



# 01. La sociedad de la información

La **sociedad de la información** es el nombre que recibe la sociedad occidental en la que vivimos, donde las tecnologías que facilitan la creación, distribución y manipulación de la información juegan un papel importante en las actividades sociales, culturales y económicas.

Este concepto nace en Japón, en el año 1960, y busca resumir los cambios que vivieron las sociedades tras implementar el uso de las TICs (tecnologías de la información y la comunicación) de forma masiva. Se consideran estas sociedades como las sucesoras de las sociedades industriales.

La popularización de internet, los avances tecnológicos, la masificación del uso de celulares con alta capacidad de procesamiento, la aparición de las redes sociales fueron propulsores de cambios en la esfera social.

A continuación, mencionamos algunas características de la sociedad de la información:

- Hay un mayor acceso a la información.
- El acceso es inmediato.
- Se ve facilitada la comunicación entre individuos.
- Globalización con transferencia de conocimiento.
- Mayor debate y diversidad de opiniones.
- Fomenta la productividad y la eficiencia, por su rapidez.

# 02. La 4ta Revolución Industrial y la Industria 4.0

De la misma forma que la Revolución Industrial, marcó un antes y un después en la forma de producir, introduciendo las máquinas como principal impulsor de la economía, la Revolución Digital transformó la forma de producir a lo largo y ancho del planeta. Actualmente, se dice que **estamos atravesando la 4ta Revolución Industrial**, donde vivimos una **transición hacia nuevos sistemas ciber-físicos que operan en forma de complejas redes**. Esta nueva era se caracteriza por la convivencia de una gran variedad de tecnologías, que borran los límites entre



lo físico, lo digital y lo biológico, generando una fusión entre estos tres planos y un cambio de paradigma.

El término “**Industria 4.0**” se comenzó a acuñar en la década de 2010, y surge de un grupo de especialistas multidisciplinarios convocados por el gobierno alemán, enfocados en diseñar un programa de mejoras para la industria manufacturera. Entre los pilares tecnológicos de la Industria 4.0 se destacan: sistemas ciber físicos de integración; máquinas y sistemas autónomos (robots); internet de las cosas (IoT); manufactura aditiva (impresión 3D); big data y análisis de macrodatos; computación en la nube; simulación de entornos virtuales; inteligencia artificial; ciberseguridad; y realidad aumentada.

La transformación más profunda se produce por la posibilidad de conectar en tiempo real a todos los actores sociales mediante Internet. La conectividad alcanza a consumidores, empresas, gobierno, organizaciones de la sociedad civil, y es posible mediante dispositivos (smartphones, computadoras, sensores, wearables, etc.), sistemas informáticos y plataformas digitales ( e-commerce , e-government , redes sociales). Pero la novedad de esta época es que la conectividad alcanza también a los objetos permitiendo la conexión en varios sentidos: máquina-máquina (M2M), máquina-producto, máquina-humano, producto-humano.

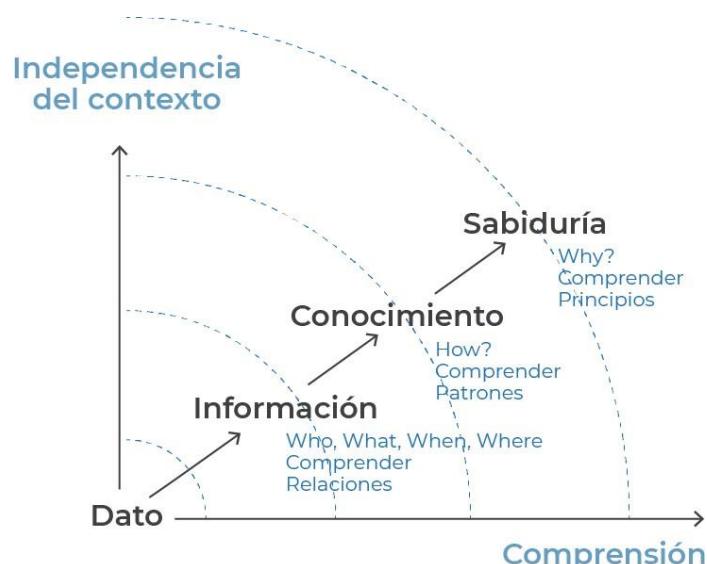
### Tendencias en Data & Analytics 2020-2021

- IA más inteligente, rápida y responsable
- Decision Intelligence
- X Analytics: Data No Estructurada
- Gestión de datos aumentada
- Cloud es una realidad
- Choques entre mundos de datos y analytics
- Data Marketplaces e Intercambios
- Blockchain en data & analytics



## 03. Datos, Información, Conocimiento y Sabiduría

Hablamos de información, pero ¿sabemos qué es la información? o ¿cómo llegan los datos a ser sabiduría? A continuación, presentaremos los distintos conceptos y veremos el proceso de transformación que los datos precisan recibir para convertirse en conocimiento y sabiduría.



Proceso de Transformación de los Datos en Conocimiento. Fuente de la imagen: Tomado y Adaptado de Bellinger (2004).

**Los datos**, son hechos estructurados o no, que no poseen una relación intrínseca. La relación es una construcción otorgada por una persona que los analiza. Podemos pensar a los datos como un subconjunto de la información, donde todos los datos se pueden convertir en información, pero no toda información es un dato.

Avanzando en las definiciones, la **información** es un conjunto de datos procesados o datos que se le ha dado sentido por medio de una relación. La información representa el cuál, el quién, el cuándo y el dónde. Son los datos en un contexto.

El **conocimiento** resulta el conjunto de representaciones entretejidas basadas en información, con análisis, interpretación y argumentación, de un determinado contexto.



con significación y conciencia de sus interrelaciones. Representa el cómo. El conocimiento puede ser tácito o explícito. El conocimiento tácito reside en las personas, no ha sido documentado ni codificado, está conformado por la experiencia, valores, ideas, emociones e información de cada individuo. El conocimiento explícito es aquel que puede ser codificado y representado articuladamente. En otras palabras, el conocimiento explícito se encuentra en una fuente estática y está codificado, como por ejemplo, en repositorios.

**La sabiduría es entendida como una abstracción significativa, basada en procedimiento y en experiencia.** La sabiduría permite la acción racional y la utilización de la experiencia. La sabiduría requiere un mayor nivel de comprensión de los datos, la información y el conocimiento, y viene representada por el por qué.

## 04. Ciclo de vida de los datos

El ciclo de vida de los datos, también llamado ciclo de vida de la información se refiere a todo el periodo de tiempo que los datos existen en su sistema. **Este ciclo de vida abarca todas las etapas por las que pasan los datos, desde su primera captura hasta su archivo.** Presentamos un ejemplo del marco del ciclo de vida de los datos:

### 1. Creación, ingesta u obtención de los datos

Tanto si se generan datos a partir de la entrada de datos, como si se adquieren datos existentes de otras fuentes o se reciben señales de dispositivos, se obtiene información de alguna manera. Esta etapa describe cuándo los valores de los datos entran en el sistema para su análisis.

### 2. Procesamiento de los datos

Son muchos los procesos que intervienen en la limpieza y preparación de los datos para su posterior análisis. Aunque el orden de las operaciones puede variar, la preparación de los datos suele incluir la integración de los datos procedentes de múltiples fuentes, la validación de los datos y la aplicación de transformaciones. A menudo, los datos se formatean, se resumen, se agrupan y se estandarizan como parte del flujo de trabajo del procesamiento de datos.

### 3. Análisis de los datos

La exploración e interpretación de los datos puede requerir una serie de análisis. Puede tratarse de análisis estadísticos o de visualización. También puede significar el uso de modelos de datos tradicionales o la aplicación de la *machine learning* (ML).

### 4. Publicación de resultados

En esta etapa es donde el conocimiento obtenido se convierte en decisiones. Los datos ofrecen todo su valor empresarial cuando se difunde la información obtenida del análisis.



## 5. Archivar

Una vez que los datos han sido recogidos, procesados, analizados y compartidos, suelen almacenarse para futuras consultas. Para que los archivos tengan algún valor futuro, es importante mantener los **metadatos** sobre cada elemento de los registros, en particular sobre la procedencia de los datos.

El ciclo de vida de los datos se desarrolla desde el último paso hasta el primero en forma circular.

# 05. La gobernanza de los datos

La gobernanza de los datos es el conjunto de procesos, funciones, políticas, normas y mediciones que garantizan el uso eficaz y eficiente de la información con el único fin de que las empresas alcancen sus objetivos.

La puesta en marcha de este tipo de estrategias empresariales logra maximizar el valor de los datos, reducir costos y gestionar mejor los posibles riesgos. Además, asegura la calidad y seguridad de la información y permite definir quién puede emprender acciones, sobre qué datos, en qué circunstancias y mediante qué métodos.

Los datos son un factor clave en cualquier entidad y la capacidad para procesarlos se ha vuelto fundamental. Es por ello que la gobernanza de datos está recogida en el artículo 24 del RGPD (Reglamento General de Protección de Datos). El **Reglamento General de Protección de Datos** (RGPD, o GDPR por sus siglas en inglés) es el nuevo marco legal en la Unión Europea.

## 05.01. ¿Cuál es el objetivo de RGPD?

El RGPD (Reglamento General de Protección de Datos) busca proteger los datos personales y la forma en la que las organizaciones los procesan, almacenan y, finalmente, destruyen, cuando **esos datos ya no son requeridos**. La ley provee control individual acerca de cómo las compañías pueden usar la información que está directa y personalmente relacionada con los individuos, y otorga ocho derechos específicos.

Además, establece normas muy estrictas, que rigen lo que sucede si se viola el acceso a datos personales y las consecuencias (penalidades) que las organizaciones pueden sufrir en tal caso.



**Datos personales** es toda información sobre una persona física identificada o identificable, cuya identidad pueda determinarse, directa o indirectamente, mediante un identificador (nombre, número de identificación, localización).

El Reglamento en materia de protección de datos también regula específicamente las categorías especiales de datos en su artículo 9, los también conocidos como **datos especialmente sensibles** que son datos personales que revelan:

- el origen étnico o racial
- las opiniones políticas
- las convicciones religiosas o filosóficas
- la afiliación sindical
- los datos genéticos
- los datos biométricos dirigidos a identificar de manera única a una persona física
- los datos relativos a la salud
- los datos relativos a la vida sexual u orientación sexual de una persona física

## 05.02. ¿A quién aplica RGPD?

EL RGPD aplica a:

- Organizaciones con presencia física en al menos algún país miembro de la Unión Europea.
- Organizaciones que procesan o almacenan datos sobre individuos que residen en la Unión Europea.
- Organizaciones que utilizan servicios de terceros que procesan o almacenan información sobre individuos que residen en la Unión Europea.

Por lo tanto, **afecta a organizaciones de la Unión Europea y a organizaciones que posean empleados o clientes en la Unión Europea.**

## 05.03. ¿Cuáles son los 8 derechos que establece el RGPD?

- **Derecho a estar informado:** Proporciona transparencia sobre cómo son utilizados los datos personales.
- **Derecho al acceso:** Provee acceso a los datos, a cómo son utilizados, y a cualquier información suplementaria que pueda ser utilizada juntos con los datos.
- **Derecho a la rectificación:** Otorga el derecho a que los datos personales sean rectificados en caso de ser incorrectos o incompletos.
- **Derecho a ser borrado (o derecho a ser olvidado):** Es el derecho a que los datos personales sean removidos de cualquier lugar si no existe una razón convincente para que estén almacenados.
- **Derecho a restringir el procesamiento:** Permite que los datos sean almacenados, pero no procesados.

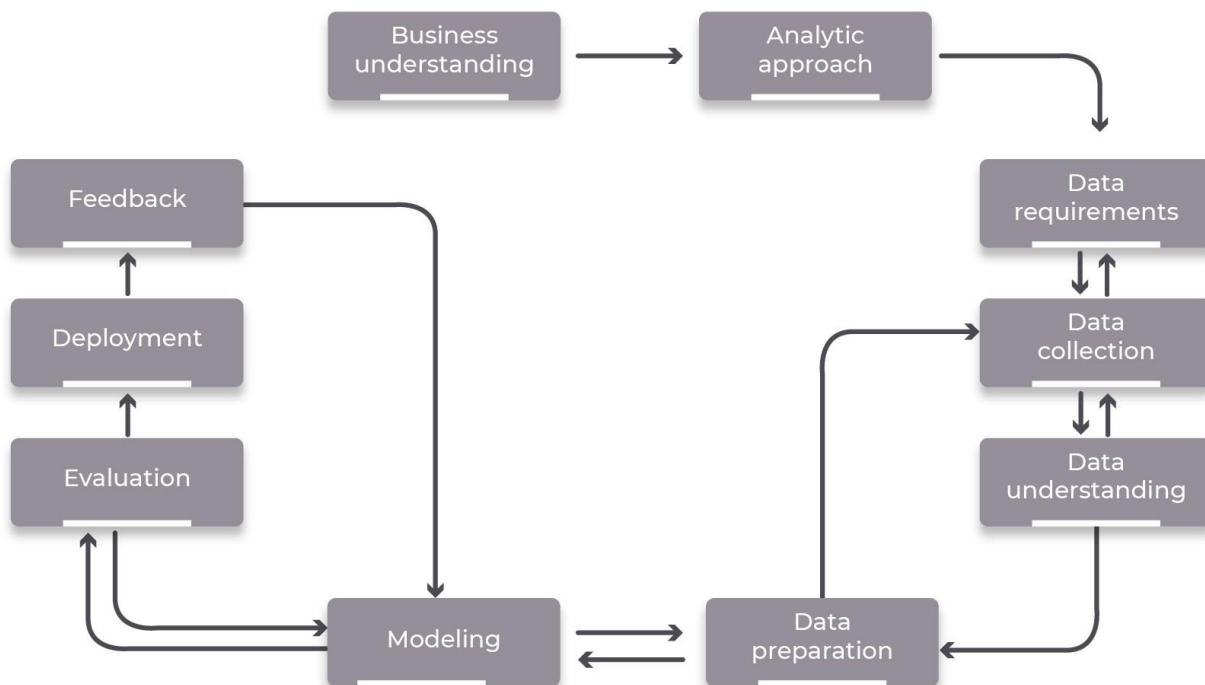


- **Derecho a la portabilidad de datos:** Podemos solicitar copias de la información almacenada sobre nosotros, para utilizar en cualquier otro lugar.
- **Derecho a objetar:** Otorga el derecho a objetar acerca del procesamiento de los datos. Un ejemplo podría ser la objeción de que nuestros datos sean utilizados por organizaciones de marketing directo.
- **Derecho sobre la toma de decisiones y creación de perfiles automáticos:** Permite objetar sobre la toma de decisiones automáticas que se hagan sobre los datos personales. "Automáticas" se refiere a sin intervención humana. Por ejemplo, la definición de determinados hábitos de compra online, en función a comportamientos previos.

## 06. Los proyectos de ciencia de datos

A medida que las capacidades de análisis de datos se vuelven más accesibles y prevalentes, los científicos de datos necesitan una metodología que funcione como guía, independientemente de las tecnologías, los volúmenes de datos o los enfoques involucrados, para desarrollar una estrategia de análisis.

A continuación, compartimos una metodología planteada por IBM para el desarrollo de un proyecto de ciencia de datos. Esta metodología consta de 10 etapas que forman un proceso iterativo para el uso de datos para descubrir información. Cada etapa desempeña un papel vital en el contexto de la metodología general.



**1. Comprensión del negocio (Business Understanding):** Cada proyecto comienza con la comprensión del negocio. Los patrocinadores del negocio definen el problema, los objetivos del proyecto y los requisitos de la solución desde una perspectiva empresarial. Esta primera etapa sienta las bases para una resolución exitosa del problema. Para ayudar a garantizar el



éxito del proyecto, los patrocinadores deben participar a lo largo del proyecto para proporcionar experiencia en el dominio, revisar los hallazgos intermedios y garantizar que el trabajo continúe en el camino correcto para generar la solución deseada.

**2. Enfoque analítico (Analytic Approach):** Una vez que el problema de negocio se ha establecido claramente, el científico de datos puede definir el enfoque analítico que utilizará para resolver el problema. Esta etapa implica expresar el problema en el contexto de las técnicas estadísticas y de aprendizaje automático, para que la organización pueda identificar las más adecuadas para el resultado deseado. Por ejemplo, si el objetivo es predecir una respuesta como “sí” o “no”, entonces el enfoque analítico podría definirse como construir, probar e implementar una modelo de clasificación.

**3. Requerimiento de datos (Data Requirements):** El enfoque analítico seleccionado determina los requerimientos de los datos. Específicamente, los métodos analíticos que se utilizarán requieren ciertos contenidos de datos, formatos y representaciones, sobre los que se deberá trabajar.

**4. Recolección de los datos (Data Collection):** En la etapa de recolección de datos, los científicos de datos identifican y reúnen los datos disponibles para el abordaje del problema. Puede suceder que sea necesario realizar una inversión para obtener nuevos datos. Resulta conveniente aplazar la decisión de inversión hasta que se sepa más sobre los datos y el modelo a utilizar. Al incorporar más datos, los modelos predictivos pueden representar mejor los eventos atípicos (outliers).

**5: Comprensión de los datos (Data Understanding):** Despues de la recopilación de los datos, los científicos de datos suelen utilizar herramientas de estadística descriptiva y técnicas de visualización para comprender las características de los datos, evaluar su calidad y descubrir información inicial sobre ellos.

**6: Preparación de los datos (Data Preparation):** Esta etapa abarca todas las actividades para construir el conjunto de datos que se utilizará en el modelado. Las actividades de preparación de datos incluyen: la limpieza de datos (tratar con valores faltantes o no válidos, eliminar duplicados, formatear adecuadamente), combinar datos de múltiples fuentes (archivos, tablas, plataformas) y transformar datos en variables más útiles.

La preparación de datos suele ser el paso más lento. Sin embargo, algunos pasos de la preparación de los datos son comunes en diferentes problemas, pudiéndose automatizar ciertos pasos.

**7. Modelado (Modeling):** A partir de la primera versión del conjunto de datos preparado, la etapa de modelado se centra en el desarrollo de modelos predictivos o descriptivos de acuerdo con el enfoque analítico previamente definido (Paso 2). Por lo general, se prueban múltiples algoritmos para encontrar el mejor modelo para las variables disponibles.



**8. Evaluación (Evaluation):** Durante el desarrollo del modelo, los modelos deben evaluarse para comprender su calidad y garantizar que aborda de manera adecuada y completa el problema del negocio. La evaluación del modelo implica el cálculo de varias medidas de diagnóstico, lo que permite interpretar la calidad del modelo y su eficacia para resolver el problema.

**9. Despliegue (Deployment):** Una vez que un modelo satisfactorio ha sido desarrollado y aprobado por los sponsors, se implementa en el entorno de producción o en un entorno de prueba comparable. Por lo general, se implementa de forma limitada hasta que su rendimiento se haya evaluado completamente. La implementación de un modelo en un proceso de negocio operativo generalmente involucra grupos, habilidades y tecnologías adicionales dentro de la empresa.

**10. Retroalimentación (Feedback):** Al recopilar los resultados del modelo implementado, la organización obtiene información sobre el rendimiento del modelo y su impacto en el entorno en el que se implementó. Por ejemplo, la retroalimentación podría tomar la forma de tasas de respuesta a una campaña promocional dirigida a un grupo de clientes identificados por el modelo como clientes de alto potencial. Al analizar esta información, se puede refinar el modelo para mejorar su precisión y utilidad.

Fuente: IBM Analytics | White paper | Foundational Modeling for Data Science

## 07. Roles en un proyecto de ciencia de datos

**Analista de datos:** es un perfil profesional que gracias a la interpretación de los datos puede establecer estrategias dentro de una empresa. Por lo tanto, debe saber recopilar datos a la vez que analizarlos de forma estadística. El analista de datos se encarga de:

- Extraer, procesar y agrupar datos
- Analizar esas agrupaciones de datos
- Generar informes

Un analista debe presentar competencias en matemáticas y estadísticas. De este modo consiguen realizar un análisis completo de los datos extraídos.

La comunicación efectiva debe ser una de sus características, ya que no solo tratará con el científico de datos, sino que su misión radica en que todas las personas comprendan los datos



con los que trabajan. Al analista de datos le mueve la curiosidad por descubrir qué hay detrás de toda la información que genera una empresa.

**Ingeniero de datos:** es el profesional encargado de sentar las bases para la adquisición, el almacenamiento, la transformación y la gestión de los datos en una organización. Este especialista asume la configuración de la infraestructura tecnológica necesaria para que grandes volúmenes de datos se conviertan en materia prima accesible para otros especialistas, como los analistas de datos o los científicos de datos. Por lo tanto, son los responsables de mantener sistemas escalables, con alta disponibilidad y rendimiento, integrando nuevas tecnologías y desarrollando el software necesario.

El ingeniero de datos se encarga de:

- Preparar los datos como parte de los procesos ETL (extracción, transformación y carga) de datos
- Proporcionar acceso al conjunto de datos para que sea usado por otros especialistas de datos
- Trabajar con un administrador de base de datos para crear y gestionar el almacenamiento de los datos.
- Garantizar la entrega de datos a escala.
- Poner el sistema en producción.

**Científico de datos:** las tareas del científico de datos son transversales a las del analista de datos y al ingeniero de datos. El científico de datos estará desarrollando estrategias para analizar los datos, acompañando a la preparación de los datos para su análisis, asesorando en la exploración, análisis y visualización de los datos, construyendo modelos predictivos e implementando los modelos en producción. Dentro de sus conocimientos específicos, el científico de datos tiene un amplio dominio de la estadística, incluyendo pruebas y distribuciones estadísticas, domina las técnicas analíticas como el machine learning, deep learning y analítica de texto, y sabe comunicarse y colaborar con las áreas de TI y de negocios.

**Investigación en ciencia de datos:** Los investigadores inventan cosas nuevas. La emoción de su carrera es demostrar que lo que antes era imposible es posible. Puede pensar en esto como el lado de la investigación teórica de la ciencia de datos, y es para personas con un doctorado en ML / AI / Estadísticas / Optimización. Crean teorías para expandir los tipos de problemas que la humanidad puede resolver.

También podemos encontrar otros perfiles, como un **analista empresarial** que ayude a definir el problema, un **arquitecto de tecnología informática** que supervise los procesos subyacentes y



la infraestructura, y un **desarrollador de aplicaciones** que implemente los modelos o las salidas del análisis en aplicaciones y productos.

### Lecturas complementarias

Industria 4.0: fabricando el futuro

<https://publications.iadb.org/publications/spanish/document/Industria-40-Fabricando-el-Futuro.pdf>

El cuarto paradigma

<https://www.noticiasdelcosmos.com/2009/10/la-ciencia-ante-el-cuarto-paradigma.html>



Autor: Sol Represa. Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirlIgual 4.0 Internacional](#). Mundos E.