



5.2 EL CICLO DE VIDA DEL DATO Y SU RELACIÓN CON EL ANÁLISIS Y LA CIENCIA DE DATOS

Introducción

En la sociedad actual el dato se ha convertido en un recurso estratégico y esencial para la toma de decisiones, la optimización de procesos y la anticipación a los problemas. Para comprender su importancia, es necesario conocer cómo se gestionan los datos a lo largo del tiempo, desde que se generan hasta que se eliminan. A dicho proceso se le denomina **el ciclo de vida del dato**.

Este ciclo describe las etapas por las que pasa un dato desde su creación hasta su eliminación definitiva. Es un proceso completo que abarca la recolección, procesamiento, almacenamiento, análisis, interpretación, uso y eliminación del dato. Además, permite que estas acciones se hagan siguiendo principios de calidad, seguridad y eficiencia, adaptándose a normas legales y éticas.



1. Captura o Generación del Dato

El ciclo comienza con la captura del dato, que puede producirse de forma manual (a través de encuestas o formularios) o automática (por sensores, aplicaciones o dispositivos digitales). Esta información puede tener origen en personas, máquinas o sistemas.

Ejemplo: Cuando alguien se registra en una página web, introduce su nombre, correo electrónico y contraseña. Ese formulario genera un conjunto de datos personales.

2. Almacenamiento

Una vez generados, los datos deben almacenarse de forma segura y accesible. Esto puede hacerse tanto en servidores *on premise* o *on cloud*. El almacenamiento debe garantizar la integridad y disponibilidad de los datos, además de cumplir con normas de protección de datos.

En esta fase es crucial considerar el cumplimiento de normativas como el RGPD (Reglamento General de Protección de Datos) en la Unión Europea o la LOPD en España.

Ejemplo: Los datos de los usuarios registrados en una aplicación se almacenan en bases de datos como MySQL o MongoDB, accesibles sólo mediante credenciales autorizadas.

3. Procesamiento

Aquí los datos son organizados, limpiados y transformados para facilitar su análisis. Esto incluye la eliminación de duplicados, el tratamiento de valores nulos o erróneos y la conversión de formatos; hablamos de la normalización de los datos. El objetivo es convertir los datos en información útil y fiable.

Las herramientas de procesamiento pueden incluir software de análisis estadístico, lenguajes como Python o R, y plataformas como Apache Spark o Excel.

Ejemplo: Una empresa de reparto analiza registros de entrega. Se eliminan datos repetidos y se corrigen errores como direcciones mal escritas para mejorar la logística.

4. Análisis

El análisis permite extraer valor de los datos, identificando patrones, tendencias o anomalías. Se utilizan técnicas estadísticas, modelos predictivos y visualizaciones para facilitar la interpretación de la información. Esta etapa es clave para la toma de decisiones basada en datos.

Ejemplo: Un supermercado analiza qué productos se venden más durante las vacaciones para planificar el inventario del año siguiente.

5. Interpretación y visualización

Los resultados del análisis deben comunicarse de forma clara a los usuarios o responsables de la toma de decisiones. Esto se hace a través de informes, dashboards, visualizaciones gráficas o

presentaciones. La comunicación efectiva del dato permite transformar la información en conocimiento.

Ejemplo: El departamento de marketing presenta en una reunión mensual un dashboard que muestra el crecimiento de usuarios por región, basado en datos analizados.

6. Toma de decisiones

La interpretación de los datos se utiliza como base para actuar. Las decisiones pueden ser estratégicas, operativas o tácticas, y deben basarse en evidencias extraídas del análisis. El concepto de "data-driven decision making" (toma de decisiones basada en datos) es clave en organizaciones modernas.

Ejemplo: Una universidad detecta que muchos estudiantes abandonan cierta asignatura, y decide revisar el plan docente y ofrecer apoyo académico adicional.

7. Mantenimiento y actualización

Los datos deben mantenerse organizados y actualizados si se quieren comparar con futuras mediciones.

8. Archivado y conservación

Una vez utilizados, los datos pueden conservarse por motivos legales, históricos o de análisis futuro. Esta fase implica establecer políticas de retención y definir durante cuánto tiempo los datos deben estar disponibles. También es importante establecer sistemas de respaldo (backups) y recuperación ante desastres, para asegurar la continuidad del acceso a los datos importantes.

Ejemplo: Un hospital guarda durante varios años los historiales médicos de sus pacientes, según la normativa sanitaria vigente.

9. Eliminación o Depuración

Cuando los datos ya no son útiles o legales de conservar, deben eliminarse de forma segura. Esto protege la privacidad de las personas y reduce el riesgo de accesos no autorizados. La eliminación puede hacerse mediante técnicas de borrado lógico o destrucción física, dependiendo del tipo de soporte.

Ejemplo: Una empresa elimina los currículos de candidatos descartados para un puesto de trabajo después de seis meses, conforme a su política de protección de datos.

El ciclo de vida del dato no es lineal: en muchas organizaciones, los datos pueden volver a etapas anteriores, ser reutilizados o transformarse en nuevos conjuntos de información. Por ello, es fundamental implementar una estrategia integral de gestión del dato que incluya normas éticas, legales y técnicas para garantizar su calidad y seguridad.

Análisis de datos y la ciencia de datos

El análisis de datos se enfoca en entender lo que ha ocurrido y por qué. A partir de los datos disponibles, se intentan explicar situaciones o responder preguntas específicas. Por ejemplo, una fábrica puede analizar cuántas unidades se producen cada día, cuánto tiempo han estado paradas las máquinas o cuántos productos han presentado defectos. Esta información permite conocer el estado actual del proceso, pero no necesariamente anticipar lo que ocurrirá en el futuro.

La ciencia de datos, en cambio, da un paso más allá. Es una disciplina interdisciplinaria al combinar la estadística, la informática y las matemáticas. Con ellas no sólo analiza, sino también predice lo que ocurrirá y sugiere acciones. La ciencia de datos se basa en técnicas como la **minería de datos, el aprendizaje automático o la inteligencia artificial** y de esta manera desarrollar modelos predictivos que ayudan a tomar decisiones complejas.

Por ejemplo, en la fábrica mencionada en el anterior ejemplo, la ciencia de datos no se limitará a describir lo que pasa sino que analizará cuándo y por qué ocurren las paradas de las máquinas, identificar variables críticas (como la temperatura de éstas) y se desarrollarán modelos que predigan cuándo podrían repetirse los fallos. Con esa información, se podrán sugerir cambios en el mantenimiento o ajustar procesos para evitar problemas futuros. Incluso se posibilita la automatización de la respuesta que el sistema debe dar ante ciertos riesgos, lo cual sería impensable con un análisis tradicional.

Por último hay que destacar que el análisis de datos forma parte de la ciencia de datos, pero esta última permite un enfoque mucho más amplio, al incluir donde etapas como la implementación, el monitoreo y la mejora continua de modelos automatizados.

Conclusión

Tanto el análisis de datos como la ciencia de datos son esenciales en el uso moderno de la información. Mientras el primero permite comprender lo que ha ocurrido, la segunda nos prepara para lo que pueda pasar. En ambos casos, es necesario la adecuada gestión del ciclo de vida del dato, ya que sin ella, ninguna de estas disciplinas podría aplicarse con eficacia.

(última actualización 20/05/2025)

Eusko Jaurlaritzaren Lanbide Heziketako Sailburuordetza. Lan honek Creative Commons Aitortu-EzKomertziala-PartekatuBerdin 4.0 Nazioarteko Baimena dauka (CC BY-NC-SA 4.0).

