



# [Capacitación Marco de Referencia de Gestión de Datos del Estado]

[Glosario]

Instrucciones: Se recomienda leer este documento antes de la capacitación del 12 de marzo para manejar ciertos conceptos claves.

Marzo 2025

# Consideraciones

Este glosario reúne los **principales términos y conceptos fundamentales en la gestión de datos**, en particular, para su administración, calidad, seguridad y uso dentro del sector público. Su objetivo es proporcionar a los funcionarios y funcionarias una referencia clara y estandarizada que facilite la comprensión de las mejores prácticas y roles clave en la gestión de datos en el Estado. Al familiarizarse con estos conceptos antes de la capacitación, los participantes podrán contar con una base común de conocimiento, promoviendo una aplicación más efectiva y clara del instrumento en el futuro.

## Conceptos

### 1. Gobernanza de datos

La gobernanza de datos es la función organizativa que implica definir, implementar y monitorear estrategias, políticas y decisiones relacionadas con la gestión y uso de los activos de datos, para lograr que estos sean correctos, fiables, seguros y útiles. También supone rendir cuentas sobre el uso eficaz, eficiente y aceptable de los datos de una organización.

***Estrategia Nacional de Datos***

### 2. Gestión de datos

La gestión de datos, por su parte, se refiere a las funciones y medidas operativas sobre los datos, orientadas a garantizar que los datos sean adecuados para cumplir con los objetivos estratégicos de una organización.

***Estrategia Nacional de Datos***

### 3. Calidad de datos

La calidad de datos se refiere al conjunto de metodologías y herramientas a través de las cuales se define, controla y mejora la calidad de los datos.

Para asegurar la calidad de los datos, existen diversos mecanismos y controles, como por ejemplo:

- Perfiles y monitoreo de datos: Evaluación continua para identificar errores y anomalías.
- Normas y estándares: Reglas establecidas para garantizar formatos, integridad y validez.
- Procesos de limpieza y enriquecimiento: Detección y corrección de errores, eliminación de duplicados y mejora de información.
- Automatización y validación: Uso de herramientas que verifican la calidad antes de ingresar o modificar datos.

Estos mecanismos permiten mantener datos confiables y precisos, reduciendo riesgos y mejorando la eficiencia operativa y el cumplimiento normativo.

***Guía Técnica MGDE***

## 4. Arquitectura de Datos

Estructura y organización de los datos y sus componentes en una institución, se encarga de establecer los dominios, inventarios y modelos de información. También los diseños de las plataformas para gestionar los datos.

**Informe 2 - Marco de Referencia de Gestión de Datos (MRD) - Elaboración de un marco de referencia de gestión de datos para los órganos de la Administración del Estado y diseño de herramientas de autoevaluación**

## 5. Data Lake

Un data lake o lago de datos es un repositorio centralizado que ingiere, almacena y permite el procesamiento de grandes volúmenes de datos en su forma original.

- Almacena datos en su **forma bruta**, sin necesidad de preprocesamiento.
- Puede contener datos **estructurados, semiestructurados y no estructurados**.
- Escalable y diseñado para **procesamiento masivo y análisis avanzado**.
- Permite análisis en tiempo real, machine learning y big data.

**Guía Técnica MGDE**

## 6. Data Warehouse

Un Data Warehouse (almacén de datos) es un repositorio centralizado que almacena datos estructurados (tablas de bases de datos, planillas de cálculos) y datos semiestructurados (archivos XML, páginas web) con fines de informes y análisis.

Su arquitectura está diseñada para consultas rápidas y eficientes, permitiendo la integración, limpieza y consolidación de datos históricos.

**Guía Técnica MGDE**

## 7. Data Mart

Un data mart es un subconjunto de un almacén de datos enfocado en una línea de negocio, departamento o área temática en particular. Contiene una parte pequeña y específica de los datos que la organización almacena en un sistema de almacenamiento más grande.

**Guía Técnica MGDE**

## 8. Dato

Representación de un atributo o variable cuantitativa o cualitativa, por medio de una secuencia de letras, números o símbolos y que es posible capturar a través de la observación y/o medición de una o más personas, objetos, hechos, juicios o fenómenos.

**NT de Interoperabilidad**

## 9. Catálogo de datos

Un catálogo de datos es una herramienta para crear y administrar el inventario de los activos de datos dentro de una organización. Maneja metadatos que permiten describir y clasificar el activo así como capacidades de búsqueda que permitan a los usuarios encontrar y acceder a la información de forma rápida y sencilla.

**Guía Técnica MGDE**

## 10. Metadatos

Datos que describen el contexto, el contenido y la estructura de un dato.

**NT de Interoperabilidad**

## 11. Datos maestros

Datos que proporcionan el contexto para los datos de la actividad del negocio en forma de conceptos comunes y abstractos que se relacionan con la actividad. Incluye los detalles (definiciones e identificadores) de objetos internos y externos involucrados en transacciones del negocio, como usuarios(as), productos, personal y dominios controlados (valores de código).

**Estrategia Nacional de Datos**

## 12. Datos de referencia

Conjuntos de datos que categorizan o relacionan otros datos por medio de códigos.

**Estrategia Nacional de Datos**

## 13. Datos abiertos

Dato digital con las características técnicas y jurídicas necesarias para que pueda ser usado, reutilizado y redistribuido libremente por cualquier persona u órgano de la Administración del Estado, en cualquier momento y lugar.

**NT de Interoperabilidad - Portal Datos Abiertos**

## 14. Seguridad de los datos

Conjunto de acciones, políticas, medidas preventivas y reactivas destinadas a la prevención, mitigación, manejo, respuesta y estudio de las amenazas y riesgos de incidentes de seguridad, a la reducción de sus efectos y el daño causado; antes, durante y después de su ocurrencia; respecto de los activos y activos de información y la continuidad de servicios, con el fin de proteger, preservar y restablecer la confidencialidad, integridad y disponibilidad de aquellos y de las plataformas electrónicas de los órganos de la Administración del Estado, aumentando su resiliencia en el tiempo.

**NT de Seguridad de la Información y ciberseguridad**

## 15. Linaje de datos

El linaje de datos es el proceso de seguimiento del flujo de datos a lo largo del tiempo, proporcionando una comprensión clara de dónde se originaron los datos, cómo han cambiado y su destino final dentro de la canalización de datos.

**IBM - Linaje de datos**

## 16. Interoperabilidad de datos

Cualquier operación que permita a dos órganos de la Administración del Estado, conectados directamente entre sí por medio de una pieza de software denominada nodo, intercambiar o transmitir datos, documentos o expedientes electrónicos.

**NT de Interoperabilidad**

## 17. Integración de datos

Procesos de movimiento, cruce y consolidación de datos en forma coherente dentro y entre almacenes de datos, aplicaciones y organizaciones.

### ***Estrategia Nacional de Datos***

#### Diferencias clave entre Integración e Interoperabilidad

- **Integración de datos** implica combinar y unificar datos de diversas fuentes en una estructura común.
- **Interoperabilidad de datos** se refiere a la capacidad de los sistemas de diferentes plataformas para intercambiar y usar datos de manera fluida, incluso si esos sistemas son diferentes tecnológicamente.

## 18. Inteligencia de Negocios

La inteligencia de negocios (BI) se encarga de todo lo referente al análisis de datos, desde las definiciones a las herramientas tanto para análisis descriptivo, diagnóstico, predictivo y prescriptivo, como apoyo a la toma de decisiones. El objetivo es apoyar la toma de decisiones, mejorando el análisis y la comprensión de los datos para facilitar decisiones más informadas y estratégicas.