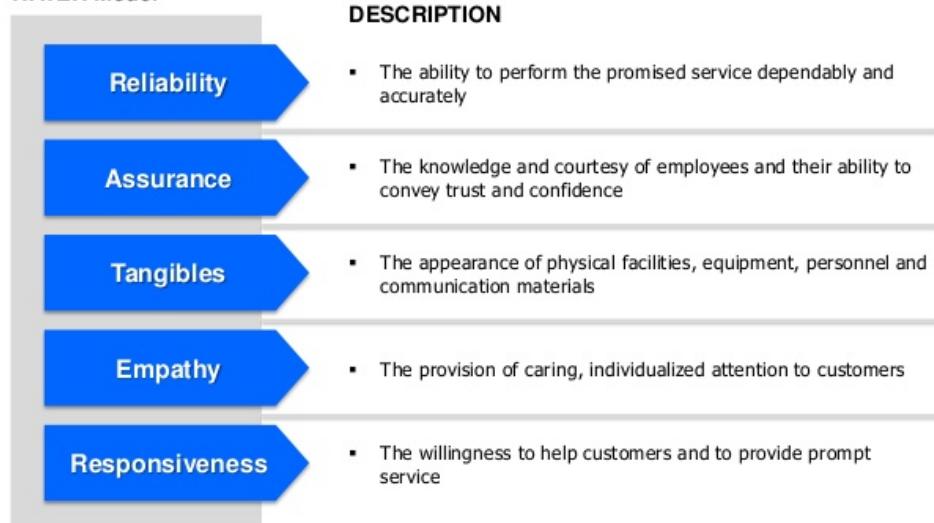
## Características (modelo RATER)

Por último, las características claves que debe de tener un buen CRM, estudiando a Zeithaml et al. (1988) y posteriormente a Barroso y Martín (2000), consisten en que el cliente perciba que el servicio global recibido por su marca es superior al de otros. Esto también se relaciona con el concepto de expectativas que el cliente tiene y la capacidad de las compañías de cumplir con esas expectativas de los clientes en cuanto a la calidad del servicio. Por tanto, es importante conocer las características determinantes para el cliente a la hora de evaluar el servicio recibido por su compañía favorita y que el sistema CRM debe de conocer.

Parasuraman et al. (1988) establecieron el modelo RATER (o Servqual) de factores determinantes de calidad del servicio para los clientes en un buen CRM, basado en las iniciales (RATER) que describimos a continuación.

## The RATER model allows customer service experiences to be explored and assessed quantitatively

**RATER Model**



Source: "SERVQUAL" by Valarie A. Zeithaml, A. Parasuraman, and Leonard L. Berry

© Operational Excellence Consulting. All rights reserved.

20

Figura 4. Modelo RATER. Fuente: Parasuraman et al. (1988).

- ▶ **Reliability:** habilidad para cumplir lo prometido.
- ▶ **Assurance:** capacidad de los empleados en dar un buen servicio.
- ▶ **Tangibles:** la presencia física del personal, material, etc. que proporciona el servicio.
- ▶ **Empathy:** la capacidad de suministrar atención y cuidado como a uno mismo.
- ▶ **Responsiveness:** disposición a atender al cliente de forma rápida y amable.

Toda estrategia CRM que pretenda conquistar al cliente deberá cumplir las características RATER para conseguir ese objetivo.

CRM: definir los comportamientos del cliente a partir del análisis de los datos que se tienen del mismo y orientar los procesos internos a captar clientes, venderles productos o servicios y mantener una relación a largo plazo con ellos.

## 7.5. Factores clave y bases para un buen CRM

### Factores clave para la implementación de un buen CRM

Es evidente que la implantación de un CRM proporcionará beneficios mutuos a empresa y cliente. Entre los factores clave para su implantación podemos mencionar:

- ▶ **La tecnología.** Sin ella no tendría sentido en la actualidad un CRM. Pero recuerda: la tecnología no es un fin sino un medio para tener los **datos y el conocimiento** a fin de llegar y mantener la relación con el cliente.
- ▶ **Comunicación proactiva e interactiva.** Una pareja no se mantiene si no hay comunicación fluida y «con algo que decir». Ejemplos:
  - Utilización multicanal (móvil, *e-mail*, RR. SS.) para estar en contacto útil con los clientes.
  - Información puntual del estado de sus pedidos (entrega, posventa, etc.).
  - Servicio 24 horas.
  - Vistas al cliente (en caso de B2B).
  - Información personalizada al cliente cuando la reclama o creemos que es el momento adecuado.
- ▶ Olvidar la orientación al producto o transacción y migrar hacia la **gestión relacional**. La transacción es un eslabón de una cadena superior de relaciones duraderas.
- ▶ **Retención de clientes.** Un cliente mantenido a satisfacción es un cliente rentable. Conseguir un cliente nuevo es cinco veces más costoso que mantener los actuales. Ejemplo: programas de banca para premiar a los clientes que domicilian su nómina en la entidad.

- ▶ Proporcionar valor a través de productos y servicios **antes, durante y posteriormente** al momento de la transacción.
- ▶ Perspectiva a **largo plazo**: es preferible perder una venta que un cliente. Además, esto permite a la compañía una visión estratégica también a largo plazo.
- ▶ Énfasis en el **servicio al cliente**. Es un compromiso indiscutible en el CRM.
- ▶ **Alto grado de compromiso** de la empresa con el cliente: debe ser su prioridad número 1 y extenderse como filosofía en toda la organización: el CRM es una estrategia de compañía, no un *software*.

## Bases para la gestión de un exitoso CRM

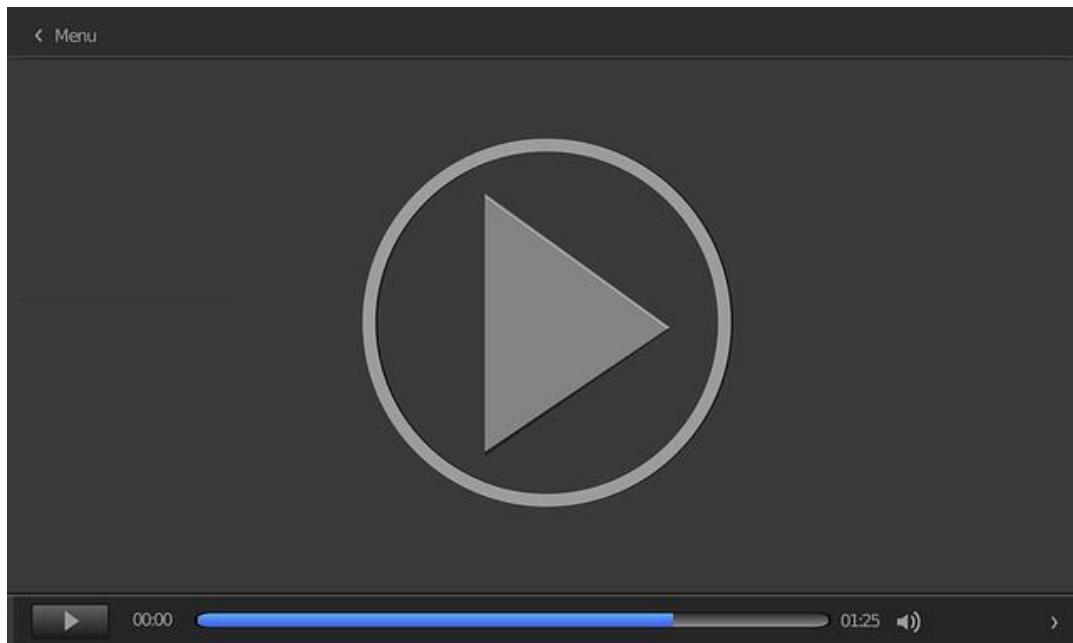
Podemos enumerar muchas premisas básicas, pero en breves palabras y siempre partiendo de la importancia del conocimiento del cliente a través del análisis de la información. Son las siguientes (Domínguez y Muñoz, 2010):

- ▶ Identificar las expectativas del cliente: qué quiere, cómo y cuándo lo quiere.
- ▶ Comparar las características del servicio prestado y las expectativas del cliente y determinar las áreas de mejora.
- ▶ Comunicación omnicanal permanente y de valor con el cliente (si no tienes nada interesante que decir, mejor no digas nada).
- ▶ La interacción con los clientes debe ser accesible en toda la organización para fluir de forma coordinada y no repetitiva.
- ▶ La satisfacción del cliente debe de ser una prioridad de toda la compañía.
- ▶ Los problemas de los clientes son de toda la organización y deben de coordinarse para su resolución en lugar de «pasarse la pelota entre departamentos».

- ▶ Las necesidades, motivaciones y deseos de los clientes deben ser conocidas y entendidas por la organización en una buena gestión del SIM, de manera que de forma interactiva el cliente siempre tenga una respuesta positiva de la organización y, en muchos casos actuales, la organización se anticipe a los deseos de los clientes.

En resumen, el *marketing* actual no se puede entender sin el conocimiento del cliente, conocimiento que nos facilita sin dudas el uso actual de la tecnología de datos. Por otro lado, una vez con esos datos, las compañías deben ser capaces de establecer una estrategia de relaciones duraderas con los clientes. Esos datos y su estrategia corporativa serán la base de un sistema CRM que tiene que ocuparse de coordinar todas las acciones necesarias en la organización para que el cliente quede con plena satisfacción. Esto es, que el cliente sepa que es «el boss» o «el rey».

Por último, accede al vídeo *La gestión del intangible más valioso.*



---

Accede al vídeo:

<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=89cbee2e-8002-4d6d-bb9b-acd400c8764f>

---

## 7.6. Referencias bibliográficas

Barroso, C., y Martín, E. (2000). Desarrollo del marketing relacional en España. *Revista Europea de Dirección y Economía de la Empresa*, 9(3), 25-46.

<https://dialnet.unirioja.es/servlet/articulo?codigo=147362>

Brunetta, H. (s.f.). Inicio [Página web]. *HugoBrunetta*. <https://hugobrunetta.com/>

Colina, A. (2015). Unidad I. Medios de Comunicación Social [Diapositivas].  
[https://www.ecotec.edu.ec/material/material\\_2015F1\\_COM349\\_11\\_50868.pdf](https://www.ecotec.edu.ec/material/material_2015F1_COM349_11_50868.pdf)

Domínguez, A., y Muñoz, G. (2010). *Métricas del marketing*. ESIC.

Kotler, P. (1999). *El marketing según Kotler*. Paidós.

Kotler, P., y Armstrong, G. (2013). *Fundamentos de marketing*. Pearson.  
[https://frrq.cvg.utn.edu.ar/pluginfile.php/14584/mod\\_resource/content/1/Fundamentos%20del%20Marketing-Kotler.pdf](https://frrq.cvg.utn.edu.ar/pluginfile.php/14584/mod_resource/content/1/Fundamentos%20del%20Marketing-Kotler.pdf)

Peppers, D., y Rogers, M. (2004). *Managing customer relationships: a strategic framework*. Wiley.

Parasuraman, A., Zeithaml, V. A., y Berry, L. L. (1988). Servqual: a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-40.

Vidal, I., (2004). *Cómo conquistar el mercado con una estrategia CRM*. FC Editorial.

Zeithaml, V., Berry, L. L., y Parasuraman, A. (1988). Communication and control processes in delivery of service quality. *Journal of Marketing*, 52(2), 35-48.

## ¿Por qué una plataforma de marketing relacional?

---

Maza, G. (6 de septiembre de 2016). ¿Por qué una plataforma de marketing relacional? *PuroMarketing*. <http://www.puromarketing.com/30/27607/plataforma-marketing-relacional.html>

---

En este interesante artículo verás cómo el autor descubre la importancia en el *marketing* digital de «sacar provecho» al *marketing* relacional.

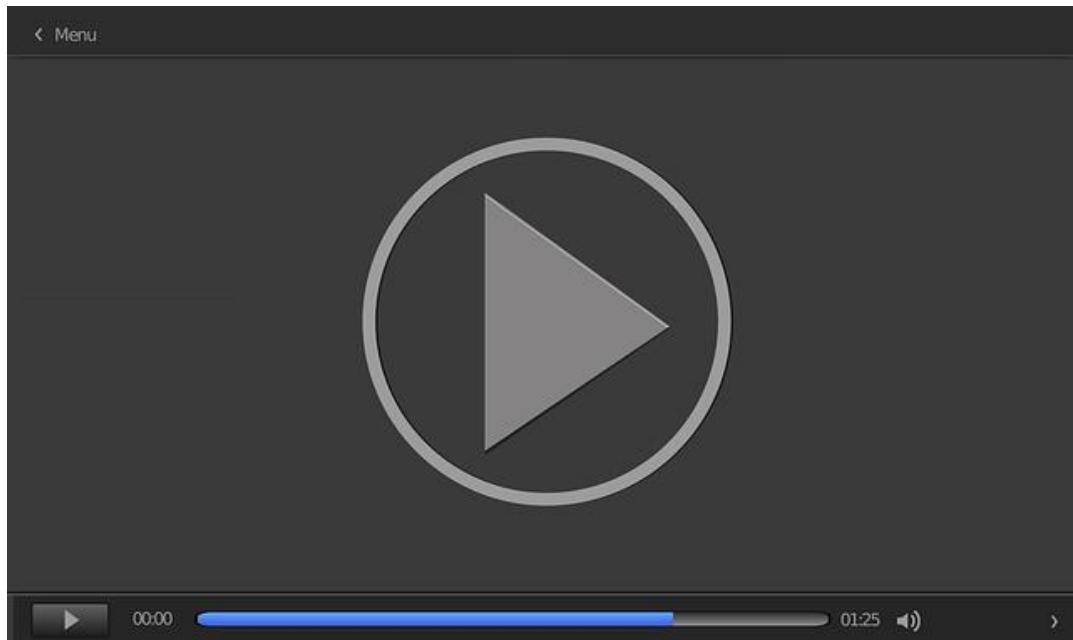
## Brain centric marketing: la evolución del customer centric para un nuevo marketing

Navarro, J. M. (8 de mayo de 2017). Brain centric marketing: la evolución del customer centric para un nuevo marketing. *PuroMarketing*. <https://www.puromarketing.com/27/28755/brain-centric-marketing-evolucion-customer-centric-para-nuevo-marketing.html>

¿Es posible relacionar lo que el cliente quiere o desea con lo que la empresa ofrece? Pues con el uso de los datos y la filosofía *customer centricity* debe poderse. El CC es una evolución y mejora del *marketing* relacional y los sistemas CRM tradicionales.

## Marketing digital vs. marketing tradicional: cinco diferencias

Solaeche, J. (15 de julio de 2020). *Marketing Digital vs Marketing Tradicional: 5 Diferencias* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=7kWI7zf15no>



Accede al vídeo:

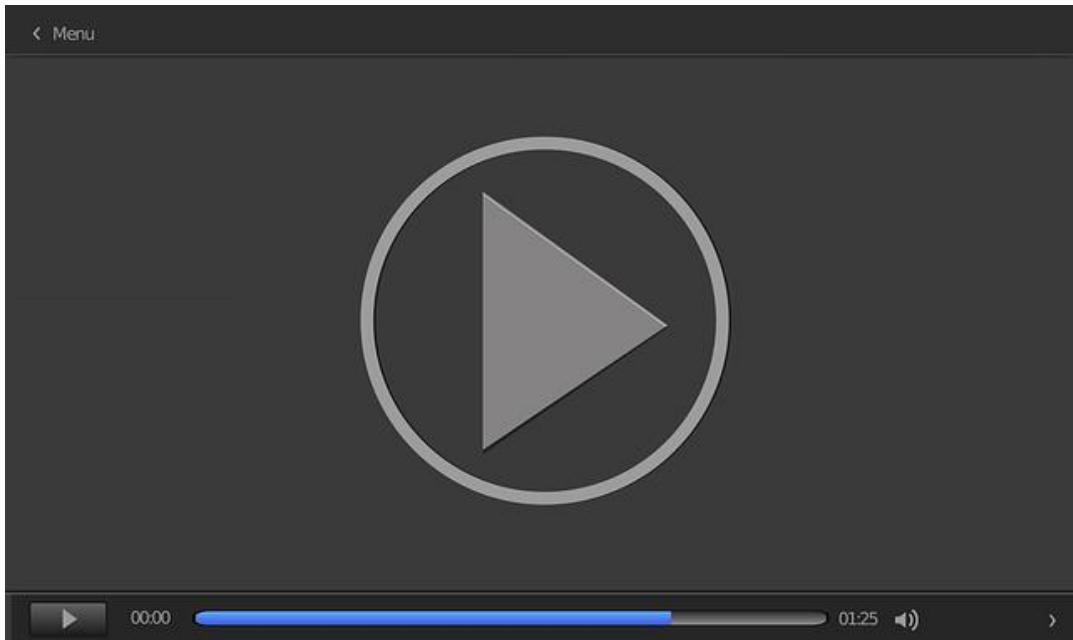
<https://www.youtube.com/embed/7kWI7zf15no>

Principales diferencias entre el *marketing* digital y el *marketing* tradicional. Cómo implementar cada una de ellas en el negocio.

## ¿Por qué necesitamos un CRM?

QuanticVision. (4 de octubre de 2010). ¿Qué es CRM? [Vídeo].

Youtube. <https://www.youtube.com/watch?v=jjHRTqPR30M>



Accede al vídeo:

<https://www.youtube.com/embed/jjHRTqPR30M>

Clarificador vídeo sobre la razón de un CRM.

1. Según Kotler, el cuarto y definitivo paso en el proceso de *marketing* es:

  - A. El diseño del programa de *marketing*.
  - B. La consecución de relaciones beneficiosas con los clientes.
  - C. Entender el mercado y las necesidades de los clientes.
  - D. Diseñar una estrategia adecuada con el foco en el cliente.
2. Según el profesor Peppers, el beneficio global obtenido a través de la gestión del conocimiento, basada en una estrategia de *marketing* relacional es del:

  - A. 42 %.
  - B. 24 %.
  - C. 30 %.
  - D. 60 %.
3. Entre los factores clave para la implantación de un buen CRM, no se encuentra:

  - A. La tecnología.
  - B. Una buena comunicación con el cliente.
  - C. Perspectiva a largo plazo.
  - D. Todos los anteriores son factores clave.
4. La A de *assurance* en el modelo RATER se traduce por:

  - A. Capacidad de los empleados en dar un buen servicio.
  - B. La presencia física del personal, material, etc. que proporciona el servicio.
  - C. La capacidad de suministrar atención y cuidado como a uno mismo.
  - D. Disposición a atender al cliente de forma rápida y amable.

5. Según Domínguez y Muñoz, entre las bases para un buen CRM no está:
  - A. Identificar las expectativas del cliente: qué quiere, cómo y cuándo lo quiere.
  - B. Comparar las características del servicio prestado y las expectativas del cliente y determinar las áreas de mejora.
  - C. Comunicación omnicanal permanente y de valor con el cliente (si no tienes nada interesante que decir, mejor no digas nada).
  - D. Todas deberían estar en un buen CRM.
6. Una palabra clave que nos recuerde qué significa CRM puede ser:
  - A. Cliente.
  - B. Transformación.
  - C. Diferencia.
  - D. Marca.
7. Según Ignasi Vidal, el máximo nivel de *marketing* relacional sería:
  - A. El estadístico.
  - B. El de socio.
  - C. El básico.
  - D. El proactivo.
8. La razón principal del *marketing* relacional es:
  - A. Mantener y fidelizar clientes.
  - B. Vender más.
  - C. Analizar el servicio al cliente.
  - D. Realizar promociones.

- 9.** Como ejemplo de comunicación proactiva e interactiva, podemos destacar:
- A. Utilización multicanal (móvil, *e-mail*, RR. SS.) para estar en contacto útil con los clientes.
  - B. Información puntual del estado de sus pedidos (entrega, posventa, etc.).
  - C. Servicio 24 horas.
  - D. Todos los anteriores son ejemplos válidos.
- 10.** La capacidad de suministrar atención y cuidado como a uno mismo es reconocida en el modelo RATER como:
- A. *Reliability*.
  - B. *Empathy*.
  - C. *Assurance*.
  - D. *Tangibles*.

Gobierno del Dato y Toma de Decisiones

---

## Tema 8. Introducción a la protección de datos

# Índice

## Esquema

### Ideas clave

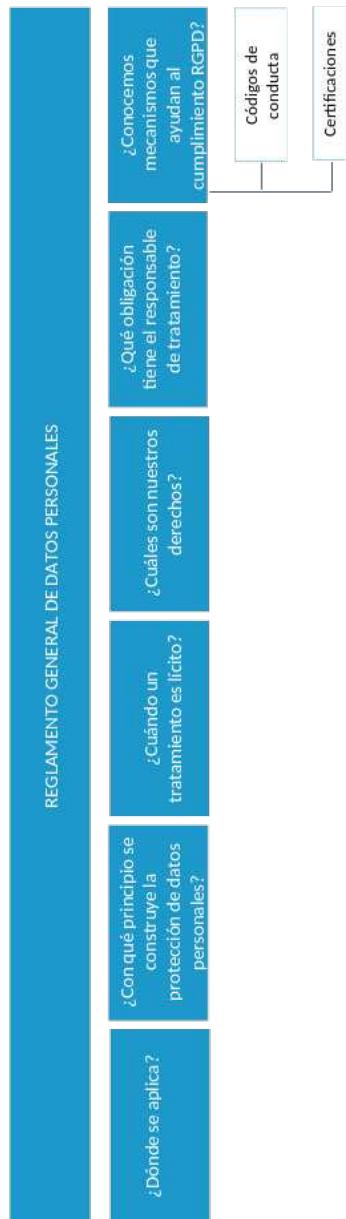
- 8.1. Introducción y objetivos
- 8.2. Conceptos
- 8.3. Principios generales de protección de datos en Europa
- 8.4. Licitud de tratamiento. El consentimiento informado
- 8.5. Derecho de información
- 8.6. El derecho de interesado
- 8.7. Obligaciones generales del responsable de tratamiento y encargado
- 8.8. Otros marcos internacionales
- 8.9. Protección de datos en EE. UU. y otros países
- 8.10. Transferencias internacionales de datos
- 8.11. Seguridad de la información y protección de datos
- 8.12. Referencias bibliográficas

### A fondo

- Hacia un nuevo modelo europeo de protección de datos
- Agencia Española de Protección de Datos
- Comisión Europea

### Test

# Esquema



## 8.1. Introducción y objetivos

Este tema desarrolla los principios de la protección de datos de carácter personal, y para ello nos introducimos en el marco internacional más exigente, que es el de la UE. Este marco se aplica a los países miembros de la UE y está influyendo notablemente en el desarrollo normativo de otros países fuera del espacio europeo. Además, extiende su aplicación fuera de las fronteras de la UE.

Para lograr soltura en la lectura de textos legales, aconsejamos consultar la información disponible en la web de la Agencia Española de Protección de Datos (AEPD) sobre el:

- ▶ Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos), en adelante RGPD.

## 8.2. Conceptos

Las diferentes legislaciones en materia de privacidad y protección de datos de carácter personal tienen por objeto garantizar y proteger, en lo que concierne al tratamiento de datos personales, las libertades públicas y los derechos fundamentales de las personas físicas, y especialmente de su honor e intimidad personal y familiar.

Es importante entender los conceptos aquí incorporados para seguir adecuadamente el contenido de la asignatura.

### Principales conceptos (art. 4, RGPD)

- ▶ «“Datos personales”: toda información sobre una persona física identificada o identificable (el interesado)» (art. 4.1, RGPD).

Y se considerará **persona física identificable** «toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona» (art. 4.1, RGPD).

Un aspecto controvertido por su carácter interpretativo se deriva del término *identifiable*, es decir, información que facilita de manera indirecta identificar a una persona. Este sería el caso de, por ejemplo: la dirección IP, el número de teléfono fijo, la cadena de ADN, la huella dactilar, datos biométricos, etc. Para entender mejor el concepto de identifiable, podemos recurrir al considerando 26 RGPD, que explica que la determinación de si una persona física es identifiable, se han de tener en consideración todos los medios, razonablemente accesibles por «alguien» que permitan identificar de una forma bien sea directa o indirecta (a través de algún

identificador) a la persona física. Se deben tener en consideración todos los medios técnicos disponibles, los costes operativos y económicos asociados y el tiempo requerido para considerar la probabilidad de que dicha identificación sea posible. Una cuestión que habrá que tener en cuenta en ámbitos de aplicaciones móviles o IoT que, a través de identificadores únicos asociados a dispositivos o una *app*, estos pueden permitir la identificación inequívoca.

- ▶ «“Tratamiento”: cualquier operación o conjunto de operaciones realizadas sobre datos personales o conjuntos de datos personales, ya sea por procedimientos automatizados o no, como la recogida, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión, limitación, supresión o destrucción» (art. 4.2, RGPD).
- ▶ «“Limitación del tratamiento”: el marcado de los datos de carácter personal conservados con el fin de limitar su tratamiento en el futuro» (art. 4.3, RGPD).
- ▶ «“Elaboración de perfiles”: toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física» (art. 4.4, RGPD).
- ▶ «“Pseudonimización”: el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identifiable» (art. 4.5, RGPD).

- ▶ «“Fichero”: todo conjunto estructurado de datos personales, accesibles con arreglo a criterios determinados, ya sea centralizado, descentralizado o repartido de forma funcional o geográfica» (art. 4.6, RGPD).
- ▶ «“Responsable del tratamiento” o “responsable”: la persona física o jurídica, autoridad, servicio u otro organismo que, solo o junto con otros, determine los fines y medios del tratamiento» (art. 4.7, RGPD).
- ▶ «“Encargado del tratamiento” o “encargado”: la persona física o jurídica, autoridad, servicio u otro organismo que trate datos personales por cuenta del responsable del tratamiento» (art. 4.8, RGPD).
- ▶ «“Destinatario”: la persona física o jurídica, autoridad, servicio u otro organismo al que se comuniquen datos personales, se trate o no de un tercero. No obstante, no se considerarán destinatarios las autoridades públicas que puedan recibir datos personales en el marco de una investigación concreta de conformidad con el Derecho de la Unión o de los Estados miembros; el tratamiento de tales datos por dichas autoridades será conforme con las normas en materia de protección de datos aplicables a los fines del tratamiento» (art. 4.9, RGPD).
- ▶ «“Tercero”: persona física o jurídica, autoridad, servicio u organismo distinto del interesado, del responsable del tratamiento, del encargado del tratamiento y de las personas autorizadas para tratar los datos personales bajo la autoridad directa del responsable o del encargado» (art. 4.10, RGPD).
- ▶ «“Consentimiento del interesado”: toda manifestación de voluntad libre, específica, informada e inequívoca por la que el interesado acepta, ya sea mediante una declaración o una clara acción afirmativa, el tratamiento de datos personales que le conciernen» (art. 4.11, RGPD).

- ▶ «“Violación de la seguridad de los datos personales”: toda violación de la seguridad que ocasione la destrucción, pérdida o alteración accidental o ilícita de datos personales transmitidos, conservados o tratados de otra forma, o la comunicación o acceso no autorizados a dichos datos» (art. 4.12, RGPD). La violación también puede ser referida como **brecha de seguridad**.
- ▶ «“Datos genéticos”: datos personales relativos a las características genéticas heredadas o adquiridas de una persona física que proporcionen una información única sobre la fisiología o la salud de esa persona, obtenidos en particular del análisis de una muestra biológica de tal persona» (art. 4.13, RGPD).
- ▶ «“Datos biométricos”: datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona, como imágenes faciales o datos dactiloscópicos» (art. 4.14, RGPD). Los datos biométricos tienen un uso creciente en métodos de autenticación de las personas. El RGPD permite a los estados miembros legislar sobre las condiciones adicionales o limitaciones del tratamiento de datos biométricos, genéticos o relativos a la salud (art. 9.4, RGPD).
- ▶ «“Datos relativos a la salud”: datos personales relativos a la salud física o mental de una persona física, incluida la prestación de servicios de atención sanitaria, que revelen información sobre su estado de salud» (art. 4.15, RGPD). Los datos de salud disponen de un marco de amparo garantista específico por las características y el impacto que dichos datos tienen en los ciudadanos, de ahí que exista una definición específica al respecto.

- ▶ «“Establecimiento principal”:
  - »En lo que se refiere a un responsable del tratamiento con establecimientos en más de un Estado miembro, el lugar de su administración central en la Unión, salvo que las decisiones sobre los fines y los medios del tratamiento se tomen en otro establecimiento del responsable en la Unión y este último establecimiento tenga el poder de hacer aplicar tales decisiones, en cuyo caso el establecimiento que haya adoptado tales decisiones se considerará establecimiento principal.
  - »En lo que se refiere a un encargado del tratamiento con establecimientos en más de un Estado miembro, el lugar de su administración central en la Unión o, si careciera de esta, el establecimiento del encargado en la Unión en el que se realicen las principales actividades de tratamiento en el contexto de las actividades de un establecimiento del encargado en la medida en que el encargado esté sujeto a obligaciones específicas con arreglo al presente Reglamento» (art. 4.16, RGPD).
- ▶ «“Representante”: persona física o jurídica establecida en la Unión que, habiendo sido designada por escrito por el responsable o el encargado del tratamiento con arreglo al artículo 27, represente al responsable o al encargado en lo que respecta a sus respectivas obligaciones en virtud del presente Reglamento» (art. 4.17, RGPD).
- ▶ «“Empresa”: persona física o jurídica dedicada a una actividad económica, independientemente de su forma jurídica, incluidas las sociedades o asociaciones que desempeñen regularmente una actividad económica» (art. 4.18, RGPD).
- ▶ «“Grupo empresarial”: grupo constituido por una empresa que ejerce el control y sus empresas controladas» (art. 4.19, RGPD).

- ▶ «“Normas corporativas vinculantes”: las políticas de protección de datos personales asumidas por un responsable o encargado del tratamiento establecido en el territorio de un Estado miembro para transferencias o un conjunto de transferencias de datos personales a un responsable o encargado en uno o más países terceros, dentro de un grupo empresarial o una unión de empresas dedicadas a una actividad económica conjunta» (art. 4.20, RGPD). Las reglas corporativas vinculantes BRC (acrónimo en inglés) se configuraron en su momento como vehículo para racionalizar las transferencias internacionales de datos en grupos corporativos siempre con la autorización de la autoridad competente.
- ▶ «“Autoridad de control”: la autoridad independiente establecida por un Estado miembro con arreglo a lo dispuesto en el artículo 51» (art. 4.21, RGPD). En términos genéricos, hablamos de agencias de protección de datos de los Estados miembros. En el caso de España, la autoridad es la Agencia Española de Protección de Datos.
- ▶ «“Autoridad de control interesada”: la autoridad de control a la que afecta el tratamiento de datos personales debido a que:
  - »El responsable o el encargado del tratamiento está establecido en el territorio del Estado miembro de esa autoridad de control.
  - »Los interesados que residen en el Estado miembro de esa autoridad de control se ven sustancialmente afectados o es probable que se vean sustancialmente afectados por el tratamiento.
  - »Se ha presentado una reclamación ante esa autoridad de control» (art. 4.22, RGPD).

Es de resaltar la posibilidad de presentar reclamaciones, por parte de los afectados, a las autoridades de control de su país de residencia, aunque el tratamiento tenga lugar en otro estado miembro.

- ▶ «Tratamiento transfronterizo»: se establece en dos circunstancias diferenciadas:
  - «El tratamiento de datos personales realizado en el contexto de las actividades de establecimientos en más de un Estado miembro de un responsable o un encargado del tratamiento en la Unión, si el responsable o el encargado está establecido en más de un Estado miembro.
  - »El tratamiento de datos personales realizado en el contexto de las actividades de un único establecimiento de un responsable o un encargado del tratamiento en la Unión, pero que afecta sustancialmente o es probable que afecte sustancialmente a interesados en más de un Estado miembro» (art. 4.23, RGPD).

Estaremos ante un tratamiento transfronterizo siempre que los datos crucen la frontera, bien porque el responsable del tratamiento o encargado está presente en diferentes Estados de la UE, bien porque afecta a ciudadanos residentes en Estados diferentes al de residencia del responsable o encargado de tratamiento.

- ▶ «“Objeción pertinente y motivada”: la objeción a una propuesta de decisión sobre la existencia o no de infracción del presente Reglamento, o sobre la conformidad con el presente Reglamento de acciones previstas en relación con el responsable o el encargado del tratamiento, que demuestre claramente la importancia de los riesgos que entraña el proyecto de decisión para los derechos y libertades fundamentales de los interesados y, en su caso, para la libre circulación de datos personales dentro de la Unión» (art. 4.24, RGPD).

- ▶ «“Servicio de la sociedad de la información”: todo servicio conforme a la definición del artículo 1, apartado 1, letra b), de la Directiva (UE) 2015/1535 del Parlamento Europeo y del Consejo» (art. 4.25, RGPD). Es decir, todo servicio de la sociedad de la información o todo servicio prestado normalmente a cambio de una remuneración, a distancia, por vía electrónica y a petición individual de un destinatario de servicios.
- ▶ «“Organización internacional”: una organización internacional y sus entes subordinados de Derecho internacional público o cualquier otro organismo creado mediante un acuerdo entre dos o más países o en virtud de tal acuerdo». (art. 4.26, RGPD).

### 8.3. Principios generales de protección de datos en Europa

El RGPD se construye sobre siete principios generales, un conjunto de reglas básicas y de obligado cumplimiento, para el tratamiento de datos de carácter personal. El RGPD consolida y amplia los principios existentes en la legislación europea anterior (la directiva 95/40/CE) (art. 5., RGPD).



Figura 1. Principios de la protección de datos personales RGPD. Fuente: elaboración propia.

► **Licitud, lealtad y transparencia.** Los datos de carácter personal deben ser tratados de manera:

- **Lícita** (art. 6, RGPD).
- **Leal:** «fidedigno, verídico y fiel, en el trato o en el desempeño de un oficio o cargo» (RAE, 2020).
- **Transparente** (art. 12, RGPD).

Es decir, el tratamiento debe ser conforme a la Ley, tal y como se establece la legitimidad en el RGPD, debe ser fiel por parte de los responsables de tratamiento frente a los afectados, no debe mediar engaño ni tratar los datos sin lealtad al consentimiento obtenido por parte del afectado. Y debe permitir al afectado conocimiento sobre el tratamiento al que están sometidos sus datos personales., en los términos que refleja el RGPD, que veremos más adelante.

► **Limitación de la finalidad.** Recabados para una finalidad concreta, que debe ser explícita y legítima, de manera que los datos no podrán ser tratados posteriormente para ninguna otra finalidad, a menos que la nueva finalidad tenga la consideración de compatible, como es el caso del tratamiento de datos para fines posteriores como son los casos de archivo en interés público, fines de investigación científica e histórica o fines estadísticos. «El tratamiento ulterior de los datos personales con fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos no se considerará incompatible con los fines iniciales (“limitación de la finalidad”)» (art. 5.1, RGPD).

El tratamiento de datos con finalidades históricas, científicas y estadísticas deben **realizarse conforme a las garantías establecidas en el propio RGPD**. El artículo 89 («Garantías y excepciones aplicables al tratamiento con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos») fija, entre otras, la necesidad de incorporar al tratamiento medidas técnicas y organizativas, garantizar el respeto al principio de minimización de los datos personales y aplicar si es posible pseudonimización.

- ▶ **Minimización de datos.** Los datos recabados deben ser adecuados, pertinentes y limitados a los estrictamente necesarios para la finalidad prevista para la que son recabados.

Esto implica que **la finalidad ha de justificar los datos que están siendo objeto de tratamiento**, de manera que la misma finalidad no pueda satisfacerse con un conjunto de datos de los afectos menor.

- ▶ **Exactitud de los datos.** Los datos deberán ser exactos y estar actualizados, y el responsable de tratamiento debe tomar las medidas necesarias para que los datos que sean inexactos con respecto a los fines del tratamiento se supriman o rectifiquen sin retrasos.
- ▶ **Limitación del plazo de conservación.** Los datos personales se mantendrán, permitiendo la identificación de los afectados, por el tiempo estrictamente necesario para los fines del tratamiento. Solo podrán conservarse si se tratan exclusivamente con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos, siempre que sea conforme al artículo 89, apartado 1, del RGPD. En cualquier caso, el responsable de tratamiento debe aplicar las medidas técnicas y organizativas necesarias establecidas en el RGPD para proteger los derechos y libertades del interesado.

- ▶ **Seguridad adecuada: confidencialidad e integridad.** Los datos serán tratados de manera que se garantice la seguridad adecuada, que debe incluir la protección frente al tratamiento no autorizado o ilícito, su pérdida, destrucción o el daño accidental. Para ello deberá aplicar las medidas técnicas y organizativas apropiadas.
- ▶ **Responsabilidad proactiva.** El responsable del tratamiento debe velar en todo momento por la licitud, lealtad y la trasparencia del tratamiento y el cumplimiento de los principios establecidos.

El principio de la responsabilidad proactiva (art. 5.2, RGPD) establece la obligación de los responsables de tratamiento (RT), de actuar con diligencia, de forma proactiva y continuada, para garantizar el cumplimiento del Reglamento General de Protección de Datos, y de este modo garantizar los derechos y libertades de los afectados.

Por ejemplo:

- ▶ Debe ser el responsable del tratamiento quien pruebe la legitimidad del tratamiento, mediante prueba de haber obtenido el consentimiento del afectado o en aplicación de otros principios de legitimación.
- ▶ Debe ser el responsable del tratamiento (RT) quien, ante una situación de violación de la seguridad, deba demostrar la improbabilidad de que dicha violación entrañe riesgo para los derechos y libertades de las personas físicas.

Exigir esta responsabilidad tiene por objeto garantizar los derechos de los afectados (considerando 85, RGPD).

## Ámbito de aplicación

El ámbito de aplicación de la legislación no diferencia los sistemas de tratamiento utilizados en el tratamiento de datos de carácter personal.

**El RGPD es aplicable a los tratamientos independientemente de que estos sean automatizados, manuales o mixtos (art. 2, RGPD).**

Pero no se aplica a todos los tratamientos de datos de carácter personal, existen exclusiones.

El RGPD no será de aplicación a los tratamientos efectuados por los ciudadanos, en el ámbito de su vida personal y doméstica.

Tampoco se aplicará en los tratamientos con fines de prevención, investigación, detección o enjuiciamiento de infracciones penales, o de ejecución de sanciones penales, incluida la de protección frente a amenazas a la seguridad pública y su prevención (art. 2, RGPD); así como los derivados de la aplicación del capítulo 2 del título V del TUE (Tratado de la UE), que aborda las políticas sobre controles en las fronteras, asilo e inmigración.

Referente la aplicación geográfica, **la norma se aplicará a los responsables de tratamiento que se encuadren fuera de la UE** en alguno de los supuestos siguientes (art. 3, RGPD):

- ▶ Si el tratamiento se realiza en el contexto de las actividades de un establecimiento del responsable o del encargado en la Unión Europea, independientemente de dónde se realiza el tratamiento y dónde reside el afectado, el tratamiento estará sometido al RGPD (art. 3.1, RGPD).
- ▶ Si el tratamiento se realiza por parte de responsables de tratamiento residentes fuera de la UE, que o bien ofrecen bienes o servicios a residentes europeos o realizan prácticas de control o seguimiento de su comportamiento (art. 3.2, RGPD).

Hay que recordar que en este segundo caso el responsable de tratamiento o el encargado deberá asignar un representante en la UE, a menos que sea el tratamiento ocasional y de bajo riesgo para los derechos libertados de los afectados porque, por ejemplo, no afecta a categorías de datos especiales como salud, origen racial, etc. (art. 27, RGPD).

Situaciones a las que se añadirán aquellas en las que el derecho de los Estados miembros se aplique en virtud del derecho internacional público como, por ejemplo, las embajadas.

De modo que la aplicación del RGPD transciende el espacio de la UE al establecer como hecho objetivo circunstancias que solo dependen de que el tratamiento se realice sobre ciudadanos residentes en la UE y no sobre la residencia del responsable de tratamiento o encargado del mismo.

**Extiende el ámbito de aplicación territorial a entidades no radicadas en la UE con actividades que se centren en la oferta de bienes o servicios dirigidas a ciudadanos residentes en la Unión Europea o al seguimiento de estos. Por ejemplo, páginas web de noticias, de venta de servicios o productos dirigidos al espacio de la UE.**

Veamos un esquema de decisión sobre cuándo el responsable de tratamiento debe quedar regulado por el Reglamento General de Protección de Datos.

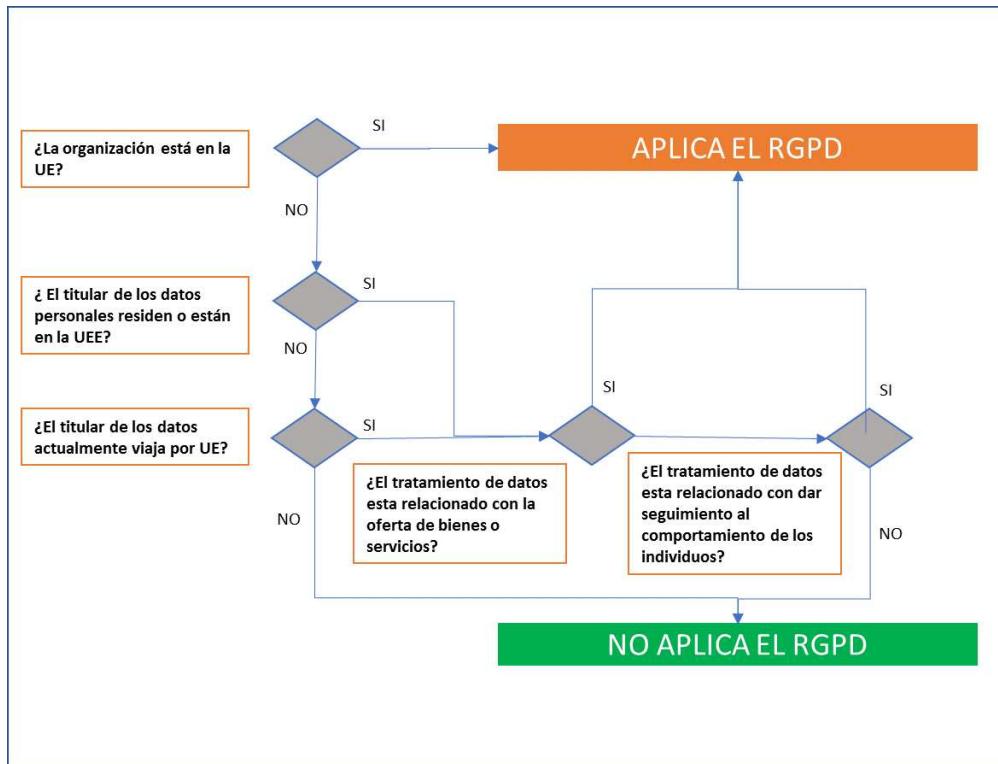


Figura 2. Diagrama de flujo para aplicar RGPD.

## 8.4. Licitud de tratamiento. El consentimiento informado

El RGPD establece bajo qué circunstancias el tratamiento tendrá la consideración de lícito, siendo responsabilidad del responsable de tratamiento que el mismo se ampare en dichas circunstancias. (art .6, RGPD).

El tratamiento será lícito, como principio general si existe consentimiento por parte del interesado. El interesado facilitó el consentimiento del tratamiento de sus datos para el fin o finalidades requeridos (art. 6.1, RGPD).

Pero existen otras circunstancias para las que no es indispensable el consentimiento del interesado, al cumplir alguna de las siguientes condiciones (art. 6.1, RGPD):

- ▶ El tratamiento es necesario para la ejecución de un contrato por parte del interesado.
- ▶ El tratamiento es **necesario para dar cumplimiento a una obligación legal** del responsable de tratamiento. Por ejemplo, esta condición da amparo al tratamiento que realizan las empresas en relación con el cumplimiento de sus obligaciones de la legislación laboral.
- ▶ El tratamiento está destinado y **es necesario para proteger un interés vital** del propio interesado o de otra persona. Como es el caso de situaciones de vida o muerte del afectado.
- ▶ Si el tratamiento es necesario **para el cumplimiento de una misión pública** o en ejercicio de poderes públicos. Un ejemplo es la publicación de listados de colegiados por parte de los colegios profesionales, en cumplimiento de su misión de ordenación de la actividad de los colegiados y obligación de publicidad para velar por los intereses de los usuarios.

- ▶ **El tratamiento es necesario para la satisfacción de intereses legítimos** perseguidos por el responsable del tratamiento o una tercera parte.

Hay situaciones en las que el responsable de tratamiento puede hacer valer su interés legítimo a la hora de efectuar un tratamiento de datos personales, sin necesidad de que exista consentimiento del afectado, siempre que no prevalezca este interés legítimo frente a los derechos de los afectados y no afecte a categorías especiales de datos.

Para entender un poco mejor de qué situaciones estamos hablando, nada mejor que recurrir al texto del RGPD, concretamente al considerando 47:

«El interés legítimo de un responsable del tratamiento, incluso el de un responsable al que se puedan comunicar datos personales, o de un tercero, puede constituir una base jurídica para el tratamiento, siempre que no prevalezcan los intereses o los derechos y libertades del interesado, teniendo en cuenta las expectativas razonables de los interesados basadas en su relación con el responsable. Tal interés legítimo podría darse, por ejemplo, cuando existe una relación pertinente y apropiada entre el interesado y el responsable, como en situaciones en las que el interesado es cliente o está al servicio del responsable. En cualquier caso, la existencia de un interés legítimo requeriría una evaluación meticulosa, inclusive si un interesado puede prever de forma razonable, en el momento y en el contexto de la recogida de datos personales, que pueda producirse el tratamiento con tal fin» (considerando 47, RGPD).

El propio considerando nos da dos ejemplos de tratamientos bajo interés legítimo: «el tratamiento de datos de carácter personal estrictamente necesario para la prevención del fraude [y] el tratamiento de datos personales con fines de mercadotecnia directa puede considerarse realizado por interés legítimo» (considerando 47, RGPD).

Para profundizar en esta cuestión está disponible el Dictamen 06/2014 sobre el concepto de interés legítimo del responsable del tratamiento de los datos en virtud del artículo 7 de la Directiva 95/46/CE, del Grupo de trabajo del artículo 29, del año 2014: [https://www.aepd.es/sites/default/files/2019-12/wp217\\_es\\_interes\\_legitimo.pdf](https://www.aepd.es/sites/default/files/2019-12/wp217_es_interes_legitimo.pdf)

Por último, hay que destacar que el RGPD prevé que los estados miembros puedan introducir disposiciones que amplíen o concreten los aspectos recabados en los apartados b) y d).

### **Tratamiento de datos de menores de edad y su consentimiento**

El tratamiento de datos de menores tiene una consideración especial, por su vulnerabilidad, como en estos casos (considerando 38, RGPD):

- ▶ Utilización de datos personales de niños con fines de mercadotecnia.
- ▶ Elaboración de perfiles de personalidad o de usuario.
- ▶ Obtención de datos personales relativos a niños cuando se utilicen servicios ofrecidos directamente a un niño.

Para ello la norma establece que el consentimiento del menor solo sea válido a partir de los 16 años (art. 8, RGPD). Ahora bien, la norma reconoce el hecho diferencial y de costumbre de los países de la UE que no atribuyen a la misma edad la capacidad de obrar en su propio nombre, por ello el RGPD reconoce a los Estados miembros de la UE la capacidad de poder determinar la edad a partir de la cual se considera con capacidad de consentir, que no podrá ser menor de 13 años. Así, en España, la actual Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LOPD-GDD) fija la edad válida para el consentimiento en los 14 años.

Para el tratamiento de los datos de los menores de 14 años que se fundamente en el consentimiento, será necesario recoger el consentimiento titular de la patria potestad o tutela del menor. El responsable de tratamiento deberá demostrar que cuenta con este consentimiento.

**Para los menores de edad, el consentimiento al tratamiento de datos personales debe ser facilitado por los padres o tutores.**

Aunque también contempla limitaciones (considerando 38, RGPD) y no debe requerirse en el acceso o utilización de los servicios preventivos o de asesoramiento ofrecidos directamente a los niños. Un ejemplo: servicios de prevención de maltrato infantil o acoso escolar.

### **Tratamiento de datos de categorías especiales**

El RGPD en su artículo 9 prohíbe de manera taxativa el tratamiento de datos personales considerados especiales. Las categorías de datos especiales son:

- ▶ Los datos que revelen el origen étnico o racial.
- ▶ Las opiniones políticas.
- ▶ Las convicciones religiosas.
- ▶ Las convicciones filosóficas.
- ▶ La afiliación sindical.
- ▶ Los tratamientos de datos genéticos.
- ▶ Los datos biométricos dirigidos que identifican unívocamente a una persona.
- ▶ Los datos relativos a la salud de las personas.
- ▶ Los datos relativos a la vida sexual de las personas o a su orientación sexual.

Será legítimo el tratamiento de los datos correspondientes a categorías especiales cuando medie **consentimiento explícito** por parte del afectado o se den alguna de las siguientes circunstancias:

- ▶ Se requiere para que el responsable del tratamiento cumpla con sus obligaciones legales en el marco laboral y de la seguridad social.
- ▶ Es necesario para proteger un interés vital del afectado o de otra persona, si el afectado no está en disposición de otorgar su consentimiento, por ejemplo, por estar incapacitado físicamente.
- ▶ El tratamiento es realizado en el marco de la actividad de una organización sin ánimo de lucro, fundación o asociación cuya finalidad sea política, filosófica, religiosa o sindical. Esta excepción solo es aplicable para el caso de que los afectados sean miembros o lo hubieran sido o se trate de simpatizantes que se comuniquen con frecuencia. Todo ello siempre que el tratamiento sea dentro de la organización y cualquier comunicación al exterior requerirá consentimiento del afectado.
- ▶ Los datos son públicos y ha sido el interesado el que los ha hecho públicos.
- ▶ Cuando el tratamiento sea necesario en las acciones de reclamaciones judiciales.
- ▶ El tratamiento es necesario por razones de interés público, avalado por la ley.

- ▶ El tratamiento es necesario en el ámbito de la asistencia sanitaria o medicina preventiva conforme al derecho de la UE.
- ▶ El tratamiento es necesario por motivos de salud pública conforme al derecho de la UE.
- ▶ El tratamiento es necesario para fines de archivo de interés público, histórico, científico o estadístico conforme al derecho de la UE.

## 8.5. Derecho de información

El RGPD establece la obligación de que el responsable de tratamiento, cuando vaya a recabar datos personales y de manera previa a su recogida o su registro, deba informar a los interesados. Este derecho se desarrolla en el RGPD en dos artículos:

- ▶ Artículo 13: «Información que deberá facilitarse cuando los datos personales se obtengan del interesado».
- ▶ Artículo 14: «Información que deberá facilitarse cuando los datos personales no se hayan obtenido del interesado».

El **objetivo** es:

- ▶ **Garantizar la transparencia** del tratamiento de manera que se conozca claramente la finalidad del tratamiento, quien va a efectuar el mismo y el tiempo durante el cual los datos serán tratados.
- ▶ **Sustentar la legitimidad del tratamiento** cuando este se base en el consentimiento del afectado, basado en la información facilitada por el responsable de tratamiento.
- ▶ **Informar de los derechos que asiste al afectado** con respecto al tratamiento de datos personales y dónde poder ejercitarlo, facilitando información de la empresa y datos de contacto.
- ▶ **Conocer si los datos serán exportados internacionalmente**, identificando los países destinatarios y la existencia de un nivel de protección equivalente al europeo.

Cuando los datos se recaban directamente del interesado, por ejemplo, en los casos en que los datos proceden de alguna comunicación o cesión legítima, la organización receptora de los datos debe informar a los interesados en un plazo razonable (art. 14.2, RGPD):

- ▶ Como máximo antes de que trascurra un mes desde que se consiguen los datos.
- ▶ En el caso de que los datos personales sean para comunicarse con el interesado, se deberá informar como tarde en la primera comunicación con el interesado (por ejemplo, en el primer envío de información publicitaria).
- ▶ Y habrá que informar antes de que los datos se comuniquen a terceros, otros destinatarios.

El responsable de tratamiento debe actuar proactivamente, garantizando el derecho de información antes de la recogida de los datos, y debe probar que cumplió con la obligación de informar.

**Hay excepciones a esta obligación:** por un lado, cuando ya el interesado disponga de dicha información; por otro, siendo los datos no recabados directamente del interesado (datos cedidos o comunicados por un tercero), cuando se de alguna de las siguientes circunstancias (art. 14.5, RGPD):

- ▶ La comunicación no sea posible o suponga un esfuerzo desproporcionado, por ejemplo, en situaciones de un volumen desproporcionado de usuarios y en particular cuando el tratamiento tenga fines de archivo en interés público, investigación científica o histórica o fines estadísticos.
- ▶ La obtención de los datos o la obligación de su comunicación esté recogida en el ordenamiento jurídico.
- ▶ Cuando por ley el tratamiento de los datos deba tener carácter confidencial.

Tal y como recoge el artículo 12 («Transparencia de la información, comunicación y modalidades de ejercicio de los derechos del interesado»):

«La información facilitada, dentro del marco de transparencia, debe ser clara, concisa, transparente, legible de fácil comprensión» (art. 12, RGPD).

En especial, en los casos en los que la información se dirija a los niños. En la práctica, nos encontramos con formas habituales de informar como son los típicos formularios en papel o electrónicos de registro de usuarios, pero también van creciendo otros puntos de recogida de datos, lo que implica establecer nuevos mecanismos para garantizar el derecho de información. Por ejemplo, se recopilan datos en aplicaciones móviles, mediante sensores IoT, durante la navegación de páginas web, en entrevistas telefónicas, etc.

En cualquiera de los casos, siempre habremos de aplicar el principio de transparencia y la capacidad de demostrar el cumplimiento del derecho de información.

**El RGPD contempla dos posibles situaciones a la hora de informar**, que no solo afectan al momento en el que se ha de facilitar la información, sino también a la información misma por facilitar. Así tenemos:

1. Los datos se recaban directamente del interesado (artículo 13).
2. Los datos no se obtienen directamente del interesado (artículo 14).

A continuación, veremos los requerimientos de información concretos para cada uno de los casos.

- ▶ Los datos se recaban directamente del interesado. En este caso la información que debe ser facilitada al interesado es (art. 13, RGPD):
  - ▶ La identificación de quien determina la finalidad del tratamiento de los datos:
    - La identidad del responsable del tratamiento y los datos de contacto y, en su caso, de su representante.
    - Los datos de contacto del delegado de protección de datos.
  - ▶ Para qué se van a tratar los datos: finalidad del tratamiento.

- ▶ La base de legitimación del tratamiento:
  - El consentimiento explícito del afectado.
  - O la base jurídica del mismo, cuando para el tratamiento no se requiere que exista consentimiento por parte del interesado por la legislación.
  - O los intereses legítimos explícitos, del responsable del tratamiento, cuando el tratamiento se legitimase en ellos.
- ▶ Los destinatarios de los datos personales, que no necesariamente debe ser solo el responsable de tratamiento.
- ▶ Sobre las transferencias internacionales:
  - Informar de la transferencia prevista, países destinatarios.
  - Indicar si existe o no el reconocimiento de adecuación por parte de la Comisión Europea.
- ▶ El plazo de conservación de los datos:
  - El tiempo de conservación establecido.
  - Criterios que determinarán el plazo de conservación.
- ▶ Los derechos del afectado:
  - Acceso, rectificación, supresión, limitación del tratamiento, oposición y portabilidad de los datos.
  - Derecho a retirar el consentimiento, para el caso de que el tratamiento se legitimase en virtud del consentimiento del afectado, sin que ello afecte al tratamiento realizado previamente.
  - Derecho a presentar reclamación frente a la autoridad de control.

- ▶ Sobre la obligación de facilitar los datos o de las consecuencias de no facilitarlos, indicando si los datos obligatorios se han de facilitar por:
  - Requisito legal o contractual.
  - Requisito para suscripción de un contrato.
- ▶ Si los datos serán utilizados para la toma de decisiones automatizadas o elaboración de perfiles, deberá informar:
  - Explicación razonable de la lógica aplicada.
  - Y las consecuencias previstas del tratamiento para el interesado.
- ▶ Los datos no se recaban directamente del interesado. En este caso, a la información anteriormente especificada habrá que añadir (art. 14, RGPD):
  - ▶ El origen de los datos.
  - ▶ Las categorías de datos personales que son objeto de tratamiento.

Debemos tener en consideración que, al no ser los datos recabados directamente del interesado, el interesado no dispone de esta información.

El RGPD recoge la **posibilidad de que la información sea facilitada por capas o niveles**. De manera que se disponga de una **primera capa con información básica resumida** que se presentaría al afectado en el mismo momento y en el mismo medio utilizado para la recogida de los datos. También se incluiría una referencia mediante la que se redirige al usuario para acceder a una segunda capa de información adicional, donde se ha incorporado de forma detallada toda la información requerida.

Por ejemplo, la Agencia Española de Protección de Datos, para el ejercicio del derecho a la información, recomienda el siguiente esquema de agrupación de la información por capa a modo de resumen:

Epígrafe	Información básica (1ª capa, resumida)	Información adicional (2ª capa, detallada)
<b>“Responsable”</b> (del tratamiento)	Identidad del Responsable del Tratamiento	Datos de contacto del Responsable
		Identidad y datos de contacto del representante
		Datos de contacto del Delegado de Protección de Datos
<b>“Finalidad”</b> (del tratamiento)	Descripción sencilla de los fines del tratamiento, incluso elaboración de perfiles	Descripción ampliada de los fines del tratamiento
		Plazos o criterios de conservación de los datos
		Decisiones automatizadas, perfiles y lógica aplicada
<b>“Legitimación”</b> (del tratamiento)	Base jurídica del tratamiento	Detalle de la base jurídica del tratamiento, en los casos de obligación legal, interés público o interés legítimo.
		Obligación o no de facilitar datos y consecuencias de no hacerlo
<b>“Destinatarios”</b> (de cesiones o transferencias)	Previsión o no de Cesiones	Destinatarios o categorías de destinatarios
	Previsión de Transferencias, o no, a terceros países	Decisiones de adecuación, garantías, normas corporativas vinculantes o situaciones específicas aplicables
<b>“Derechos”</b> (de las personas interesadas)	Referencia al ejercicio de derechos.	Cómo ejercer los derechos de acceso, rectificación, supresión y portabilidad de sus datos, y la limitación u oposición a su tratamiento
		Derecho a retirar el consentimiento prestado
		Derecho a reclamar ante la Autoridad de Control
<b>“Procedencia”</b> (de los datos)	Fuente de los datos (cuando no proceden del interesado)	Información detallada del origen de los datos, incluso si proceden de fuentes de acceso público
		Categorías de datos que se traten

Figura 3. Esquema de agrupación de la información. Fuente: AEPD, s.f.

El RGPD posibilita que la información que se facilita en la primera capa se articule en torno a la combinación de iconos normalizados. Se pretende facilitar de un modo sencillo, fácilmente visible y legible una primera información sobre las características del tratamiento. Información que será ampliable conforme a lo anteriormente expuesto.

De esta manera, se trata de dar respuesta al habitual desistimiento por parte de los interesados de la lectura de cláusulas informativas y la tendencia a otorgar el consentimiento sin reparar en la información del tratamiento al que se está consintiendo. De este modo, unos iconos facilitan una primera aproximación que,

aunque no exhaustiva, sí sirve al usuario para focalizar la atención e incluso diferenciar unos escenarios de otros de forma sencilla y visual. Por ejemplo, si existen cesiones o comunicaciones de datos o si los datos serán objeto de transferencia internacional. El establecimiento de este conjunto normalizado de iconos es potestad de la Comisión Europea.

ICON	ESSENTIAL INFORMATION	FULFILLED
	No personal data is <b>collected</b> beyond the minimum necessary for each specific purpose of the processing	
	No personal data is <b>retained</b> beyond the minimum necessary for each specific purpose of the processing	
	No personal data is <b>processed</b> for purposes other than the purpose it was provided for	
	No personal data is <b>disseminated</b> to private third parties for purposes other than the purpose it was provided for	
	No personal data is <b>sold</b>	
	No personal data is retained in <b>unencrypted</b> form	

Figura 4. Ejemplo de los iconos propuestos en la fase legislativa del RGPD. Fuente: Hoskins, 2012.

Ejemplo de cláusula de información para el cumplimiento con el derecho de información

- ▶ Información básica (primera capa):

Responsable	Editorial de publicidad S. A.
Finalidad	Gestión de la suscripción a la publicación semanal de la revista.
Legitimación	Ejecución de un contrato.
Destinatarios	No se contemplan cesiones más que las legalmente exigidas.
Derechos	Acceder, rectificar y suprimir los datos, para conocer todos sus derechos; acceder a la información adicional.
Información adicional	Puede consultar la información adicional en detalle en relación con el tratamiento de datos de carácter personal en la página web <a href="http://www.edipublisa.com/IRGPD">http://www.edipublisa.com/IRGPD</a> (link simulado para el ejemplo).

Tabla 1. Ejemplo información básica (primera capa). Fuente: elaboración propia.

- ▶ Información adicional (segunda capa):

## **Responsable del tratamiento**

Identidad: Editorial de publicidad S. A. CIF: A99855669.

Dirección: C/LOPD n.o 22, Madrid, 28055.

Teléfono: 91999999.

Email: rgpd@edipublisa.com

## **Delegado de protección de datos**

Contacto: dpd@edipublisa.com

## **Finalidad**

Le informamos de que el tratamiento de sus datos es necesario para cumplir con nuestras obligaciones del contrato de suscripción, para el envío de nuestro ejemplar semanal y para proceder al giro de los recibos bancarios para el pago de las cuotas mensuales y facilitarle la información que nos solicite. También trataremos la información para mejorar su experiencia de usuario, para ello elaboraremos un perfil comercial con sus preferencias sobre los contenidos de la revista. Su perfil no será utilizado en la toma de decisiones automatizadas.

En el caso de que haya marcado la casilla correspondiente a aceptar ser informado sobre ofertas de productos y servicios de su interés, procederemos al tratamiento de sus datos a tal fin enviándole periódicamente información comercial.

Le informamos de que conservaremos sus datos durante la duración del contrato manteniendo la información el tiempo indispensable para garantizar las obligaciones legales.

Y en el caso de que hubiera manifestado su interés en ser informado sobre nuestra oferta de productos y servicios de su interés, mantendremos su información hasta que nos indique lo contrario.

## **Legitimación**

El tratamiento de datos se efectúa en virtud del contrato de suscripción que usted mantiene con nuestra organización.

## **Destinatarios**

Se informa de que sus datos serán comunicados con el único fin de facilitar el envío de la publicación a entidades de mensajería. También procederemos a la comunicación de sus datos a entidades bancarias para proceder al giro del recibo de suscripción.

No está prevista ninguna comunicación o cesión adicional de información.

## **Derechos**

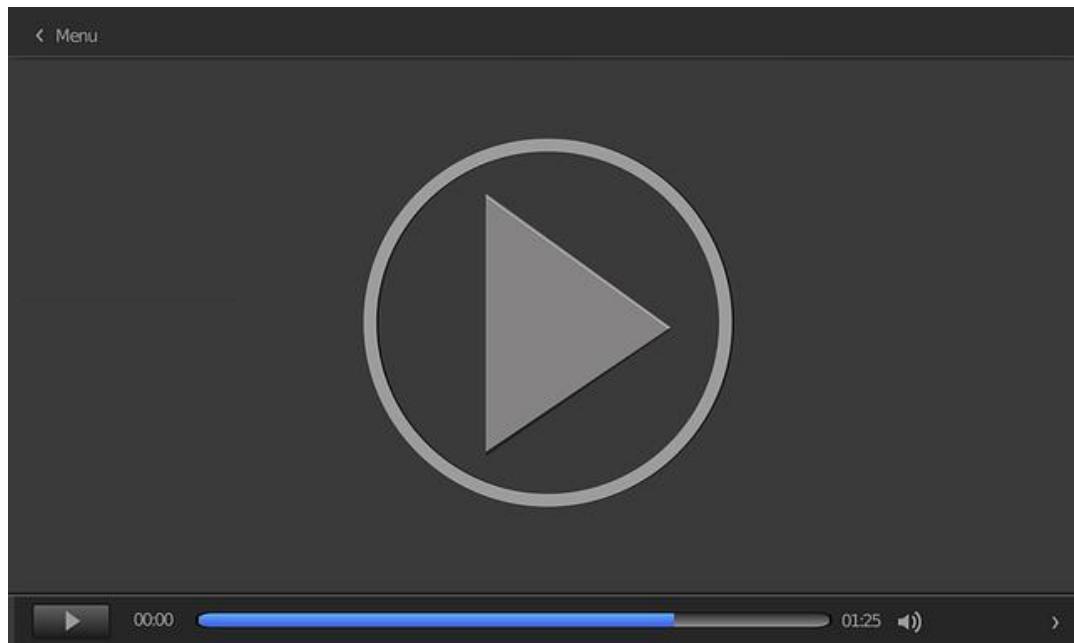
Le informamos de que puede ejercitar sus derechos, ante nuestro delegado de protección de datos, en la dirección [dpd@edipublisa.com](mailto:dpd@edipublisa.com) o mediante el teléfono 91999999, donde se le facilitará el formulario correspondiente para su solicitud. Se le informa de que deberá acompañar la solución de su DNI como prueba de su identidad.

Podrá ejercer su derecho:

1. Derecho a solicitar el acceso a los datos personales relativos al interesado.
2. Derecho a solicitar su rectificación o supresión.
3. Derecho a solicitar la limitación de su tratamiento de sus datos, si bien los conservaremos para el ejercicio o defensa de reclamaciones.
4. Derecho a oponerse al tratamiento como, por ejemplo, a la elaboración de perfiles comerciales.
5. Derecho a la portabilidad de los datos.

Para más información sobre sus derechos, puede consultar con nuestro DPD o en la Agencia Española de Protección de Datos.

Ahora, accede al vídeo *Concienciación en protección de datos*.



---

Accede al vídeo:<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=7d02c3a9-9215-40d8-ab0e-abd800c247c7>

---

## 8.6. El derecho de interesado

En este aspecto, el RGPD incorpora nuevos derechos que extienden los reconocidos en las legislaciones previas.

Al derecho de información se le unen otros seis derechos.

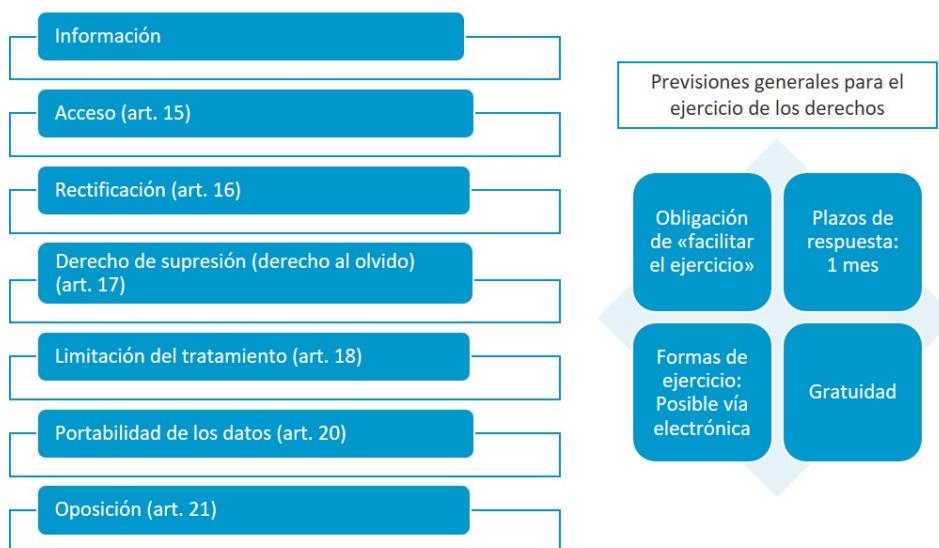


Figura 5. Derechos de interesado. Fuente: elaboración propia.

### Derecho de acceso (art. 15, RGPD)

Derecho del titular de los datos (ciudadano) a conocer de la mano del responsable de tratamiento **si se están tratando sus datos o no, y en caso afirmativo tener acceso a sus datos de carácter personal.**

Además, tendrá derecho a:

- ▶ Ser informado de las **finalidades del tratamiento** de datos personales.
- ▶ El tipo de datos que son recabados (categorías de datos).
- ▶ Los **destinatarios de los datos** o categorías de destinatarios a lo que son o serán comunicados los datos.
- ▶ Ser informado del derecho a solicitar: la rectificación o supresión de los datos, la limitación del tratamiento, la posibilidad de oponerse al tratamiento.
- ▶ Ser **informado del origen de los datos** cuando estos no hayan sido recabados directamente del interesado por parte del responsable de tratamiento.
- ▶ La **facultad de poder presentar una reclamación** ante la autoridad de control. Por ejemplo, en España ante la Agencia Española de Protección de Datos.
- ▶ En el caso de decisiones automatizadas, **derecho a conocer la lógica y las posibles consecuencias para el afectado** derivadas de esta decisión (art. 22, RGPD).
- ▶ En el caso de que los datos sean objeto de **transferencia internacional de datos**, el afectado tendrá **derecho a conocer las garantías que ofrece el país de destino**. (art. 46, RGPD).

El interesado tendrá derecho a acceder a una copia de los datos que son objeto del tratamiento, ahora bien, para copias adicionales el responsable del tratamiento podrá repercutir un coste.

## Derecho de rectificación (art. 16, RGPD)

Derecho del titular de los datos (ciudadano) a rectificar los datos inexactos, completar los datos existentes y a que el responsable del tratamiento proceda a la rectificación sin dilación indebida. El retraso puede ser motivo de sanción.

## Derecho de supresión (derecho al olvido) (art. 17, RGPD)

Derecho del titular de los datos (ciudadano) a solicitar y obtener, sin retraso no justificado, la eliminación (supresión) de sus datos de carácter personal.

El responsable del tratamiento estará obligado a atender la solicitud de supresión en los siguientes casos:

- ▶ Los datos personales ya no son necesarios para los fines que fueron recabados o tratados.
- ▶ El interesado retira el consentimiento habilitante del tratamiento (base legal).
- ▶ Existe un tratamiento ilícito de los datos personales.
- ▶ Por imperativo legal.
- ▶ Son datos personales obtenidos en relación con la oferta por parte del responsable del tratamiento de servicios de la sociedad de la información a menores.

En los casos de los datos personales que se hayan hecho públicos, el RGPD introduce una interesante novedad: la obligación del responsable del tratamiento de tomar medidas para informar a aquellos que estén tratando dichos datos personales, sobre la solicitud de supresión de:

- ▶ Todo enlace a esos datos personales.
- ▶ Cualquier copia o réplica de los datos.

En cualquier caso, la supresión no será aplicable en todas las situaciones, el derecho de supresión está limitado en los casos siguientes:

- ▶ El tratamiento es necesario para ejercer el derecho de información o de libertad de expresión.
- ▶ Existe obligación legal aplicable que exige el tratamiento.

- ▶ En el ámbito de presentación de reclamaciones (formulación, ejercicio o defensa).
- ▶ Motivos de interés público en el ámbito de la salud pública, que es uno de los ámbitos que el art. 9 («Tratamiento de categorías especiales de datos personales») recoge como lícito.
- ▶ Con la finalidad de investigación científica o histórica, fines de archivo en interés público o con finalidad estadística.

## Derecho a la limitación del tratamiento (art. 18, RGPD)

El derecho a la limitación en el tratamiento de datos que realiza un responsable de tratamiento en determinados casos. Podrá conservar los datos para un uso limitado. Será de aplicación:

- ▶ Cuando exista una impugnación por el interesado sobre la exactitud de los datos personales, y será aplicable durante el plazo razonable de verificación de la exactitud por parte del responsable de tratamiento.
- ▶ Ante un tratamiento ilícito, que obliga la supresión de los datos, el interesado prefiera que no se suprima, sino que se limite el uso de estos.
- ▶ Los datos no son ya necesarios para el responsable del tratamiento, lo que obligaría a la supresión de los mismos, pero sí son datos necesarios por parte del interesado en procesos de reclamaciones.
- ▶ Durante el tiempo en el que se evalúa la legitimidad por parte del interesado de oponerse a un tratamiento.

Como norma general para los derechos de rectificación, supresión o limitación del tratamiento, el responsable de tratamiento debe comunicar a los destinatarios de los datos el ejercicio de derechos efectuado (salvo que esto no sea posible o requiera medios desproporcionados).

En cualquier caso, el interesado tendrá derecho a conocer los destinatarios según el artículo 19 («Obligación de notificación relativa a la rectificación o supresión de datos personales o la limitación del tratamiento») del RGPD.

### **Derecho de portabilidad de los datos (art. 20, RGPD)**

Derecho a **recibir** por parte del responsable del tratamiento **los datos personales en un formato estructurado, para trasmitirlos a un tercero** o a que los datos se comuniquen directamente a un tercero seleccionado por el usuario. Esto será posible cuando:

- ▶ El tratamiento esté basado en el consentimiento del afectado o en un contrato.
- ▶ El tratamiento tenga lugar con medios automatizados.

Este derecho facilita la transición de los datos de una forma más eficaz entre los operadores de un sector, motivada por el cambio de cliente de operador, por ejemplo, en telefonía, energía, etc.

### **Derecho de oposición (art. 21, RGPD)**

El derecho a que el interesado pueda oponerse al tratamiento de sus datos personales.

El interesado podrá instar al cese del tratamiento de sus datos personales cuando este tratamiento no esté basado en el consentimiento de afectado y sí en el interés público o el ejercicio de poderes públicos o el interés legítimo del responsable de tratamiento.

En los casos en que el tratamiento de datos tenga una finalidad científica o histórica o estadística, y quede amparado por art. 89 del RGPD, el interesado tiene el derecho a oponerse a que sus datos sean tratados. Derecho que queda limitado si el tratamiento obedece a un interés público.

Un ejemplo de tratamiento basado en un interés legítimo es el **tratamiento de los datos cuya finalidad sea mercadotecnia directa o la elaboración de perfiles**. En este escenario, aunque al responsable le asiste el interés legítimo que le habilitaría para el tratamiento sin el consentimiento del interesado, este tiene el derecho a oponerse al tratamiento de forma total o parcial.

Como expone el RGPD, en el caso de existir un conflicto entre interés legítimo y el derecho del afectado, debe ser el responsable el que «demuestre que sus intereses legítimos imperiosos prevalecen sobre los intereses o los derechos y libertades fundamentales del interesado» (considerando 69, RGPD).

## **Decisiones individuales automatizadas, incluida la elaboración de perfiles (art. 22, RGPD)**

El derecho a no ser objeto de decisiones que tengan un efecto jurídico o le puedan afectar significativamente y que se basen en tratamientos automatizados de datos, lo que incluye la elaboración de perfiles, a menos que:

- ▶ La decisión sea necesaria para el establecimiento o ejecución de un contrato entre los interesados y el responsable de tratamiento.
- ▶ Esté autorizada por la ley.
- ▶ Exista un consentimiento explícito del afectado.

En cualquier caso, estas decisiones no pueden basarse en categorías especiales de datos, como «datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera única a una persona física, datos relativos a la salud o datos relativos a la vida sexual o las orientación sexuales de una persona física» (art. 9.1, RGPD), salvo que sea aplicable alguna excepción de las previstas en el artículo 9.2 como, por ejemplo, el consentimiento explícito del afectado.

En todo caso, será necesario establecer medidas adecuadas para la salvaguarda de los derechos y libertades de los afectados, y también del interés legítimo del interesado.

## **Limitaciones a los derechos (art. 23, RGPD)**

El RGPD reconoce el derecho a que las leyes de la Unión Europea y de los Estados miembros puedan limitar los derechos de los ciudadanos en los casos en los que esto sea necesario para salvaguardar en democracia:

«a) La seguridad del Estado.b) La defensa.c) La seguridad pública.d) La prevención, investigación, detección o enjuiciamiento de infracciones penales o la ejecución de sanciones penales, incluida la protección frente a amenazas a la seguridad pública y su prevención.e) Otros objetivos importantes de interés público general de la Unión o de un Estado miembro, en particular un interés económico o financiero importante de la Unión o de un Estado miembro, inclusive en los ámbitos fiscal, presupuestario y monetario, la sanidad pública y la seguridad social.f) La protección de la independencia judicial y de los procedimientos judiciales.g) La prevención, la investigación, la detección y el enjuiciamiento de infracciones de normas deontológicas en las profesiones reguladas» (art. 23, RGPD).

Las limitaciones a los derechos deberán tener su correspondiente soporte legal. Su aplicación debe seguir preservando el respeto a los derechos y libertades de los ciudadanos, de manera que la limitación al derecho debe estar motivada y debe ser proporcional al fin (bien) que se persigue.

## 8.7. Obligaciones generales del responsable de tratamiento y encargado

En este epígrafe vamos a aproximarnos a las obligaciones desde la perspectiva de aquellas que deben cumplir las organizaciones que realicen tratamientos de datos de carácter personal, bien por que actúen como **responsable de tratamiento, decidiendo sobre la finalidad** de este, bien como **encargado de tratamiento siguiendo las instrucciones del responsable**.

Es importante que repases las definiciones del apartado Conceptos.

El reglamento tasa las obligaciones de los participantes en el tratamiento y la posible sanción derivada del incumplimiento de estas. De manera que el responsable de tratamiento está obligado a aplicar una serie de medidas que se han identificado como medidas de responsabilidad activa, pensadas para desarrollarse incluso antes de iniciar el tratamiento y durante toda la duración de este, de manera que se garantice que los tratamientos son conformes a la normativa RGPD.

Uno de los aspectos novedosos es el requerimiento de considerar, como base para determinar las medidas técnicas y organizativas destinadas a garantizar el cumplimiento del RGPD, una aproximación basada en el riesgo que el tratamiento específico puede suponer para los derechos y libertades de los afectados por el tratamiento y, lo que es muy relevante, poder demostrarlo (art. 24.1, RGPD).

Por lo tanto, será necesaria la realización de una valoración del riesgo que determinará las medidas más adecuadas para ser aplicadas. Para este análisis habrá que tener en consideración: los tipos de tratamiento efectuados (automatizados, no automatizados), la naturaleza de los datos o categorías de datos para ser tratados, el número de personas afectadas y el número de tratamientos que una organización realice.

En las grandes organizaciones, la AEPD (Agencia Española de Protección de Datos) recomienda que la aproximación a este análisis se efectúe aplicando alguna metodología de riesgos, mientras que, en organizaciones pequeñas, se pueda aproximar con un análisis cualitativo. En este sentido, la AEPD dispone de herramientas para ayudar a la pyme en esta labor. Esta aproximación requiere revisar periódicamente dichas medidas para su evaluación y mejora, y actualizarlas si es preciso (art. 24.2, RGPD), lo que habrá que abordar siempre que la valoración del riesgo del tratamiento pueda verse afectada.

Además, el responsable y encargado de tratamiento deberán llevar registro de las actividades de tratamiento de datos personales que realizan, debiendo estar identificados y documentados.

Uno de los objetivos clave del RGPD es garantizar la protección de los datos desde el diseño (art. 25.1, RGPD), aplicar medidas técnicas y organizativas para dar cumplimiento al RGPD, ya desde el momento de elección de los medios de tratamiento para darles continuidad durante toda la duración de este. Un ejemplo, la pseudonimización:

«La aplicación de la pseudonimización a los datos personales puede reducir los riesgos para los interesados afectados y ayudar a los responsables y a los encargados del tratamiento a cumplir sus obligaciones de protección de los datos» (considerando 28, RGPD).

La introducción de la pseudonimización como medida no excluye la aplicación de otras medidas destinadas a reducir los riesgos del tratamiento, pero sí se pretende generalizar su uso como medida de no difícil implementación y que facilita la mitigación del riesgo en mayor o menor medida en función del tratamiento de datos personales, los mecanismos de reversibilidad existentes y la seguridad de estos.

El RGPD incorpora, así mismo, la protección de datos por defecto (art. 25.2, RGPD), la aplicación de medidas técnicas y organizativas que garanticen por defecto que solo sean tratados los datos mínimos estrictamente necesarios para la finalidad del tratamiento. Esto obliga a la minimización de datos personales recogidos, la extensión de su tratamiento, la minimización del plazo de su conservación y de su accesibilidad.

El artículo 25 del RGPD recoge la posibilidad de que el responsable de tratamiento pueda utilizar como elemento de prueba del cumplimiento de sus obligaciones el modelo de certificación que recoge el RGPD en su artículo 42, que veremos más adelante.

Por ejemplo, en el caso de una red social donde los datos personales no se pongan a disposición de un número amplio de personas, a menos que el afectado así lo establezca. Es decir, que el usuario pueda, por un lado, decidir la información a la que tendrá acceso la comunidad de la red social con sus diferentes segmentos y que, por otro, al crear un perfil este se configure por defecto sin acceso público.

Por otra parte, los roles de los participantes en el tratamiento específico deben estar claros y definidos, por ejemplo, cuando un tratamiento cuenta con más de un responsable debido a que hay más de una organización y conjuntamente determinan los objetivos del tratamiento y sus medios. En estos casos, el RGPD requiere que el reparto de responsabilidades sea acordado entre las partes formalmente, en lo referente a las obligaciones del RGPD y el derecho a información, teniendo los afectados derecho a acceder a los aspectos esenciales del acuerdo. (art. 26, RGPD).

En estos casos los interesados tendrán derecho a acudir a cualquiera de los corresponsables de tratamiento.

Además, las prestaciones de encargados a responsables de tratamiento serán contractualmente establecidas de conformidad a los requerimientos de la norma. Es decir, habrá contrato formal y especificación de responsabilidades.

A continuación, se presenta un cuadro resumen, de **las medidas de responsabilidad activa** a las que está obligado a aplicar el responsable de tratamiento (art. 27, RGPD).

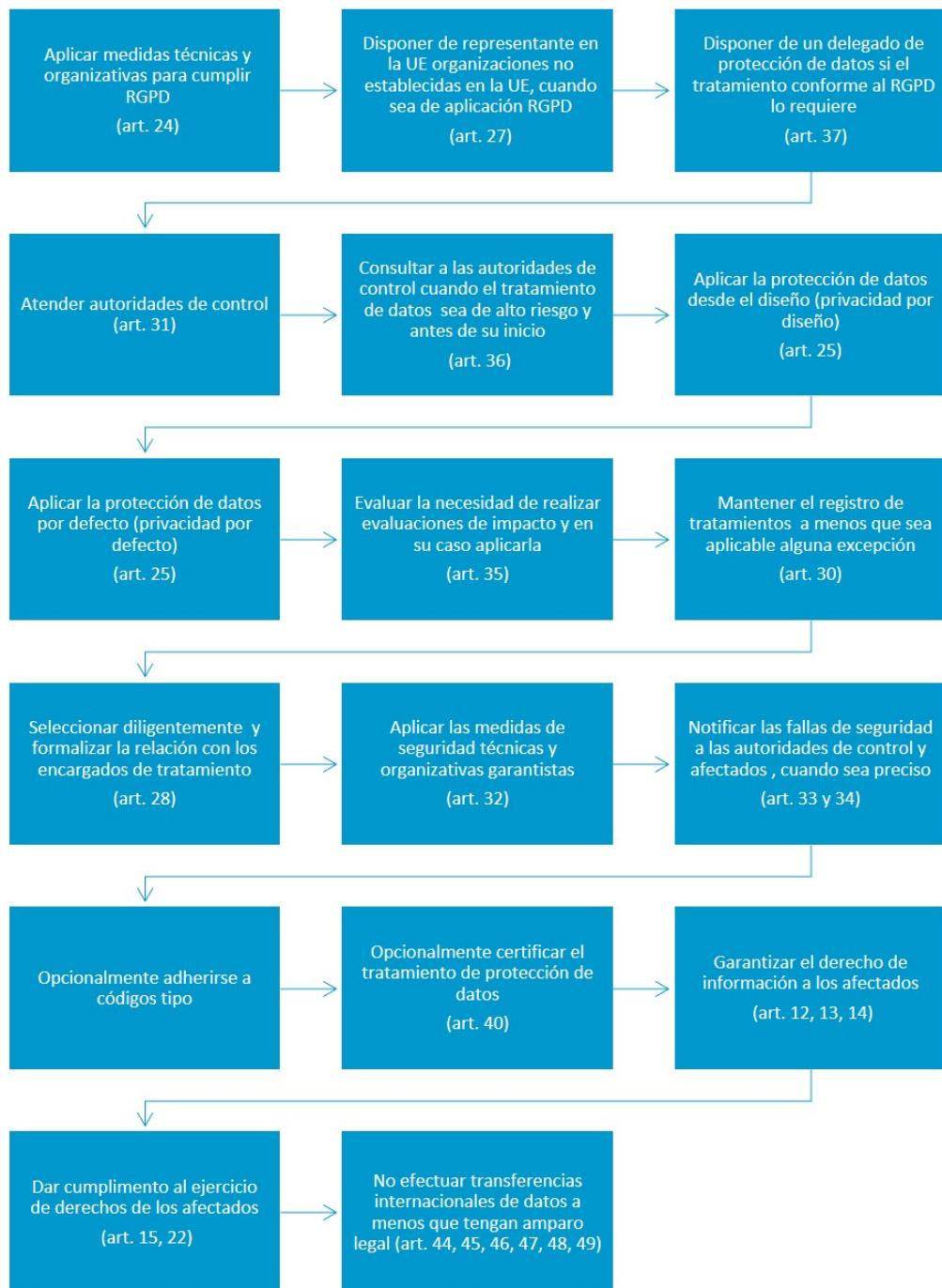


Figura 6. Medidas de responsabilidad activa. Fuente: elaboración propia.

## Responsabilidad del encargado de tratamiento

El RGPD transfiere al responsable de tratamiento (RT) la **obligación de actuar diligentemente** en los casos en que necesite para llevar a cabo el tratamiento de datos personales a un tercero, a un encargado de tratamiento (ET). De hecho, el responsable del tratamiento tiene la responsabilidad de que el encargado de tratamiento disponga y aplique las medidas técnicas y organizativas requeridas por el RGPD. Aunque el tratamiento se encargue a un tercero (encargado de tratamiento), **el responsable de tratamiento debe velar por el cumplimiento del RGPD.**

Además, **deberá existir entre las partes un contrato o equivalente** que fije:

- ▶ El objeto del tratamiento.
- ▶ La duración.
- ▶ La finalidad y naturaleza de este.
- ▶ El tipo y categorías de datos personales.
- ▶ Las categorías de interesados.
- ▶ Las obligaciones del responsable de tratamiento.
- ▶ Las obligaciones del encargado de tratamiento, como mínimo:
  - Obligación de tratar los datos únicamente siguiendo instrucciones del RT.
  - La obligación de confidencialidad del personal participante en el tratamiento
  - La aplicación de las medidas seguridad (art. 32, RGPD).
  - Autorización para subcontratación, y esta incorpora las mismas obligaciones ya establecidas entre RT y ET (principal).
  - La supresión o devolución de datos al finalizar el tratamiento a elección del RT.

- Facilitar al RT su capacidad de control y supervisión del cumplimiento de sus obligaciones —por ejemplo, mediante auditorías— y de prueba de cumplimiento por parte del ET de sus obligaciones.

Para facilitar la relación entre RT y ET, el RGPD recoge la posibilidad de que la relación se fundamente con base en cláusulas tipo a establecer por las autoridades de control.

Además, se activan mecanismos de certificación en el cumplimiento del RGPD y la adhesión a códigos de conducta como elementos probatorios de la capacidad de ofrecer garantías suficientes para el cumplimiento de las obligaciones establecidas en el RGPD (arts. 42 y 43, RGPD).

Si un ET infringe el RGPD, al determinar los medios y fines de tratamiento (que es obligación del RT), será considerado a todos los efectos RT y, por consiguiente, le serán exigidas las responsabilidades correspondientes como, por ejemplo, la prueba de disponer del consentimiento de los afectados para el tratamiento de datos personales.

## Registro de actividades de tratamiento (RAT) (art. 30, RGPD)

El RGPD determina la obligación de que el RT o representante legal lleven a cabo el registro documentado de las actividades de tratamiento de datos que se realizan. Es decir, **las actividades de tratamiento deben estar documentadas**. Este registro debe incluir:

- ▶ Nombre, datos de contacto del RT y si aplica corresponsable de datos, representante o delegado de protección de datos (DPO).
- ▶ La finalidad o finalidades del tratamiento de datos.
- ▶ Descripción de las categorías de interesados: clientes, trabajadores, alumnos, etc.
- ▶ Descripción de las categorías de datos personales tratados: identificación, salud, financieros, etc.
- ▶ Categorías de destinatarios a los que se comunicarán los datos: sector de actividad o grupos de actividad a los que pertenecen.
- ▶ En caso de que existan transferencias internacionales de datos, documentación de las garantías que presenta el tratamiento para las libertades y protección de datos que ampara la transferencia en el marco del RGPD incorporando la descripción del principio aplicado para la legitimación de las transferencias internacionales, así como la justificación de la aplicación de excepciones del artículo 49 del RGPD (que veremos más adelante).
- ▶ Plazos previstos para la supresión de los datos personales (de las categorías de datos): máximo tiempo que se mantendrán los datos o criterios para su cancelación.
- ▶ Descripción general de medidas técnicas y organizativas de seguridad aplicadas (art. 32.1, RGPD).

Además, la obligación de mantenimiento de un registro se extiende también al encargado de tratamiento:

- ▶ Nombre, datos de contacto del encargado de tratamiento, del responsable del tratamiento por cuenta del cual se efectúa el tratamiento de datos (quién contrata), representantes y delegado de protección de datos (DPO) y si aplica categorías de tratamientos realizados por cuenta del RT.
- ▶ En caso de que existan transferencias internacionales de datos, incluirá la documentación de garantías para las libertades y protección de datos que ampara la transferencia internacional en el marco del RGPD. En situaciones en las que el encargado de tratamiento realiza la exportación internacional de los datos para que estos sean tratados por una filial del grupo, se debe justificar el amparo legal a la misma.
- ▶ Descripción general de medidas técnicas y organizativas de seguridad aplicadas conforme al artículo 32.1 del RGPD.

Si la empresa u organización emplea a **menos de 250 trabajadores** no se le aplicarán la obligación de registro, siempre y cuando:

- ▶ El tratamiento de datos por realizar no entraña riesgo para los derechos y libertades de los ciudadanos.
- ▶ Es ocasional.
- ▶ No incluye categorías de datos especialmente protegidas o datos relativos a condenas e infracciones penales (art. 10, RGPD).
- ▶ Esto facilitara el tratamiento de datos a pequeñas empresas de sectores como los servicios no sociosanitarios.

- ▶ La notificación previa de los tratamientos a las autoridades de control en materia de protección de datos se había desarrollado de manera desigual en los países miembros de la UE, tanto en alcance de los tratamientos de obligada notificación, así como en el coste administrativo para los responsables de tratamiento. Con la entrada en vigor del RGPD, esta previsión desaparece. Por tanto, la necesidad de llevar este registro implicará la desaparición de la preceptiva inscripción de ficheros, previa al inicio del tratamiento (como era el caso de España ante la Agencia Española de Protección de Datos).

## **Cooperación con las autoridades de control (art. 31, RGPD)**

Los ET y RT están obligados a colaborar directamente o a través de sus representantes con las agencias de protección de datos.

## **Obligaciones sobre la seguridad de los datos**

El RGPD también incorpora la seguridad como principio de protección de datos, como ya veíamos en el artículo 5, donde se especifica este principio sobre el tratamiento de datos de carácter personal, que deberán ser «tratados de tal manera que se garantice una seguridad adecuada de los datos personales, incluida la protección contra el tratamiento no autorizado o ilícito y contra su pérdida, destrucción o daño accidental, mediante la aplicación de medidas técnicas u organizativas apropiadas (integridad y confidencialidad)» (art. 5.1.f, RGPD).

La aplicación de este principio necesariamente se traduce en **obligaciones para el responsable de tratamiento y para el encargado de tratamiento**, como se establece en el artículo 32 («Seguridad del tratamiento»), donde se requiere a ambos la **aplicación de medidas técnicas y organizativas apropiadas para garantizar un nivel de seguridad adecuado**. Para ello se tendrán en cuenta:

- ▶ El estado de la tecnología.
- ▶ Los costes de aplicación.
- ▶ La naturaleza del tratamiento y de los datos objeto de tratamiento.
- ▶ El alcance (el número de afectados, la duración...).
- ▶ El contexto.
- ▶ Los fines del tratamiento.
- ▶ Los riesgos, probabilidad y gravedad, variables para los derechos y libertades de las personas físicas.

Estableciendo que al menos deberán aplicarse las siguientes medidas técnicas y organizativas (art. 32, RGPD):

- ▶ La **pseudonimización** de los datos personales, ya comentada anteriormente.
- ▶ El **cifrado** de datos personales.
- ▶ **Medidas para garantizar la confidencialidad, integridad, disponibilidad y resiliencia** (la capacidad de un sistema para recuperar su estado inicial cuando ha cesado la perturbación a la que había estado sometido tanto de los sistemas como de los servicios de tratamiento).

- ▶ **Medidas para restaurar la disponibilidad de los sistemas y servicios de tratamiento** y el acceso a los datos de carácter personal, de un modo rápido en los casos de incidentes físicos (incendio, avería, etc.) o técnicos (corrupción de sistemas, etc.).
- ▶ Aplicar un proceso regular para la **verificación, evaluación y valoración de las medidas** tanto técnicas como organizativas para garantizar la seguridad del tratamiento.

Es importante resaltar que deberán tomar medidas para que cualquier individuo que participe en el tratamiento de datos personales solo pueda tratar dichos datos a partir de las instrucciones del responsable de tratamiento (art. 32.4, RGPD).

Un aspecto de potencial dificultad es la evaluación de la adecuación del nivel de seguridad aplicado teniendo en cuenta los riesgos que presenta el tratamiento para los derechos y libertades de los afectados. Para ello el artículo 32 introduce como criterio de evaluación «los riesgos que presente el tratamiento de datos, en particular como consecuencia de la destrucción, pérdida o alteración accidental o ilícita de datos personales transmitidos, conservados o tratados de otra forma, o la comunicación o acceso no autorizados a dichos datos» (art. 32.2, RGPD). En cualquier caso, no será fácil.

Para facilitar esta aproximación, el tercer punto del artículo 32 facilita la adhesión por parte del responsable de tratamiento a un **código de conducta** que haya sido aprobado por las autoridades de control (art. 40, RGPD) o la posibilidad de someter su sistema a un proceso de certificación, como prueba de cumplimiento de los requerimientos en materia de seguridad y del resto de sus obligaciones por parte del responsable de tratamiento de sus obligaciones en el marco del RGPD.

De manera que tanto el cumplimiento de códigos de conducta y la certificación opcional serán mecanismos que permitan mostrar por parte del responsable de tratamiento o encargado de tratamiento el cumplimiento de sus obligaciones.

Ante un incidente o evento de seguridad, no le eximirán de potenciales responsabilidades, pero si serán prueba de descargo ante un procedimiento sancionador.

Por consiguiente, los aspectos que veremos posteriormente tanto de certificación como de establecimiento de códigos de conducta adquieren un central protagonismo por la capacidad que estos deben aportar para objetivar los criterios de evaluación de la adecuación de las medidas técnicas y organizativas.

### **Notificaciones de violaciones de seguridad de los datos personales a la autoridad de control**

El reglamento establece como medida de seguridad adicional y de transparencia la obligación para los responsables de tratamiento **de notificar a la autoridad de control que le corresponda la violación de seguridad en un plazo de 72 horas**.

En el caso de que no se efectúe esta notificación en el plazo establecido por un retraso, deberá argumentarse los motivos del retraso (hay que recordar que la autoridad de control se determina en virtud de la aplicación del artículo 55). La notificación deberá incluir la siguiente información:

- ▶ Una descripción de la violación de seguridad y, si es posible, las categorías de datos afectados, así como dimensionar el número de interesados afectados y de registros de datos de carácter personal.
- ▶ Datos de contacto para ampliar información. Generalmente, serán los datos del delegado de protección de datos.
- ▶ Describir los riesgos o consecuencias de la violación sufrida.
- ▶ Declarar las medidas técnicas y organizativas adoptadas o propuestas para finalizar con la violación y, si procede, para reducir los riesgos para los afectados.

Como no siempre será posible disponer de toda la información indicada de un modo inmediato y en el plazo establecido, el reglamento permite que se pueda facilitar dicha información de manera progresiva (art. 33.4, RGPD).

**Hay que recordar que la no notificación de la violación puede ser objeto de sanción.**

Como obligación general, el RT deberá llevar un registro de violaciones incorporando todos los elementos relevantes relacionados con la violación, los efectos ocasionados y las medidas correctoras aplicadas. Este registro quedará a disposición de la autoridad de control.

### **Comunicación de violaciones de seguridad a los interesados**

Como medida adicional de transparencia, según el artículo 34 del RGPD, cuando exista un alto riesgo para los derechos y libertades de los ciudadanos, el responsable de tratamiento estará obligado a comunicar la situación a los afectados, informando de las características de la violación de seguridad y de las medidas adoptadas.

Esta comunicación no será necesaria siempre que se den alguna de las siguientes condiciones:

- ▶ La violación afecte a datos cifrados, por lo que el riesgo de acceso a datos personales sea despreciable.
- ▶ Que se tomaran medidas inmediatamente después de la violación que permitan considerar que no existe alta probabilidad de que potenciales perjuicios se materialicen para los afectados.
- ▶ Que suponga un esfuerzo ímparo, en cuyo caso bastará con una notificación pública.

## Evaluaciones de impacto

En los casos en los que es probable que un tratamiento entrañe un alto riesgo para los derechos y libertades de las personas físicas, el responsable de tratamiento realizará, antes del tratamiento, una evaluación del impacto de las operaciones de tratamiento en la protección de datos personales (PIA: *privacy impact assessment*) (art. 35, RGPD). ¿Cuándo habrá que realizar un PIA, atendiendo al RGPD? Cuando estemos en los siguientes escenarios:

- ▶ Evaluación sistemática y exhaustiva de aspectos personales de personas físicas que se base en un tratamiento automatizado.
- ▶ Elaboración de perfiles sobre cuya base se tomen decisiones que produzcan efectos jurídicos para las personas físicas o que les afecten significativamente.
- ▶ Tratamiento a gran escala de categorías especiales de datos o datos penales.
- ▶ Observación sistemática a gran escala de una zona de acceso público.  
Videovigilancia.

El RGPD establece el marco general de aplicación del PIA, pero reconoce a las autoridades de control la capacidad de determinar qué tratamientos (arts. 35.4 y 35.5, RGPD), por su naturaleza, deberán ser sometidos a un PIA. Por lo que progresivamente se irán concretando tratamientos en los que será requerido y aquellos en los que se considere de no aplicación.

A modo de resumen, hay que indicar que una evaluación de impacto, en línea con el RGPD, deberá incluir como mínimo:

- ▶ Descripción de los tratamientos de datos previstos, de sus finalidades y, si el tratamiento se sustenta en un interés legítimo, la especificación de este.
- ▶ Una evaluación de la necesidad del tratamiento propuesto y la proporcionalidad de dicho tratamiento con respecto las finalidades perseguidas.

- ▶ Una evaluación de los riesgos que el tratamiento pueda implicar para los interesados en lo referente a sus derechos y libertades.
- ▶ Las medidas que serán de aplicación para mitigar los riesgos mencionados, medidas para garantizar la seguridad y justificación de conformidad del tratamiento con el RGPD.

Siempre que exista un cambio en el perfil de riesgo del tratamiento, será preciso efectuar una revisión del PIA.

Si el resultado del PIA identificara que el tratamiento de datos personales es de alto riesgo, el responsable de tratamiento deberá consultar ante la autoridad del control, que se pronunciará por escrito para asistir al responsable para incorporar medidas adicionales requeridas y dar viabilidad al tratamiento (art. 36, RGPD).

### **Delegado de protección de datos**

El RGPD establece la obligación de nombrar un delegado de protección de datos (DPD). Esta es una figura nueva en el ordenamiento jurídico, aunque converge con las tendencias corporativas de identificación del *data privacy officer*.

Las funciones que ha de desempeñar el DPD incluyen:

- ▶ **Informar o asesorar** al responsable de tratamiento (en su caso, al encargado de tratamiento) y a sus empleados de las obligaciones que establece el Reglamento General de Protección de Datos.
- ▶ **Supervisar** que se da un adecuado **cumplimiento de la normativa** en los tratamientos de datos realizados por la organización y de sus políticas internas en materia de protección de datos personales, supervisando las asignaciones de responsabilidades, la concienciación y formación del personal y de las auditorías.
- ▶ Ser responsable de **supervisar la aplicación de evaluaciones de impacto** y asesorar al respecto, en los casos en que la organización esté obligada.

- ▶ Ser el **punto de contacto con las autoridades de control** y cooperar con ellas.
- ▶ Los datos de contacto del delegado de protección de datos deben ser públicos y, además, ser notificados a la autoridad de control correspondiente.
- ▶ Estar **a disposición de los interesados para los procesos de ejercicio de derechos por parte de los interesados**.
- ▶ Actuar **con independencia y reportando a la dirección** de la organización.

El perfil de un delegado de protección de datos (DPD) o *data protection officer* (DPO) necesita amplios conocimientos en derecho y en la práctica de la protección de datos personales, pero el reglamento no determina exactamente qué titulación debe ostentar el profesional para su desempeño como DPD.

Para dar respuesta a esta cuestión y con objeto de incorporar elementos tangibles y de confianza en el ejercicio profesional, desde la Agencia Española de Protección de Datos **se ha desarrollado el marco para la certificación del DPD**, una iniciativa pionera que desarrolla junto con la Entidad Nacional de Acreditación (ENAC) conforme a la norma UNE-EN ISO/IEC 17024:2012. La certificación está llamada a ser garante de las capacidades del DPD con relación a sus conocimientos y experiencia frente a los responsables de tratamiento, es decir, es una certificación profesional individual.

Las entidades certificadoras serán las acreditadas por parte de la ENAC y la certificación es voluntaria. Es decir, no hay obligación por parte del DPO de estar certificado.

---

Se puede acceder al esquema de certificación desde la web de la AEPD:

<https://www.aepd.es/sites/default/files/2020-07/esquema-aepd-dpd.pdf>

---

El nombramiento del DPO, que puede ser personal interno o externo, **será obligatorio** siempre que:

- ▶ El **tratamiento lo lleve a cabo una autoridad u organismo público** (excepto los tribunales que actúen en ejercicio de su función judicial).
- ▶ Las actividades principales del responsable o del encargado de tratamiento consistan en **operaciones de tratamiento que**, debido a su naturaleza, alcance o fines, **requieran una observación habitual y sistemática de interesados a gran escala**, como los casos de videovigilancia o seguimiento de actividad en Internet.
- ▶ Las actividades principales del responsable o del encargado consistan en el **tratamiento a gran escala de categorías especiales de datos personales**, como son los datos de salud.

Debemos resaltar que los **grupos empresariales podrán disponer de un único DPD**.

### **Responsabilidades civiles y penales de los responsables o encargados de tratamiento**

El incumplimiento de sus obligaciones por parte de responsables y encargados de tratamiento puede implicar responsabilidades de diversa índole:

- ▶ **Responsabilidad penal:** depende del código penal de cada uno de los países miembros de la UE, no existiendo una homogeneización en este sentido para el supuesto penal. A modo de ejemplo en la legislación española, el artículo 197 del Código Penal (capítulo I: «Del descubrimiento y revelación de secretos»):

SUPUESTOS	PENAS
Aquellos que descubran secretos o vulneren la intimidad de los interesados sin el consentimiento de estos, accediendo al correo o interceptando comunicaciones, realizando escuchas.	De 1 a 4 años, multas de 12 a 24 meses.
Aquellos que se apoderen, modifiquen o utilicen datos de carácter personal en perjuicio de un tercero (por ejemplo, la revelación o divulgación de imágenes personales).	De 1 a 4 años, multas de 12 a 24 meses.
Aquellos que, sin haber participado directamente en la obtención de la información personal «robada», sí participan en su divulgación.	Entre 2 y 4 años.
Las actividades ilícitas mencionadas son llevadas a cabo por el responsable de tratamiento.	De 3 a 5 años.
Si existe ánimo de lucro.	Hasta 7 años.
Aquellos que divulguen a terceros imágenes o grabaciones captadas inicialmente con el consentimiento del afectado y que atenten gravemente a la intimidad personal de esa persona.	<ul style="list-style-type: none"> <li>- De 3 meses a 1 año o multa de 6 a 12 meses.</li> <li>- De 2 a 6 años o multa de 12 a 24 meses en caso de que exista relación afectiva, la víctima sea menor de edad o exista fin lucrativo.</li> </ul>

Tabla 2. Supuestos y penas. Fuente: art. 197, Ley Orgánica 10/1995, de 23 de noviembre.

- ▶ **Responsabilidad administrativa:** deriva en sanciones administrativas impuestas por la autoridad competente. Como es el caso de las sanciones de las autoridades de control de protección de datos (en España, la Agencia Española de Protección de Datos).
- ▶ **Responsabilidad civil:** puede suponer el pago de indemnizaciones por los perjuicios sobre los afectados derivados del incumplimiento de la normativa en protección de datos.

## 8.8. Otros marcos internacionales

Son numerosas las iniciativas establecidas en formalizar marcos de referencia para garantizar la privacidad en los tratamientos de datos de carácter personal y que no tienen un componente legal. Entre las más interesantes tenemos:

- ▶ Principios Generales de la Protección de Datos de ODCE.
- ▶ ISO/IEC 29100:2011. Information technology. Security techniques. Privacy framework.
- ▶ ISO/IEC 27701:2019. Security techniques. Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management. Requirements and guideline.

Entre ellos no es difícil encontrar amplias similitudes respecto a los principios que establecen para el tratamiento de datos personales, si bien los desarrollos de estos marcan las diferencias.

Así, no es difícil establecer un **mapeo** entre los diferentes marcos de protección de la privacidad, mostrando su convergencia.

Principios de privacidad OCDE	Principios de privacidad ISO/IEC
1. Limitación a la recolección	1. Consentimiento y elección 3. Limitación a la recolección
2. Calidad de datos	6. Exactitud y calidad
3. Especificación del propósito	2. Legitimidad de propósito y especificación
4. Limitación en el uso	4. Minimización de datos 5. Limitación de uso, retención y divulgación:
5. Seguridad y salvaguardas	10. Seguridad de la información
6. Apertura	7. Apertura, transparencia y notificación
7. Participación individual	8. Participación individual y acceso
8. Responsabilidad ( <i>accountability</i> )	9. Rendición de cuentas 11. Cumplimiento de la privacidad

Tabla 3. Mapeo de marco de privacidad. Fuente: Brandon y de Souza, 2016.

### ISO/IEC 27701:2019, extensión de privacidad de la ISO/IEC 27001:2013

En 2019, se publica la ISO/IEC 27701:2019 (Security techniques. Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management. Requirements and guideline) con el objetivo de mejorar el sistema de gestión de seguridad de la información (SGSI) vinculado a la ISO/IEC 27001:2013, con requisitos adicionales para establecer, implementar, mantener y mejorar un sistema de gestión de información de privacidad (PIMS). La ISO/IEC 27701:2019 Es una norma certificable.

## 8.9. Protección de datos en EE. UU. y otros países

La regulación en materia de protección de datos personales es muy dispar a nivel internacional, existiendo notables diferencias entre los marcos regulatorios, tanto desde el punto de vista del ámbito de aplicación como de los derechos reconocidos.

Analizando la legislación de diferentes estados se identifican dos aproximaciones legislativas muy diferentes atendiendo a su tradición jurídica:

- ▶ Estados con un marco regulatorio general y completo, con leyes de tipo ómnibus que en algunos casos se complementan con regulaciones sectoriales.
- ▶ Estados que cuentan con múltiples normas jurídicas sectoriales o territoriales, por lo que en muchos casos resulta complejo identificar todas y cada una de las normas aplicables.

También se encuentra una amplia disparidad en lo referente a los derechos y obligaciones que establece la ley y el amparo que las mismas dan a los ciudadanos en relación con el tratamiento de sus datos personales.

En la siguiente infografía se recoge esta visión por parte de un relevante bufete internacional en el ámbito de la protección de datos, DLA Piper.

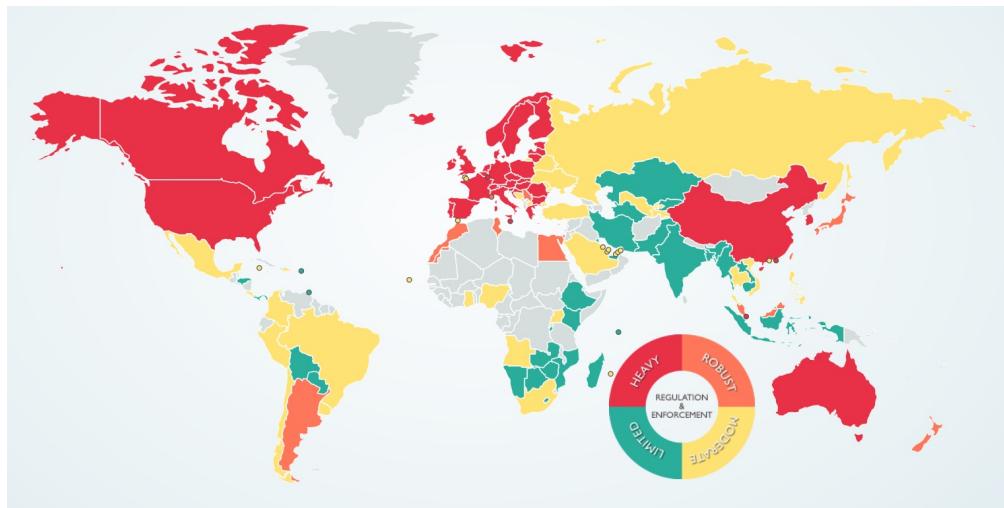


Figura 7. Protección de datos alrededor del mundo. Fuente: Avram et al., 2020.

---

Puedes acceder a la información de DLA Piper sobre la situación legislativa internacional en la web:<https://www.dlapiperdataprotection.com/>

---

En Iberoamérica, cabe resaltar la situación de Argentina, que desde el 30 de junio de 2003 tiene el reconocimiento de la Comisión Europea de ofrecer un nivel adecuado de protección de datos personales.

### **Protección de datos en EE. UU.**

Son muchos los que encuentran en EE. UU. los antecedentes de la privacidad y la protección de datos personales y el carácter inspirador de esta para la normativa europea en sus orígenes.

En 1798, la cuarta enmienda de la Constitución de EE. UU. reconoció el derecho de los ciudadanos a la inviolabilidad del domicilio, pero no sería hasta el siglo XIX cuando el concepto de privacidad moderno tomaría cuerpo en la sociedad estadounidense, a partir del debate social emergente que cuestionaba el amparo sobre la vida privada que ofrecía la ley.

De hecho, a finales del siglo, los juristas norteamericanos Samuel Warren y Louis Brandeis, los que lideran el concepto jurídico de **derecho a la vida privada** (*privacy law*), en un artículo publicado en la revista de la Universidad de Harvard con el título «The right to privacy» introdujeron por primera vez la privacidad como acción civil y promovieron la creación de un nuevo derecho civil de amparo al espacio personal frente a su difusión al público no autorizada.

Pero deberían pasar años hasta la consolidación del derecho a la privacidad, que en EE. UU. integra cuatro ámbitos: la esfera privada, la apropiación del nombre, la distorsión de la imagen y la difusión pública de hechos privados; para que a través de un conjunto de normas se fuera consolidando dicho derecho:

- ▶ *Privacy Act* (Ley de protección a la Intimidad de 1974).
- ▶ *Freedom of Information Act* (Ley de Libertad de la Información, FOIA).
- ▶ *Fair Credit Reporting Act* (Ley de Equidad Financiera de 1978).
- ▶ *Privacy Act* (Ley de protección de la Intimidad de 1974).

La Ley de Protección a la Intimidad de 1974 tiene por objeto la protección de la intimidad de las personas, cuyos datos personales figuran en bancos de datos del gobierno, incorporando los siguientes principios:

- ▶ La prohibición de la existencia de bancos de datos personales secretos.
- ▶ El derecho del ciudadano a conocer la información tratada y la finalidad de dicho tratamiento.
- ▶ El derecho a corregir la información registrada.
- ▶ La prohibición del tratamiento de datos personales para otras finalidades para las que fueron recabados sin consentimiento del titular.

La *Privacy Act*, que sería posteriormente modificada entre otros por las enmiendas surgidas de las leyes de Libertad de la Información (*Freedom of Information Act*), es de aplicación a las bases de datos del Gobierno federal con ciertas limitaciones, como son los archivos de los servicios de inteligencia, inmigraciones, lucha con el narcotráfico, etc., y reconoce el derecho de acceso a la información personal.

En este punto podemos encontrar cierta similitud con los principios que rigen la protección de datos en la UE. Pero, como indicaba, es una ley que afecta solo al Gobierno federal y es que el sistema jurídico anglosajón se desarrolla en torno a leyes sectoriales en lugar del establecer un marco general regulador, tipo ómnibus, como es el caso de la UE y de España.

Por ello no es difícil encontrar un conglomerado de normas a nivel federal o incluso estatal que versan sobre el derecho a la privacidad. EE. UU. dispone del entorno de 20 normas específicas o que abordan ampliamente la privacidad o la seguridad de los datos personales como, por ejemplo, *Fair Credit Reporting Act* de 1978, que regula aspectos del tratamiento de datos de los individuos por parte de las entidades financieras, exigiendo ciertas garantías de confidencialidad sobre el acceso por parte del Gobierno y por las agencias de información de crédito.

A ello se suman cientos de desarrollos normativos específicos de los Estados miembros. Por ejemplo, California dispone de más de 25 normas sobre privacidad y seguridad de datos.

En general, la normativa exige que sea preceptiva la información previa a la recolección de los datos, especialmente cuando se trata de datos considerados sensibles (la que versa sobre la salud, información financiera, la correspondiente a menores de 13 años...) y la utilización de los datos para finalidades permitidas.

En relación con la existencia de requerimientos de seguridad, las leyes tanto en el ámbito estatal como federal incorporan obligaciones destinadas a garantizar la confidencialidad de la información y no son pocos los Estados en los que es obligatorio comunicar la existencia de brechas de seguridad que hayan supuesto el acceso a datos de personales.

En EE. UU. no existe un organismo de control general en materia de protección de datos, como sí ocurre en los Estados miembros de la UE, pero sí dispone de instituciones o autoridades de vigilancia y control sectoriales como, por ejemplo, la Federal Trade Commission (FTC), entidad con la que se desarrolla los principios de *privacy shield*.

## Protección de datos en EE. UU. y la UE

Aunque existen amplios puntos de convergencia entre las normativas de EE. UU. y la UE, las primeras ideas sobre protección de datos a ambos lados del Atlántico nacen en la década de los 70 y uno de los fundamentos de los principios recogidos en el Convenio de Protección de Datos del Consejo de Europa (1981) tenía como inspiración los principios de información justa (*Report of the Secretary's Advisory Committee on Automated Personal Data Systems: Records, Computers and the Rights of Citizens*, 1973), principios que también inspiraron las directrices de privacidad de la OCDE (1980).

Sin embargo, estos principios evolucionaron de forma diferente: en EE. UU. apostaron por la autorregulación, mientras que en la UE llegaron a la consideración de derecho fundamental, vinculante en el Tratado de Lisboa de 2009.

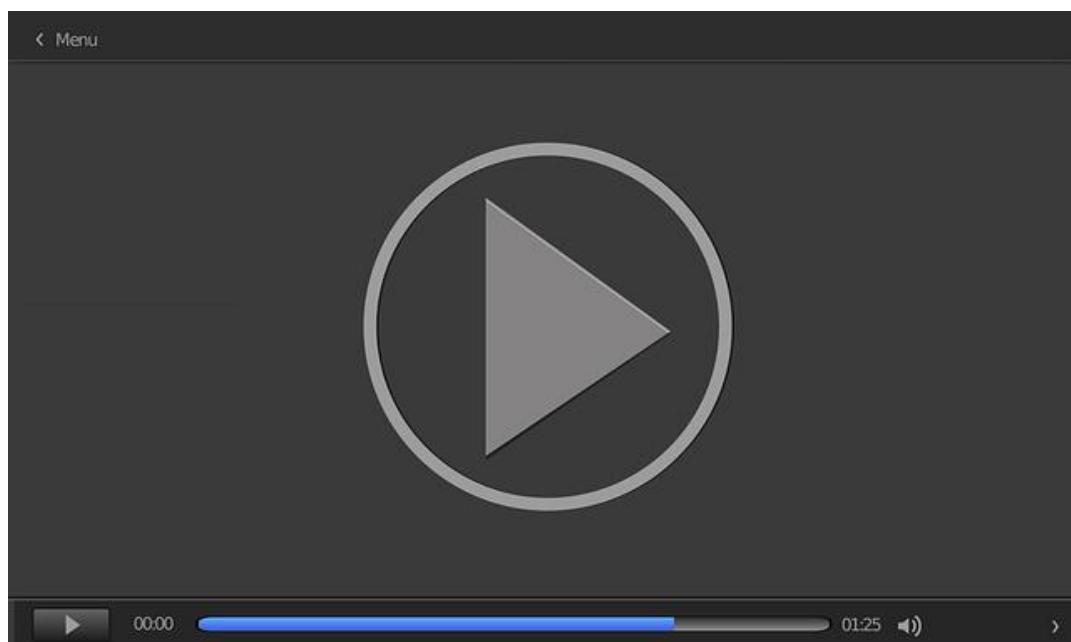
Por ello, el modo y alcance en el que se reconocen los derechos en torno a la protección de datos presentan notables divergencias:

- ▶ En la propia concepción del derecho que ya se manifiesta en la cuarta enmienda de la Constitución de EE. UU., con un alcance más limitado que el derecho a la vida privada que reconoce el artículo 7 de la Carta de la UE, donde «toda persona tiene derecho al respeto de su vida privada y familiar, de su domicilio y de sus comunicaciones» (art. 7, Carta de los Derechos Fundamentales de la Unión Europea).
- ▶ En la sustancial diferencia en la concepción de datos personales. En la UE, como hemos visto, abarca toda información que identifica o hace identifiable a una persona. Una definición mucho más amplia que la extendida en EE. UU., que se centra más en el dato que identifica y no en el dato que hace posible la identificación de las personas. También es posible encontrar alguna excepción a esto. Por ejemplo, en la nueva regulación COPPA (*Children's online privacy protection rule*) del FTC se reconoce como dato personal el identificador persistente que pueda ser utilizado para reconocer a un usuario en el tiempo a lo largo de diferentes webs o servicios *online*.

Esta concepción hace que la ley de EE. UU. se centre en reparar el daño al consumidor y la búsqueda de un adecuado equilibrio entre privacidad y las necesidades relacionadas con las transacciones comerciales que fundamenta la no existencia de restricciones en EE. UU. en relación con la exportación de datos (excepto para ciertas informaciones gubernamentales), con un impacto significativo en los flujos de datos internacionales. Un escenario en el que las organizaciones y las empresas de la UE se ven sometidas a muchas más restricciones que las de EE. UU.

Las implicaciones de estas divergencias y su impacto están detrás del debate dentro de la comunidad en busca de una mitigación de las barreras existentes a través de una aproximación en el concepto de dato personal en EE. UU. Por ejemplo, de PII 2.0. (*personal identification information*), que incorporaría al concepto de dato personal en EE. UU. los datos de personas identificables y la evaluación continua del riesgo de identificación.

Para terminar con este apartado, accede al vídeo *Las evaluaciones de impacto*.



---

Accede al vídeo:<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=df57beee-ccdd-4c6c-a6c1-abd800c247ee>

---

## 8.10. Transferencias internacionales de datos

Una de las motivaciones principales de establecer un marco común de protección de los datos de carácter personal es la de facilitar el movimiento transfronterizo de datos. En esta línea no solo tenemos el Reglamento General de Protección de Datos, sino también otras iniciativas como las de la OCDE (*OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*).

En este sentido, el RGPD sigue los mismos principios consolidados en la UE, que garantiza que el **movimiento de datos entre países miembros de la UE no está limitado**. Pero también se van a permitir aquellos movimientos internacionales de datos fuera de la UE, siempre que se cumplan con los requerimientos establecidos.

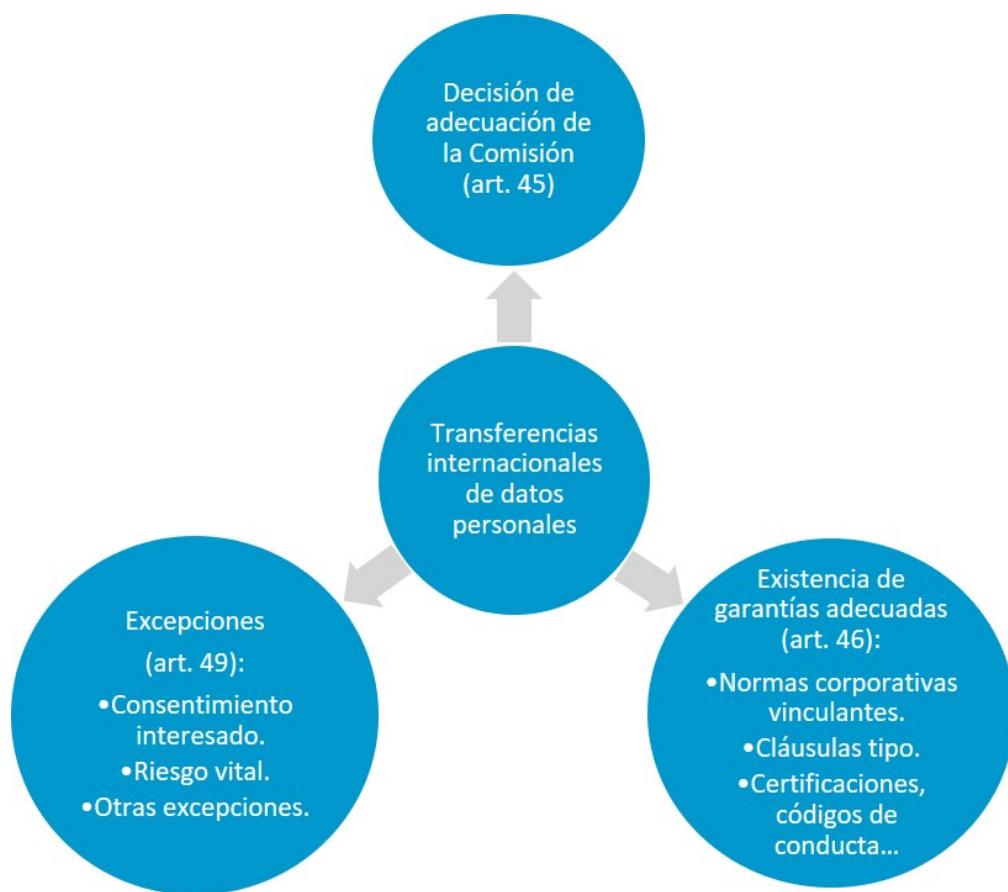


Figura 8. Transferencia de datos personales. Fuente: elaboración propia.

- ▶ **Transferencias basadas en una decisión de adecuación de la Comisión** (art. 45, RGPD): identifica a los países que tienen un marco de protección de datos de garantías equivalentes al de la UE. Para ello, la Comisión, según el RGPD, tendrá en consideración: el estado de derecho en relación con los derechos humanos y las libertades fundamentales, la existencia de mecanismos de control mediante autoridades independientes y los compromisos internacionales.

En la actualidad, la lista de países que ostenta esta consideración y a los que se puede realizar una exportación de datos legal bajo este principio son países con un nivel adecuado de protección, declarados hasta la fecha son los siguientes:

- ▶ **Suiza.** Decisión 2000/518/CE de la Comisión, en 2000.
- ▶ **Andorra.** Decisión 2010/625/UE de la Comisión, en 2010.
- ▶ **Canadá.** Decisión 2002/2/CE de la Comisión, en 2001, respecto de las entidades sujetas al ámbito de aplicación de la ley canadiense de protección de datos.
- ▶ **Argentina.** Decisión 2003/490/CE de la Comisión, en 2003.
- ▶ **Uruguay.** Decisión 2012/484/UE de la Comisión, en 2012.
- ▶ **Guernsey.** Decisión 2003/821/CE de la Comisión, en 2003.
- ▶ **Isla de Man.** Decisión 2004/411/CE de la Comisión, en 2004.
- ▶ **Jersey.** Decisión 2008/393/CE de la Comisión, en 2008.
- ▶ **Islas Feroe.** Decisión 2010/146/UE de la Comisión, en 2010.
- ▶ **Israel.** Decisión 2011/61/UE de la Comisión, en 2011.
- ▶ **Nueva Zelanda.** Decisión 2013/65/UE de la Comisión, en 2012.
- ▶ **Japón.** En 2019.

- ▶ **Transferencias internacionales mediante garantías adecuadas** (art. 46, RGPD): cuando no existe una decisión por parte de la Comisión de adecuación, se puede recurrir a otros mecanismos que ofrecen garantías adecuadas, reconocidos por el RGPD:
  - Instrumentos que vinculen jurídicamente y exigibles entre autoridades u organismos públicos.
  - Normas corporativas vinculantes (art. 47, RGPD).
  - Las cláusulas tipo de protección de datos, adoptadas por la Comisión.
  - Las cláusulas tipo de protección de datos personales, adoptadas por una autoridad de control y que hayan sido aprobadas por la Comisión.
  - Con base en el código de conducta junto a compromisos vinculantes y exigibles en el país destinatario.
  - Con base en la certificación junto a compromisos vinculantes y exigibles en el país destinatario.
  - Cláusulas contractuales entre el responsable o el encargado y el responsable, encargado o destinatario de los datos personales en el tercer país u organización internacional que hayan sido autorizadas por la autoridad de control competente.
  - Disposiciones incluidas en acuerdos administrativos entre las autoridades u organismos públicos que incluyan derechos efectivos y exigibles para los interesados y sobre los que se haya pronunciado positivamente la autoridad de control.
- ▶ Fuera de estos casos, la legitimación de una transferencia internacional debe estar sustentada por alguna de las siguientes **excepciones** (art. 49, RGPD):

- **Consentimiento explícito** del interesado a la transferencia de datos, tras haber sido informado de los posibles riesgos para él de dichas transferencias debido a la ausencia de una decisión de adecuación y de garantías adecuadas.
- **Necesaria para la ejecución de un contrato** entre el interesado y el responsable del tratamiento.
- Necesaria por razones importantes de interés público.
- La transferencia es requerida para la formulación, el ejercicio o la defensa de reclamaciones.
- Necesaria para proteger los **intereses vitales** del interesado o de otras personas.
- La transferencia se efectúa desde un registro público, destinado a facilitar información al público en general.

Además, si no es aplicable nada de lo anterior, se introduce una excepción, aplicable para las transferencias de datos, en los casos en que estas sean esporádicas, no repetitivas y para satisfacer de manera urgente el interés legítimo del responsable de tratamiento, siempre y cuando, como ya hemos visto con relación al interés legítimo, este prevalezca sobre los derechos y libertades de los afectados; y además debe ser notificada a las autoridades de control.

## Acuerdos

La UE y EE. UU. han buscado en los últimos 25 años mecanismos para salvar las diferencias regulatorias y facilitar las transferencias internacionales de datos de la UE hacia EE. UU. mediante acuerdos bilaterales. Primero fue el acuerdo Safe Harbor, al que le sucedió, tras su anulación en 2015, el acuerdo Privacy Shield, que también ha sido anulado, en julio de 2020, por el Tribunal de Justicia de la Unión Europea (TJUE).

Al cierre de este temario, para las transferencias internacionales de datos desde UE a los EE. UU., hay que tener en consideración que:

EE. UU. no goza de ningún acuerdo específico, por lo que debe regirse por los principios de la transferencia internacional de datos a cualquier país fuera de la UE y que no cuente con una decisión de adecuación de la Comisión.

## 8.11. Seguridad de la información y protección de datos

Como sabemos, la seguridad es uno de los principios fundamentales de la protección de datos personales, y así se contempla en el RGPD.

Los datos serán «tratados de tal manera que se garantice una seguridad adecuada de los datos personales, incluida la protección contra el tratamiento no autorizado o ilícito y contra su pérdida, destrucción o daño accidental, mediante la aplicación de medidas técnicas u organizativas apropiadas (“integridad y confidencialidad”)» (art. 5.1, RGPD).

Objetivos alineados con la definición de la seguridad de la información que hacen los principales estándares de la materia:

- ▶ **Seguridad de la información:** «preservación de la confidencialidad, integridad y disponibilidad de la información» (ISO/IEC 27000:2014).

Término al que también es habitual asociar otras propiedades como son autenticidad, responsabilidad, *accountability*, no repudiación o confiabilidad.

Hay que remarcar que **sin seguridad no sería posible garantizar la privacidad de los datos personales ni los derechos y libertades de los ciudadanos**. Ahora bien, la seguridad es un camino en búsqueda de un concepto absoluto, pero imposible de alcanzar, cual paradoja de Zenón.

**La seguridad absoluta no existe.** Se abordan los riesgos sobre la disponibilidad, integridad y confidencialidad aplicando medidas técnicas y organizativas, con objeto de mitigarlos, transferirlos o aceptarlos; buscando el equilibrio entre el coste de las medidas y su efecto sobre los riesgos.

Por ello es fundamental entender que la seguridad debe ser un proceso gestionado y continuo, orientado a la mejora continua, un ciclo de Deming, más conocido como PDCA (planificar-hacer-verificar-actuar permanentemente), destinado a dotar a los tratamientos de datos de carácter personal del entorno de seguridad adecuado y robusto para minimizar los riesgos sobre las personas.

El RGPD integra la seguridad dentro de las obligaciones de **responsabilidad proactiva**, tanto de responsables de tratamiento como de encargados de tratamiento. No detalla un conjunto de medidas de seguridad, específicamente establecidas, aunque el reglamento contempla aproximaciones que pueden ayudar a las organizaciones a demostrar su alineamiento con la responsabilidad exigida en el RGPD, como son las certificaciones y los códigos de conducta.

Aunque el RGPD prevé el desarrollo por parte de la autoridad de control de la especificación de medidas concretas de seguridad aplicables a escenarios de riesgo de tratamiento de datos personales específicos.

**El RGPD incorpora una aproximación a la seguridad desde la responsabilidad exigible, superando las aproximaciones tuteladas por las autoridades previas al RGPD.**

Este tutelaje por parte de la autoridad se mostraba de forma clara en el marco legislativo español, que ha incorporado un conjunto de medidas de seguridad concretas para aplicar a los tratamientos, medidas que eran obligatorias y con carácter de mínimos requeridos. El RD 1720/2007 (derogado) clasificaba las medidas de seguridad que aplicar en función de tres niveles de seguridad diferenciados y que se establecen en función del tipo de datos de carácter personal (DCP) manejados (art. 81, RDLOPD).

En España, estos reglamentos fueron una oportunidad para **promover el concepto de seguridad de la información** entre empresas y organizaciones.

De hecho, en el mundo de la mediana y pequeña empresa, donde la concienciación en seguridad era realmente baja, de la mano de las adaptaciones a la normativa en protección de datos personales se fueron incorporando conceptos como: uso de contraseñas, realización de copias de seguridad, personal autorizado, etc. Se introducían medidas de seguridad en las organizaciones para dar respuesta a una obligación y se beneficiaban las empresas al incluir dentro del dominio de aplicación de las medidas información vital para las empresas: clientes, RR. HH., etc. Seguramente, esto fue posible por la concreción de controles y medidas de seguridad.

## Buenas prácticas en la seguridad de la información

En estos años se han venido desarrollando formalmente conjuntos de buenas prácticas para la gestión de la seguridad de la información, alcanzando una madurez significativa y cuya aproximación a la seguridad siempre se basa en el riesgo y en la eficacia de los controles de seguridad en su reducción.

Entre los numerosos marcos existentes, a título informativo tenemos:

- ▶ [NIST](#). Las normas de seguridad del National Institute of Standards and Technology de EE. UU. Establece las bases de seguridad de las AA. PP. e integra publicaciones.
- ▶ Serie ISO/IEC 27000. Integra, entre otras, la conocida norma ISO/IEC 27001:2013 (Information technology. Security techniques. Information security management systems. Requirements).
- ▶ COBIT 2019. Marco de referencia para la gestión integral del gobierno de las TI en las organizaciones que incluye la seguridad de la información. Desarrollado por ISACA.
- ▶ Common Criteria (ISO/IEC 15408). Orientada a la armonización de los requisitos de seguridad de productos de *hardware, software, firmware*.

- [ENS](#): Esquema Nacional de Seguridad. Ofrece un marco de seguridad para las administraciones públicas españolas.

## ISO/IEC 27001

Veamos con un poco más de detenimiento la ISO/IEC 27001.

- ISO/IEC 27001/2
  - A brief history

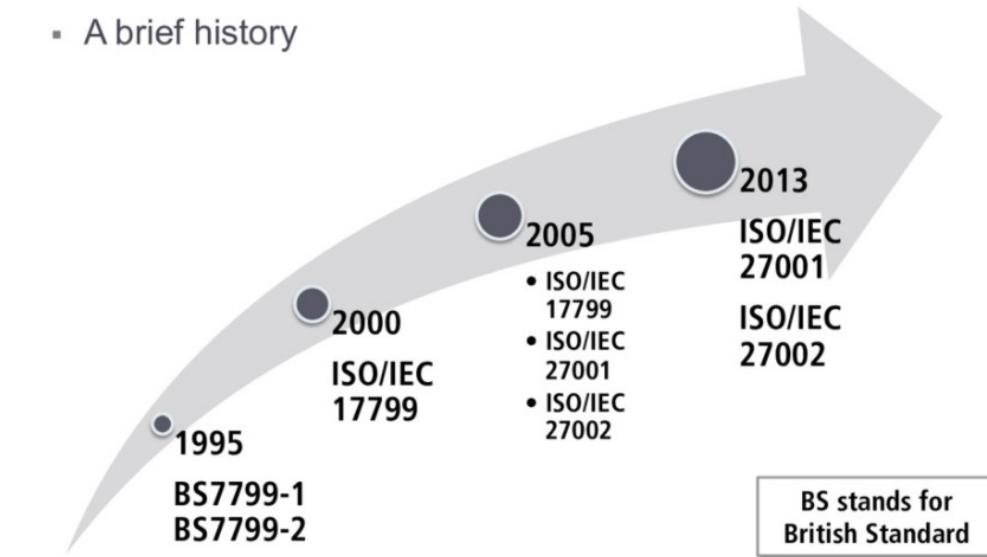


Figura 9. Breve historia ISO/IEC 270001. Fuente: Martins, 2018.

Consolida más de 18 años de trabajo en el desarrollo del estándar cuyos orígenes son desarrollados por British Standard Institution (BSI).

Establece el marco de referencia para un sistema de gestión de la seguridad de la información (SGSI) y es certificable IEC/ISO 27001:2013 (Information technology. Security techniques. Information security management systems. Requirements) y la norma ISO/IEC 27002:2013 (Information technology. Security techniques. Code of practice for information security controls) incorporaba un marco de buenas prácticas de controles de seguridad de la información.

Las dos normas forman parte de la serie ISO/IEC 27000, que desarrolla controles de seguridad, metodologías de evaluación de riesgos, etc. Es decir, desarrolla un marco de gestión de la seguridad de la información, que además es certificable.

Para familiarizarnos con los controles de seguridad que habitualmente se deben disponer en las organizaciones para salvaguardar los activos de información, nos introducimos brevemente en la norma de seguridad ISO/IEC 27002.

La ISO/IEC 27002 **incorpora 14 dominios de seguridad, sobre los que identifica 35 objetivos de control de seguridad, con un total de 114 controles de seguridad.**

En el siguiente esquema se recogen los dominios de seguridad y los objetivos de control de la norma.

## ISO/IEC 27001

5. Políticas de seguridad de las organizaciones.

6. Aspectos organizativos de los sistemas de información.

7. Recursos humanos.

8. Gestión de activos de información.

9. Control de acceso a la información.

10. Cifrado.

11. Seguridad física y del entorno.

12. Seguridad de la operativa.

13. Telecomunicaciones.

14. Adquisición, desarrollo y mantenimiento de los sistemas de información.

15. Suministradores.

16. Gestión de incidentes.

17. Continuidad de negocio.

18. Cumplimiento.

Tabla 4. Dominios de seguridad y objetivos de control ISO/IEC 27001. Fuente: elaboración propia.

Nota: en la Biblioteca virtual de UNIR al cierre de este tema se dispone de acceso a la biblioteca AENOR, donde está disponible la UNE-EN ISO/IEC 27001:2017 (con origen en la ISO/IEC 27001:2013).

## 8.12. Referencias bibliográficas

AEPD. (s.f.). Guía para el cumplimiento del deber de informar.<https://www.aepd.es/es/documento/guia-modelo-clausula-informativa.pdf-0>

Avram, R., Daoud, S., y Atme, J. (2020). Oracle public cloud [Diapositivas]. Oracle.[https://clubutilisateursoracle.org/wp-content/uploads/2020/09/AUFO\\_Cycle\\_Securite\\_Session\\_2\\_Oracle\\_220920.pdf](https://clubutilisateursoracle.org/wp-content/uploads/2020/09/AUFO_Cycle_Securite_Session_2_Oracle_220920.pdf)

Brandon, H., y de Souza, L. (2016). *Big Data Analytics and privacy & data protection*. Atos Codex.<https://atos.net/wp-content/uploads/2017/01/atos-big-data-analytics-privacy-data-protection-whitepaper.pdf>

Carta de los Derechos Fundamentales de la Unión Europea. Diario Oficial de las Comunidades Europeas C 364, 18 de diciembre de 2000, pp. 1-22.[https://www.europarl.europa.eu/charter/pdf/text\\_es.pdf](https://www.europarl.europa.eu/charter/pdf/text_es.pdf)

Hoskins, M. (18 octubre de 2012). Hooray - more data protection compliance diagrams [Mensaje en un blog]. Blogger.<http://dataprotector.blogspot.com/2012/10/hooray-more-data-protection-compliance.html>

Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal. Boletín Oficial del Estado, 24 de noviembre de 1995, núm. 281, pp. 33987-34058.<https://www.boe.es/eli/es/lo/1995/11/23/10/con>

Martins, M. (2018). Information Security Strategic Management [Diapositivas]. Speaker Deck.<https://speakerdeck.com/mmartins000/information-security-strategic-management>

Real Academia Española. (2020). Leal. En *Diccionario de la lengua española* (23a ed.). <https://dle.rae.es/leal>

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento General de Protección de Datos). Diario Oficial de la Unión Europea L 119, 4 de mayo de 2016, pp. 1-88. <https://www.boe.es/doue/2016/119/L00001-00088.pdf>

## Hacia un nuevo modelo europeo de protección de datos

Piñar, J. L. (dir.) (2007). *Reglamento general de protección de datos. Hacia un nuevo modelo europeo de protección de datos*. Reus.

Profundo análisis sobre cada uno de los aspectos relevantes del nuevo marco europeo de protección de datos. Una extensa guía de referencia.

## Agencia Española de Protección de Datos

AEPD. Página web oficial. <https://www.aepd.es/es>

Encontrarás preguntas frecuentes, documentación de soporte e información sobre procedimientos sancionadores.

## Comisión Europea

Comisión Europea. (s.f.). Protección de datos [Página web]. [https://ec.europa.eu/info/law/law-topic/data-protection\\_es](https://ec.europa.eu/info/law/law-topic/data-protection_es)

En esta página encontrarás información sobre la Comisión Europea para el nuevo Reglamento de Protección de Datos.

- 1.** De entre los siguientes, ¿cuáles tienen la consideración de datos de carácter personal por parte del RGPD?

  - A. Un nombre.
  - B. Un número de identificación.
  - C. Información genética.
  - D. Todos los anteriores.
  
- 2.** De entre los siguientes datos, ¿cuáles no tienen la consideración de datos de categorías especiales?

  - A. Las opiniones políticas.
  - B. Los datos de ingresos mensuales.
  - C. Los tratamientos de datos genéticos.
  - D. Los datos relativos a la salud de las personas.
  
- 3.** Cuando nuestros datos de carácter personal se quieren incorporar a un tratamiento por una organización, está obligada a:

  - A. Informar sobre quién o quiénes son los responsables del tratamiento.
  - B. Informar sobre la finalidad por la que recaban los datos y se van a realizar los tratamientos.
  - C. Informar si procede de los destinatarios o grupos de destinatarios de las comunicaciones de datos que se vayan a efectuar.
  - D. Todas las respuestas anteriores son ciertas.
  
- 4.** De las operaciones siguientes, indica cuáles se consideran tratamiento de datos:

  - A. Recogida de datos de clientes en un formulario *online*.
  - B. Conservación de datos en copias de seguridad.
  - C. Consulta de datos de forma telemática.
  - D. Todas las respuestas anteriores son ciertas.

**5.** De entre los siguientes, indica cuáles son principios de la protección de datos conforme al RGPD.

- A. Minimización de datos.
- B. Exactitud de los datos.
- C. Responsabilidad proactiva.
- D. Todos los anteriores.

**6.** De entre los siguientes, indica cuáles son obligaciones del responsable de tratamiento conforme al RGPD.

- A. Atender a las autoridades de control.
- B. Notificar las fallas de seguridad a las autoridades de control y afectados, cuando sea preciso.
- C. Seleccionar diligentemente y formalizar la relación con los encargados de tratamiento.
- D. Todas las anteriores.

**7.** Respecto a las transferencias internacionales a EE. UU., indica qué afirmación es cierta:

- A. Si existe el consentimiento del afectado, se puede realizar dicha transferencia.
- B. Si EE. UU. dispone de una decisión de adecuación de la Comisión que reconoce un nivel de protección en materia de protección de datos equivalente al de la UE.
- C. Las organizaciones de EE. UU. certificadas de acuerdo con el Escudo de Privacidad (Privacy Shield) tienen reconocido un nivel de protección equivalente al de la UE.
- D. En ningún caso se pueden transferir datos a organizaciones residentes en EE. UU.

**8.** Indica qué afirmación es verdadera:

- A. Las legislaciones en protección de datos de EE. UU. y la UE son legislaciones tipo ómnibus y equivalentes.
- B. La exportación de datos desde la UE a EE. UU. es libre.
- C. A nivel internacional, la mayoría de los países ofrecen un nivel fuerte de protección en materia de protección de datos equivalente al de la UE.
- D. Existe disparidad entre los marcos legales de la UE y de EE. UU.

**9.** Las trasferencias internacionales de datos son una práctica habitual, pero ¿son legales?

- A. Sí, siempre que el usuario adecuadamente informado consienta a la transferencia.
- B. Sí, cuando es necesaria para ejecutar un contrato entre el afectado y el responsable del tratamiento.
- C. Sí, si existe una decisión de la Comisión que reconoce, sobre el país destinatario, un nivel de protección equivalente al de la UE.
- D. En todos los casos anteriores son legales.

**10.** Respecto a los derechos que el RGPD reconoce a los afectados, ¿cuál de entre los siguientes no es cierto?

- A. Acceso.
- B. Rectificación.
- C. Oposición.
- D. Legitimación.

Gobierno del Dato y Toma de Decisiones

---

# Tema 9. Big data y protección de datos personales

# Índice

## Esquema

### Ideas clave

- 9.1. Introducción y objetivos
- 9.2. ¿Amenaza el big data a la privacidad?
- 9.3. Cómo cumplir con la protección de datos en el big data
- 9.4. Privacidad por diseño
- 9.5. Evaluaciones de impacto (PIA/EIPD)
- 9.6. Referencias bibliográficas

### A fondo

Opinion 03/2013 on purpose limitation

Privacidad por defecto: los siete principios fundamentales

Guía de evaluación de impacto de la Agencia Española de Protección de Datos

Guía de evaluación de impacto de la autoridad en protección de datos de Reino Unido

## Test



## 9.1. Introducción y objetivos

En este tema vamos a analizar los retos que el *big data* plantea a la privacidad y protección de datos y conocer cuáles son las tendencias que se están articulando para dar respuesta a dichos retos.

Este tema también aborda las implicaciones que desde el punto de vista de la privacidad y la protección de datos tienen los tratamientos de datos masivos dentro del entorno *big data analytics*, para ello establecemos los objetivos siguientes:

- ▶ Identificar brevemente los retos para las empresas.
- ▶ Describir las posibles implicaciones que el *big data* puede tener sobre la privacidad de las personas y qué aspectos de la normativa hay que tener en consideración en un proyecto *big data*.
- ▶ Analizar el concepto de seguridad por diseño, iniciativa que pretende dar respuesta a los retos planteados y facilitar un equilibrio entre los requerimientos del negocio y los derechos de los ciudadanos.
- ▶ Introducirnos en las evaluaciones de impacto, herramienta que ayuda a evaluar los riesgos que para la privacidad tienen los productos y servicios, y a identificar controles mitigadores.

## 9.2. ¿Amenaza el big data a la privacidad?

*Big data:* crecimiento exponencial de la disponibilidad y el uso automatizado de la información, se refiere a conjuntos de datos digitales gigantes en poder de las corporaciones, los gobiernos y otras grandes organizaciones, que luego son analizados exhaustivamente utilizando algoritmos informáticos.

El *big data* se basa tanto en la creciente capacidad de la tecnología para apoyar la recogida y el almacenamiento de grandes cantidades de datos como en la capacidad de analizar, conocer y sacar el máximo partido de todo el valor de los datos (en particular, el uso de aplicaciones de análisis).

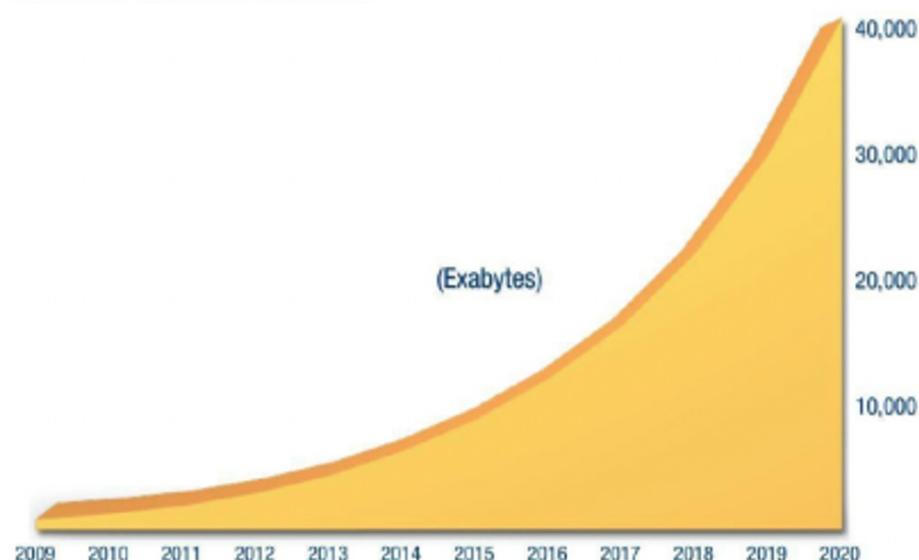
Es evidente el gran potencial que presenta el *big data* para la innovación, algunos lo han considerado como el *oil* del siglo XXI. Gracias al avance tecnológico existente, unido a la facilidad de recolectar información, será fuente de importante desarrollo del conocimiento al poder aplicar la correlación sobre grandes volúmenes de datos, de manera que el valor de la información que se pueda llegar a obtener sea inimaginable.

Dentro de este debate, no son pocos los que reconocen en el análisis de datos y el conocimiento basado en la correlación un nuevo paradigma en el pensamiento humano que desplazaría a un segundo plano la búsqueda de relaciones de causa y efecto. Sea o no una revolución, lo cierto es que está aquí y existe un amplio consenso sobre las posibilidades que el *big data* ofrece y ofrecerá en el futuro, que hoy ya se pueden ver en su amplia aplicación en la investigación, entre otras, la investigación de la física de partículas o la investigación genética.

La expectativa de que el *big data* pueda, en última instancia, conducir a decisiones mejores y más informadas ya ha facilitado su incorporación en diferentes sectores muy diversos, entre ellos, la salud, las comunicaciones móviles, las redes inteligentes, la gestión del tráfico, la detección de fraudes o la comercialización y venta al por menor, tanto *online* como *offline*.

Hoy en día existe un gran desconocimiento del volumen real de datos que pueden estar almacenados por las diferentes organizaciones, gobiernos, empresas... y del uso y finalidades que se le pueda dar a esta información; pero existe el consenso de considerar que su incremento será exponencial.

**The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020**



*This IDC graph predicts exponential growth of data from around 3 zettabytes in 2013 to approximately 40 zettabytes by 2020. An exabyte equals 1,000,000,000,000,000 bytes and 1,000 exabytes equals one zettabyte. Source: IDC's Digital Universe Study, December 2012, <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>.*

Figura 1. Crecimiento del universo digital 2010-2020. Fuente: Zhenhua's Wiki, s.f.

Se dispone de tecnologías con capacidad de analizar ingentes cantidades de datos no estructurados de diferentes fuentes a altas velocidades de procesamiento. A lo que se habrá de añadir el Internet de las cosas, facilitado por el nuevo estándar de Internet Protocol versión 6 (IPv6) y una potencial conexión IPv6 que, tal y como señala Wikipedia, admite 340 282 366 920 938 463 463 374 607 431 768 211 456 direcciones (2<sup>128</sup> o 340 sextillones) —cerca de  $6,7 \times 10^{17}$  (670 mil billones)— por cada milímetro cuadrado de la superficie de la Tierra.

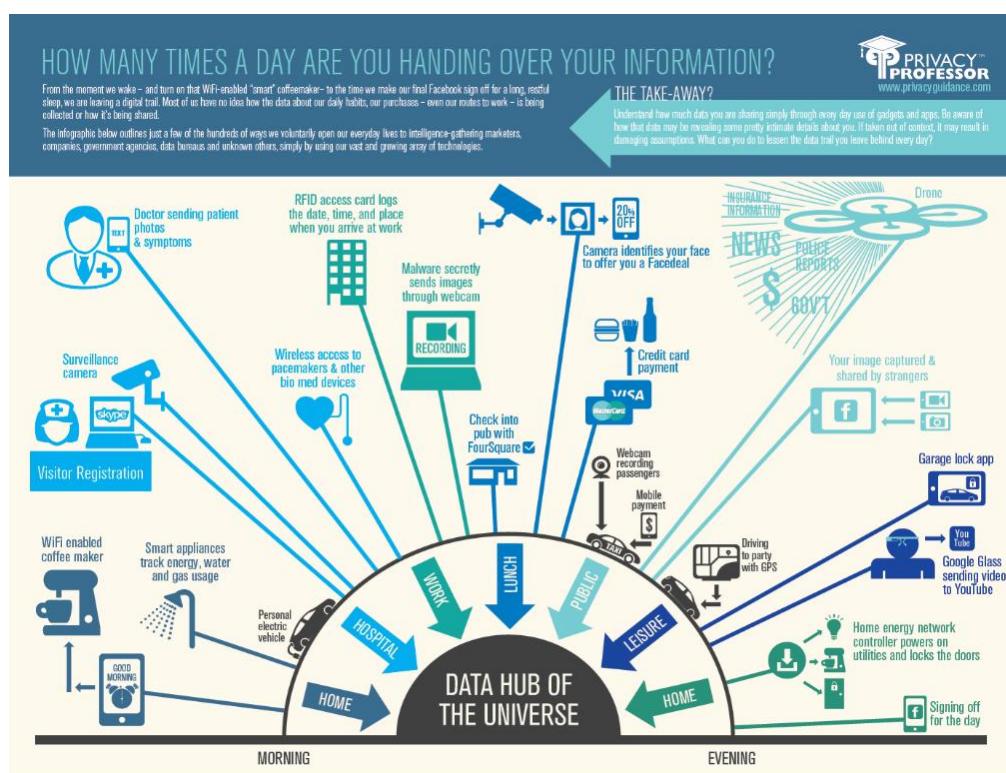


Figura 2. Centro de datos del universo. Fuente: Protiviti Chile, 2018.

## Del *big data* al *big brother*

Actualmente, hay un amplio debate sobre la potencial amenaza que, desde el punto de vista de la protección de datos y el derecho a la intimidad, presenta esta acumulación ingente de datos y sus potenciales usos, vislumbrándose grandes y significativos riesgos.

No se cuestiona una tecnología que ya está ofreciendo importantes avances en áreas de amplio consenso, en relación con el beneficio que genera a la sociedad, como es el caso de la ciencia; pero sí se cuestiona el uso que de la tecnología se pueda derivar y la afectación que esto pueda tener para los ciudadanos.

Es en este marco donde la presencia de gobiernos y empresas con ingentes cantidades de datos sobre los ciudadanos se perciben como una amenaza no solo sobre la privacidad, sino también sobre su libertad.

Potencialmente podrían facilitar un conocimiento amplio e íntimo de los individuos, a los que, además, se les podría someter a decisiones automatizadas con base en dicho conocimiento. La amenaza es un potencial *big brother* que sepa todo sobre nosotros, facilitado por un mundo hiperconectado en el que el acceso a multitud de fuentes de datos y su tratamiento permita descubrir el perfil más íntimo de las personas —el yo más íntimo— en el que quedarían integrados: pensamientos, deseos, ilusiones, conductas, etc.

La amenaza se fundamentaría, en primer lugar, en la **trazabilidad de nuestros actos y conductas en Internet**, donde existe una continua captación de datos sobre nuestras interacciones, que además podría reflejar nuestras conductas más íntimas, gracias a una falsa percepción de anonimato por parte del individuo. En este contexto sería posible **ligar las acciones del yo virtual con el yo real**, algo que con la incorporación de altas capacidades de tratamiento y fuentes adicionales de información se presupone muy posible.

Esta amenaza se materializaría en las decisiones que, con base en este conocimiento de las personas, se podrían fundamentar, bien a través de un conocimiento particular de cada individuo o bien a través de la aplicación de patrones o perfiles. Una decisión que podría tener carácter predictivo, basada en un histórico de datos y su acomodo en perfiles establecidos. Incluso se podrían aplicar decisiones sobre hechos no realizados pero que por inferencia sí se producirían, lo que podríamos llamar una «seguridad preventiva», tal y como se trataba, por ejemplo, en la película *Minority Report*.

Este debate mediático se puede seguir fácilmente por los recurrentes artículos o entradas en blogs presentes en Internet.

**Al margen de este debate apocalíptico, la discusión habría que centrarla en cómo crear un equilibrio que garantice la privacidad y protección de datos de los ciudadanos al mismo tiempo que da respuesta a las necesidades de las empresas.**

Hoy por hoy, en las empresas el *big data* está siendo utilizado para identificar tendencias generales y correlaciones. Por ejemplo, en el campo del *marketing* y la publicidad el *big data* puede ser utilizado para analizar o predecir preferencias personales, el comportamiento de las personas y las actitudes de los clientes individuales y así traducir estos en medidas o decisiones que se toman sobre los clientes, como pueden ser descuentos personalizados, ofertas especiales y anuncios dirigidos basados en perfiles.

El reto está en que estos usos se desarrollen sin comprometer los derechos de los ciudadanos, el derecho a la privacidad y a la protección de datos personales; es decir, sin atentar contra los derechos y libertades de los individuos.

## Retos para las empresas y organizaciones

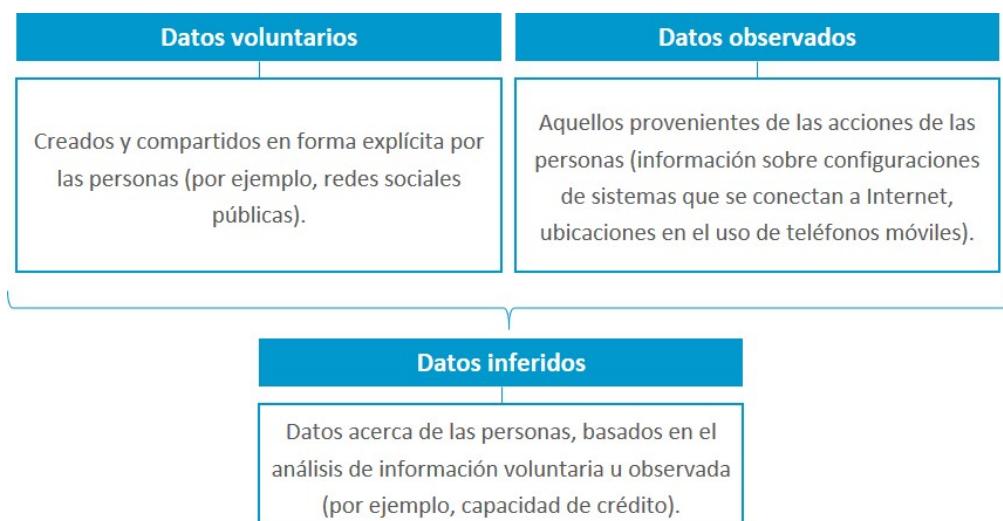
Las tecnologías de análisis de grandes datos son, sin duda, una gran oportunidad para las empresas, al facilitar la búsqueda de la obtención de patrones y correlaciones que pueden no ser evidentes y sí útiles para el negocio.

Un ámbito que adquiere especial protagonismo, como ya hemos comentado, es el área de *marketing*, en la búsqueda no solo de conocer mejor a los clientes en lo referente a sus preferencias y comportamientos de compra, sino también para facilitar la toma de decisiones sobre los mismos. Esto puede llegar a suponer una **ventaja competitiva** con relación a:



Figura 3. Ámbitos beneficiados. Fuente: elaboración propia.

En este contexto los datos personales que manejan las empresas son:



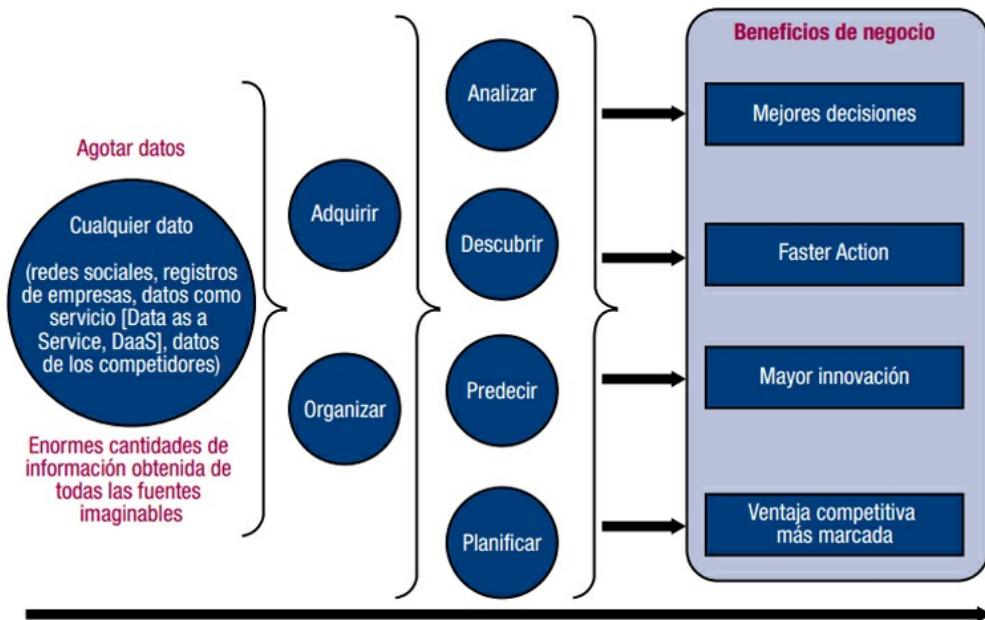


Figura 4. *Big data*, impactos y beneficios. Fuente: ISACA, 2013.

Estas fuentes dispares fácilmente nos van a llevar a pensar que el tratamiento de estos datos no es indiferente para las regulaciones en protección de datos y que cumplir con las obligaciones de la normativa debe ser un objetivo clave en los procesos de *big data*.

El incumplimiento de la normativa no solo puede tener repercusiones económicas, **un riesgo económico** importante mediante sanciones, sino también un **riesgo reputacional** que convierta una ventaja competitiva en una amenaza al ser una «ventaja» ilícita. Las empresas deberán, por tanto, gestionar los riesgos derivados del cumplimiento normativo en protección de datos.

Pero, es más, vivimos en un mundo que está ampliamente globalizado, Internet facilita que ciudadanos de un extremo del mundo interactúen con entidades ubicadas en el polo opuesto del planeta y, seguramente, sometidos a regulaciones en materia de privacidad y protección de datos muy diferentes. Por tanto, las empresas deberán gestionar esta disparidad legislativa y la potencial inseguridad jurídica que pueda introducir en sus operaciones.

Esta disparidad de regulación en un ámbito capaz de generar ventajas competitivas sin fronteras somete a obligaciones bien diferentes para sus actores, en función de la legislación aplicable. Esta desigualdad, en cuanto las reglas de juego, puede incidir negativamente sobre aquellas entidades que estén sometidas a marcos regulatorios más estrictos. Esta es una de las críticas importantes que se vierten sobre la legislación de la Unión Europea en materia de protección de datos.

La normativa de la UE es especialmente garantista en cuanto a los derechos de los ciudadanos se refiere y reconoce el derecho de autodeterminación informativa como un derecho fundamental, un derecho instrumental, básico para la libertad del individuo.

Este hecho, frente a legislaciones más laxas como la de EE. UU., puede ser considerado como una barrera para entidades europeas en la lucha por un mercado global y todo un **riesgo para su competitividad** en Internet (comercio electrónico, redes sociales, etc.). Además, esta disparidad internacional también puede ser un **riesgo jurídico** para las grandes corporaciones de carácter multinacional.

Por último, asistimos a una creciente necesidad por parte de las empresas de datos y de su análisis para la toma de decisiones estratégicas y la evaluación de sus resultados, lo que puede tensar los procesos internos de autorregulación en la materia y la asunción de comportamientos no éticos o no legales en pro de beneficios económicos. Dos ejemplos de ello:

- ▶ Sucumbir a la tentación inmediata de proceder a la captación de todo tipo de dato sin finalidad concreta alguna y con la expectativa de un uso futuro, bajo la expectativa de que se pueda extraer información y conocimiento y ayudar en la toma de decisiones.
- ▶ La utilización de datos de carácter personal para finalidades diferentes de aquella para la que los datos fueron recabados, no amparadas por la ley.

A ello también habría que añadir un **riesgo social**, de manera que la generalización de determinadas prácticas poco transparentes pudiera desencadenar un rechazo social a determinados usos, que *a priori* no debería tener una connotación negativa, y a un fomento de legislación más restrictiva. Para ello, solo debemos reflexionar sobre las noticias que en los últimos tiempos han aparecido en los medios sobre Google o sobre el espionaje indiscriminado de gobiernos a sus ciudadanos.

El caso de Google es especialmente interesante, ya que es uno de los líderes en servicios de *marketing* basados en técnicas de *analytics*. Es posible que sus prácticas, debido a su posición dominante en el mercado, no tengan un impacto directo sobre el uso que los usuarios hacen de sus servicios, pero seguramente se intensifica la preocupación de las autoridades de protección de datos, que buscarán reforzar los mecanismos para garantizar el cumplimiento de la normativa. Esta circunstancia, entre otras, ha contribuido al nuevo marco regulador de la protección de datos personales en la UE. Veamos dos ejemplos:

- ▶ Google se enfrenta a una demanda por haber espiado el contenido de millones de mensajes enviados y recibidos por estudiantes estadounidenses que utilizan aplicaciones de la compañía para la *suite* de educación. Además, está acusada de haber utilizado esa información para construir perfiles encubiertos utilizados para enviar publicidad dirigida a los estudiantes.

Al hilo de la cuestión en Estados Unidos, se plantean nuevas preguntas acerca de la compatibilidad entre las leyes de protecciones de los niños en Estados Unidos y el uso de *big data*.

- ▶ El reconocimiento de estas prácticas se deja entrever en la modificación de los términos de servicios (ToS).

---

Puedes consultar los términos de servicio de Google en:

<https://policies.google.com/terms?hl=es>

---

Los términos del servicio de Google indican, por ejemplo, que:

«Google usará los derechos que le confiere esta licencia únicamente con el fin de:

- ▶ Gestionar y mejorar los servicios [...]. Esto incluye el uso de sistemas automatizados y de algoritmos para analizar tu contenido y poder hacer lo siguiente:
- ▶ Comprobar si hay spam, software malicioso o contenido ilegal.
- ▶ Ofrecerte servicios personalizados, como recomendaciones, resultados de búsqueda, contenido y anuncios personalizados» (Google, s.f.).

A lo que se suman las actuaciones de las agencias de protección de datos de los Estados miembros de la Unión Europea, que se saldaron con sanciones

En el caso de España, en diciembre de 2013, Google fue objeto de apertura de expediente sancionador por parte de la Agencia Española de Protección de Datos por vulnerar gravemente los derechos de los ciudadanos. En el procedimiento se declaran ilegales los tratamientos de datos personales realizados por Google con su nueva política de privacidad, donde los argumentos motivadores de la sanción serían los siguientes:

- ▶ Se considera constatado que Google no da a los usuarios información suficiente sobre qué datos recoge y para qué fines los utiliza, combina los obtenidos a través de distintos servicios, los conserva durante un tiempo indefinido y obstaculiza el ejercicio de los derechos ARCO (acceso, rectificación, cancelación-supresión y oposición).
- ▶ Esta combinación de los datos que recoge por medio de diferentes servicios excede ampliamente de las expectativas razonables de la mayoría de los usuarios, que no son conscientes de ello y pierden el control de su propia información personal.

- ▶ La resolución considera que se recoge y trata ilegítimamente información con datos de carácter personal de todos los usuarios que acceden a sus servicios (usuarios dados de alta y visitantes). Estaríamos hablando de más de un centenar de servicios dirigidos a usuarios residentes en España. Además, no se informa con claridad a los usuarios de Gmail de que se realiza un filtrado del contenido del correo y de los ficheros anexos para insertar publicidad.

Por tres infracciones a la LOPD se impuso a Google una sanción de 300 000 euros por cada una, requiriéndole que cumpla con la ley sin dilación.

## **¿Cuáles son los riesgos y los desafíos planteados por el *big data* para el derecho a la protección de los datos personales y la privacidad?**

Desde el punto de vista de la protección de datos y el derecho a la intimidad, presenta grandes y significativos riesgos.

El grupo de trabajo WP29 (en adelante, WP29) es un órgano consultivo independiente de la UE sobre protección de los datos y la vida privada, creado en virtud del artículo 29 de la Directiva 95/46/CE. En él están presentes las autoridades de protección de datos de los países miembros de la Unión Europea, el Supervisor Europeo de Protección de Datos y la Comisión Europea, que con carácter consultivo es independiente. Este órgano y sus competencias ampliadas con la entrada en vigor del RGPD pasa a denominarse Comité. Tiene entre sus funciones:

- ▶ Aconsejar a los Estados en relación con la protección de datos.
- ▶ Promover la aplicación de la directiva de protección de datos en todos los Estados miembros de la UE.
- ▶ Facilitar a la Comisión dictamen sobre las leyes comunitarias que afectan al derecho a la protección de datos personales (órgano consultivo).

En el documento «Opinion 03/2013 on purpose limitation» elaborado por el WP29, identifica cinco importantes amenazas en materia de protección de datos derivadas del *big data*:



Figura 5. Amenazas en materia de protección de datos. Fuente: Art. 29 WP, 2013.

- ▶ **La magnitud de la recopilación de datos**, la capacidad de seguimiento y elaboración de perfiles de los individuos, teniendo en cuenta la variedad y el detalle de los datos recopilados y la combinación de los datos con diferentes fuentes pueden ser una importante amenaza a la privacidad de las personas y el derecho a la determinación informativa.
- ▶ **La seguridad de los datos**, que no siempre responde a los riesgos de la información que es tratada, configurándose niveles de protección muy a la zaga de la expansión en el volumen; facilitando potenciales situaciones de gran impacto sobre los ciudadanos como, por ejemplo, en el caso accesos no autorizados por atacantes cibernéticos.
- ▶ **La transparencia**, la posible desprotección del ciudadano frente a usos sobre los que no se le proporciona suficiente información para poder ejercitar de manera adecuada sus derechos, el sometimiento de los ciudadanos a decisiones que no entienden y sobre las que no tienen ningún control.
- ▶ **La inexactitud, la discriminación, la exclusión y el desequilibrio económico** entre las grandes corporaciones y el ciudadano, que limitan las capacidades de este en la defensa de sus intereses.
- ▶ **El dramático aumento de las posibilidades de vigilancia del gobierno**. Que facilita una capacidad de supervisión y acceso a información sobre el comportamiento de los individuos, con la capacidad de generar inferencias sobre aspectos políticos, filosóficos y religiosos, nos disponibles hasta ahora.

En el texto se considera, además, que el tipo de técnicas de análisis empleado puede llevar a resultados que son inexacos, discriminatorios o ilegítimos. Por ejemplo: un algoritmo podría detectar una correlación errónea y, a continuación, dibujar una inferencia estadística, de manera que, cuando se aplique para informar en acciones comerciales o en la toma de decisiones, estas podrían ser injustas y discriminatorias. Esto puede perpetuar los prejuicios y los estereotipos existentes y agravar los problemas de la exclusión social y la estratificación de clases.

Además, y en términos más generales, la disponibilidad de grandes conjuntos de datos y las herramientas de análisis sofisticadas utilizadas para examinarlos no están al alcance de todos y puede potenciar el desequilibrio económico entre las grandes empresas, por un lado, y los consumidores, por el otro. Este desequilibrio económico puede dar lugar, por ejemplo, a una discriminación de precios injusta con respecto a los productos y servicios que se ofrecen, así como a anuncios y ofertas dirigidas altamente intrusivas, perturbadoras y personalizadas.

También podría dar lugar a otros efectos significativos adversos en las personas, por ejemplo, con respecto a las oportunidades de empleo, préstamos bancarios u opciones de seguro de salud.

## 9.3. Cómo cumplir con la protección de datos en el big data

Todo tratamiento de datos personales ha de cumplir con los principios de la protección de datos establecidos en la legislación en protección de datos personales vigentes y *big data analytics*. Si trata datos de carácter personal, no es una excepción.

### **Concepto de dato personal identificado o identifiable**

Revisemos rápidamente el concepto de dato de carácter personal. En el Reglamento General de Protección de Datos, se definen los datos personales como «toda información sobre una persona física identificada o identifiable (“el interesado”)» (art. 4.1, RGPD).

Y se considerará persona física identifiable: «toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona» (art. 4.1, RGPD).

No es una definición que esté lejos de la que nos facilitaba la Directiva 95/46/CE, pero sí es actualizada, debido a la relevancia de los avances en los últimos años, como es la consideración de la identidad genética en el concepto de **identifiable**.

Dato personal: cualquier información concerniente a las personas físicas identificadas y/o identificables, tanto relativo a su identidad (nombre y apellidos, domicilio, filiación, NIF...), así como relativo a su existencia y ocupaciones (trabajo, estudios, enfermedades, actividades de ocio, etc.)

Es fácil encontrar un consenso a la hora de determinar qué información hace que las personas físicas sean identificadas en el marco de las legislaciones de la UE. Ahora bien, la cuestión no resulta tan trivial si nos preguntamos: ¿qué información concreta hace a una persona identifiable? Pregunta cuya respuesta no siempre se resuelve de manera uniforme.

Ejemplos de este carácter interpretativo amplio lo tenemos en la consideración de la dirección IP como dato personal. La Agencia Española de Protección de Datos lo considera como dato de carácter personal al hacer identificables a las personas, mientras que en Francia ya son dos las sentencias que restarían ese carácter de identifiable a la IP. Lejos de pretender abundar en los argumentos de un caso concreto, que no pretende más que mostrarse como ejemplo, sí debemos resaltar aquello en lo que existe cierta unanimidad: **la particularidad de cada caso.**

No obstante, el devenir interpretativo viene marcado claramente por la jurisprudencia europea, y tras más de veinte años de directiva europea, han existido adaptaciones o correcciones interpretativas al hilo de las sentencias de los tribunales y, en especial, del Tribunal de Justicia Europeo.

Hoy estaríamos considerando la IP como dato de carácter personal, incluso las IP dinámicas.

El considerando 26 del RGPD establece que, **para determinar si una persona es identifiable**, deberán tenerse en cuenta todos los medios que puedan ser utilizados por el responsable de tratamiento o cualquier otra persona para identificar directa o indirectamente a un individuo y estos medios deben estar objetivamente al alcance. Para ello se deberían tener en consideración los costes asociados, la cantidad de tiempo necesaria, la tecnología disponible y el desarrollo tecnológico. Es decir, no tendrá dicha consideración si son requeridos medios desproporcionados para proceder a la identificación de los individuos.

Además, el considerando clarifica dos aspectos (considerando 26, RGPD):

- ▶ La pseudonimización, como técnica primaria para reducir el riesgo de identificación, y la consideración de los datos pseudonimizados como datos de carácter personal.
- ▶ La anonimización de los datos como mecanismo de obtención de conjuntos de datos a los que no les son aplicables la normativa en protección de datos personales RGPD.

Ahora bien, del propio considerando se puede concluir que los datos que **no** tengan la consideración de datos identificables —y, por tanto, no están sujetos a la normativa en protección de datos— más adelante —con los avances tecnológicos y la facilidad de acceso a los mismos, la reducción de costes e incluso la facilidad de uso de otras fuentes alternativas de información— pueden elevar su consideración a dato identificable y, por tanto, estar sometidos a la normativa de protección de datos.

En cualquier caso, siempre que tratemos **datos de carácter personal** en el ámbito de *big data* deberemos aplicar desde el inicio la normativa en protección de datos personales. Esto implicará cumplir con los principios de la protección de datos y evaluar y aplicar técnicas que permitan sustraer los datos de las obligaciones legales, como la anonimización.

Por tanto, de inicio, el tratamiento de datos debe alinearse con los principios de la protección de datos personales, tal y como son entendidos en el RGPD (art. 5) los principios relativos al tratamiento. Recordémoslos brevemente:

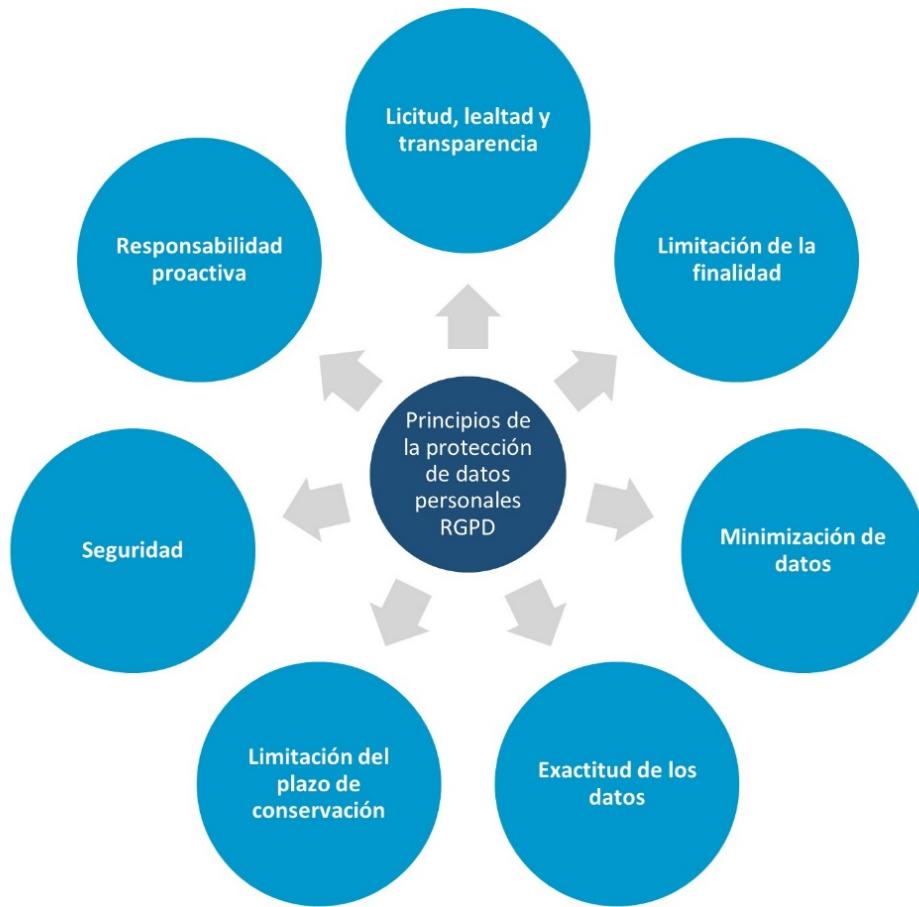


Figura 6. Principios de la protección de datos personales RGPD. Fuente: elaboración propia.

- ▶ **Licitud, lealtad y transparencia:** con el consentimiento informado como base de la legitimación del tratamiento, a falta de la aplicabilidad de otras vías legales de legitimación (art 6., RGPD).
- ▶ **Limitación de la finalidad:** para la que se van a tratar los datos establecida de forma concreta.
- ▶ **Minimización de datos:** la reducción de los datos tratados a los estrictamente necesarios para el fin perseguido por el tratamiento.
- ▶ **Exactitud de los datos:** como obligación requerida para el mantenimiento de los datos y, en su defecto, proceder a su supresión.

- ▶ **Limitación del plazo de conservación:** durante el tiempo estrictamente necesario para cumplir con la finalidad del tratamiento.
- ▶ **Confidencialidad e integridad (seguridad):** garantizando la seguridad de los datos personales durante todo el ciclo de vida de los datos.
- ▶ **Responsabilidad proactiva:** el conjunto de obligaciones establecidas al responsable de tratamiento o encargado de tratamiento, que obliga a que los tratamientos sean conformes al resto de principios y a garantizar la transparencia y al ejercicio de derechos de los ciudadanos.

## Retos de aplicación de la protección de datos en *big data*

Consentimiento del afectado: el principio de responsabilidad proactiva exige al responsable de tratamiento la prueba de la legalidad de este, y uno de los mecanismos establecidos para la legitimación (art 6., RGPD) es el consentimiento por parte del afectado, siendo el responsable el que debe probar que efectivamente dispone del consentimiento del afectado y que este otorgó respetando los términos establecidos en el reglamento (art. 7, RGPD).

El consentimiento debe ser concreto, para una finalidad específicamente informada, conforme al principio de finalidad, no pudiéndose tratar los datos para otros fines.

Los usos potenciales para los que se van a utilizar los datos a menudo no pueden ser totalmente identificados en la fase de captación de datos de usuarios, por ejemplo. Esto podría implicar que los usuarios no hayan sido informados de dichas finalidades y abriría la necesidad de obtener el correspondiente consentimiento para el tratamiento de datos para otras finalidades o que estas tuvieran la consideración de compatibles (art 89., RGPD) con la finalidad informada al afectado en el proceso de captación de la información (habría que tener la capacidad de prueba) o esté amparado por ley.

Además, en los procesos de *analytics* pueden aparecer nuevos datos que tengan la consideración de dato personal, mediante el enriquecimiento con fuentes externas, que faciliten los procesos de reidentificación de las personas, luego su tratamiento implicaría la legitimación del mismo.

Habrá que prestar especial atención a los tratamientos de datos personales de categorías especiales y a los datos de los menores.

### **Interés legítimo como base legal del tratamiento en lugar del consentimiento**

Es de resaltar que **confiar en la legitimación con base en intereses legítimos no es una opción flexible**, que no permite cualquier cosa. Las organizaciones de *big data* siempre deben equilibrar sus propios intereses con los de las personas involucradas.

Hay que resaltar que no es posible recurrir al interés legítimo cuando dichos intereses sean anulados por los intereses o derechos y libertades fundamentales del interesado que requieren protección de datos personales, en particular, cuando el sujeto de datos es un niño. Esto lleva a la necesidad de:

- ▶ **Ponderación de intereses:** evaluar, previamente a su aplicación, los intereses de la organización frente a los intereses de los individuos.
- ▶ **Asunción de responsabilidad:** poder justificar que efectivamente existe un verdadero interés legítimo que, adecuadamente ponderado, no menoscaba los intereses de los afectados.
- ▶ **Informar:** la base del tratamiento en el interés legítimo no exime de informar al afectado.

Algunos ejemplos de intereses legítimos:

- ▶ El ejercicio del derecho de libertad de expresión o información, incluidas las situaciones en las que se ejerza dicho derecho en los medios de comunicación y en las artes.
- ▶ La prospección convencional y otras formas de comercialización o publicidad.
- ▶ La ejecución de derechos reconocidos en procedimientos judiciales, incluido el cobro de deudas mediante procedimientos extrajudiciales.
- ▶ La prevención del fraude, el uso indebido de servicios o el blanqueo de dinero.
- ▶ La supervisión de los empleados con fines de seguridad o de gestión.
- ▶ Los regímenes internos de denuncia de irregularidades.
- ▶ La seguridad física, la tecnología de la información y la seguridad en la red.
- ▶ El tratamiento con fines históricos, científicos o estadísticos.
- ▶ El tratamiento con fines de investigación.

## Comunicación de datos

La analítica *big data* a menudo dibuja conjuntos de datos de múltiples fuentes para la obtención de relaciones o correlaciones que puedan generar conocimientos útiles. Este acceso a diferentes fuentes de datos para recopilar datos de carácter personal para ser objeto de tratamiento debe ser legitimado en el marco del RGPD.

Para estas dos hipotéticas situaciones no es difícil pensar en el esfuerzo que puede suponer recabar el consentimiento poscaptación para el caso de grandes masas de datos. Además, en muchas ocasiones no sería posible realizarla ante la dificultad de contactar con el afectado, pues no se cuenta con la información de contacto o esta no está actualizada.

El artículo 14 del RGPD («Información que deberá facilitarse cuando los datos personales no se hayan obtenido del interesado») reconoce las situaciones en las que la comunicación de dicha información no es posible o suponga un esfuerzo desproporcionado, y articula una salida para el caso de los tratamientos con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos, de manera que no se exija la comunicación siempre y cuando el hecho de tratar de cumplir con la obligación de información pueda obstaculizar gravemente los objetivos del tratamiento.

Para ello se debe cumplir con las garantías del artículo 89 del RGPD («Garantías y excepciones aplicables al tratamiento con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos»). Estas garantías exigen, por ejemplo, procesos específicos para garantizar el respeto al principio de minimización de los datos personales.

## Ejercicio de derechos

El acceso a fuentes dispares de datos en un mundo tan interconectado sometido a diferentes regulaciones puede imposibilitar el ejercicio eficaz de los derechos de acceso, rectificación, supresión y oposición.

Además, el fortalecimiento a los derechos que supone el RGPD (en virtud del cual se obliga a los responsables de tratamiento a comunicar el derecho ejercido a aquellas entidades a las que hubiera comunicados los datos, para que se actúe en consecuencia) añade una complejidad adicional: «obligación de notificación relativa a la rectificación o supresión de datos personales o la limitación del tratamiento» (art. 19, RGPD). Cabe recordar que esta obligación es la base del llamado **derecho al olvido**.

## **Acceso a datos por cuenta de terceros**

Los servicios en la nube y transfronterizos pueden imposibilitar la regularización de la relación responsable del tratamiento y encargado de tratamiento en los términos establecidos en el artículo 28 del RGPD («Encargado del tratamiento»), máxime en un ecosistema de servicios en la nube en los que la superposición de diferentes capas de servicio y prestadores haga prácticamente imposible conocer los agentes implicados o sus ubicaciones.

## **Transferencias internacionales**

La deslocalización de servicios y servicios en la nube pueden llevar asociados encargos de datos transfronterizos que, dadas las sustanciales y diferentes regulaciones en materia de protección de datos, no fueran legítimos conforme a la ley al ser realizados en espacios donde los marcos de protección no son equivalentes a los de la UE (capítulo V, RGPD: «Transferencias de datos personales a terceros países u organizaciones internacionales»).

Sobre este tema, cabe resaltar que las diferencias normativas de la UE y EE. UU. implican que las transferencias solo sean legítimas si están amparadas por alguna de las excepciones del RGPD o si están autorizadas por la Agencia Española de Protección de Datos Personales. Obtener el consentimiento del afectado también sería una opción, pero ya hemos comentado sus dificultades.

## Derecho de oposición y decisiones automatizadas de datos

El RGPD reconoce a los ciudadanos el derecho a no verse sometidos a una decisión con efectos jurídicos, sobre ellos o que les afecte de manera significativa, que se base exclusivamente en un tratamiento automatizado, incluida la elaboración de perfiles. Además, reconoce el derecho del afectado a obtener información sobre los criterios de valoración y el programa utilizado en el tratamiento para adoptar la decisión (arts. 13 y 14, RGPD). Esto tiene implicaciones sobre las finalidades de los tratamientos en *big data* y la obligación de transparencia frente a las impugnaciones, algo complejo por la propia naturaleza de los procesos empleados.

## Principios de la protección de datos, por ejemplo:

- ▶ **Principio de limitación del plazo de conservación (art. 5, RGPD):** implica que los datos solo serán retenidos por el tiempo razonablemente necesario conforme a la finalidad de los mismos, lo que parece estar en la antítesis del *big data*, donde la acumulación y el enriquecimiento mediante la adición de otros conjuntos de datos para uso prospectivo y una posible utilización futura desconocida pueden ser considerados una necesidad.
- ▶ **Principio de minimización de datos:** los datos recabados deben ser adecuados y pertinentes para la finalidad específica; ahora bien, la tendencia del *big data* al *overcollecting* puede atentar contra este principio que exige tratar los datos estrictamente necesarios.
- ▶ **Principio de limitación de la finalidad:** referido a la utilización de datos personales recabados para una finalidad con otros fines no siempre compatibles o no amparados por la ley. Esta tendencia a utilizar los datos que se tienen para otros fines sin considerar la finalidad concreta para la que se recabaron y sin analizar si la nueva finalidad es compatible o es admitida por la ley no es conforme a la normativa.

## Limitación de la finalidad

Pretender recurrir a expresiones de finalidades ambiguas o excesivamente genéricas sobre las que recabar el consentimiento pondría en cuestión el principio de calidad e información y atentaría contra el principio de autodeterminación informativa de los afectados, ya que debemos recordar que, para que un usuario pueda ejercer su derecho al control sobre sus datos, el consentimiento debe estar basado en la transparencia y claridad en la finalidad.

**Como norma debemos considerar que los datos de carácter personal** objeto de tratamiento **no podrán usarse para finalidades diferentes de aquellas para las que los datos se hubieran recogido**, a menos que:

- ▶ Exista alguna habilitación legal aplicable.
- ▶ Esta nueva finalidad sea compatible con la finalidad con la que fueron recabados los datos.

El principio de limitación de la finalidad no necesariamente crea una barrera para el análisis de *big data*, pero significa que se debe hacer una evaluación de la compatibilidad de los propósitos de procesamiento para poder usar los datos con otros fines.

**Los datos pueden usarse para otros fines si la finalidad es compatible, tras un análisis de compatibilidad de finalidades.**

El WP29 (Art. 29 WP, 2013) identifica los aspectos que se deberían considerar a la hora de evaluar la compatibilidad de finalidades, entre ellos establece:

- ▶ La relación entre los fines para los cuales fueron recabados los datos y los fines de su posterior procesamiento.
- ▶ El contexto en el que los datos han sido recogidos y las expectativas razonables de los titulares de los datos en cuanto a su uso posterior.
- ▶ la naturaleza de los datos y el impacto que el nuevo tratamiento puede tener sobre los titulares de los datos.
- ▶ las medidas aplicadas por el responsable de tratamiento para garantizar que el tratamiento es leal y para evitar cualquier impacto indebido sobre los titulares de los datos.

**La información que las personas han puesto en las redes sociales se va a utilizar para evaluar sus riesgos para la salud o su solvencia, o para comercializar ciertos productos para ellos; a menos que se les informe de esto y se les pida que den su consentimiento, es poco probable que sea justo o compatible.** Con el fin de identificar las medidas necesarias, el documento propone diferenciar dos escenarios de ejemplo:

- 1.** Las organizaciones de procesamiento de los datos quieren detectar tendencias y correlaciones en la información.
- 2.** Las organizaciones están interesadas en las personas: perfiles.

A continuación, veremos estos escenarios con más detalle.

## Tratamiento de datos personales para el análisis de tendencias y correlaciones en la información

Aquí el concepto de separación funcional desempeña un papel clave. La separación funcional supone que cada función o departamento que participa en el tratamiento de datos personales solo tiene acceso a aquellos atributos estrictamente necesarios para su función.

Por ejemplo, el área de *telemarketing* accede a los datos de contacto telefónico de los clientes, pero no al resto de información de estos, aunque su finalidad se circunscriba dentro de un habitual tratamiento de datos para fines comerciales (legítimo para las empresas con respecto a sus clientes).

La separación funcional no supone una disociación, técnicamente hablando, pues de la aplicación de esta no se obtiene necesariamente un conjunto de datos que no permite la identificación de las personas, sino más bien un conjunto de datos limitado en cuanto a los atributos que contiene.

Pues bien, en la medida en que esta separación funcional sea posible, podría ser un factor importante para decidir si el uso adicional de los datos para análisis de *marketing* y otros pueden ser considerados compatibles.



Figura 7. Deberes responsables del tratamiento. Fuente: elaboración propia.

De manera que no exista duda de que los datos utilizados para fines estadísticos u otros fines de investigación no estarán a disposición de tratamientos que tomen decisiones en relación con los titulares de los datos de que se trate (salvo autorización expresa de las personas interesadas). La organización deberá aplicar las medidas técnicas y organizativas adecuadas para ello.

Como es lógico en este punto, la disociación de datos puede ser una buena herramienta para facilitar el análisis y, al mismo tiempo, para garantizar que no se toman decisiones sobre individuos concretos con base en el análisis específico de sus datos o comportamientos.

**Para estos casos, la anonimización total o parcial puede ser de gran ayuda.**

### **Tratamiento de datos personales para análisis de comportamiento de las personas**

Este es un posible escenario en el que una organización quiere específicamente analizar o predecir las preferencias personales, el comportamiento y las actitudes de los clientes individuales y, con base en dicho análisis, posteriormente tomar medidas o decisiones.

En estos casos casi siempre se requiere el consentimiento libre, específico, informado e inequívoco del afectado. De lo contrario, su uso posterior no puede ser considerado compatible con la finalidad que motivó la recogida de datos. Veamos un ejemplo:

Un supermercado se plantea mejorar el conocimiento que tiene de sus clientes con objeto de mejorar su experiencia y personalizar la oferta de productos. En su momento, facilitó a sus clientes una tarjeta de descuento, de manera que aquellos que la poseían accedían a un descuento sobre sus compras que no dependía de hábitos ni volúmenes de compra, por lo que no se asociaban a los clientes las compras realizadas. Es decir, la empresa no conocía qué compraba cada uno de sus clientes.

Ahora han pensado en registrar todas las compras que realiza un cliente, para lo que se servirán de la tarjeta de fidelización a la que asociará cada compra. Esto permitirá saber qué compró y cuándo cada uno de los clientes que poseen la tarjeta. Analizar esta información facilitará un mejor conocimiento de gustos y hábitos de consumos, por lo que se le podrán dirigir ofertas específicas atendiendo a su histórico de compra.

Por ejemplo, si compra pañales, se le comunicarán ofertas personalizadas de productos infantiles o, atendiendo a la frecuencia de compra de pañales, se le ofrecerá una oferta exclusiva para que se acerque al supermercado.

Este sería uno de los casos en los que el nuevo tratamiento de datos no podría considerarse compatible con el tratamiento que se realizaba anteriormente sobre las tarjetas de fidelización. Antes de activar este sistema, deberán previamente recabar el consentimiento de los afectados.

En opinión del WP29, es importante destacar que, en caso de ser necesario tal consentimiento —por ejemplo, para el seguimiento y elaboración de perfiles para fines de venta directa, de anuncios basados en comportamientos, intermediación de datos, publicidad basada en la localización o investigación de mercados basada en seguimiento—, se debe garantizar la transparencia y el consentimiento informado de los interesados o consumidores. Para ello, los interesados deberían tener acceso a sus perfiles, así como a la lógica de la toma de decisiones (algoritmo) que llevó a la elaboración del perfil.

**En otras palabras, las organizaciones deberían revelar sus criterios de toma de decisiones. Esta es una medida fundamental y de gran calado en el mundo del *big data*.**

---

Para ahondar más en la cuestión se puede consultar la Recomendación CM/Rec(2010)13 del Comité de Ministros a los Estados miembros sobre la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal en el contexto de la creación de perfiles, disponible en:

[https://search.coe.int/cm/Pages/result\\_details.aspx?ObjectId=09000016805cdd2a#\\_ftn1](https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016805cdd2a#_ftn1)

---

También hay que considerar que la mayoría de las veces no será la información recabada en sí la que es sensible, sino más bien —y lo que es motivo de consideración— son las inferencias que se extraen de la información y la forma en que se dibujan esas inferencias.

El WP29 considera, además, que el potencial riesgo de inferencias erróneas se podría reducir con la transparencia. De hecho, si los titulares de los datos pueden acceder y corregir o actualizar sus perfiles y deciden hacerlo, facilitaría que la información sobre la que se base el tratamiento sea más precisa.

De hecho, en muchas situaciones, medidas como permitir el acceso directo a los datos de los interesados en un formato fácil de usar y legible ayudaría a mejorar los mismos datos, además de corregir el desequilibrio económico entre las grandes empresas, por un lado, y los interesados, por otro.

También se podrían compartir los beneficios con los individuos, por ejemplo, el acceso a la información sobre el consumo de energía en un formato fácil de usar podría hacer más fácil a las familias obtener las mejores tarifas de gas y electricidad, así como controlar su consumo de energía y modificar su estilo de vida para reducir sus facturas y, en consecuencia, su impacto ambiental. Un detalle que parece haber pasado desapercibido en España en la incorporación de lectores inteligentes de electricidad y el despliegue previsto de los mismos.

Por otro lado, permitir la portabilidad de datos (que se regula en la propuesta del RGPD) podría facilitar a las empresas y consumidores maximizar los beneficios del *big data* de una manera más equilibrada y transparente. También puede ayudar a minimizar las prácticas injustas o discriminatorias y reducir los riesgos del uso de los datos inexactos para la toma de decisiones, lo cual beneficiaría tanto a empresas como a personas o consumidores.

Sin duda, el WP29 considera la transparencia como un elemento clave, no solo a la hora de garantizar los derechos de los ciudadanos, sino también como vehículo enriquecedor de los datos y, por consiguiente, de los resultados.

## Consecuencias de la incompatibilidad de finalidades

El WP29 también considera que el procesamiento incompatible no se puede remediar con solo adoptar un nuevo fundamento jurídico.

No cumplir con el requisito de compatibilidad tiene graves consecuencias: **el tratamiento de los datos personales de manera incompatible con los fines especificados en la captación de datos es ilegal** y, por lo tanto, no está permitido. El RGPD lo considera una infracción.

**La legalización** del tratamiento de datos personales para **finalidades incompatibles** no basta con solo cambiar los términos de un contrato con el **interesado** o mediante la identificación de un interés legítimo adicional del responsable del tratamiento. Esto va en contra del espíritu del principio de limitación de finalidades y elimina su fundamento (Art. 29 WP, 2013).

En otras palabras, el responsable del tratamiento no puede limitarse a considerar el tratamiento posterior como una nueva actividad de procesamiento desconectado de la anterior y eludir esa prohibición.

## **Uso de los datos personales con motivo de la investigación histórica, estadística o científica**

El artículo 6.1.b de la Directiva 95/46/C contenía una disposición específica sobre el tratamiento posterior de los datos de carácter personal con fines históricos, estadísticos o científicos. Esta disposición, en relación con los considerandos pertinentes, permite el tratamiento posterior de los datos con fines históricos, estadísticos y científicos, siempre y cuando el responsable del tratamiento aplique garantías apropiadas y, en particular, cuando garantice que los datos no se utilizarán para apoyar las medidas o decisiones relativas a personas concretas.

En este sentido, el artículo 5.2 del RGPD establece que la legitimidad del tratamiento de datos personales para los fines de la investigación histórica, estadística o científica (y según las condiciones establecidas en el artículo 86) **solo es posible mediante un tratamiento de datos que no permita la identificación de los sujetos.**

Es decir, **se recomienda siempre que sea posible** el tratamiento de datos que no permitan la identificación de los sujetos (disociados); **en este caso será necesario separar la información que permite la identificación de los sujetos del resto de información (separación funcional).** La aplicación de pseudonimización o, si es posible, el trabajo sobre datos anonimizados.

El WP29 recomienda que las publicaciones con motivo de investigaciones históricas, estadísticas o científicas **solo podrán incorporar datos de carácter personal si hay consentimiento por parte del interesado**, el interesado ha hecho públicos los datos o la publicación de los datos personales es necesaria para presentar los resultados de una investigación, siempre que los intereses o los derechos o libertades fundamentales del afectado no prevalezcan sobre tales objetivos. También abre la puerta a que la comisión pueda adoptar decisiones a fin de especificar los criterios y requisitos del tratamiento de los datos personales implicados, así como las limitaciones a los derechos de información y acceso por parte de los interesados.

Al respecto, el WP29 considera necesaria la incorporación de garantías adicionales, como las medidas técnicas u organizativas para garantizar la separación funcional u otras garantías que contribuyan a la transparencia o elección. Además, esta disposición no se implica ni deja claro que su uso ulterior con fines de investigación históricos, estadísticos o científicos esté sujeto a la misma evaluación de la compatibilidad de las finalidades de los tratamientos de datos de carácter personal.

En España, la LOPD contemplaba este tipo de tratamientos en línea con lo establecido en la Directiva 95/46/C. Además, desarrollos específicos legislativos posteriores tratan de dar cumplida respuesta en ámbitos de investigación específica, como es el caso de la investigación biomédica: Ley 14/2007, de 3 de julio, de Investigación biomédica.

En resumen, no se deberán tratar datos de carácter personal en estas áreas, a menos que sea estrictamente necesario para los fines de la investigación y no sea posible aplicar procesos de disociación o separación funcional entre datos identificativos y el resto de los datos, aplicando en este caso las medidas de seguridad adecuadas y organizativas destinadas a limitar la posible identificación de las personas.

Para terminar con este asunto, hagamos una breve pausa en la *Resolución big data* (de la 36a edición de la International Conference of Data Protection and Privacy Commissioner), pues seguir sus recomendaciones seguramente nos facilite obtener los beneficios derivados del *big data* sin hipotecar los principios de la protección de datos:

«- Respetar el principio de especificación de finalidad.

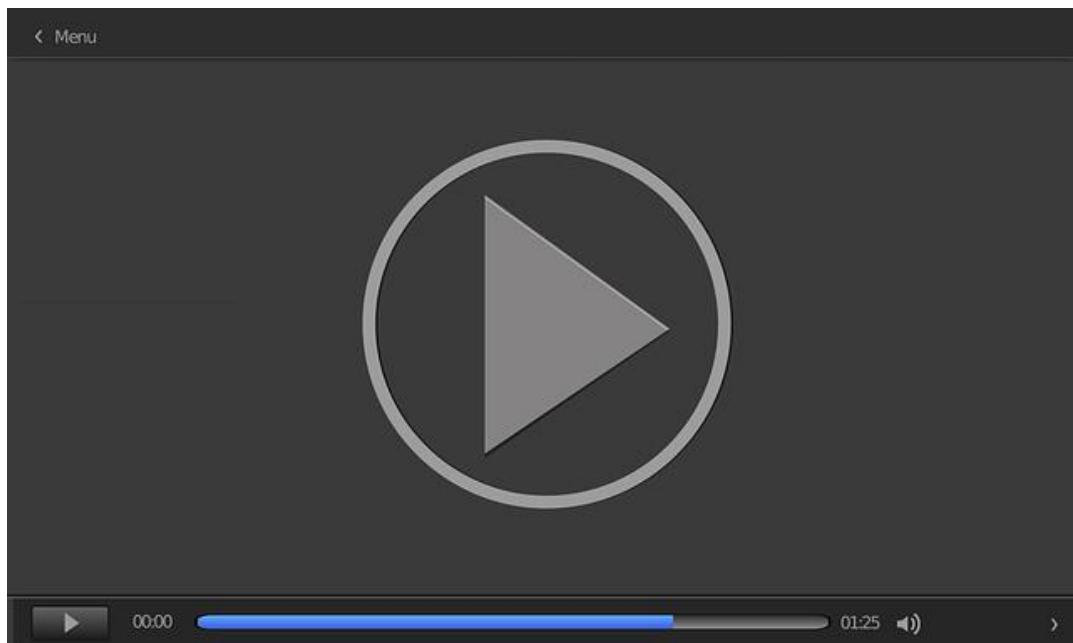
- ▶ Limitar la cantidad de información recolectada y almacenada a un nivel que sea necesario para el propósito legítimo que pretende.
- ▶ Obtener, cuando sea apropiado, el consentimiento válido del titular de los datos en relación con el uso de información personal para fines de análisis y de creación de perfiles.
- ▶ Ser transparentes acerca de qué información se recolecta, cómo se procesa, con qué propósito serán utilizados y si será transferida a terceros.
- ▶ Dar a las personas acceso apropiado a los datos que han sido recolectados sobre ellas y a la información y decisiones que se han tomado con esos datos. Las personas deben ser avisadas de la fuente de sus datos personales y, cuando sea apropiado, de su derecho a corregir su información, así como de las herramientas para controlar esta información.
- ▶ Ofrecer a las personas cuando sea apropiado acceso a la información sobre los insumos principales y los criterios para la toma de decisiones (algoritmos) que se han utilizado como base para el desarrollo del perfil. La información debe presentarse en un formato claro y comprensible.
- ▶ Llevar a cabo una evaluación de impacto en la privacidad, especialmente cuando el análisis del *big data* implica usos novedosos o inesperados de los datos personales.
- ▶ Desarrollar y utilizar tecnologías del *big data* de acuerdo con los principios de la Privacidad por Diseño.

- ▶ Considerar cuándo los datos anónimos mejorarán la protección de la privacidad. La anonimización puede ayudar a mitigar los riesgos para la privacidad asociados con el análisis del *big data*, pero solo si la anonimización está diseñada y gestionada apropiadamente. La solución óptima para anonimizar los datos debe decidirse caso por caso, posiblemente utilizando una combinación de técnicas.
- ▶ Tener mucho cuidado y actuar cumpliendo la legislación aplicable en materia de protección de datos, cuando se comparten o se publican conjuntos de datos con seudónimos o que pueden ser identificables indirectamente. El acceso debe ser limitado y controlado cuidadosamente si los datos contienen suficientes detalles, es decir, que pueden vincularse con otros conjuntos de datos o contienen datos personales.
- ▶ Demostrar que las decisiones respecto al uso del *big data* son justas, transparentes y responsables. Relacionado con el uso de datos para fines de creación de perfiles. Tanto estos como los algoritmos en que están basados requieren una valoración continua. Este necesita revisiones regulares para verificar si los resultados de la creación de perfiles son responsables, justos y éticos y si son compatibles y proporcionados con el propósito para el cual los perfiles son usados. Debe evitarse la injusticia con las personas debido a resultados completamente automatizados que arrojen un falso positivo o un falso negativo. Siempre debe estar disponible una valoración manual de resultados, con efectos significativos para los individuos» (Norwegian Data Protection Authority, 2014).

## Usar fuentes de datos externas

También deberemos prestar atención al uso de datos de fuentes externas, facilitados por terceros. Debemos conocer si el tratamiento de datos que se va a realizar tiene una finalidad compatible o cuenta con la base legal adecuada como, por ejemplo, el consentimiento del afectado. Para evaluar estas cuestiones, es necesario plantear un proceso análisis específicos.

Para terminar este apartado, accede al vídeo *RGPD y el cloud computing*.



---

Accede al vídeo:<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=4108f7ec-71e6-4f80-93ff-abd800c247a2>

---

## 9.4. Privacidad por diseño

La privacidad por diseño (PbD) es un concepto desarrollado por Ann Cavoukian, Ph, Comisionada de Información y Privacidad de Ontario, Canadá, en los años 90. Concepto con el objeto de responder a los retos de los sistemas de datos en red y a gran escala.



Figura 8. Privacidad por diseño. Fuente: elaboración propia.

Nace como evolución al concepto de **tecnologías de mejora de la privacidad** y promueve garantizar la privacidad en todo el ciclo de vida de los datos, no solo desde el desarrollo de marcos regulatorios, sino también integrándola en el centro de las actividades de las organizaciones y englobando:

- ▶ Sistemas de tecnologías de la información.
- ▶ Prácticas de negocio responsables.
- ▶ Diseño físico e infraestructura en red.

Y de este modo poder asegurar la privacidad y obtener control de las personas de su propia información, facilitando el principio de autodeterminación informativa de los afectados y ofreciendo, además, a las organizaciones una ventaja competitiva sostenible.

La privacidad por diseño (PbD) desarrolla siete principios que pueden ser aplicados a todo tipo de información personal y, con especial atención, a datos de significada sensibilidad como es la información médica y datos financieros.

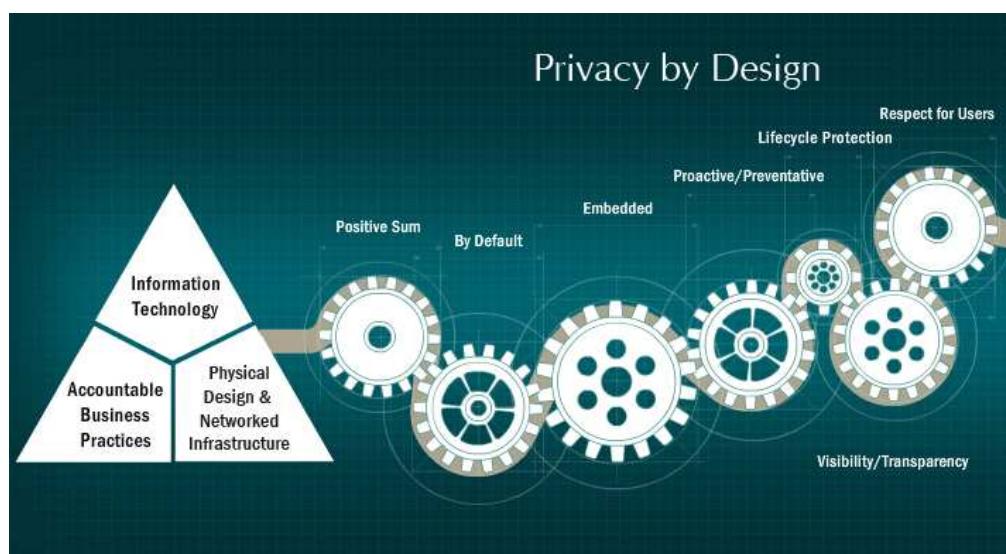


Figura 9. Los siete principios fundamentales de la privacidad por diseño. Fuente: MediaBUZZ, 2013.

Estos siete principios fundamentales, son:

- ▶ **Proactivo, no reactivo; preventivo, no correctivo.** El enfoque basado en medidas proactivas, en lugar de reactivas. Anticipar y prevenir eventos de invasión de la privacidad antes de que estos ocurran.
- ▶ **Privacidad como la configuración predeterminada (privacidad por defecto).** Asegurándose de que los datos personales estén protegidos automáticamente y por defecto en cualquier sistema de TI o en cualquier práctica de negocios, actuando como una configuración predeterminada de alta privacidad que requeriría la acción de la persona o el usuario para reducir esa configuración de garantía y facilitar acceso a su información. Dentro de esta configuración determinada y como buenas prácticas se incluiría:
  - Finalidad: la información recolectada debe ser una finalidad concreta.
  - Información: el usuario debe ser informado de la finalidad de la recolección.
  - Limitar los datos a la finalidad especificada.
  - Anonimizar los datos, limitar el uso de los datos, su retención, divulgación a la finalidad con la que fueron recabados y a las regulaciones legales.
- ▶ **Privacidad incrustada en el diseño.** La privacidad por diseño se incrusta en el diseño y la arquitectura de los sistemas de información y en las prácticas de negocios. Se convierte en un componente esencial de la funcionalidad central que está siendo entregada. La privacidad es parte integral del sistema, sin disminuir su funcionalidad.
- ▶ **Funcionalidad total «todos ganan», no «si alguien gana, otro pierde».** La privacidad por diseño busca acomodar todos los intereses y objetivos legítimos de una forma **ganar-ganar** y evita las falsas dualidades, tales como privacidad versus seguridad, partiendo de la idea de que sí es posible tener ambas al mismo tiempo.

► **Seguridad de extremo a extremo. Protección de ciclo de vida completo .**

Protección de dato en todo el ciclo de vida, desde la recolección del dato pasando por su almacenamiento y retención y, la fase final, su destrucción adecuada. Para ello se recurriría a la aplicación de medidas de seguridad eficaces.

► **Visibilidad y transparencia. Mantenerlo abierto.** Que las partes involucradas, negocio y tecnologías operen de conformidad a objetivos claros y declarados de forma transparente y bajo el principio de que todas las operaciones van a ser realizadas de conformidad con las premisas y objetivos fijados y van a estar sujetos a esquemas de verificación independientes.

► **Respeto por la privacidad de los usuarios. Mantener un enfoque centrado en el usuario.** Por encima de todo, requiere que los arquitectos y operadores prioricen el interés de las personas ofreciendo servicios robustos de privacidad por defecto, información adecuada y opciones amigables para el usuario, facilitándonos el control sobre sus datos.

## 9.5. Evaluaciones de impacto (PIA/EIPD)

Una evaluación de impacto es un análisis de los riesgos que sobre el derecho fundamental a la protección de datos personales de los afectados puede tener un tratamiento de datos determinado, ya sea como producto o como servicio.

Las evaluaciones de impacto forman parte de la respuesta que desde la perspectiva de la **privacidad por diseño** se está articulando desde las autoridades de protección de datos de la UE y que se incorporan en el articulado del reglamento de protección de datos de la UE.

En este epígrafe nos centramos en la guía desarrollada por la AEPD al entender que se adapta mejor a la legislación española, si bien es recomendable consultar otras como la elaborada por la Information Commissioner's Office (*Conducting privacy impact assessment code of practice*).

### **Antecedentes de la evaluación de impacto**

Encontramos el origen de estas evaluaciones en el ámbito anglosajón, de ahí que sean ampliamente conocidas como PIA (*privacy impact assessments*) y no es difícil encontrar metodologías para desarrollar de forma completa dicha evaluación. En concreto, la Agencia Española de Protección de Datos nos ofrece una guía de cómo realizar una EIPD.

Las evaluaciones de impacto de protección de datos no eran obligatorias en la Directiva 96/45/CE, ni tampoco existe una metodología única o exigida, pero son incorporadas en el marco del reglamento de protección de datos de la UE y actualmente son ampliamente recomendados para los entornos *big data*.

En otros países sí se anticipó obligatoriedad, como el caso de Reino Unido, donde desde julio de 2008 es obligatorio realizar estas evaluaciones en el entorno de la administración de Estado y sus agencias.

En este tema desarrollaremos el concepto de un PIA/EIPD para facilitar su entendimiento e incorporaremos una aproximación metodológica desde una perspectiva didáctica. Como comentábamos, a la hora de meternos en harina existen múltiples metodologías a las que recurrir, aunque no todas tienen el mismo alcance, por ejemplo:

- ▶ Guía de evaluación de impacto de la AEPD (Agencia Española de Protección de Datos).
- ▶ *Conducting privacy impact assessments code of practice*, ICO (Information Commissioner's Office).
- ▶ ISO 22307:2008. Financial services. Privacy impact assessment (última revisión: 2012).
- ▶ ISO/IEC WD 29134:2017. Information Technology. Security Techniques. Guidelines for privacy impact assessment.

Uno de los primeros países en el que se dio forma a las EIPD ha sido el Reino Unido, que publicó la primera guía en 2007 (*PIA Handbook*), revisada posteriormente en 2009.

La metodología intenta facilitar la integración de la evaluación de impacto dentro de los proyectos y las herramientas de gestión de riesgos y define el PIA como una herramienta más de gestión de riesgos con un foco que va más allá de asegurar a las organizaciones el mero cumplimiento normativo en protección de datos personales.

Las metodologías PIA en muchas ocasiones han tomado como inspiración metodológica las metodologías de evaluación de riesgos más conocidas, como:

► Gestión de riesgos:

- ISO 31000:2009. Risk management. Principles and guidelines.
- Combined Code and Turnbull Guidance.
- UK Treasury's The Orange Book: Management of Risk.
- ENISA's approach to risk management.

► Seguridad de la información:

- ISO/IEC 27005:2011. Information security risk management.
- IT-Grundschutz.
- NIST SP 800-39. Managing Information Security Risk.
- ISACA and COBIT.

► Metodologías de análisis de riesgos:

- CRAMM.
- EBIOS.
- OCTAVE.
- NIST SP 800-30. Guide for Conducting Risk Assessments.

## En qué situaciones se debe realizar una evaluación de impacto

El RGPD recoge la obligación de la realización de evaluaciones de impacto en los siguientes casos (art. 35, RGPD):

- ▶ **Evaluación sistemática y exhaustiva de aspectos personales** de personas físicas que se base en un tratamiento automatizado.
- ▶ **Elaboración de perfiles** sobre cuya base se tomen decisiones que produzcan efectos jurídicos para las personas físicas o que les afecten significativamente.
- ▶ **Tratamiento a gran escala de categorías especiales de datos** o datos penales.
- ▶ Observación sistemática a gran escala de una zona de acceso público (por ejemplo, videovigilancia).

El RGPD recoge la capacidad de la autoridad de control de concretar una lista con los tratamientos que serán objeto de PIA, además de los ya contemplados en el RGPD.

La evaluación de impacto, por tanto, será obligatoria en general cuando las operaciones de tratamiento entrañen riesgos específicos para los derechos y libertades de los interesados debido a su tecnología, naturaleza, alcance, contexto o fines.

Una evaluación de impacto no siempre es necesaria y dependerá tanto del volumen de datos tratados, finalidades, naturaleza de los datos y tecnologías utilizadas.

Como ejemplo, veamos dos posibles casos:

- ▶ Caso 1. Diferencias en el volumen de afectados:
  - Fichero de clientes de un autónomo que alcanza, por ejemplo, a unos pocos cientos de ciudadanos.
  - Fichero de clientes de una operadora de comunicaciones que alcanza millones.
- ▶ Caso 2. Naturaleza de los datos tratados:
  - Los datos tratados por un pequeño servicio de reparaciones.
  - Los datos tratados por un consultorio médico.

En ambos escenarios es fácil prever un mayor perjuicio sobre los ciudadanos en el caso de, por ejemplo, un evento de brecha de seguridad que facilite el acceso a los datos personales a personal no autorizado para las situaciones *b* de los casos propuestos que para las situaciones *a*.

El alcance del PIA/EIPD como la formalidad de este se entiende que es poco generalizable y requiere significativos esfuerzos. Por ello, la **AEPD ha emitido una lista.**

---

Accede a la lista de tipos de tratamiento de datos que requieren evaluación de impacto relativa a protección de datos (art. 35.4) desde el siguiente enlace: <https://www.aepd.es/sites/default/files/2019-09/listas-dpia-es-35-4.pdf>

---

En esta lista se recogen características de tratamientos que deben ser objeto de una EIPD. Para ello bastara con reunir **dos o más características** de entre las incluidas.

A modo de resumen, las características son:

- ▶ Uso de datos a gran escala.
- ▶ Tratamientos que impliquen la asociación, combinación o enlace de registros de bases de datos de dos o más tratamientos con finalidades diferentes o por responsables distintos.
- ▶ Perfilado o valoración de sujetos:
  - Toma de decisiones automatizadas.
  - Tratamientos que impliquen la observación, monitorización, supervisión, geolocalización o control del interesado de forma sistemática y exhaustiva.
  - El uso de categorías especiales, datos relativos a condenas o infracciones penales o datos que permitan determinar la situación financiera o de solvencia patrimonial o deducir información sobre las personas relacionada con categorías especiales de datos.
  - Uso de datos biométricos con el propósito de identificar de manera única a una persona física.
  - Datos genéticos.
  - Tratamientos de datos de sujetos vulnerables o en riesgo de exclusión social.
  - Tratamientos que impliquen la utilización de nuevas tecnologías o un uso innovador de tecnologías consolidadas.
  - Tratamientos de datos que impidan a los interesados ejercer sus derechos, utilizar un servicio o ejecutar un contrato.

También la AEPD ha emitido un listado de tratamientos que no tienen que ser objeto de una evaluación de impacto, aunque en algún caso puedan reunir dos o más características de las comentadas en el documento anterior.

---

Accede a la lista orientativa de tipos de tratamientos que no requieren una evaluación de impacto relativa a la protección de datos según el artículo 35.5 RGPD desde el siguiente enlace: <https://www.aepd.es/sites/default/files/2019-09>ListasDPIA-35.5I.pdf>

---

En este listado, a modo de resumen, se incluyen:

1. Realizados bajo las directrices de las autoridades de control.
2. Realizados estrictamente bajo las directrices de códigos de conducta aprobados.
3. Necesarios para el cumplimiento de una obligación legal.
4. Realizados por autónomos que ejerzan de forma individual.
5. Realizados por pymes para la gestión interna con finalidad de contabilidad, gestión de recursos humanos y nóminas, seguridad social y salud laboral, requeridos por ley.
6. Realizados por comunidades.
7. Realizados por colegios profesionales y asociaciones sin ánimo de lucro.

A la hora de proceder a realizar el EIPD, es importante resaltar que es mejor cuando más temprano tenga lugar. Se recomienda su incorporación en la fase de definición del servicio o producto. Cuanto antes se realice antes se podrán:

- ▶ Evaluar los posibles riesgos y establecer medidas o controles compensatorios que mitiguen los mismos.
- ▶ Una mayor eficacia en la aplicación de los controles e incluso a un menor coste.
- ▶ La implantación de controles antes de que se genere un perjuicio sobre los derechos de los ciudadanos.

## En qué consiste EIPD o PIA

La EIPD, tal y como es entendida por parte de la AEPD, es una evaluación de riesgos que se caracteriza por:

- ▶ Ser un **proceso más amplio que la simple comprobación del cumplimiento de la normativa** en protección de datos personales (la expectativa es que vaya más lejos que una evaluación y que, con base en los riesgos identificados, formule controles con independencia de que sean o no explícitamente requeridos por la norma, por ejemplo, medidas de seguridad).
- ▶ Se debe realizar **antes de implantar un nuevo producto, servicio, sistema de información** o cuando uno existente sufra significativos cambios. Totalmente alineado con los principios de privacidad por diseño.
- ▶ Es un **proceso sistemático y reproducible**. Debe sustentarse en una aproximación metodológica.
- ▶ Ha de orientarse en **revisar procesos**. La revisión de procesos mejora el entendimiento de los flujos de datos y de sus finalidades.
- ▶ Debe permitir una identificación clara de los responsables de cada una de las fases del EIPD.

- ▶ Debe tener una **aproximación abierta** que invite a participar constructivamente a todos los afectados: departamentos, grupos de interés, entidades externas, agentes sociales; el enriquecimiento de la evaluación desde la participación de todos los afectados se considera clave a la hora alcanzar unos buenos resultados.
- ▶ El resultado del EIPD debe formalizarse en un documento.
- ▶ Debe ser un **proceso transparente** cuyo resultado sea conocido por miembros y grupos de interés de la organización.
- ▶ Debe ser un **proceso periódico** que evalúe la eficacia de la aplicación de las medidas previamente definidas y aplicadas y proponga correcciones o la aplicación de medidas adicionales en pro del objetivo de la reducción del riesgo. La orientación como un proceso de mejora continua que permita responder adecuadamente a los riesgos identificados y mejorar la eficacia de las medidas. Una aproximación basada en un modelo PDCA (*plan-do-check-act*).

## Fases del EIPD

La metodología para desarrollar la EIPD que define la guía desarrollada por la AEPD contempla las siguientes fases:

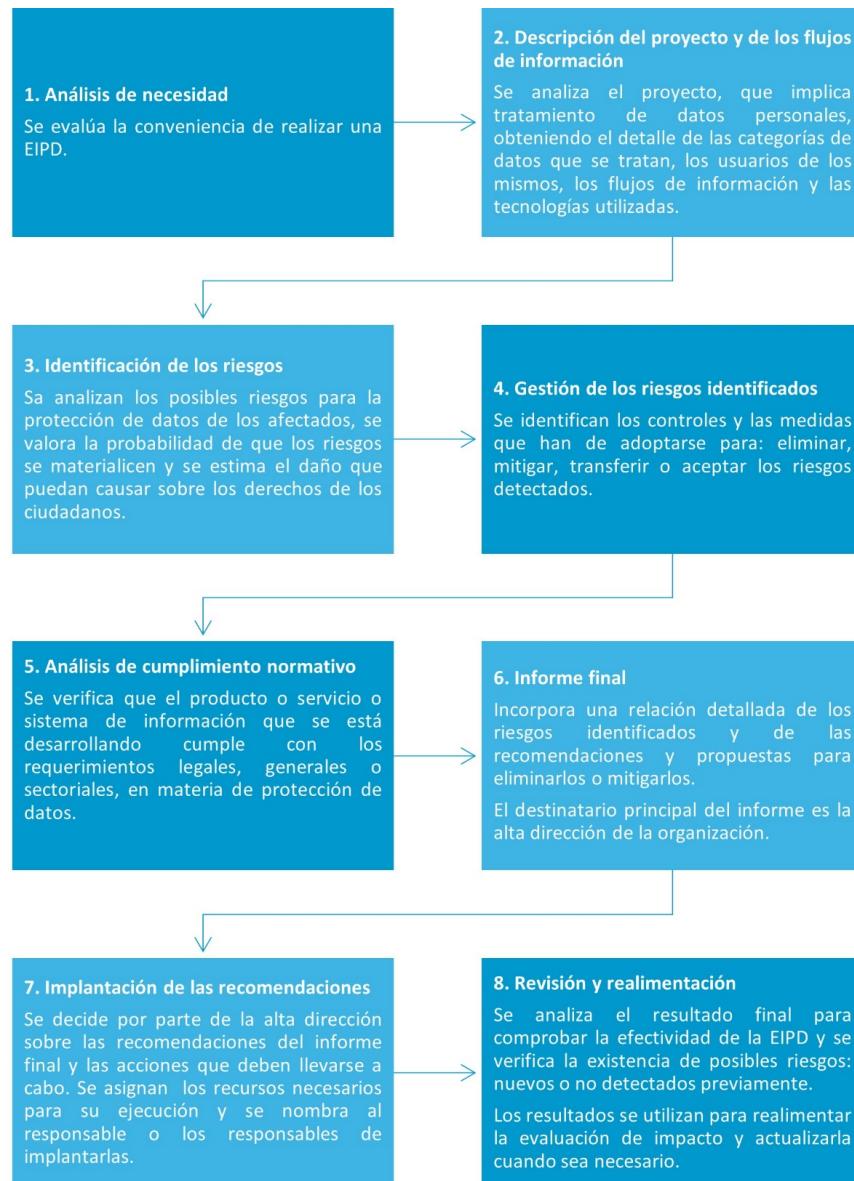


Figura 10. Metodología para desarrollar la EIPD. Fuente: elaboración propia.

Es una aproximación cíclica de mejoramiento que en sus correspondientes fases integra a las partes afectadas (internas o externas a la organización) para alcanzar una correcta identificación de los riesgos.

A continuación, profundizaremos en algunas de las fases identificadas (siguiendo la guía de evaluación de impacto de la Agencia Española de Protección de Datos).

## **El equipo de trabajo para desarrollar el EIPD**

La guía no establece reglas sobre quién ha de participar en la evaluación, esto lógicamente variará de una organización a otra. Ahora bien, por la propia naturaleza del proceso parece lógico que en el equipo estén integrados:

- ▶ Delegado de protección de datos con conocimientos de obligaciones legales en el ámbito de la protección de datos.
- ▶ Representantes del departamento TIC.
- ▶ Responsables de seguridad.
- ▶ Representantes de las áreas de negocio o departamentos que patrocinen el proyecto o se vean afectados por el mismo.

El éxito del EIPD va a depender también de la implicación de la alta dirección, de su compromiso y apoyo para liderar y garantizar la disponibilidad de los recursos necesarios para su realización. Y también dependerá del alcance que se le pretenda dar de la capacidad de su concreción y de lo ambicioso del mismo. Cuanto más ambicioso sea, más riesgos tendrá de no responder a los objetivos definidos.

## Análisis de necesidad

Ya hemos mencionado los ámbitos en los que la AEPD considera aplicable el EIPD. Por lo que en esta fase procederíamos a verificar si estamos en alguna de dichas circunstancias que nos indiquen la necesidad de realizar el EIPD.

## Descripción del proyecto y los flujos de datos personales

Como mencionábamos, es conveniente realizar el EIPD al inicio del proyecto. Esto puede suponer que la descripción, como los flujos identificados, sea un tanto inmadura. El grado de definición en las fases iniciales siempre es más difuso, por lo que será conveniente, si se da el caso, proceder a una revisión posterior de estos aspectos.

En esta fase de descripción debemos incorporar en **la documentación del EIPD**:

- ▶ Un **resumen del proyecto y características** incluyendo descripción de necesidad u oportunidad de este.
- ▶ Explicitar **los aspectos que resulten ser más relevantes**, que desde la perspectiva de la protección de datos personales tendrá el proyecto y que *a priori* sean susceptibles de presentar un mayor riesgo o que presentan una mayor dificultad a la hora de cumplir con la normativa.
- ▶ Además, es aconsejable incorporar una descripción detallada de:
  - Medios de tratamiento y tecnologías que se utilizan.
  - Categorías de datos personales que se tratan (datos de identificación...), las finalidades de su tratamiento, los colectivos afectados y justificación de la necesidad de la utilización de estos datos.
  - Criterios de acceso a cada categoría de datos personales, justificando convenientemente dicha necesidad. Es decir, quién accede.

- Los flujos de información en el ciclo de vida de los datos:
  - Recogida de datos.
  - Circulación de los datos dentro de la organización.
  - Las comunicaciones de los datos.
  - Las posibles cesiones provenientes de terceros.
  - Las transferencias internacionales previstas.
  - Los encargos de tratamientos previstos.

## Identificación y evaluación de riesgos para la protección de datos

La utilización inadecuada o no lícita de datos personales es un riesgo inherente al tratamiento de datos personales, veamos algunos ejemplos:

- ▶ El uso de la información accesible. Por ejemplo, el uso de Internet, ubicaciones para fines comerciales (correo no deseado, publicidad dirigida...).
- ▶ La utilización de los hábitos o información de los empleados recogidos y utilizados por sus superiores para dirigir la búsqueda de evidencias que sustenten posibles despidos procedentes.
- ▶ Falseamiento de identidades para realizar actividades ilegales en nombre de los interesados, este último frente persecución penal.
- ▶ La toma de decisiones sobre información **alterada** de forma **no deseada** de los datos de salud de manera que los pacientes son inadecuadamente atendidos, con el empeoramiento de su condición e incluso generando efectos de discapacidad o la muerte.

El objetivo del PIA/EIPD es identificar los riesgos existentes y gestionar este riesgo: evitándolo o eliminándolo, mitigándolo, transferirlo o aceptarlo. Ello implica el establecimiento de medidas para actuar frente a los riesgos identificados, evaluar la eficacia de la implementación de dichas medidas e incorporar acciones correctoras o compensatorias, dentro de una visión cíclica de la gestión de riesgos.

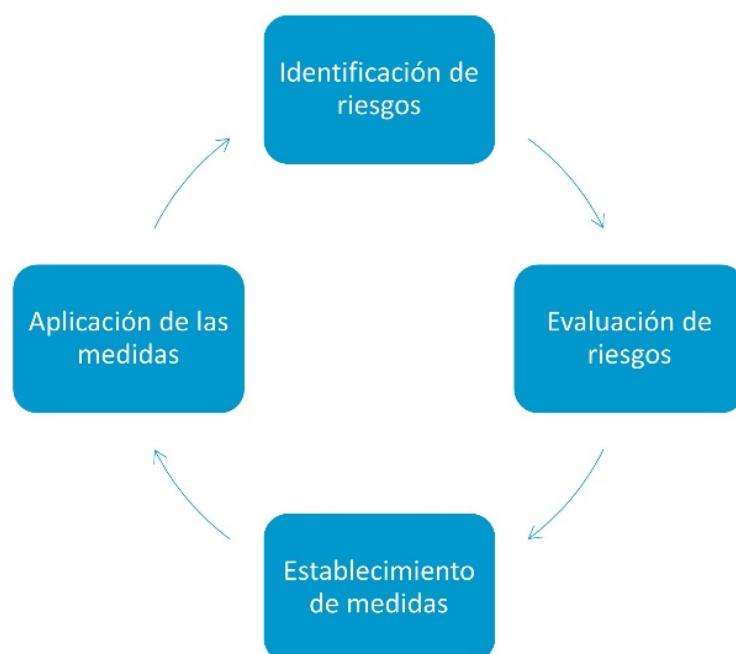


Figura 11. Objetivo del PIA/EIPD. Fuente: elaboración propia.

Los riesgos sobre los que se orienta principalmente la evaluación son los riesgos que afectan a los derechos de los ciudadanos, y esto implica evaluar aspectos técnicos, organizativos y jurídicos. Pero, al mismo tiempo del modo en que a las organizaciones realizan los tratamientos de datos personales y de cómo estas respetan los derechos de los afectados, se derivan potenciales riesgos para las propias organizaciones que trascienden eventuales sanciones que convienen tener también en consideración en la evaluación, aunque habitualmente las metodologías no los contemplen:

- ▶ **Pérdida de ventas:** una respuesta negativa por parte de los consumidores hacia un producto o servicio si se identifica una falla referente a la protección de datos. La pérdida de confianza por el consumidor origina una baja de ventas o la retirada del producto o servicio.
- ▶ **Incremento de costes:** la incorporación de costes adicionales para subsanar las deficiencias que sin los principios de la protección de datos hubieran sido incorporados en las fases de diseño y desarrollo. Incremento de los costes de puesta en el mercado de los productos o servicios o incluso de obligaciones indemnizatorias afectan a la viabilidad del proyecto.
- ▶ **Imagen corporativa:** impacto sobre la imagen de la organización y sobre la confianza de accionistas o terceras partes interesadas.

Por tanto, en el EIPD se **consideran dos tipos de riesgos:**

- ▶ Los que afectan a las personas:
  - La posible violación de los derechos de los ciudadanos directamente afectados por el tratamiento, por ejemplo:
    - La disponibilidad del consentimiento para el tratamiento de los datos personales.
    - El deber de secreto.
    - Bloqueo y cancelación de los datos efectiva tras el requerimiento por parte del afectado.
    - Procesos de rectificación de datos no eficaces. No comunicación a cesionarios.
  - La pérdida de información, por ejemplo:
    - Desastre que afecta a los entornos de tratamiento de datos.

- Falta de eficacia en las copias de seguridad.
- Procesos que atentan contra la integridad de la información.
- La pérdida de información, por ejemplo:
  - Decisiones que afectan a los ciudadanos.
  - Revelación de información.
- ▶ Los que afectan a las organizaciones, entre otros:
  - La negativa percepción del producto o servicio puede originar una disminución en su uso.
  - Costes por el rediseño.
  - La posible retirada del mismo.
  - Reputación e imagen.
  - Sanciones.

Para el análisis de riesgos podemos recurrir a diferentes metodologías que nos pueden ayudar en este punto. Algunas en las que encontraremos especial acomodo para la realización del EIPD son:

- ▶ *Methodology for privacy risk management* de la Commission Nationale de l'Informatique et des Libertés (CNIL). Muy recomendable.
- ▶ MAGERIT (Metodología de Análisis y Gestión de Riesgos de los Sistemas de Información), herramienta creada por el Consejo Superior de Administración Electrónica.
- ▶ Risk IT de ISACA.
- ▶ ISO/IEC 27005. Gestión de riesgos de la seguridad la información.

- ▶ ISO 31000:2009. Gestión de riesgos, principios y directrices que da nombre a una familia de normas sobre principios y directrices de gestión del riesgo.
- ▶ ISO/IEC 31010, Gestión de riesgos, técnicas de evaluación del riesgo, en la que se detallan diversos métodos que pueden ayudar a identificar y detectar los riesgos de un nuevo producto o servicio.

De la misma manera que existen diferentes metodologías, también nos encontramos con diferentes formas de definir el riesgo, una de ellas es definir el concepto de riesgo con base en los componentes que los configuran:

- ▶ Amenazas a las que se está sometido (¿qué puede pasar?).
- ▶ Las vulnerabilidades existentes (¿qué facilita que si algo pasa afecte o genere un mayor perjuicio?).
- ▶ La probabilidad de que un evento tenga lugar, es decir, que una amenaza se materialice y explote una vulnerabilidad (¿con qué frecuencia se puede dar?).
- ▶ El impacto que dicho evento tendrá sobre los afectados en el caso de materializarse: que la amenaza aproveche una vulnerabilidad y que se puede establecer con base en la sensibilidad de la información divulgada (salud, orientación sexual, orientación política, etc.) y número de personas afectadas.

De manera que se define así:

El **riesgo** es la posibilidad de que una amenaza se materialice aprovechando una vulnerabilidad.

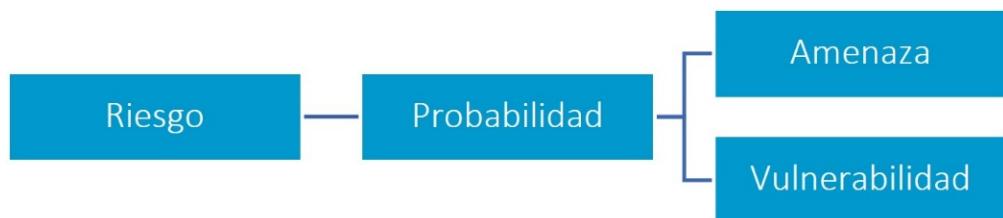


Figura 12. Concepto de riesgo. Fuente: elaboración propia.

La cuantificación del riesgo, en términos de cálculo (económico o cualitativo de daños), se puede obtener a partir del potencial impacto y con base en la probabilidad de que una amenaza explote una vulnerabilidad.

### Riesgo = Probabilidad x Impacto.

Vemos un par de ejemplos para entenderlo:

► Ejemplo A:

- **Amenaza:** en un centro sanitario conocido por la prestación asistencial a personas de gran proyección pública, podemos identificar como posible amenaza a la privacidad de los enfermos, la curiosidad de los profesionales del centro o de terceros interesados con motivaciones económicas.
- **Vulnerabilidad:** el centro dispone de un sistema de información clínica, es decir, la historia clínica está informatizada. Por tanto, el acceso a la historia clínica de un paciente le facilita toda la información sobre su estado de salud. Este sistema está en una fase de mejoramiento, el sistema de control de acceso lógico al sistema informático dispone de un usuario/clave (ADMIN/ADMIN) que aún no ha sido eliminado y que permite el acceso global a toda la información.

- **Impacto:** se ha filtrado información sobre el estado de salud de la persona pública citada junto con un listado de cientos de pacientes, esto ha perjudicado tanto a los afectados como a la entidad que ha sido sancionada y ha perdido reputación.

En este mero ejemplo podemos considerar que la amenaza está ampliamente motivada (la motivación es un factor relevante a la hora de evaluar la amenaza que puede suponer la acción de un atacante o del adversario) y la vulnerabilidad existente es muy fácil de explotar, lo que nos llevaría a estimar la probabilidad de que dicho evento, amenaza-vulnerabilidad, tuviera lugar como alta-muy alta.

► Ejemplo B:

- **Amenaza:** una trama de estafadores quiere verificar el número de cuenta de un anciano para proceder al giro fraudulento de recibos con cargo a la cuenta del anciano.
- **Vulnerabilidad:** en la oficina bancaria, por baja del personal, se ha incorporado un nuevo profesional temporal el cual no ha recibido formación sobre confidencialidad ni protección de datos. Los estafadores suplantan la personalidad de un hijo y obtienen la cuenta del anciano y el saldo de la cuenta girando un recibo por falsos servicios prestados, que el anciano no devuelve.
- **Impacto:** la revelación le ocasiona un quebranto económico al anciano, sin que exista repercusión mediata ni denuncia ante las instancias jurisdiccionales correspondientes.

Como decíamos, existen diferentes metodologías de evaluación de riesgos y para proceder a la misma de una forma sencilla y cualitativa podemos evaluar los diferentes riesgos atendiendo a una clasificación sencilla:

► Probabilidad de que suceda:

- Muy alta (81 %-100 %): es casi seguro que suceda.
- Alta (61 %-80 %): es muy probable.
- Media (41 %-60 %): es probable.
- Baja (21 %-40 %): es poco probable.
- Muy baja (0 %-20 %): es muy poco probable.

► Impacto que tiene en caso de que suceda:

- Muy alto (81 %-100 %): información sensible con un volumen alto de afectados.
- Alto (61 %-80 %): información sensible con un volumen alto de afectados.
- Medio (41 %-60 %): muchos afectados, informaciones no sensibles.
- Bajo (21 %-40 %): muy pocos afectados, pero información sensible.
- Muy bajo (0 %-20 %): pocos o muy pocos afectados, información no sensible.

Si estimamos un mapa de calor del impacto atendiendo a la sensibilidad de los datos y alcance en el número de afectados, tendremos:

N.º afectados	Muy bajo	Bajo	Medio	Alto	Muy alto
Sensibilidad					
No sensible	Verde	Amarillo	Amarillo	Naranja	Rojo
Sensible	Amarillo	Naranja	Naranja	Rojo	Rojo

Figura 13. Mapa de calor sensibilidad/afectados. Fuente: elaboración propia.

De esta manera, si evaluamos el riesgo considerando la ecuación anterior (Riesgo = Probabilidad x Impacto), podemos establecer un mapa de riesgo cualitativo:

Impacto	Muy bajo	Bajo	Medio	Alto	Muy alto
Probabilidad					
Muy baja	Verde	Verde	Verde	Amarillo	Naranja
Baja	Verde	Verde	Amarillo	Naranja	Naranja
Media	Verde	Amarillo	Amarillo	Naranja	Rojo
Alta	Amarillo	Naranja	Naranja	Rojo	Rojo
Muy alta	Naranja	Naranja	Rojo	Rojo	Rojo

Figura 14. Mapa de calor probabilidad/impacto. Fuente: elaboración propia.

De este modo, eventos con muy alta probabilidad de ocurrencia pero que tienen un muy bajo efecto sobre los derechos a la protección de datos personales o la organización tendrán la misma consideración (valor) de aquellos eventos que, aun teniendo una baja probabilidad de ocurrencia, su impacto puede ser muy severo.

De la propia definición de riesgo se entiende que es preciso extender el análisis a todo el mapa de amenazas por vulnerabilidades y, con base en su probabilidad, cualificar el riesgo derivado. Este valor de riesgo obtenido nos guía a la hora de identificar qué vulnerabilidades debemos corregir, empezando por:

- ▶ Corregir aquellas vulnerabilidades que implican un mayor riesgo, eliminando la vulnerabilidad o incorporando controles que reduzcan la probabilidad de que sea explotada por una amenaza. Veamos unos ejemplos:
  - En el acceso lógico a la base de datos, exige la incorporación de un identificador y una clave. Incorporar procedimientos de distribución de usuarios/claves adecuados y robustez en las contraseñas no elimina el riesgo, pero lo reduce al disminuir la probabilidad de que un **agente motivado** acceda fácilmente mediante, por ejemplo, un ataque de diccionario.
  - El establecimiento de procedimientos adecuados y la formación del personal de atención al cliente para identificar y actuar ante un intento de suplantación de identidad que pretenda acceder a información de un ciudadano. Tampoco elimina el riesgo, pero lo reduce al hacerlo también su probabilidad.
  - Aplicar los controles más inmediatos en coste y esfuerzo que reducen el perfil de riesgo de manera más significativa. Por ejemplo: eliminar un dato sensible del tratamiento de datos por realizar, que por el mayor impacto en la protección de datos incorpora un riesgo mayor (por mayor impacto en los afectados).

El objetivo del análisis de riesgos es evaluar los riesgos existentes y orientar la adopción de controles para la disminución de este. Como los componentes del riesgo son dinámicos, las evaluaciones deben ser periódicas, dentro de un proceso de gestión del riesgo que permita evaluar la eficacia de las medidas aplicadas, su efecto sobre el perfil de riesgo y su mejora.

La guía de EIPD de la AEPD ya incorpora una serie de riesgos que debemos evaluar en un producto o servicio que realice un tratamiento de datos personales, y en su Anexo II incluye un posible modelo para ayudar a sistematizar y gestionar las fases de identificación y gestión de riesgos. En la guía de EIPD los riesgos que se proponen que sean evaluados son, por ejemplo:

► **Riesgos de carácter general:**

- Pérdidas económicas y daños reputacionales derivados de:
  - Incumplimiento de la legislación sobre protección de datos personales.
  - Incumplimiento de legislaciones sectoriales aplicables con incidencia en la protección de datos personales.
  - Falta de medidas de seguridad adecuadas que generen pérdida de datos.
- Pérdida de competitividad del producto o servicio, por impacto en la reputación.
- No disponer de conocimiento experto en protección de datos desde la fase de definición del servicio o producto.

► **Legitimación de los tratamientos y cesiones de datos personales:**

- Incurrir en tratamiento o cesiones de datos no adecuados para las finalidades del tratamiento.
- No disponer de legitimidad.

- Trabajar sobre consentimientos de dudosa legalidad.
- No atender a las revocaciones de consentimiento.
- Incorporar datos de categorías especialmente protegidas sin velar adecuadamente por la legitimidad de dicho tratamiento.
- Incurrir en el enriquecimiento de los datos personales no previstos en las finalidades, no informando adecuadamente a los afectados.
- Incorporar prácticas de reidentificación de personas incorporando fuentes adicionales de información.
- Utilización de *cookies* de seguimiento de manera no legítima.
- Impedir la utilización del producto o servicio de manera anónima cuando la identificación no es indispensable para el producto o servicio.

► **Transferencias internacionales:**

- Acceso secreto de las autoridades de terceros países.
- No asistir a los ciudadanos en el ejercicio de sus derechos ante los importadores de datos.

► **Notificación administrativa de los tratamientos:**

- No disponer de procedimientos internos para mantener actualizadas las notificaciones a la AEPD.

► **Transparencia de los tratamientos ante los afectados:**

- No informar adecuadamente de los datos recabados (*cookies*, ubicación geográfica en dispositivos móviles, comportamiento en Internet y hábitos de navegación, etc.).

- Dificultar el acceso a la política de protección de datos o la redacción de estas de forma confusa que no permite conocer a los afectados la finalidad y uso de sus datos personales.

► **Calidad de los datos. Cumplimiento de los principios**

- Incorporar categorías de datos no necesarias.
- Errores que facilitan la pérdida de integridad de la información.
- Conculcar garantías en el uso de datos personales con fines históricos, científicos o estadísticos.
- Utilizar los datos para otros fines, no informados y para los que no se tiene consentimiento. Por ejemplo, la toma de decisiones que puede ser adversa o discriminatoria, como la oferta de precios diferentes.
- Riesgo de realizar inferencias erróneas usando, por ejemplo, técnicas de minería de datos, reconocimiento facial y datos biométricos.
- No disponer de procedimientos internos que garanticen la cancelación de datos de oficio una vez que estos ya no son necesarios para la finalidad para la que se recabaron.

► **Datos especialmente protegidos, con especiales requerimientos:**

- Fallas en el consentimiento.
- Disociaciones deficientes en procesos de investigación que solo requieren información anonimizada y que facilitan la reidentificación de los individuos.

► **Deber de secreto al que están obligados los participantes en el tratamiento:**

- Acceso no autorizado.
- Violaciones de la confidencialidad por parte de los empleados.

► **Tratamiento por encargo a terceras partes:**

- No existencia de contrato o no conformidad con la ley.
- La ausencia de protocolos para garantizar la debida diligencia en la elección del encargado de tratamiento.
- Ausencia de control sobre subcontrataciones.
- No considerar la cadena de encargos dentro de los procesos de ejercicio de derechos de los afectados.
- Ausencia de disposiciones concretas sobre los datos personales una vez finalizada la prestación del encargo de tratamiento: destrucción de los datos, devolución de estos y su portabilidad a otros sistemas.

► **Derechos de los usuarios:**

- Ausencia de procedimientos para la respuesta eficaz al ejercicio de derechos.
- Obstrucción del ejercicio para su dilación o limitación del derecho.

► **Seguridad de la información:**

- Deficiencias en la definición de funciones y obligaciones en el tratamiento de datos personales.
- Inexistencia de política de seguridad.

## Consulta a las partes afectadas

Estas fases siempre son las más delicadas en toda metodología por las posibles controversias que se pueden generar tras realizar la consulta, mayor si el tratamiento afecta a amplios sectores de la población, y no siempre es factible su realización. Pero en el caso de que sea posible realizarlo, nos permite:

- ▶ Obtener la opinión de grupos o personas que no están viciados por el proyecto (habrá que tener en consideración el interés de las partes).
- ▶ Dotar al proyecto de transparencia y conocer las incertidumbres o preocupaciones que puede generar entre los afectados con objeto de responder a las mismas de forma adecuada.

La consulta a agentes externos normalmente se realizará a asociaciones de usuarios, consumidores, ONG dedicadas a la defensa de la privacidad, organizaciones sectoriales, sindicales..., es decir, a agentes con cierto carácter de representación sobre los colectivos afectados o directamente sobre grupos de consumidores o posibles usuarios. Por ejemplo, si afecta a los trabajadores de una corporación, en esta fase se debería entablar la consulta con, entre otros, el comité de empresa.

También esta fase contempla la realización de consultas internas incorporando, según el tamaño de la organización, a la dirección, los departamentos implicados e incluso a proveedores que vayan a participar en el tratamiento de datos personales.

Se recomienda que esta fase se aborde cuanto antes, siempre y cuando se disponga de suficiente grado de definición sobre el tratamiento de datos por realizar, su finalidad y el proyecto de definición del producto o servicio, para que su exposición a los grupos de interés sea efectiva.

En general, **los principios que debe cubrir esta fase** son:

- ▶ **Planificar adecuadamente:** realizarla en el momento más adecuado y con tiempo suficiente para poder disponer de las respuestas dentro del proceso de evaluación que se está realizando. Es una fase que se suele demorar, esto hay que tenerlo muy en cuenta.
- ▶ **Concreción y claridad:** concreción en el alcance y claridad de las cuestiones planteadas son claves en el resultado adecuado de esta fase. Es importante obtener la mejor respuesta y para ello hay que prestar especial atención a que las cuestiones sean claras y concretas.
- ▶ **Extensión y diversidad:** la extensión de los consultados y su diversidad dotará en un mayor enriquecimiento en las opiniones obtenidas.
- ▶ **Transparencia:** es conveniente facilitar información sobre los resultados a las partes intervenientes. También suele ser un tema delicado al que habrá que prestar atención, a la mejor forma de abordar este aspecto.

## El informe de la evaluación de impacto

El informe final deberá hacerse público de forma parcial o en su totalidad (dependerá de las posibles restricciones que sean aplicables: legales, comerciales o de seguridad) y para ello se podrá hacer uso del sitio web de la organización. Cabe recordar que uno de los aspectos que motivan la EIPD es la transparencia.

No existe un modelo único de informe, pero la guía establece los aspectos que dicho informe debe integrar:

1. Identificación del proyecto, la persona o personas responsables de la EIPD, sus datos de contacto, la fecha de realización del informe y número de versión del mismo.
2. Resumen del informe con los resultados esenciales.
3. Introducción y descripción general del proceso de evaluación.
4. Resultado del análisis de necesidad de la evaluación y su justificación.
5. Descripción general del proyecto.
6. Descripción detallada de los flujos de datos personales.
7. Riesgos identificados.
8. Identificación de partes interesadas o a las que afecta el proyecto, tanto internas como externas a la organización y resultados de las consultas llevadas a cabo.

Figura 15. Guía informe evaluación impacto. Fuente: elaboración propia.

## 9.6. Referencias bibliográficas

Article 29 Working Party. (2013). Opinion 03/2013 on purpose limitation [Archivo PDF].[https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203\\_en.pdf](https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf)

Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. Diario Oficial de las Comunidades Europeas L 281, 23 de noviembre de 1995, pp. 31-50.

Google. (s.f.). Términos del servicio de Google [Página web].<https://policies.google.com/terms?hl=es>

ISACA. (2013). *Big data: impactos y beneficios*. ISACA.

MediaBuzz. (agosto de 2013). The 7 foundational principles of privacy by design [Página web].<https://www.mediabuzz.com.sg/best-practices-aug-13/the-7-foundational-principles-of-privacy-by-design>

Norwegian Data Protection Authority. (13 de octubre de 2014). *Resolution Big Data*. 36th International Conference of Data Protection and Privacy Commissioners, Mauricio.<http://globalprivacyassembly.org/wp-content/uploads/2015/02/Resolution-Big-Data.pdf>

Protiviti Chile [@ProtivitiChile]. (8 de febrero de 2018). En un mundo cada vez más digital, los auditores internos deben ser observadores entusiastas de todos los cambios tecnológicos que [Imagen adjunta] [Tuit]. Twitter.<https://twitter.com/ProtivitiChile/status/961606275443326977>

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento

de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento General de Protección de Datos). Diario Oficial de la Unión Europea L 119, 4 de mayo de 2016, pp. 1-88.<https://www.boe.es/DOUE/2016/119/L00001-00088.pdf>

Zhenhua's Wiki. (s.f.). Big Data [Página web].<https://techlarry.github.io/Data%20Science/Big%20Data/>

## Opinion 03/2013 on purpose limitation

---

Article 29 Working Party. (2013). Opinion 03/2013 on purpose limitation [Archivo PDF].[https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203\\_en.pdf](https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf)

---

Documento del grupo de trabajo del artículo 29 (WP29) que analiza el principio de limitación de la finalidad en el tratamiento de datos personales y realiza aclaraciones sobre la consideración de finalidades compatibles.

## Privacidad por defecto: los siete principios fundamentales

Cavoukian, A. (2011). *Privacy by design: the 7 foundational principles*. Information and Privacy Commissioner of Ontario [Archivo PDF]. [https://iapp.org/media/pdf/resource\\_center/pbd\\_implement\\_7found\\_principles.pdf](https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf)

Documento que describe los principios de la privacidad.

## Guía de evaluación de impacto de la Agencia Española de Protección de Datos

AEPD. (s.f.). Guía práctica para las evaluaciones de impacto en la protección de los datos sujetas al RGPD [Archivo PDF]. <https://www.aepd.es/sites/default/files/2019-09/guia-evaluaciones-de-impacto-rgpd.pdf>

Desarrollada por la AEPD, guía en el proceso de la evaluación de impacto.

### Guía de evaluación de impacto de la autoridad en protección de datos de Reino Unido

ICO. (s.f.) Conducting privacy impact assessment code of practice [Archivo PDF].<https://ico.org.uk/media/about-the-ico/consultations/2052/draft-conducting-privacy-impact-assessments-code-of-practice.pdf>

Incorpora *templates* útiles en el proceso de evaluación.

1. Identifica cuáles de los siguientes son principios de la privacidad por diseño:

  - A. Privacidad por defecto.
  - B. Privacidad incrustada en el diseño.
  - C. Seguridad extremo a extremo.
  - D. Todos los anteriores son correctos.
  
2. ¿Qué países forman el Comité de Protección de Datos RGPD, antiguo Grupo de trabajo del artículo 29 (WP29)?

  - A. Los miembros de la UE.
  - B. Los miembros de la OCDE.
  - C. Los países miembros del G8.
  - D. Alemania, Francia, Italia, España y Reino Unido.
  
3. Identifica cuáles de los siguientes aspectos se corresponden con las amenazas del *big data* definidas por el WP29:

  - A. Magnitud de la recopilación.
  - B. Seguridad de los datos.
  - C. Aumento dramático de las posibilidades de vigilancia del gobierno.
  - D. Todas las anteriores son correctas.
  
4. ¿Qué afirmación de las siguientes respecto a las evaluaciones de impacto es falsa?

  - A. La ley vigente obliga a la realización de evaluaciones de impacto cada seis meses de manera obligatoria.
  - B. La evaluación de impacto es un proceso sistemático y reproducible.
  - C. El resultado de la evaluación se debe plasmar en un documento.
  - D. La evaluación de impacto debe ser un proceso periódico.

- 5.** ¿En qué casos de los siguientes no se recomienda la realización de una evaluación de impacto?
- A. Cuando el tratamiento afecta a un número elevado de personas.
  - B. Cuando se traten grandes volúmenes de datos personales.
  - C. Cuando existan transferencias internacionales.
  - D. Cuando el tratamiento sea esporádico, limitado y afecte a personas de interés público.
- 6.** ¿Cuáles de los siguientes tratamientos no requieren la realización de una evaluación de impacto?
- A. Los realizados bajo las directrices de las autoridades de control.
  - B. Los necesarios para el cumplimiento de una obligación legal.
  - C. Los realizados por autónomos que ejerzan de forma individual.
  - E. Ninguno de los anteriores requiere la realización de una evaluación de impacto.
- 7.** De entre los siguientes, ¿cuál no es un principio de la privacidad por diseño?
- A. Privacidad por defecto.
  - B. Privacidad por diseño (incrustada en el diseño).
  - C. Seguridad de extremo a extremo.
  - D. Privacidad global.
- 8.** En *big data*, ¿qué principio de entre los siguientes se ve principalmente afectado por el *overcollecting*?
- A. Principio de minimización de datos.
  - B. Principio de limitación de la finalidad.
  - C. Exactitud de los datos.
  - D. Confidencialidad e integridad.

**9.** El interés legítimo puede dar base legal al tratamiento de datos personales. De entre las siguientes, indica qué afirmación al respecto es incorrecta:

- A. Si un tratamiento se basa en el interés legítimo, no requiere el consentimiento del afectado.
- B. El interés legítimo se aplica bajo la responsabilidad del responsable de tratamiento, que en su caso deberá demostrar que dicho interés no es anulado por los intereses o derechos y libertades fundamentales del interesado.
- C. El interés legítimo permite justificar cualquier tratamiento de datos de carácter personal.
- D. El derecho de información se debe garantizar, aunque la base legal del tratamiento sea el interés legítimo.

**10.** Una universidad va a realizar un tratamiento de datos biométricos a gran escala que afectará a todos los miembros de la comunidad educativa. En este contexto, ¿cuál de las siguientes afirmaciones no es correcta?

- A. Es necesario realizar una EIPD, al ser datos biométricos destinados a la identificación unívoca de las personas a gran escala.
- B. Los datos biométricos no son datos de carácter personal.
- C. Los datos biométricos pertenecen a categorías especiales de datos.
- D. No es necesario realizar una EIPD porque es una pyme.

Gobierno del Dato y Toma de Decisiones

---

# Tema 10. La disociación de datos personales y técnicas de anonimización

# Índice

## Esquema

### Ideas clave

- 10.1. Introducción y objetivos
- 10.2. Definiciones
- 10.3. La disociación y anonimización de datos
- 10.4. Técnicas de anonimización
- 10.5. K-anonimato y sus variantes
- 10.6. Herramientas de software
- 10.7. Riesgos asociados a las técnicas de anonimización
- 10.8. Principios a la hora de construir un data warehouse
- 10.9. Referencias bibliográficas

### A fondo

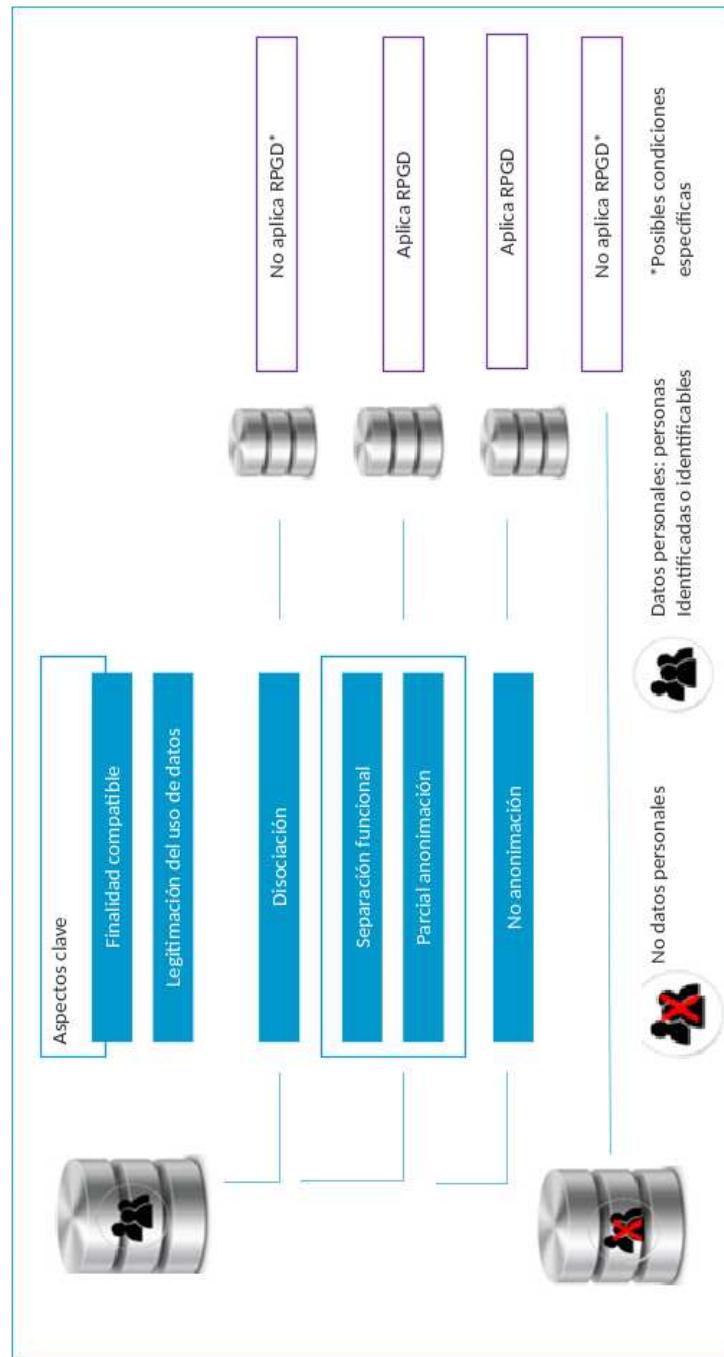
Opinion 05/2014 on anonymisation techniques

Código de buenas prácticas para la gestión de riesgos derivados de los procesos de disociación

Página web del grupo de trabajo del artículo 29

## Test

# Esquema



## 10.1. Introducción y objetivos

En este tema vamos a introducir algunas técnicas de anonimización de datos y conoceremos los riesgos de dichas técnicas y, en general, de los procesos de anonimización, frente a la reidentificación de los individuos.

Este tema nos introduce en el mundo de la anonimización de datos, como respuesta a las necesidades del tratamiento de grandes volúmenes de datos, y en cómo es en el ámbito científico y estadístico o del *marketing* y de las implicaciones en materia de protección de datos.

Como herramienta fundamental, hablaremos de la disociación de datos, proceso que facilita la obtención de conjuntos de datos totalmente anonimizados (datos disociados, a partir de datos personales que no permiten la reidentificación de los individuos y sobre los que no es aplicable la normativa de protección de datos).

Hablaremos de técnicas de anonimización, de los riesgos asociados de reidentificación sobre un conjunto de datos, parcialmente anonimizados. También introduciremos las recomendaciones del **grupo de trabajo del artículo 29** sobre los conjuntos de datos disociados y las técnicas de anonimización.

## 10.2. Definiciones

- ▶ **Procedimiento de disociación:** «todo tratamiento de datos personales de modo que la información que se obtenga no pueda asociarse a persona identificada o identifiable» (art. 3.f, LOPD 15/1999, de 13 de diciembre).
- ▶ **Seudonimización:** «el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identifiable» (art. 4, RGPD).
- ▶ **Anonimización:** «proceso por el cual deja de ser posible establecer por medios razonables el nexo entre un dato y el sujeto al que se refiere. Es aplicable también a la muestra biológica» (art. 3.c, Ley 14/2007, de 3 de julio).
- ▶ **Dato agregado:** son datos tratados de manera que los valores representan acumulados o variables estadísticas como medias o han sido agrupados en rangos de datos.
- ▶ **Microdatos:** datos a nivel de registro que no han sido objeto de tratamiento de agregación.
- ▶ **Identificadores indirectos:** variables que combinadas pueden llegar a permitir la identificación directa, pero no son consideradas identificadores directos.
- ▶ **Seudónimos:** son identificadores personales que identifican a una persona, pero diferentes de los que habitualmente utiliza esa persona.
- ▶ **Cuasiidentificadores:** un conjunto de identificadores indirectos que combinados pueden identificar individuos, pero no lo hace tomados de manera independiente.
- ▶ **Supresión:** técnica que permite obtener un conjunto k-anonimato eliminando una

celda (valor) o un registro.

- ▶ **Deidentificación:** proceso de eliminación de identificadores, cuasiidentificadores, de manera que se dificulta la identificación de los individuos a menos que se cuenten con información adicional.
- ▶ **Riesgo de singularización:** la posibilidad de extraer de un conjunto de datos algunos registros (o todos los registros) que identifican a una persona.
- ▶ **Riesgo de vinculabilidad:** la capacidad de vincular como mínimo dos registros de un único interesado o de un grupo de interesados, ya sea en la misma base de datos o en dos bases de datos distintas. Si el atacante puede determinar (por ejemplo, mediante un análisis de correlación) que dos registros están asignados al mismo grupo de personas, pero no puede singularizar a las personas en este grupo, entonces la técnica es resistente a la singularización, pero no a la vinculabilidad.
- ▶ **Riesgo de inferencia:** la posibilidad de deducir con una probabilidad significativa el valor de un atributo a partir de los valores de un conjunto de otros atributos.

## 10.3. La disociación y anonimización de datos

El RGPD, en su considerando 26, indica que no es aplicable sobre datos anónimos.

«Los principios de protección de datos no deben aplicarse a la información anónima, es decir información que no guarda relación con una persona física identificada o identificable, ni a los datos convertidos en anónimos de forma que el interesado no sea identificable, o deje de serlo. En consecuencia, el presente Reglamento no afecta al tratamiento de dicha información anónima, inclusive con fines estadísticos o de investigación» (considerando 26, RGPD).

La disociación de datos es una herramienta útil para abordar los procesos de *analytics* sin las limitaciones que incorpora la normativa en protección de datos personales, siempre y cuando se controlen razonablemente los potenciales riesgos de reidentificación.

Deberemos considerar que, en el caso de que dicho riesgo se materialice, volveremos a estar sujetos a todas las obligaciones derivadas de la normativa.

La justificación de esto nos la da propia normativa en la definición de su objeto de aplicación, es decir, si no tratamos datos personales.

El término **disociación** es un concepto definido en la Ley Orgánica de Protección de Datos 15/1999 como:

«Todo tratamiento de datos personales de modo que la información que se obtenga no pueda asociarse a persona identificada o identificable»  
(art. 3.f, LOPD, de 13 de diciembre)».

Esta definición se reitera en el Real Decreto 1720/2007: «todo tratamiento de datos personales que permita la obtención de datos disociados» (art. 5.p, RD 1720/2007, de 21 de diciembre), siendo un

dato disociado «aquél que no permite la identificación de un afectado o interesado» (art. 5.e, RD 1720/2007, de 21 de diciembre).

Por tanto, si atendemos a la aplicación de la LOPD, en su artículo 2: «la presente Ley Orgánica será de aplicación a los datos de carácter personal registrados en soporte físico, que los haga susceptibles de tratamiento y a toda modalidad de uso posterior de estos datos por los sectores público y privado» (art. 2.1, LOPD, de 13 de diciembre).

Tendremos como resultado que, **aplicando procesos de disociación, al no contar con datos que identifiquen o faciliten la identificación de los individuos, no estarían sujetos al amparo de las normas de protección de datos** y, por consiguiente, ese conjunto de datos podría tratarse sin las limitaciones que de la normativa en tratamientos de datos personales se deriva: estaríamos tratando datos disociados.

En opinión del WP29, en su «Opinion 05/2014 on Anonymisation Techniques», opinión adoptada el 10 de abril de 2014, resalta el hecho de que diferentes estudios versan sobre la dificultad de la creación de verdaderos conjuntos de datos anonimizados a partir de un conjunto de datos personales extensos, mientras se intenta mantener un conjunto de datos suficientes para el objeto de la tarea o procesamiento.

Los procesos de anonimización pueden restar información que sea necesaria para el procesamiento. Además, un conjunto de datos considerado anónimo, combinado con otros, puede originar que uno o varios individuos sean identificados.

Además, existe un factor de riesgo inherente a la anonimización: es cada vez más difícil de lograr auténticos conjuntos disociados a partir de conjuntos de datos personales por el avance de la tecnología informática moderna y la disponibilidad ubicua de la información. Hay autores que cuestionan esta opinión, considerando que los riesgos son realmente bajos.

Este factor de riesgo ha de considerarse en la evaluación de la conveniencia de cualquier técnica de anonimización (incluyendo los posibles usos de los datos que son anonimizados a través de estas técnicas de anonimización) y el potencial impacto sobre los afectados y la probabilidad de ocurrencia debe ser evaluada.

Un proceso de disociación para alcanzar una anonimización completa requerirá que se evite cualquier posibilidad razonable de establecer un vínculo con los datos de otras fuentes con el fin de reidentificar a los individuos, pero en la práctica hay unas zonas grises muy significativas, donde un responsable de tratamiento puede creer que un conjunto de datos está disociado (anonimizado), pero un tercero motivado será todavía capaz de identificar al menos algunos de los individuos de la información dada a conocer.

Gestionar y revisar periódicamente el riesgo de reidentificación, incluyendo la identificación de los riesgos residuales, es un elemento importante de cualquier sólido enfoque en esta área.

Es de reseñar que, cuando un responsable del tratamiento no elimina los datos originales (identificables) a nivel de evento y el responsable del tratamiento maneja parte de este conjunto de datos (por ejemplo, después de la eliminación o enmascaramiento de los datos de identificación), el conjunto de datos resultante todavía será de datos personales.

Para el responsable de tratamiento existe la posibilidad de proceder a la reidentificación mediante medios razonables.

Mientras que, si un tercero procesa un conjunto de datos tratados con un proceso de disociación adecuado (contará con datos anónimos y facilitados por el encargado de tratamiento original), pueden hacerlo legalmente sin necesidad de tomar en cuenta las necesidades de protección de datos, ya que no podrá (directa o indirectamente) identificar los titulares de los datos en el conjunto de datos original, que no está a su disposición.

## Anonimización parcial

No siempre será posible la disociación completa o anonimización debido a la naturaleza del procesamiento (por ejemplo, donde puede haber una necesidad de volver a identificar a los sujetos de datos o la necesidad de utilizar los datos más granulares que, como efecto secundario, puede permitir la identificación indirecta).

En estos casos, una ayuda será la de un proceso de anonimización parcial, en el que **el conjunto resultante no puede tener la consideración de anonimizado**, pero sí se habrá dificultado la identificación de los individuos.

Para ello nos asisten diversas técnicas que pueden ser aplicadas (incluyendo seudoanonimización, *keycoding*, *keyed hashing*, la eliminación de identificadores y valores atípicos, sustitución de identificadores únicos, introducción de ruido y otros). La selección de la técnica más adecuada deberá poner el foco en que el resultado sea lo más difícil posible de proceder a una identificación.

Además, al no contar con información totalmente anonimizada deberemos aplicar medidas adicionales de seguridad entre las que, en el caso de España, estarán las correspondientes al nivel de seguridad aplicable a los datos personales del fichero origen.

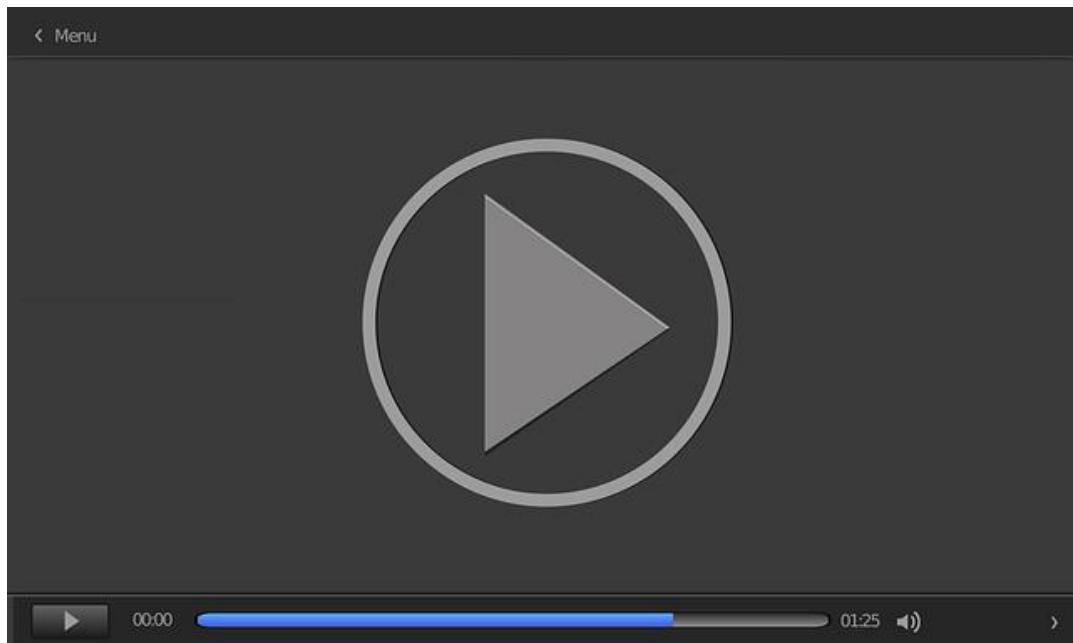
En este sentido, el WP29 recomienda la aplicación de medidas adicionales como, por ejemplo:

- ▶ La adopción de medidas adicionales de seguridad específicas (como el cifrado).
- ▶ En caso de seudoanonymización, asegurarse de que los datos que permite la vinculación de la información a un objeto de datos (las claves) que los identifica han sido codificados o encriptados y almacenados por separado.
- ▶ La incorporación de un tercero de confianza en situaciones en las que una serie de organizaciones quieran anonimizar los datos personales que tienen y se quieran utilizar en un proyecto en colaboración.
- ▶ Restringir el acceso a los datos personales aplicando el principio de la **necesidad de saber**, equilibrando cuidadosamente los beneficios de una mayor difusión contra los riesgos de la divulgación inadvertida de datos personales a personas no autorizadas. Esto puede incluir, por ejemplo, permitir acceso de solo lectura en locales controlados o la aplicación de medidas para permitir solo su acceso en entornos seguros y comunidades cerradas.

También es importante establecer obligaciones legalmente exigibles de confidencialidad a los beneficiarios de los datos, incluyendo la prohibición de la publicación de información. Siempre hay que considerar que la publicación de información tiene un alto riesgo por diferentes razones:

- ▶ Errores de divulgación involuntaria.
- ▶ Procesos de disociación erróneos que facilitan la identificación de los individuos.
- ▶ La anonimización considerada como absolutamente irreversible que potencie la publicación de datos ante la falsa sensación de seguridad.

Para acabar este apartado, accede al vídeo *Tratamiento de datos de menores*.



---

Accede al vídeo:

<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=f5ff9972-9396-4d49-95a0-abd800c24758>

---

## 10.4. Técnicas de anonimización

Existen diferentes técnicas de anonimización y en los últimos años, desde la comunidad científica, espoleados por los riesgos asociados de reidentificación vinculados a las técnicas existentes, se han ido mejorando o introduciendo nuevas técnicas.

Por lo general, el proceso de anonimización de un conjunto de datos va a requerir la aplicación conjunta de estas técnicas con objeto de reducir las debilidades que cada una de ellas presenta frente al riesgo de reidentificación.

Elegir la técnica o técnicas más adecuadas para aplicar a cada caso va a depender de aspectos relacionados con la sensibilidad de la información, la necesidad de mantener una alta correspondencia con los datos originales para el fin que se les vaya a dar y el riesgo de reidentificación comprometido.

### Clasificación de las técnicas de anonimización

Las técnicas de anonimización se pueden clasificar en función de su naturaleza, el efecto que tienen sobre los datos o en función de su aplicación.

- ▶ **Clasificación en función de su naturaleza:** existen dos familias de técnicas principales:

- Aleatorización: modifican la veracidad que tienen los datos reduciendo el vínculo de los datos y las personas reduciendo probabilidad de inferencia entre los datos.
- Generalización: su aplicación generaliza un valor en un atributo o columna o diluye los atributos (por ejemplo, reduce el código postal a los dos primeros dígitos). Reducen la significación de un individuo a partir de un valor de atributo singular.

- ▶ Clasificación en función del efecto que tienen sobre los datos:
  - Técnicas de **reducción de atributos**: se reducen valores en las tablas de datos eliminando aquellos que facilitan la reidentificación:
    - Supresión de identificadores directos.
    - Agregación: reducción del nivel de detalle de la información.
    - Muestreo: selección de datos de entre una amplia muestra.
  - Técnicas de **modificación de los datos**: los datos son alterados para reducir la posibilidad de reidentificación:
    - Generalización.
    - Adición de ruido.
    - Asignación al azar (aleatorización de los valores de los datos).
    - Permutación o intercambio de datos.
    - Privacidad diferencial.
    - Supresión de datos.
    - Seudoanonymización.
  - Métodos **basados en restricción**: se introducen restricciones en el conjunto de datos para eliminar atributos que facilitan la reidentificación:
    - Supresión de celdas.
    - Cambio del esquema de clasificación.

- **Clasificación en función de su aplicación** a microdatos o macrodatos, así tenemos:

- Técnicas para reducir los riesgos de identificación en microdatos.
- Técnicas para reducir los riesgos de identificación en macrodatos.

## Técnicas para reducir los riesgos de identificación en microdatos

Las aproximaciones habituales para reducir los riesgos de reidentificación en el ámbito de la estadística y en el tratamiento de microdatos son:

- **Técnicas que implican la reducción de atributos:** la reducción de datos contenidos en la muestra es más frecuentemente utilizada que la modificación de los datos y generalmente esto consiste en:
- **Supresión de identificadores directos:** eliminar del conjunto de datos los identificadores que de manera directa identifican a los individuos. Por ejemplo, eliminación de nombres y apellidos, DNI y número de seguridad social.
  - **Agregación y reducción del nivel de detalle de la información :** una aproximación muy común es una forma de redondeo de datos o agrupación de los mismos en torno a grandes categorías, llevando el conjunto a un riesgo aceptable de reidentificación. Para ello se reduce el número de registros que presentan atributos con una única combinación y que hace que el riesgo de reidentificación sea elevado. Por ejemplo:

- La fecha de nacimiento se reduce al año.
  - El código postal se deja en los dos primeros dígitos. Entre estas técnicas destacan: las técnicas de agregación o k-anonimato, diversidad  $\ell$  y proximidad  $t$ .
  - **Muestreo:** si el conjunto de datos tiene más registros de los necesarios, una selección muestral de registros nos puede ayudar a limitar los riesgos. Hace años era una práctica muy habitual por las limitaciones en procesamiento. La amplia superación de estas limitaciones en la era *big data* ha llevado a que esta técnica sea menos utilizada.
- **Técnicas que implican la modificación de los datos:** la modificación de datos es una aproximación más radical, pero útil a la hora de reducir la capacidad informativa de los datos. Algunas de estas técnicas son:
- **Generalización:** consiste en repetir el mismo valor en todas las filas del conjunto de datos. Para ello es frecuente escoger el atributo menos singular o específico y con sentido semántico. En la práctica, la generalización de un valor en un atributo se aproxima a la eliminación del atributo de la tabla. Ejemplo:

Edad	Sexo	Puntuación
12	M	5
12	M	7
12	H	4
12	H	8
13	M	5

Tabla 1. Ejemplo de generalización I. Fuente: elaboración propia.

Generalizamos el dato del sexo:

Edad	Sexo	Puntuación
12	Z	5
12	Z	7
12	Z	4
12	Z	8
13	Z	5

Tabla 2. Ejemplo de generalización II. Fuente: elaboración propia.

Entre las técnicas de generalización destacan las técnicas de agregación o k-anonimato, diversidad  $l$  y proximidad  $t$ , que veremos más detenidamente en el siguiente capítulo.

- ▶ **Adición de ruido a los datos:** es una técnica de aleatorización que suele ser útil cuando los atributos son altamente sensibles y pueden causar un impacto significativo sobre los individuos. Es un método que se aplica a la información numérica, consistente en la adición de ruido aleatorio a todos los valores de la columna (variable) que debe ser protegida. Y se hace de manera que perturbe lo menos posible la propia variable, de modo que la suma de los valores aplicados sea cero o tienda a cero. Es decir, en una fila suma y en otra resta manteniendo el valor de la distribución sin alteración. Para ello la varianza del valor adicionada debe tender a cero. Frente a una alta varianza que implicaría que la perturbación incluida es significativa. Supongamos que un estudio de prevalencia del cáncer en zonas geográficas requiere de alta fidelidad del dato de **localidad** de residencia del individuo o del **código postal**, por lo que es preciso mantener las cuatro cifras del código postal. Una solución sería cambiar de forma aleatoria las cinco cifras de estos datos dentro de un aceptable rango. Normalmente, los paquetes estadísticos facilitan este tipo de transformaciones. Otro ejemplo, en un estudio de ingresos y gastos personales se dispone de los siguientes datos:

Edad	Género	Código postal	Ingresos	Gastos/mes
24	F	28001	20 000	1100
34	M	28001	32 000	1800
45	F	28004	68 000	2800
32	F	28056	31 500	2000
28	F	28044	22 000	1300
73	M	28024	28 000	1200

Tabla 3. Ejemplo de adición de ruido I. Fuente: elaboración propia.

Si seleccionamos añadir **ruido** de modo que la suma de los valores introducidos sea cero y la variancia sea solo de 1, la tabla quedará como sigue:

Edad	Género	Código postal	Ingresos	Gastos/mes
24	F	28001	19 828	1100
34	M	28001	32 960	1800
45	F	28004	66 951	2800
32	F	28056	28 957	2000
28	F	28044	23 862	1300
73	M	28024	28 942	1200

Tabla 4. Ejemplo de adición de ruido II. Fuente: elaboración propia.

Hemos introducido ruido sobre la variable «ingresos», de manera que hemos restado información que facilite la reidentificación de los individuos.

La adición de ruido también es una técnica muy utilizada en las comunicaciones, de manera que estas se distorsionan y el observador accede a la comunicación con ruido añadido, lo que dificulta su análisis. La parte receptora reinvierte el algoritmo generador de ruido accediendo a la información.

Es una técnica que ofrece buenos resultados, pero normalmente debe aplicarse junto a otras como la eliminación de cuasiidentificadores y de atributos obvios; de otro modo nos vemos obligados a incorporar una distorsión importante para alcanzar los objetivos de anonimización deseados.

Muchos de los riesgos en la utilización de la técnica sobre la reidentificación vienen por la no adición de ruido suficiente o porque este no es consistente, orientando al adversario hacia el algoritmo de adición de ruido. La clave es que el adversario piense que se trata de valores ciertos sin serlo.

- ▶ **Asignación al azar: aleatorización de los valores de los datos :** esta técnica se utiliza con profusión en los entornos de pruebas de desarrollo de *software*. Por ejemplo, identidades de personas (de prueba) para los que el nombre y los apellidos se obtienen aleatoriamente a partir de un conjunto de nombre y apellidos reales.
- ▶ Los nombres se asignan con la misma distribución de la realidad, sin incluir aquellos extremos de frecuencia menor (los menos comunes). También se puede aplicar a números de identificación: DNI, número de la seguridad social. Existen programas que ayudan a realizar estas funciones generando también los códigos de control habituales de estos identificadores.

- **Permutación intercambio de datos:** es una técnica que puede ser considerada como de adición de ruido. Para su aplicación se identifican parejas de registros correspondientes a individuos con ciertas similitudes y se intercambian los identificadores quedando vinculados a distintos interesados. El conjunto final de datos está alterado frente al conjunto inicial, pero para la realización de análisis de datos ofrecen los mismos resultados. Tiene un uso extendido para la elaboración de estadísticas de datos agregados, pero es poco útil en análisis de regresiones multivariadas.
- El objetivo, en definitiva, es el de introducir cierta confusión ante un posible ataque y habitualmente los registros afectados dentro de los conjuntos de datos son del orden del 1 al 10 %.

Edad	Género	Código postal	Ingresos	Gastos/mes
24	F	28001	20 000	1100
34	M	28001	32 000	1800
45	F	28004	68 000	2800
32	F	28056	31 500	2000
28	F	28044	22 000	1300
73	M	28024	28 000	1200

Tabla 5. Ejemplo de permutación I. Fuente: elaboración propia.

La aplicación del método en este ejemplo nos permitiría intercambiar los valores de los datos de edad 28 y 24.

Edad	Género	Código postal	Ingresos	Gastos/mes
24	F	28044	19 828	1100
34	M	28001	32 960	1800
45	F	28004	66 951	2800
32	F	28056	28 957	2000
28	F	28001	23 862	1300
73	M	28024	28 942	1200

Tabla 6. Ejemplo de permutación II. Fuente: elaboración propia.

En este ejemplo intercambiamos las variables de código postal, ingresos y gastos, de manera que, en el caso de realizar un análisis por grupos de edad, el resultado no se ve alterado, pero sí hemos generado una limitación para la identificación del individuo, pues la edad no se correspondería ni el código postal.

Vemos que la permutación no altera los datos estadísticos de la columna sobre la que se realiza. En este caso mantiene la media aritmética pero también mantiene la moda, la mediana y la desviación típica. Por eso es una técnica fácil de aplicar que no distorsiona los análisis y sí introduce una dificultad de reidentificar al individuo al que corresponden los datos. Ahora bien, se debe analizar su aplicación para que sea más efectiva. Veamos otro ejemplo.

En una empresa se está realizando un estudio de ingresos por categoría profesional, sexo, edad y estudios.

Año	Sexo	Cargo	Ingresos
1998	M	Ingeniero	40 000
1997	M	Médico	45 000
2005	M	Operador	15 000
1965	M	Gerente	70 000
1957	M	Director ejecutivo	200 000
		<b>Media</b>	72 000

Tabla 7. Ejemplo de permutación III. Fuente: elaboración propia.

Con el objeto de anonimizar los datos se aplica una permutación a los datos salariales con la pretensión de que no se pueda asociar salarios a las personas.

Año	Sexo	Cargo	Ingresos
1998	M	Ingeniero	40 000
1997	M	Médico	45 000
2005	M	Operador	200 000
1965	M	Gerente	70 000
1957	M	Director ejecutivo	15 000
		<b>Media</b>	72 000

Tabla 8. Ejemplo de permutación IV. Fuente: elaboración propia.

Ahora, en el ejemplo propuesto vemos que la permutación al mantener el cargo tiene un efecto limitado, pues no es difícil inferir que el director ejecutivo y operador de las categorías existentes serán los que ocupen los extremos salariales de la muestra y, por consiguiente, es fácil determinar cuánto gana cada uno. En este caso, la permutación no es válida. En general, nos ofrece malos resultados cuando existe una fuerte correlación entre dos atributos (categoría laboral y salario).

- ▶ **Privacidad diferencial:** es una técnica que genera la alteración de los datos y pertenece a la familia de técnicas de aleatorización. Consiste en la adición de ruido selectivo a la salida de los datos, generando vistas anonimizadas del conjunto de los datos. Diferente en este sentido a las técnicas anteriores que se aplican directamente sobre el conjunto de datos base para su anonimización. En este caso, no se produce alteración de los datos, solo se muestran alterados. Tienen amplio predicamento en bancos de datos que facilitan acceso a terceros. La técnica facilita que, mediante la asignación de ruido, la misma consulta presente diferentes datos para distintos operadores. La debilidad que presenta es la potencial revelación del algoritmo de adición de ruido mediante el análisis de un volumen significativo de búsquedas. Esto se puede mitigar limitando las consultas posibles. El resto de las debilidades que presenta viene más de la deficiente aplicación de la técnica que por la naturaleza de la misma. Por ejemplo: poco ruido.

- **Técnicas que implican la supresión de datos:** cuando las técnicas de reducción o modificación no ofrecen un resultado óptimo sobre el riesgo residual de reidentificación, se pueden aplicar técnicas de supresión: eliminar los registros que ofrecen un mayor riesgo en la reidentificación de los individuos. Obviamente, esto puede introducir distorsión en los estudios, pues la supresión no es aleatoria al eliminar la fila completa del registro problemático.
- Calificaciones de matemáticas de los alumnos de la Comunidad de Madrid. Rendimiento académico de los niños:

Edad	Sexo	Puntuación
12	M	5
12	M	7
12	H	4
12	H	8
13	M	5

Tabla 9. Ejemplo de supresión I. Fuente: elaboración propia.

Se suprime el registro que frente al resto tiene mayor riesgo de reidentificación:

Edad	Sexo	Puntuación
12	M	5
12	M	7
12	H	4
12	H	8

Tabla 10. Ejemplo de supresión II. Fuente: elaboración propia.

- ▶ **Técnicas de pseudoanonymización:** es una técnica de desidentificación de un conjunto de datos, pero facilita la reidentificación de los mismos. El seudónimo permite vincular todos los datos relacionados, sin facilitar la identificación del individuo de una forma directa. Es una técnica muy extendida, existiendo aplicaciones comerciales que facilitan su aplicación.

Esta técnica reduce la vinculabilidad que podemos tener entre los datos y la identidad del individuo, pero no la elimina e incluso en algunos casos se mantiene vía un proceso reversible.

Los seudónimos pueden ser de dos tipos, reversibles o irreversibles, y se suelen obtener de diferentes maneras, por ejemplo:

- ▶ La pseudoanonymización reversible, que se puede obtener por:
  - Vía codificación del identificador: con una técnica de codificación/criptación reversible.
  - Estableciendo tablas de correspondencia entre identificador seudónimo: en este caso, por ejemplo, se asigna a cada identificador un seudónimo (números, caracteres), manteniendo en un conjunto separado de datos la correspondencia entre el seudónimo y los identificadores reemplazados.
- ▶ La pseudoanonymización irreversible, que se puede obtener mediante función *hash*: es una función que siempre devuelve una cadena de igual tamaño, independientemente del valor que se pase a la función. Es irreversible y vulnerable a la hora de poder obtener el identificador que seudoanimiza. Por ejemplo, si en un sistema hemos aplicado la función *hash* sobre el DNI.

La seudonimización en el ámbito científico o empresarial habitualmente es abordada con diferentes aproximaciones que limitan el riesgo de reidentificación:

- ▶ Los seudónimos y los identificadores son mantenidos dentro de la organización, limitándose el acceso a las claves criptográficas o las tablas de correspondencia para facilitar el anonimato de los datos.
- ▶ La correspondencia de seudónimos e identificadores son facilitados para su custodia a un tercero, ofreciendo en caso de necesidad la posibilidad de reidentificar el seudónimo. Esto es habitual en la investigación sanitaria.

Existe una especificación técnica para la seudonimización recogida en la ISO/TS 25237 (Health informatics. Pseudonymisation) y desarrollada para el ámbito sanitario, pero que es extensible a otros ámbitos de la estadística o investigación.

## Técnicas para reducir los riesgos de identificación en macrodatos

Las técnicas más extendidas en este ámbito son:

- ▶ Métodos basados en **restricciones**. Los más comunes son los siguientes:
- ▶ **Supresión de celdas**: consiste directamente en la eliminación de las filas que contienen celdas que son problemáticas. En su estudio:

Año	Sexo	Cargo	Ingresos
1998	M	Ingeniero	40 000
1997	M	Médico	45 000
2005	M	Desempleado	5 000
1965	M	Gerente	70 000
1957	M	Director ejecutivo	200 000

Tabla 11. Supresión de celdas. Fuente: elaboración propia.

- ▶ **Cambio del esquema de clasificación**: se puede realizar modificando para cada categoría los puntos de corte (es decir, los extremos de la categoría) o mediante la agregación de celdas. Una aproximación frecuente es la denominada codificación de arriba a abajo. Por ejemplo, en un conjunto de datos podemos agrupar los extremos de las edades, agrupando todas las edades en torno a un valor o bajo un determinado valor. El objetivo es eliminar los grupos menores.

Veamos un ejemplo: se está procediendo a realizar un estudio en el ámbito sanitario; junto a la información requerida del estudio se incluye el género, la edad y la patología, los datos que se tienen.

sexo	edad	Enf Tiroídes	sexo	edad	Enf Tiroídes
M	11	Fibromialgia	M	44	Cáncer de colon
M	12	Lupus	M	44	Lupus
M	12	Neumonia	M	44	Lupus
H	12	Lupus	H	44	Cáncer de colon
H	12	Enf Tiroídes	H	44	Lupus
M	13	Cáncer de colon	M	45	Enf Tiroídes
M	13	Fibromialgia	M	45	Cáncer de colon
H	14	Cáncer de colon	H	45	Enf Tiroídes
H	14	Enf Tiroídes	M	46	Lupus
H	14	Enf Tiroídes	H	46	Cáncer de colon
M	15	Hipertensión	M	47	Enf Tiroídes
M	15	Lupus	H	47	Lupus
M	15	Neumonia	H	52	Cáncer de colon
M	17	Enf Tiroídes	H	52	Hipertensión
H	17	Cáncer de colon	M	53	Enf Tiroídes
M	18	Hipertensión	H	53	Enf Tiroídes
H	18	Enf Tiroídes	H	53	Hipertensión
M	19	Cáncer de colon	H	53	Hipertensión
M	19	Enf Tiroídes	M	54	Cáncer de colon
H	19	Lupus	M	54	Enf Tiroídes
H	19	Enf Tiroídes	M	54	Enf Tiroídes
H	23	Enf Tiroídes	M	54	Enf Tiroídes
H	23	Enf Tiroídes	M	55	Enf Tiroídes
H	24	Enf Tiroídes	M	55	Enf Tiroídes
H	24	Cáncer de colon	H	55	Lupus
H	24	Enf Tiroídes	H	55	Hipertensión
H	25	Enf Tiroídes	M	56	Cáncer de colon
M	26	Cáncer de colon	M	57	Enf Tiroídes
M	27	Fibromialgia	H	59	Lupus
M	27	Lupus	H	91	Hipertensión
M	27	Cáncer de colon	H	99	Hipertensión
M	27	Lupus	M	39	Cáncer de colon
M	28	Lupus	M	39	Enf Tiroídes
M	28	Enf Tiroídes	M	39	Cáncer de colon
M	28	Enf Tiroídes	M	39	Lupus
H	30	Enf Tiroídes	M	39	Enf Tiroídes
M	31	Fibromialgia	M	39	Cáncer de colon
M	32	Fibromialgia	H	39	Enf Tiroídes
M	32	Fibromialgia	H	39	Cáncer de colon
M	32	Fibromialgia	H	39	Lupus
M	32	Fibromialgia	H	39	Enf Tiroídes
M	32	Lupus	H	39	Cáncer de colon
M	32	Lupus	H	39	Lupus
M	33	Fibromialgia	M	40	Lupus
M	33	Fibromialgia	M	40	Enf Tiroídes
M	33	Enf Tiroídes	M	40	Cáncer de colon
M	33	Cáncer de colon	M	42	Enf Tiroídes
M	33	Lupus	H	42	Lupus
M	33	Enf Tiroídes	M	43	Cáncer de colon
M	33	Enf Tiroídes	M	43	Lupus
M	33	Enf Tiroídes	M	43	Enf Tiroídes
M	33	Enf Tiroídes	H	43	Cáncer de colon
M	33	Enf Tiroídes	H	43	Enf Tiroídes
M	33	Cáncer de colon	H	37	Lupus
M	33	Enf Tiroídes	H	37	Enf Tiroídes
M	33	Cáncer de colon	H	37	Cáncer de colon
M	33	Lupus	H	37	Lupus
M	33	Enf Tiroídes	H	37	Enf Tiroídes
M	33	Enf Tiroídes	M	38	Cáncer de colon
M	33	Cáncer de colon	H	38	Lupus
M	33	Cáncer de colon	M	39	Enf Tiroídes
M	34	Neumonia	M	36	Cáncer de colon
M	34	Hipertensión	H	36	Enf Tiroídes
M	34	Lupus	M	37	Lupus
M	35	Enf Tiroídes	M	37	Enf Tiroídes
M	35	Cáncer de colon	H	37	Enf Tiroídes
M	35	Fibromialgia	H	37	Cáncer de colon
M	35	Enf Tiroídes			

Tabla 12. Ejemplo conjunto de datos I.

Con el objeto de identificar el riesgo de reidentificación, se realiza una clasificación de los registros, para lo cual se establece unas categorías por intervalos de fechas. El resultado referente al número de registros que obtenemos para cada categoría de

edad para hombres y mujeres nos muestra que hay una categoría para la que solamente tendríamos un registro. Lo que generaría que para este registro único tengamos un riesgo de reidentificación más elevado que si en la misma categoría coincidieran números registros.

	Menos de 25 años	De 26 a 35 años	De 36 a 45 años	Más de 45 años	Total
Mujeres	12	40	23	11	86
Hombres	15	1	21	12	49
Total	27	41	44	21	135

Tabla 13. Ejemplo cambio en la clasificación I. Fuente: elaboración propia.

Para reducir esto podemos modificar el intervalo de las categorías, por ejemplo:

	Menos de 20 años	De 21 a 35 años	De 36 a 45 años	Más de 45 años	Total
Mujeres	12	40	23	11	86
Hombres	9	7	21	12	49
Total	21	47	44	21	135

Tabla 14. Ejemplo cambio en la clasificación II. Fuente: elaboración propia.

Esta nueva categorización lleva a que la menor categoría de datos sea de 7 (7 registros). Al considerar que es un número insuficiente podemos eliminar una categoría uniéndola a otra:

	Hasta 35 años	De 36 a 45 años	Más de 45 años	Total
Mujeres	52	23	11	86
Hombres	16	21	12	49
Total	68	44	21	135

Tabla 15. Ejemplo cambio en la clasificación III. Fuente: elaboración propia.

Es evidente que la elección de los intervalos más adecuados o incluso la agrupación de los mismos es un proceso laborioso que siempre tendrá que tener en consideración la finalidad del estudio que se realiza. Existen herramientas que ayudan a realizar este proceso. Pero en cualquier caso tendrá un componente de análisis previo muy importante.

En nuestro ejemplo y tras aplicar sobre los datos el análisis realizado obtenemos:

Tabla 16. Ejemplo conjunto de datos II.

- ▶ **Heurística:** a veces se toman decisiones con el objeto de reducir el riesgo, a partir de la experiencia y cierto sentido lógico.
  - ▶ **Métodos basados en permutaciones:** coincidente con la aproximación de intercambio de atributos entre registros similares, pero en este caso actuamos sobre

los macrodatos.

## 10.5. K-anonimato y sus variantes

Técnica de anonimización de datos desarrollada por Latanya Sweeney (profesora de gobierno y tecnología en la Universidad de Harvard y directora del Data Privacy Lab) en un artículo publicado en el 2002: *k-Anonymity: a model for protecting privacy*. Sweeney tiene patentado varios procedimientos y programas para generar conjuntos k-anónimos.

Un conjunto de datos es k-anónimo frente al conjunto de datos representado si la información contenida para cada individuo no puede distinguirse al menos de otros  $k - 1$  individuos cuya información también aparece en el conjunto de datos. Será 4-anónimo si para cada atributo del individuo que está en la muestra hay al menos otros 3 individuos que tienen el mismo atributo.

Pongamos un ejemplo:  $k = 3$ .

Seudoidentificador	Edad	Sexo	Diagnóstico
1	0 a 10	M	Lupus
2	20 a 35	F	Lupus
3	0 a 10	M	Lupus
4	51 a 65	F	Lupus
5	20 a 35	M	Lupus
6	51 a 65	F	Lupus
7	0 a 10	M	Lupus
8	20 a 35	F	Lupus
9	51 a 65	F	Lupus
10	0 a 10	F	Lupus
11	20 a 35	M	Lupus
12	51 a 65	M	Lupus
13	0 a 10	M	Lupus
14	0 a 10	F	Lupus
15	20 a 35	M	Lupus
16	51 a 65	M	Lupus
17	0 a 10	F	Lupus
18	51 a 65	M	Lupus
19	0 a 10	F	Lupus
20	20 a 35	F	Lupus
21	20 a 35	F	Lupus
22	0 a 10	M	Lupus

Tabla 17. Ejemplo k-anonimato I. Fuente: elaboración propia.

En el ejemplo hemos marcado la clase de coincidencia de edad, sexo y diagnóstico, y vemos que como mínimo cada clase contiene 3 individuos, por lo que la muestra podemos decir que es 3-anónimo.

Ordenemos el conjunto de datos para que se vea más fácilmente:

Seudoidentificador	Edad	Sexo	Diagnóstico
10	0 a 10	F	Lupus
14	0 a 10	F	Lupus
17	0 a 10	F	Lupus
19	0 a 10	F	Lupus
1	0 a 10	M	Lupus
3	0 a 10	M	Lupus
7	0 a 10	M	Lupus
13	0 a 10	M	Lupus
22	0 a 10	M	Lupus
2	20 a 35	F	Lupus
8	20 a 35	F	Lupus
20	20 a 35	F	Lupus
21	20 a 35	F	Lupus
5	20 a 35	M	Lupus
11	20 a 35	M	Lupus
15	20 a 35	M	Lupus
4	51 a 65	F	Lupus
6	51 a 65	F	Lupus
9	51 a 65	F	Lupus
12	51 a 65	M	Lupus
16	51 a 65	M	Lupus
18	51 a 65	M	Lupus

Tabla 18. Ejemplo k-anonimato II. Fuente: elaboración propia.

Los errores habituales a la hora de aplicar la técnica vienen dados fundamentalmente por un número  $k$  insuficiente. Cuanto más alto es el valor  $k$ , menos riesgos de inferencia existen.

k-anonimato no impide los ataques de inferencia, pero el modo en que se aplique

puede reducir esta debilidad. Por el contrario, tienen un comportamiento muy satisfactorio frente a los ataques de singularización, dado que al menos  $k$  elementos compartirán los mismos atributos.

El riesgo de vinculabilidad es escaso y, en el caso de que esta tenga lugar, lo hará con  $k$  registros, por lo que la probabilidad es de  $1/k$ .

## I-Diversity

Una mejora sobre k-anonimato es I-Diversity, que consiste en aplicar el método k-anonimato, pero garantizando que cada clase tiene como mínimo  $l$  valores, de manera que se evita que existan conjuntos de datos que tengan una variabilidad escasa de atributos.

En el ejemplo anterior, estaríamos ante una situación 1-Diversity, pues no existe variabilidad del atributo de enfermedad. Es decir, si se compromete la reidentificación, sería evidente concluir que el individuo singularizado padece lupus. Habrá que distribuir el conjunto de datos de manera que al menos en cada clase k-anonimato se disponga de valores diferentes en el atributo objetivo (en este caso, enfermedad). De este modo, mejora el comportamiento del conjunto de datos sobre potenciales ataques de inferencia.

Si tuviéramos una mayor diversidad en los diagnósticos en cada clase edad/sexo, sería más difícil concluir el diagnóstico. Este es un ejemplo de 4-anonimato y 2-diversidad (como mínimo cada clase tiene dos valores diferentes en diagnóstico).

Seudoidentificador	Edad	Sexo	Diagnóstico
10	0 a 10	F	Lupus
14	0 a 10	F	Lupus
17	0 a 10	F	ELA
19	0 a 10	F	ELA
1	0 a 10	M	Lupus
3	0 a 10	M	Lupus
7	0 a 10	M	Lupus
13	0 a 10	M	ELA
22	0 a 10	M	ELA
2	20 a 35	F	Lupus
8	20 a 35	F	Lupus
20	20 a 35	F	ELA
21	20 a 35	F	ELA
5	20 a 35	M	Lupus
11	20 a 35	M	Lupus
15	20 a 35	M	ELA
4	51 a 65	F	Lupus
6	51 a 65	F	Lupus
9	51 a 65	F	ELA
12	51 a 65	M	Lupus
16	51 a 65	M	Lupus
18	51 a 65	M	ELA

Tabla 19. Ejemplo I-Diversity. Fuente: elaboración propia.

### t-Closeness

Es un perfeccionamiento de I-Diversity, consistente en la creación de clases de equivalencia de manera que la distribución de los atributos en toda la muestra no sea

alterada.

No exige que en cada clase existan / valores diferentes, sino que, además, la distribución de los valores en el atributo en el conjunto total de datos no debe verse alterada. Se aplica sobre todo cuando es muy importante para el estudio que se vaya a realizar con los datos mantener el atributo lo más próximo a la realidad, es decir, a los datos originales.

## 10.6. Herramientas de software

El proceso de reducción del riesgo es un tanto artesanal, pero existen herramientas que ayudan en este proceso. Podemos clasificar estas herramientas según su orientación:

- ▶ Manejo de identificadores directos a nivel de microdatos, herramientas que eliminan o enmascaran identificadores directos:
- ▶ Reducción del riesgo de reidentificación a partir de identificadores indirectos a nivel de microdatos:
- ▶ Manejo de datos agregados:

## 10.7. Riesgos asociados a las técnicas de anonimización

### Antecedentes

Se ha demostrado que las técnicas de deidentificación para la anonimización de datos son insuficientes a la hora de garantizar la confidencialidad, pues la reidentificación de los datos es posible.

De hecho, Paul Ohm (profesor asociado de la University of Colorado Law School), entre otros científicos, defiende como axioma que es incompatible disponer de datos utilizables y al mismo tiempo que estos datos estén perfectamente disociados, teniendo en consideración los avances realizados en la investigación en el ámbito de la anonimización de datos.

De hecho, no hay dudas de que, a partir de datos que no necesariamente de forma directa identifican a los individuos, es posible identificar a los mismos.

Tras exponer algunos casos de la literatura científica, veremos con más claridad los riesgos que se ciernen sobre las iniciativas de datos abiertos y lo importante que es analizarla en cada caso antes de proceder a facilitar los datos al público o a los grupos de investigación.

Veamos algunos casos significativos que fortalecen esta convicción.

## El caso AOL

En agosto de 2006 AOL anunció una nueva iniciativa llamada *AOL Research*, sumándose a las iniciativas de abrir los datos a la comunidad científica. Publicaron 20 millones de búsquedas realizadas por 650 000 usuarios en el buscador de AOL, correspondientes a tres meses de actividad. Esta fue una iniciativa ampliamente aplaudida por la comunidad de investigadores del comportamiento social en Internet.

Antes de hacer pública esta información, AOL trató de anonimizar los datos con objeto de proteger la privacidad de los usuarios y para ello había suprimido los datos más obvios de identificación como, por ejemplo: nombre de usuario y dirección IP. El fin era facilitar a los investigadores poder correlacionar búsqueda necesaria para el análisis de comportamiento y procedieron a sustituir estos datos por una cadena de números.

El análisis de la información permitió la identificación de individuos, lo cual supuso un gran escándalo en EE. UU. y la dimisión del CTO. Fue silenciada la iniciativa *AOL Research*.

## El caso de código postal, fecha de nacimiento y sexo

En Massachusetts, una agencia gubernamental llamada Comisión Insurance Group (GIC), que años antes había comprado un seguro de salud para los empleados del estado, decidió (en la década de los 90) liberar registros de visitas al hospital para cualquier investigador que los solicitara sin coste alguno.

Para anonimizar los datos se habían eliminado los campos que contienen el nombre, dirección, número de la seguridad social y otros identificadores. GIC asumía que con estos se garantizaba la privacidad de los pacientes, a pesar de que en los datos se incluían prácticamente un centenar de atributos por paciente. Junto a la visita al hospital se mantuvieron el código postal, la fecha de nacimiento y el sexo.

Cuando se liberaron los datos se garantizaba por las instituciones que estos estaban perfectamente anonimizados y la privacidad de los empleados públicos estaba garantizada. Una estudiante sabía que el gobernador residía en un municipio del estado denominado Cambridge. Cambridge entonces era una ciudad de más de 50 000 habitantes y con 7 códigos postales. Latanya Sweeney, que así se llamaba la estudiante, compró por solo 20 USD el censo electoral de la ciudad de Cambridge, que contenía, entre otros, la información del nombre, dirección, código postal, fecha de nacimiento y sexo de cada votante.

Cambiando los datos del censo electoral y los datos facilitados por el GIC no le resultó difícil acceder al conjunto de visitas hospitalarias y diagnósticos que había publicado GIC, referidos al gobernador del estado. En la ciudad solo seis personas compartían fecha de nacimiento, de ellos solo tres eran hombres y solo uno vivía en el código postal donde residía el gobernador. Sweeney envió registros de salud del gobernador (incluyendo diagnósticos y prescripciones) a su despacho. Para ello solo necesitó cruzar una fuente pública con datos de identificación, con una fuente de datos teóricamente anonimizada.

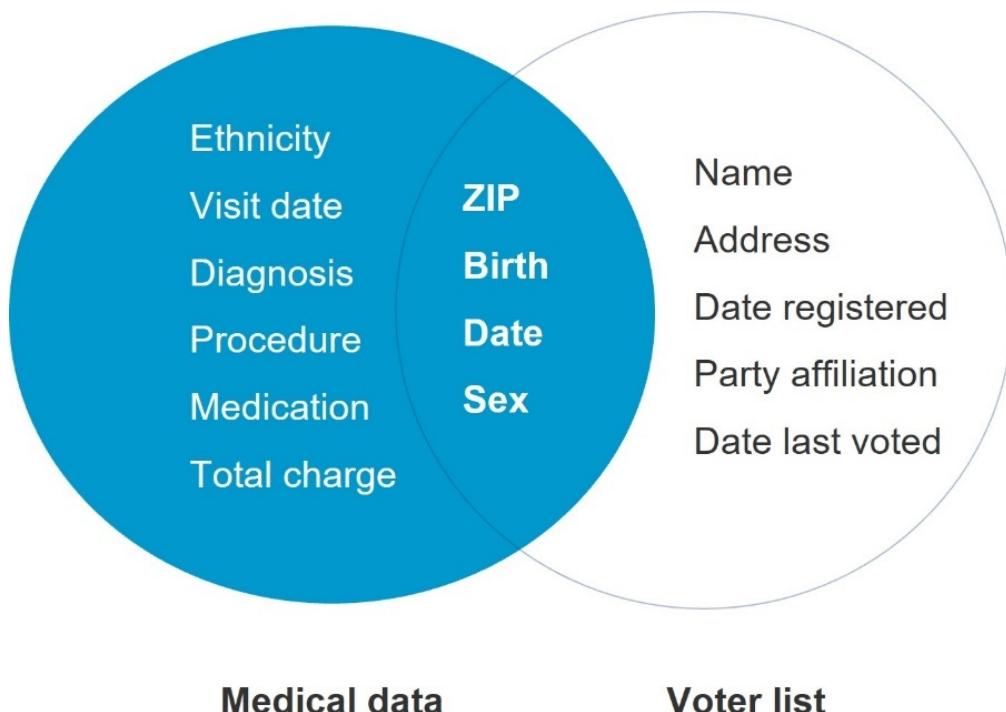


Figura 1. Cruce de datos. Fuente: elaboración propia.

Latanya Sweeney, profesora de Ciencias de la Computación, Tecnología y Política en la Universidad Carnegie Mellon, analizó el censo de los EE. UU. de 1996. En el año 2000 presentó un estudio para demostrar que en los EE. UU., simplemente con tres datos aparentemente inofensivos desde el punto de vista de su capacidad de identificar a las personas: código postal completo y la fecha nacimiento con el año incluido y el género o sexo, era posible identificar únicamente al 87 % de la población de los EE. UU.

Para ello solo necesitó cruzar una fuente pública con datos de identificación con una fuente de datos teóricamente anonimizada.

Es decir, en un 87 % de los casos cualquier entidad u organización que tuviera estos datos podría identificar al individuo a partir de los datos del censo. Pero iba más lejos, el 57 % de la población era identificable con datos menos específicos como ciudad, fecha de nacimiento y sexo. Y hasta el 18 % con los datos de estado de residencia, fecha de nacimiento y sexo.

Los datos obtenidos de su estudio fueron:

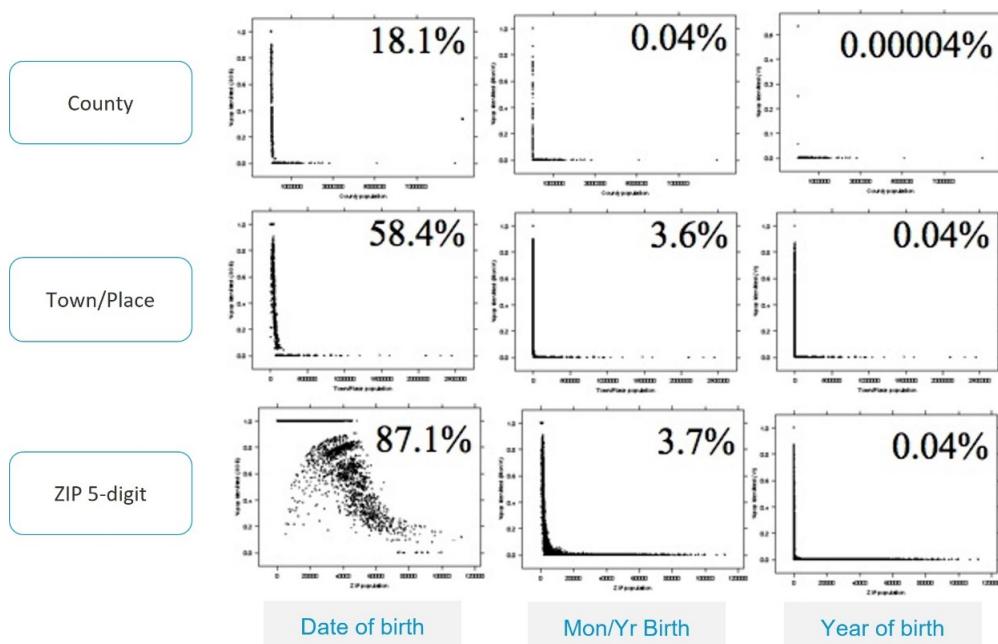


Figura 2. Resultados de estudios. Fuente: Sweeney, 2019.

Más tarde, Philippe Golle revisó el estudio de Latanya. Recalculó las estadísticas sobre el censo del año 2000 y contó solo con el código postal, la fecha de nacimiento y el sexo. La probabilidad de reidentificación era del 63 %; para el censo de 1996 (base del estudio de Latanya) era del 61 %.

	5 digit ZIP code	County
Year of birth	0,2 %	0,0 %
Year and month of birth	4,2 %	0,2 %
Year, month and day of birth	63,3 %	14,8 %

Tabla 20. Fraction of the US population uniquely identifiable by gender, location, date of birth.

Detengámonos un momento en estos datos perfectamente extrapolables a cualquier región que disponga de un censo que incorpore, junto a los datos de identificación de los individuos, los tres datos referidos (no en los porcentajes, pero sí en la posibilidad de reidentificación, pues un componente que tiene un impacto significativo es el grado de dispersión de la población).

CP	Fecha de nacimiento	Sexo	Enfermedad
28001	12/01/2005	M	Hepatitis C
28002	14/03/1987	H	Lupus
28005	25/05/1936	M	Fibromialgia
28004	22/01/1950	H	Hipertensión
28001	11/07/1947	M	Cáncer de colon

Tabla 21. Ejemplo conjunto datos (CP, nacimiento y sexo). Fuente: elaboración propia.

Si pensamos en el ámbito hospitalario y en un conjunto de datos deidentificado, tenemos que, lejos de considerar este conjunto de datos anonimizado, estaremos pensado en cuál es la probabilidad de identificar a un individuo, concluyendo:

- ▶ Que la desidentificación de los datos no origina un conjunto de datos anonimizado.
- ▶ Que es preciso aplicar técnicas adicionales para reducir o eliminar este riesgo.

## El caso Netflix Prize

El 2 de octubre de 2006 Netflix, que en aquel momento era el más importante operador de alquiler de películas *online*, publicó un millón de documentos que revelaban la calificación de las películas realizadas por medio millón de clientes entre 1999 y 2005.

Cada registro contenía la siguiente información: la película, la puntuación asignada (de 0 a 5 estrellas) y la fecha de la valoración.

Antes de publicar los datos habían procedido a anonimizarlos, manteniendo exclusivamente un identificador para poder enlazar las valoraciones de un mismo usuario. Todo el resto de información había sido eliminada, por lo que en Netflix consideraron que el conjunto de los datos garantizaba la privacidad de los usuarios.

La finalidad de los datos era facilitar a los investigadores que participaban en un concurso que tenía por objeto definir un algoritmo de recomendaciones de películas para los clientes de Netflix para facilitar en su página una mejora de la experiencia del usuario con información adicional de recomendaciones del tipo «a quienes les ha gustado una película». Por ejemplo, a quienes les ha gustado *Minority Report* también les ha gustado *Matrix* (algo que vemos con frecuencia en los grandes portales de venta de productos o servicios). Los premios ofrecidos eran importantes,

por lo que captó el interés de la comunidad científica y estudiantil.

Un grupo de investigadores de la Universidad de Tejas liderado por Arvind Narayanan y el Profesor Vitaly Shmatikov identificaron que un atacante que solo conociera un poco sobre un suscriptor individual podría identificar fácilmente el registro de ese suscriptor si este estaba presente en el conjunto de datos y sus valoraciones, o al menos identificar un conjunto de datos dentro del cual estarían las recomendaciones de ese suscriptor.

El trabajo de investigación que publicaron incluía innumerables ejemplos de lo fácil que era identificar a los subscriptores. Para ello solo sería necesario conocer sus puntuaciones para seis películas no incluidas en el *top 5*.

Para mostrar estos resultados abstractos en ejemplos concretos, Narayanan y Shmatikov compararon los datos de calificación de Netflix obtenidos con datos de bases de datos de películas similares disponibles en Internet que publican también valoraciones (Internet Movie Database), lo que facilitaba la identificación de los subscriptores. En este contexto era más fácil identificar a aquellos cuyos gustos eran más singulares.

Posteriormente, la compañía lanzó un segundo concurso que incluía datos demográficos y de comportamiento con información como: edad, género, código postal, las calificaciones realizadas y las películas elegidas. A finales de 2009 los clientes de Netflix formalizaron una demanda colectiva contra la empresa por violación de su privacidad, lo que llevó a la desestimación del segundo concurso.

## Técnicas de reidentificación de datos

Los casos vistos como ejemplo tienen un denominador común. En todos los casos la reidentificación de las personas se produjo gracias a contar con información adicional: vinculación de dos conjuntos de datos. Es posible vincular datos incluso en los casos en los que se eliminan los datos que identifican directamente y aquellos que podemos considerar cuasiidentificadores. Donde se encuentra un importante caladero de información es en Internet y las redes sociales.

En un proceso de reidentificación, tendremos:

- ▶ **Adversario:** los modelos de anonimización que se utilizan en el ámbito de la ciencia computacional se nutren, entre otros, de la teoría de juegos. Es un agente motivado, es decir, suficientemente interesado por la reidentificación, y actúa con base en el principio de esfuerzo-beneficio.
- ▶ **Información externa:** siempre que el adversario accede a una fuente de información, busca información externa con la que puede vincular la información.
- ▶ **Cruce de datos (*inner-joins*):** se cruza la información de diferentes fuentes de datos, por lo que se incrementan los datos disponibles y las posibilidades de reidentificación. Se trata de conectar filas de un conjunto de datos con filas de otro conjunto de datos, casando datos que comparten o coinciden en los dos conjuntos de datos.

Por ejemplo, veamos este conjunto de datos:

CP	Fecha de nacimiento	Sexo	Enfermedad
28001	12/01/2005	M	Hepatitis C
28002	25/05/1936	H	Lupus
28005	25/05/1936	M	Fibromialgia
28004	22/01/1950	H	Hipertensión
28001	11/07/1947	M	Cáncer de colon

Tabla 22. Ejemplo reidentificación I. Fuente: elaboración propia.

Estarían deidentificados, pero si disponemos de un conjunto de datos como el siguiente:

Nombre	Dirección	Fecha de nacimiento
Juan	xxxxxxxxxx	13/12/1955
Pedro	xxxxxxxxxx	12/02/1955
Daniel	xxxxxxxxxx	10/01/1955
María	xxxxxxxxxx	25/05/1936
Jorge	xxxxxxxxxx	21/12/1955
Juan	xxxxxxxxxx	25/05/1936

Tabla 23. Ejemplo reidentificación II. Fuente: elaboración propia.

Vemos que la fecha de nacimiento está presente en los dos conjuntos de datos.

Cruzando las tablas tenemos *inner-join* por **fecha**.

Nombre	Dirección	Fecha de nacimiento	CP	Fecha de nacimiento	Sexo	Enfermedad
Juan	xxxxxxxxxx	13/12/1955				
Pedro	xxxxxxxxxx	12/02/1955				
Daniel	xxxxxxxxxx	10/01/1955				
María	xxxxxxxxxx	25/05/1936	28002	25/05/1936	M	Fibromialgia
María	xxxxxxxxxx	25/05/1936	28005	25/05/1936	H	Lupus
Jorge	xxxxxxxxxx	21/12/1955				
Juan	xxxxxxxxxx	25/05/1936				
			28001	12/01/2005	M	Hepatitis C
			28004	22/01/1950	H	Hipertensión
			28001	11/07/1947	M	Cáncer de colon

Tabla 24. Ejemplo reidentificación III. Fuente: elaboración propia.

Nos identifica dos posibles candidatos. Si además sabemos que María es un nombre fundamentalmente de mujer y disponemos del sexo, podremos afirmar con alta probabilidad que María padece fibromialgia.

Este modelo expuesto de reidentificación responde al siguiente esquema:

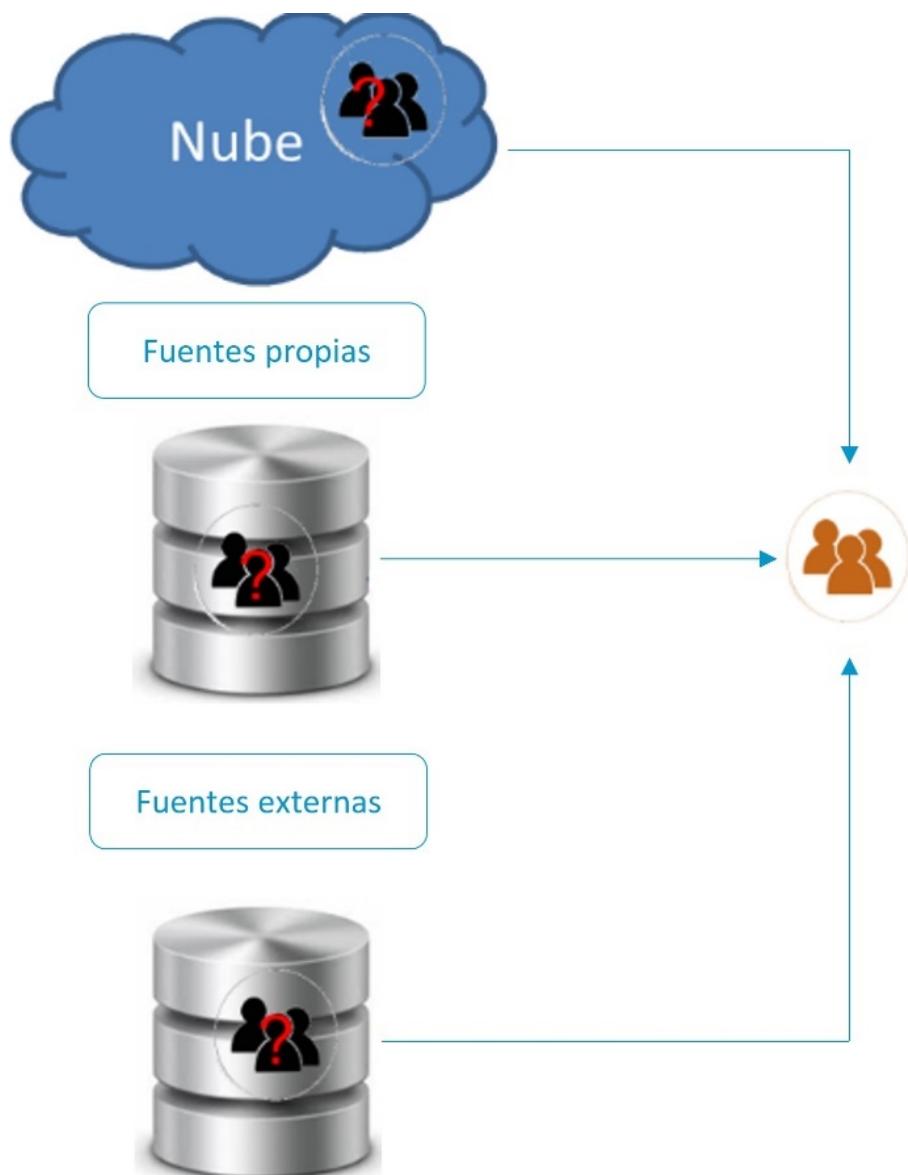


Figura 3. Esquema diferentes fuentes de datos. Fuente: elaboración propia.

Es fácil determinar que el riesgo inherente de reidentificación de cualquier conjunto de datos hoy en día es muy superior al que el mismo conjunto de datos estaría sometido en los años 90 y muy probablemente muy inferior al que tendrá en los próximos años, si tenemos en cuenta que disponemos de:

- ▶ Tecnología que es capaz de tratar información no estructurada, facilitando vinculaciones de fuentes diversas.
- ▶ Elevadas capacidades de tratamiento para realizar operaciones.
- ▶ Internet y el uso extensivo de las redes sociales como fuentes de información.
- ▶ Hiperconectividad.
- ▶ Las iniciativas de datos abiertos.

Por otra parte, técnicas que en su momento fueron consideradas suficientes como la deidentificación de los datos hoy dejan de ser válidas aplicadas de forma independiente, por lo que requieren la combinación de otras que mitiguen sus riesgos inherentes.

A continuación, identificaremos los riesgos de algunas de las técnicas de anonimización más utilizadas.

## Riesgos de las técnicas de anonimización

En opinión del WP29, en su «Opinion 05/2014 on Anonymisation Techniques», adoptada el 10 de abril de 2014, todas las técnicas tienen debilidades que se deben considerar seriamente antes de que una determinada técnica se utilice para diseñar un proceso de anonimización por parte del responsable de tratamiento.

Se deben tener en cuenta los **fines** que se persiguen con la anonimización, tales como la protección de la privacidad de los individuos cuando se hace público o accesible a terceros un conjunto de datos o el hecho de no facilitar el acceso a determinada pieza de información cuando se accede a un conjunto de datos.

Tras la evaluación determinaremos cuál es la mejor solución, que seguramente no será única, sino que tendremos que aplicar una combinación de técnicas que respondan al problema de una forma más eficaz atendiendo a las limitaciones de las técnicas de anonimización.

Según el grupo de trabajo del artículo 29 (WP29), hay tres cuestiones clave a la hora de valorar la robustez de una técnica:

- ▶ ¿Es posible destacar un individuo dentro del conjunto de datos, por ejemplo, por un valor característico de uno de sus atributos?
- ▶ ¿Sigue siendo posible vincular los registros relativos a un individuo con otras fuentes u obtener la identidad del individuo?
- ▶ ¿Se puede inferir la información relativa a una persona desde el conjunto de los datos?

Atendiendo a estas cuestiones, los **tres riesgos** de las técnicas de anonimización que propone el WP29 son:

- ▶ Señalamiento o singularización (*singling out*): corresponde a la posibilidad de aislar algunos o todos los registros que identifican a un individuo en el conjunto de datos.
- ▶ Vinculabilidad (*linkability*): la capacidad de vincular al menos dos registros referentes al mismo interesado o a un grupo de interesados (ya sea en la misma base de datos o en dos diferentes). Si un atacante puede establecer (por ejemplo, por medio de un análisis de correlación) que dos registros se asignan a un mismo grupo de personas, pero no pueden distinguir individuos en el grupo, la técnica proporciona resistencia contra *singling out*, pero no contra *linkability*.
- ▶ Inferencia (*inference*): deducir con significativa probabilidad el valor de un atributo de entre los valores de un conjunto de otros atributos.

Cada técnica no cumple con total garantía los criterios de eficacia de la anonimización, sino que presenta riesgos. Prácticamente podemos afirmar que habiendo microdatos hay riesgo de reidentificación.

Para facilitar una primera aproximación a esta visión de riesgo y conveniencia de las diferentes técnicas, se propone la siguiente tabla.

	Riesgo de singularización	Riesgo de vinculabilidad	Riesgo de inferencia
Seudonimización	Sí	Sí	Sí
	Sí	Puede que no	Puede que no
	Sí	Sí	Puede que no
	No	Sí	Sí
<i>I-Diversity/t-Closeness</i>	No	Sí	Puede que no
Privacidad diferencial	Puede que no	Puede que no	Puede que no
<i>Hash/Tokens</i>	Sí	Sí	Puede que no

Tabla 25. Riesgo de reidentificación. Fuente: Art. 29 WP, 2014.

## Consejos para reducir el riesgo de reidentificación

Desde el WP29, en su «Opinion 05/2014 on Anonymisation Techniques», se nos proponen un conjunto de recomendaciones para aplicar con independencia de la técnica utilizada para reducir el riesgo de identificación. Se aconseja, en general:

- ▶ No confiar en un enfoque de «liberar y olvidar», considerando siempre el riesgo residual de identificación que los datos tratados pueden tener, manteniendo por parte de los responsables de tratamiento las siguientes prácticas:
  - Identificar nuevos riesgos o reevaluar los riesgos periódicamente.
  - Evaluar si los controles para los riesgos identificados son suficientes y se ajustan en consecuencia.
  - Monitorizar y controlar el riesgo.
- ▶ Como parte de estos riesgos residuales es importante tener en cuenta la potencial identificación de la parte no anonimizada de un conjunto de datos (si la hay), especialmente cuando se combina con la parte anónima, además de las posibles correlaciones entre los atributos (por ejemplo, entre la ubicación geográfica y los datos del nivel de riqueza).

Estas recomendaciones se enlazan perfectamente con la **evaluación de impacto** en la protección de datos personales. Integra la incorporación de un proceso de evaluación continuo del riesgo inherente y de las medidas implantadas, lo que sería aplicable tanto a los tratamientos de datos personales como a los datos anonimizados o disociados.

Siguiendo con las recomendaciones del WP29, se deben establecer claramente los objetivos que deben alcanzarse a través del conjunto de datos anónimos, ya que desempeñan un papel clave en la determinación del riesgo de identificación.

Esto se une a la necesaria consideración de todos los elementos relevantes del contexto. Por ejemplo: la naturaleza de los datos originales, los mecanismos de control establecidos (incluidas las medidas de seguridad para restringir el acceso a las bases de datos), el tamaño de la muestra (características cuantitativas), la disponibilidad de recursos de información públicos (que puedan ser utilizados por los destinatarios), la previsión de liberación de datos a terceros (limitado, ilimitado en Internet, etc.).

Además, se debe prestar atención a los posibles atacantes, teniendo en cuenta el atractivo de los datos para los ataques dirigidos (de nuevo, la sensibilidad de la información y la naturaleza de los datos serán factores clave en este sentido).

En resumen, podemos concluir:

- ▶ La dificultad de conseguir conjuntos disociados de datos (perfectamente anonimizados).
- ▶ Los riesgos de un conjunto de datos anonimizados pueden verse alterados en el futuro, por ello se recomienda la reevaluación periódica de los mismos.

## 10.8. Principios a la hora de construir un data warehouse

Por último, y para terminar este tema, vamos a proponer un conjunto de principios que podemos seguir a la hora de construir un *data warehouse* que reduzca la exposición a un evento de revelación de datos de una forma sencilla. Estos principios son:

- ▶ **Separaciónfuncional:** facilita que el acceso a los datos se realice exclusivamente a aquellos para los que se está autorizado de acuerdo con las funciones desarrolladas, quedando imposibilitado el acceso a información adicional.
- ▶ **Agregación de datos:** siempre que sea posible y no interfiera en la finalidad del conjunto de datos es conveniente proceder a la agregación de datos de manera que se complique la posibilidad de reidentificar al individuo. Siempre debemos considerar que los registros no agregados suelen hacer más fácil la reidentificación de registro de datos agregados.
- ▶ **Deidentificación de los registros:** eliminación de identificadores directos e indirectos en la medida de lo posible. Por su parte, sustituir los identificadores por seudónimos introducirá una barrera adicional. Las expectativas sobre el nivel de anonimización de un conjunto de datos después de eliminar los identificadores directos o indirectos o sustituir estos por seudónimos no deben ser maximalistas, pues son evidentes los riesgos que aún tendrá el conjunto de datos sobre la potencial reidentificación del individuo. Por ejemplo, a través de la combinación de un conjunto de atributos. Estos son los casos de:
  - ▶ Datos de edad.
  - ▶ Geolocalizadores: código postal, datos de ubicaciones en mapas.
  - ▶ Sexo.

- ▶ Información adicional: en una historia clínica podrías tener códigos de diagnósticos, sobre todo los menos frecuentes (como enfermedades raras).
- ▶ Información sobre educación: formación o titulaciones poco usuales.
- ▶ Información sobre profesión: profesiones poco usuales.
- ▶ Raza.
- ▶ Ingresos.
- ▶ Religión.
- ▶ Indicadores socioeconómicos.
- ▶ Y otros.

Para algunos de estos valores contaremos con una distribución relativamente uniforme para el conjunto de la población. La presencia de este dato no facilita la identificación del individuo, pero combinados disponen de una mayor capacidad en ese sentido.

Por ejemplo, el dato sobre el sexo no aporta mucho. Sin embargo, sexo con edad más geolocalización y diagnóstico clínico puede generar una mayor concreción sobre el potencial del individuo al que se refiere.

Requerirá un esfuerzo significativo eliminar esta información del conjunto de datos, analizando en cada caso la posibilidad de su eliminación o agregación, atendiendo al objeto del estudio y cómo esta acción puede perturbar las conclusiones del estudio y en qué manera. Aquí adquiere especial relevancia la finalidad del uso al que están destinado estos datos.

- ▶ **Aplicación de técnicas de anonimización** sobre microdatos: combinar las diferentes técnicas de anonimización con el objeto de reducir riesgos.

Para orientar qué técnica aplicar, podemos consultar la siguiente tabla.

Técnica	Cuándo usar	Tipo de atributo
Supresión de atributos.	Cuando los atributos no son necesarios.	Todos.
Supresión de registros.	Presencia de registros atípicos.	Aplica a todo el registro.
Enmascaramiento de caracteres.	Enmascarar algunos caracteres en un atributo proporciona suficiente anonimato. Conservamos partes del valor original.	Identificador directo.
Pseudonimización.	Debemos poder distinguir los registros, pero no debemos conservar ninguna parte del valor del atributo original.	Identificador directo.
Generalización.	Modificamos los registros para ser menos precisos pero útiles.	Todos.
Permutación de datos.	Aplicable cuando no es necesario analizar las relaciones entre los atributos en el nivel de registro.	Todos.
Perturbación de datos (aleatorización-adición de ruido).	Ligeras modificaciones en los atributos no alteran el estudio, reducen riesgo de reidentificación a través del acceso a valores reales.	Identificadores directos o datos de análisis.
Agregación de datos.	Los datos agregados son suficientes para la utilidad del estudio. No se necesitan presentar registros individualizados.	Identificadores indirectos.

Tabla 26. Técnicas de anonimización. Fuente: elaboración propia.

Un **identificador directo** es un atributo que por sí solo identifica a un individuo (DNI, huella dactilar...); mientras que un **identificador indirecto** o cuasiidentificador es un atributo que por sí mismo no identifica, pero combinado con otra información puede que sí.

Debemos tener presente que aplicar técnicas de anonimización reduce la utilidad de los datos, se pierde carga de información.

Por tanto, elegir la técnica más adecuada va a requerir determinar qué utilidad necesitamos que mantengan los datos para el análisis o estudio que se vaya a realizar.

Por ejemplo, si un estudio requiere amplia precisión geográfica, tal vez no podamos generalizar el CP, incrementando posiblemente el riesgo de reidentificación.

## 10.9. Referencias bibliográficas

Article 29 Working Party. (2013). Opinion 05/2014 on anonymisation techniques [Archivo PDF].

[https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216\\_en.pdf](https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf)

Ley 14/2007, de 3 de julio, de Investigación biomédica. Boletín Oficial del Estado, 4 de julio de 2007, núm. 159, pp. 28826-28848.

<https://www.boe.es/eli/es/l/2007/07/03/14>

Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal. Boletín Oficial del Estado, 14 de diciembre de 1999, núm. 298.

<https://www.boe.es/eli/es/lo/1999/12/13/15/con>

Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal. Boletín Oficial del Estado, 19 de enero de 2008, núm. 17.

<https://www.boe.es/eli/es/rd/2007/12/21/1720/con>

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento General de Protección de Datos). Diario Oficial de la Unión Europea L 119, 4 de mayo de 2016, pp. 1-88.

<https://www.boe.es/doue/2016/119/L00001-00088.pdf>

Sweeney, L. (2019). Latanya Sweeney, Ph.D. [Página web].

<http://latanyasweeney.org/>

## Opinion 05/2014 on anonymisation techniques

Article 29 Working Party. (2013). Opinion 05/2014 on anonymisation techniques [Archivo PDF].

[https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216\\_en.pdf](https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf)

Documento del grupo de trabajo del artículo 29 (WP29) con consideraciones sobre las técnicas de anonimización de datos que facilita un entendimiento sobre los riesgos asociados a las mismas e incorpora recomendaciones para el tratamiento del riesgo residual.

## Código de buenas prácticas para la gestión de riesgos derivados de los procesos de disociación

ICO. (2012). Anonymisation: managing data protection risk code of practice [Archivo PDF].

<https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf>

Documento de la ICO en inglés.

## Página web del grupo de trabajo del artículo 29

European Comission. (s.f.). Data protection [Página web].

[http://ec.europa.eu/justice/data-protection/index\\_en.htm](http://ec.europa.eu/justice/data-protection/index_en.htm)

En la página web del grupo de trabajo del artículo 29 se puede acceder a toda la documentación. Son especialmente interesantes las opiniones que consolidan la visión interpretativa de la autoridad de protección de datos de la UE.

1. La ley de protección de datos no se aplica sobre los datos disociados obtenidos tras un proceso de disociación.

  - A. Verdadero.
  - B. Falso.
  
2. ¿Cuáles de las siguientes técnicas no es de disociación-anonimización?

  - A. Asignación al azar.
  - B. Generalización.
  - C. Adición de ruido.
  - D. T-Combinación.
  
3. ¿Cuáles de los siguientes son riesgos asociados a las técnicas de anonimización?

  - A. Singularización.
  - B. Vinculabilidad.
  - C. Inferencia.
  - D. Todos los anteriores son correctos.
  
4. La generalización es una técnica del tipo:

  - A. Técnicas que implican la modificación de datos.
  - B. Técnicas que implican la reducción de atributos.
  - C. Técnicas que implican la supresión de datos.
  - D. Técnica de pseudoanonimización.

- 5.** La Dra. Sweeney concluyó que en EE. UU. es posible identificar al 87 % de los ciudadanos únicamente, a partir de los datos de:
- A. El número de la seguridad social.
  - B. El número de licencia de conducir y el estado emisor.
  - C. El código postal completo, sexo y fecha de nacimiento con el año incluido.
  - D. El código postal completo, sexo, fecha de nacimiento con el año incluido y raza.
- 6.** ¿Cuál de los siguientes riesgos no es de reidentificación?
- A. Riesgo de singularización.
  - B. Riesgo de vinculabilidad.
  - C. Riesgo de inferencia.
  - D. Riesgo de polarización.
- 7.** De entre las siguientes técnicas, ¿cuál ofrece mejores resultados en términos de menores riesgos de reidentificación?
- A. Seudonimización.
  - B. Adición de ruido.
  - C. K-anonimato.
  - D. I-Diversity/t-Closeness.
- 8.** De entre las siguientes técnicas, ¿cuál ofrece los peores resultados en términos de menores riesgos de reidentificación?
- A. Seudonimización.
  - B. Adición de ruido.
  - C. K-anonimato.
  - D. I-Diversity/t-Closeness.

- 9.** Teniendo en consideración el siguiente conjunto de datos, el resultado de aplicar k-anonimato, indica cuál es valor  $k$  del conjunto:
- A.  $k = 2$ .
  - B.  $k = 3$ .
  - C.  $k = 4$ .
  - D.  $k = 5$ .
- 10.** Teniendo en consideración el siguiente conjunto de datos, el resultado de aplicar l-Diversity, indica cuál es el valor  $l$  del conjunto:
- A.  $l = 1$ .
  - B.  $l = 2$ .
  - C.  $l = 3$ .
  - D.  $l = 4$ .