



METADATOS

Proceso
Gestión Estratégica de Tecnologías
de la Información
Versión 01
25/04/2024

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

TABLA DE CONTENIDO

1.	INTRODUCCIÓN	3
2.	OBJETIVO	3
3.	ALCANCE	3
4.	DOCUMENTO ASOCIADO A LA GUÍA	4
5.	NORMATIVA.....	4
6.	DEFINICIONES	4
7.	QUÉ SON LOS METADATOS	4
8.	ADMINISTRACIÓN DE METADATOS Y MODELO DE MADUREZ	5
9.	FUENTES DE METADATOS	7
10.	PREGUNTAS BÁSICAS QUE DEBEMOS RESPONDER	8
11.	ENTORNO DE METADATOS GESTIONADOS.....	10
12.	CATEGORÍA DE LOS METADATOS	15
13.	GESTIÓN DE METADATOS.....	16
14.	BIBLIOGRAFÍA.....	20

LISTA DE IMÁGENES

Ilustración 1 Administración de Metadatos y Modelo de Madurez	7
Ilustración 2 Metadata Environment (MME)	11
Ilustración 3 Clasificación de los metadatos en función de la estructura de las capas	14
Ilustración 4 Modelo ampliado de arquitectura de calidad de datos.....	15



MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

1. INTRODUCCIÓN

En la era de la información y los datos masivos, el papel que juegan los metadatos ha cobrado cada vez más relevancia. Los metadatos, en esencia, son datos que describen otros datos, proporcionando información esencial sobre el contenido, el origen, la ubicación y las relaciones entre los datos. Esta información es fundamental para la gestión, uso y comprensión eficiente de los datos en una organización. Sin embargo, la gestión de metadatos puede resultar un desafío, ya que abarca diversas disciplinas, desde la arquitectura de la información y la administración de bases de datos hasta la gobernanza de datos y la administración de sistemas.

Este documento profundiza en varios aspectos esenciales de los metadatos, incluyendo su definición, las diversas fuentes de metadatos, las categorías de metadatos y las estrategias de administración de metadatos. El texto también presenta el concepto del modelo de madurez de metadatos, que a su vez proporciona una estructura para entender los requisitos de los metadatos.

Además, se examinará el papel que desempeña el linaje de los datos, un aspecto crucial de la administración de metadatos que permite rastrear el ciclo de vida completo de los datos. Finalmente, se contrastarán dos enfoques de almacenamiento de metadatos: el repositorio de metadatos centralizado y el federado, analizando sus ventajas y desventajas.

En este análisis, se consultará una serie de fuentes académicas y profesionales relevantes, incluyendo trabajos de autores reconocidos en el campo como Godínez et al., Graham, y la DAMA-DMBOK, además de recursos de Oracle. Este abordaje permitirá tener una visión completa e integral sobre los metadatos, su administración y su importancia en el mundo actual de los datos masivos.

2. OBJETIVO

Esta guía identifica los conceptos, las fuentes, el entorno, la categoría y la gestión que se deben tener en cuenta para describir información y servicios por medio de los metadatos, con el objeto de mostrar de manera lógica las relaciones entre los diferentes componentes del conjunto de metadatos.

3. ALCANCE

Es una propuesta orientada a la gestión de documentos electrónicos de archivo, con miras a garantizar la cadena digital de custodia y asegurar su autenticidad, integridad, fiabilidad y disponibilidad. Aplica para documentos electrónicos (texto, imagen, audio, video, otros) y puede ser aplicable a los documentos en soportes analógicos (físicos).

Es una guía para que cada área elabore su propio esquema, teniendo en cuenta que su uso no implica exoneración en el cumplimiento de Esquema de Metadatos.



MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

4. DOCUMENTO ASOCIADO A LA GUÍA

Procedimiento de Gestionar la Arquitectura de Información.

5. NORMATIVA

- ISO 19139:2007 Metadatos.
- NTC-ISO 23081-1 Información y documentación. Procesos para la gestión de registros. Metadatos para los registros: Parte 1: principios.
- NTC-ISO 23081-2. Información y documentación. Gestión de metadatos para los registros. Parte 2: aspectos conceptuales y de implementación
- Al elaborar la Guía, se buscó cumplir con el núcleo de Norma ISO 19115-1: 2014
 - La norma busca satisfacer los requerimientos de los usuarios de datos
 - documentación de los datos, conjuntos de datos y productos geográficos.

6. DEFINICIONES

- Metadatos: La definición más sencilla es aquella que especifica que los metadatos son datos acerca de los datos, es decir, aportan información sobre la identificación, alcance, calidad, esquema espacial y temporal, referencia espacial y distribución de datos espaciales digitales, permitiendo con ello el uso correcto de los datos para un fin determinado.
- Capas: Representación visual de un dataset geográfico en cualquier entorno de mapa digital. Conceptualmente, una capa es una porción o estrato de una realidad geográfica en una zona determinada, y es más o menos equivalente a un elemento de leyenda en un mapa de papel. En un mapa de carreteras, por ejemplo, las carreteras, parques nacionales, fronteras políticas y ríos se pueden considerar como capas diferentes

7. QUÉ SON LOS METADATOS

Los metadatos han sido ampliamente discutidos en la literatura y su definición ha evolucionado a lo largo del tiempo, adaptándose a los avances tecnológicos y a las necesidades cambiantes de la sociedad.

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

En un documento titulado "Metadata Demystified", publicado por NISO (National Information Standards Organization), se definen los metadatos como "datos estructurados que describen las características de los recursos de información" (NISO, 2019). Esta definición destaca el hecho de que los metadatos no son meramente datos, sino que están estructurados de tal manera que proporcionan información útil sobre otros datos.

Además, en el documento "What is LOM" (¿Qué es LOM?), publicado por la Universidad Roma Tre, se ofrece una visión más detallada de los metadatos, enfocándose en los metadatos de objetos de aprendizaje. Este documento describe los metadatos como "información estructurada que describe, explica, ubica, o hace de alguna manera más fácil recuperar, usar o gestionar un recurso de información" (Sciarra, 2005). Esta definición resalta el papel crucial que desempeñan los metadatos en la mejora de la recuperación y el uso de los recursos de información.

En "The Art of Enterprise Information Architecture", los autores Godínez et al., describen los metadatos como una forma de describir y catalogar los datos para que sean fácilmente localizables y utilizables (Godínez et al., 2010). De manera similar, Andy Graham en "The Enterprise Data Model" describe los metadatos como esenciales para entender la estructura, el propósito y el uso de los datos (Graham, 2012).

En resumen, la definición de metadatos ha evolucionado desde una descripción simple de los datos hasta una comprensión más profunda de su papel en la mejora de la recuperación, el uso y la gestión de los recursos de información. A pesar de las diferencias en las definiciones específicas, hay un acuerdo general de que los metadatos son esenciales para una gestión eficaz de la información en la era digital.

8. ADMINISTRACIÓN DE METADATOS Y MODELO DE MADUREZ

La administración de metadatos se refiere al manejo, organización, y utilización de metadatos para garantizar la coherencia, calidad y optimización del uso de los datos en una organización. Esta práctica abarca una variedad de actividades, incluyendo la creación, almacenamiento, búsqueda y eliminación de metadatos, además de la implementación de políticas y procesos para el manejo de estos.

En este contexto, el modelo de madurez de metadatos proporciona un marco útil para evaluar la capacidad de una organización para administrar eficientemente sus metadatos. Este modelo se estructura en cinco etapas, cada una representando un nivel creciente de sofisticación en la administración de metadatos.

1. Describir la tecnología con Metadatos (Metadatos Técnicos también llamado metadatos estructurales): Esta etapa inicial consiste en registrar y documentar los detalles técnicos de los



MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

sistemas y procesos que generan, almacenan y manejan datos. Los metadatos técnicos pueden incluir detalles sobre la estructura de las bases de datos, los esquemas de datos, las especificaciones de las interfaces de los sistemas y el software utilizado. Estos metadatos son cruciales para los equipos técnicos que implementan y mantienen las infraestructuras de TI y los sistemas de datos. En esta etapa, la administración de metadatos puede ser más reactiva que proactiva, con esfuerzos enfocados en el mantenimiento del día a día más que en la planificación estratégica.

2. Describir el Negocio con Metadatos (Metadatos de Negocio): La siguiente etapa se centra en los metadatos de negocio, que proporcionan un contexto para entender cómo los datos se relacionan con las operaciones y objetivos de la organización. Los metadatos de negocio pueden incluir definiciones de términos de negocio, reglas de negocio, procesos de negocio y relaciones entre entidades de negocio. Estos metadatos ayudan a garantizar que todos en la organización tengan una comprensión común de los datos y puedan utilizarlos eficazmente.

3. Metadatos de Negocio y Técnicos alineados: Esta etapa representa un hito en la madurez de la administración de metadatos. Los metadatos técnicos y de negocio se integran para proporcionar una visión unificada de los datos de la organización. Esto permite una mejor comunicación y colaboración entre los equipos técnicos y de negocio y mejora la capacidad de la organización para tomar decisiones basadas en datos. Además, la administración de metadatos en este punto puede empezar a ser más estratégica, con esfuerzos proactivos para mejorar la calidad de los metadatos y optimizar su uso.

4. Metadatos para Administrar el Negocio: En esta etapa, los metadatos se convierten en una herramienta clave para la toma de decisiones y la gestión de la organización. Los metadatos pueden usarse para identificar tendencias, medir el rendimiento, evaluar el impacto de los cambios y tomar decisiones informadas. La administración de metadatos puede incluir la implementación de políticas y procesos para garantizar la calidad y coherencia de los metadatos y para promover su uso en toda la organización.

5. Metadatos para Habilitar la Innovación en el Negocio: Esta es la etapa final de la madurez de la administración de metadatos. Los metadatos se utilizan no sólo para mantener y gestionar las operaciones actuales, sino también para impulsar la innovación y el crecimiento. Los metadatos pueden utilizarse para identificar oportunidades, desarrollar nuevas estrategias y modelos de negocio, y mejorar la eficiencia y la calidad. La administración de metadatos en este punto es altamente proactiva y estratégica, con un fuerte enfoque en el uso de metadatos para crear valor para la organización.

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45



9. FUENTES DE METADATOS

Los metadatos son esenciales para la comprensión y el uso efectivo de los datos. Se pueden clasificar según las fuentes de donde provienen. Estas fuentes varían dependiendo de la industria, la naturaleza de los datos y las necesidades específicas de una organización. Sin embargo, podemos agruparlas en tres categorías generales:

Metadatos generados por el sistema: Son metadatos producidos automáticamente por sistemas y aplicaciones de TI. Pueden incluir información como fechas de creación y modificación, identificadores de archivo, propietarios de archivos, etc. Por ejemplo, cuando se crea un documento en Microsoft Word, el sistema genera automáticamente metadatos sobre la fecha de creación, el autor, la última fecha de modificación y otros detalles similares.

Metadatos generados por el usuario: Son metadatos creados por los usuarios, a menudo para proporcionar contexto o información adicional sobre los datos. Por ejemplo, un usuario puede agregar etiquetas a una foto para describir su contenido, o un científico puede agregar metadatos a un conjunto de datos para explicar cómo se recogieron los datos y qué variables se midieron.

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

Metadatos importados de fuentes externas: Los metadatos también pueden obtenerse de fuentes externas, como bases de datos de terceros, bibliotecas de metadatos o proveedores de datos. Por ejemplo, un sistema de biblioteca puede importar metadatos de un catálogo de la biblioteca nacional para proporcionar información más detallada sobre los libros en su colección.

10. PREGUNTAS BÁSICAS QUE DEBEMOS RESPONDER

Para comprender y utilizar adecuadamente los metadatos, es crucial considerar algunas preguntas esenciales que estos deben responder. Este proceso está directamente ligado a la *Tabla 1 Preguntas básicas* (quién, qué, cuándo, dónde, por qué, cómo), que proporciona definiciones claras para cada tipo de conocimiento que se busca obtener de los metadatos.

Knowledge Type	Definition
Who	Se refiere a la(s) persona(s)
What	Usado como petición de información específica; para preguntar sobre el carácter, ocupación, etc. de una persona o cosa; para preguntar sobre el origen, identidad, etc. de algo; para preguntar sobre el valor, utilidad, fuerza o importancia de algo; o para referirse a cuánto es algo.
When	Se refiere al tiempo, a un periodo de tiempo, cuánto tiempo hace, cuándo, en qué circunstancias o en qué ocasión.
Where	Se refiere a una cosa o persona que está en o en un lugar, parte, punto, etc.
Why	Se utiliza para preguntar para qué, por qué razón, causa o propósito.
How	Usado para referirse a de qué modo o manera; por qué medios; en qué medida, grado, etc.; en qué estado o condición; por qué razón; por qué; a qué efecto; con qué significado; una cuestión relativa al modo o manera en que algo se hace se lleva a cabo, etc.; un modo o manera de hacer algo.

Tabla 1 Preguntas básicas

<https://www.ewsolutions.com/metadata-management-fundamentals/>

La gestión de metadatos es una parte integral de cualquier organización y cumple con dos roles cruciales, ambos de suma importancia. Uno de ellos es el papel que desempeña dentro del departamento de tecnología de la información (TI), mientras que el otro es su relevancia en la perspectiva empresarial de la organización.

Desde la perspectiva de TI, los metadatos actúan como una guía, dibujando un mapa detallado de cómo los datos se transfieren y se modifican a través de las infraestructuras y sistemas de TI. Esto

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

facilita el entendimiento de las relaciones interdependientes y ayuda a prever las implicaciones de cualquier cambio en los sistemas de TI. La información que los metadatos aportan es vital para la gestión de las bases de datos, ya que provee datos cruciales sobre su estructura, organización y características.

Por otra parte, en el contexto empresarial, los metadatos juegan un rol significativo al ofrecer el contexto necesario para interpretar y usar los datos de manera efectiva. Ayudan a los responsables de la toma de decisiones a comprender de dónde provienen los datos, cuán confiables son, cuál es su relevancia para diferentes aspectos del negocio y cómo pueden ser utilizados para obtener información valiosa. Los metadatos son un pilar fundamental de la gobernanza de datos, facilitando la comprensión y gestión de los datos a través de la organización.

Como lo explica Ralph Kimball en el libro "The Data Warehouse Toolkit" (tabla 2), los metadatos son esenciales para la administración de las bases de datos, ya que proporcionan información crucial para su estructura, organización y características.

Knowledge Type	Technical Example	Business Example
Who	¿Quién es el programador responsable de un determinado proceso de movimiento de datos?	¿Quién es el responsable principal del área temática CLIENTE?
What	¿Cuál es el linaje de datos entre nuestro sistema de clientes y nuestro almacén de datos de empresa?	¿Qué campo de nuestra analítica muestra la rentabilidad de nuestros productos?
What	¿Cuándo se ejecutan los trabajos de extracción, transformación y carga (ETL) y cuáles son las dependencias de cada trabajo?	¿Cuándo se actualizaron por última vez los datos que estoy analizando?
Where	¿En qué parte de nuestro entorno informático hay servidores que funcionan a menos del 40% de su capacidad?	¿Dónde tenemos un informe que muestre nuestros análisis de redes sociales por campaña de marketing?
How	¿Cómo se configuran los privilegios de seguridad de un nuevo analista?	¿Cómo calculamos los campos de nuestro informe principal?
Why	¿Por qué estamos experimentando más errores en la calidad de nuestros datos?	¿Por qué faltan algunos clientes en nuestros informes analíticos?

Tabla 2 Ejemplo Metadatos

<https://www.ewsolutions.com/metadata-management-fundamentals/>

Para ampliar aún más la complejidad y el valor de los metadatos, es esencial adentrarse en el universo del almacén de datos. En este entorno, los metadatos adquieren una variedad de formas y funciones para apoyar las necesidades variadas de los usuarios técnicos, administrativos y de negocios del almacén de datos.

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

Los metadatos operacionales de los sistemas fuente, incluyendo esquemas fuente y copybooks, facilitan el proceso de extracción de datos. Una vez que los datos se encuentran en el área de staging, los metadatos de staging guían los procesos de transformación y carga, incluyendo layouts de archivos de staging y tablas objetivo, reglas de transformación y limpieza, y resultados de logs de ejecución. En esencia, incluso el código de programación personalizado que se escribe en el área de staging se considera metadato.

Los metadatos que rodean el Sistema de Gestión de Bases de Datos (DBMS) del almacén de datos pueden contener detalles como las tablas del sistema, configuraciones de particiones, índices, definiciones de vistas y privilegios de seguridad a nivel de DBMS. Los metadatos de la herramienta de acceso a datos, por su parte, identifican nombres y definiciones de negocio para las tablas y columnas del área de presentación, así como restricciones de filtros, especificaciones de plantillas de aplicaciones, estadísticas de acceso y uso, y otra documentación de usuario.

La meta final es reunir, catalogar, integrar y luego aprovechar todas estas formas dispares de metadatos, similar a cómo se manejarían los recursos de una biblioteca. Aunque el esfuerzo para construir modelos dimensionales puede parecer menor en comparación, el desarrollo de un marco de metadatos para el almacén de datos no debe ser ignorado. Se debe elaborar un plan general para la gestión de metadatos, priorizando las entregas a corto plazo y considerando la necesidad de un repositorio para rastrear todos los metadatos (Kimball & Ross, 2013).

11. ENTORNO DE METADATOS GESTIONADOS

El entorno de metadatos gestionados (MME) representa los componentes arquitectónicos y los procesos necesarios para recopilar, retener y difundir metadatos en toda la empresa de manera sistemática y efectiva. Esta concepción incluye los conceptos de repositorios de metadatos, catálogos, diccionarios de datos y otros términos referidos al manejo organizado de metadatos.

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

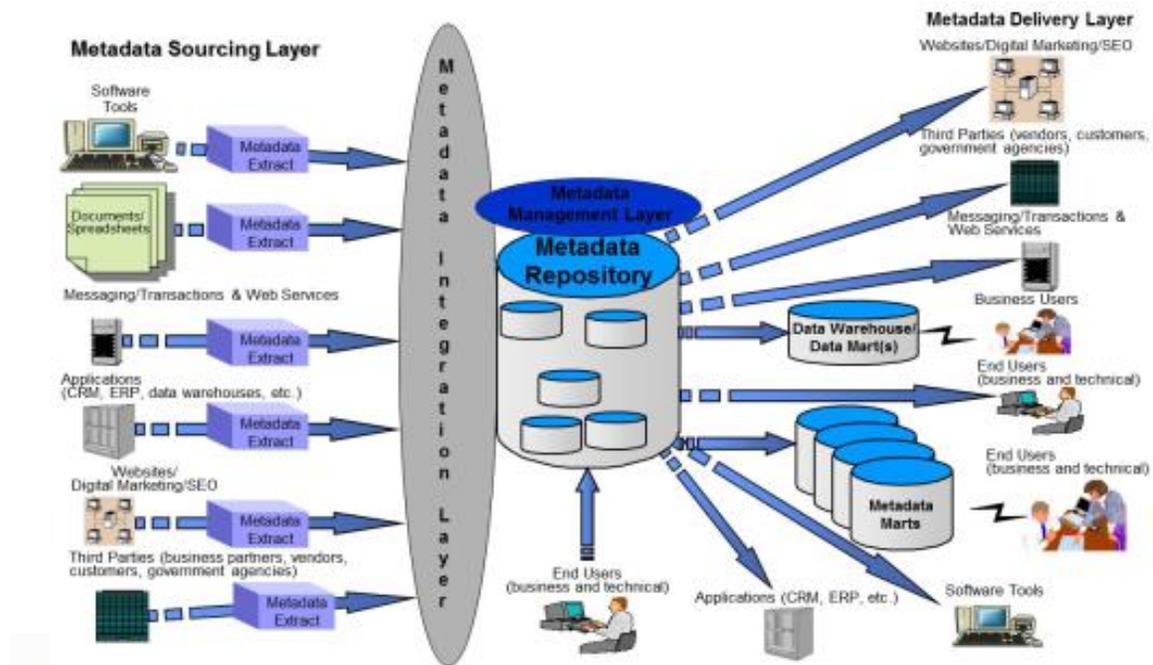


Ilustración 2 Metadata Environment (MME)

Una parte fundamental de este MME es el metamodelo, o modelo de metadatos. Estos términos son sinónimos y se refieren al modelo físico diseñado para almacenar metadatos. Un metamodelo se parece mucho a los modelos de datos que la mayoría de nosotros conoce, con elementos, tablas y relaciones. En este sentido, muchas de las mejores prácticas de modelado de datos son igualmente aplicables al modelado de metadatos, aunque presenta algunos desafíos específicos, el modelado de metadatos es una pieza clave para lograr un MME efectivo y eficiente que sirva para comprender y administrar la información en una organización.

Un metamodelo efectivo presenta cuatro características clave: es genérico, integrado, actual e histórico:

1. Genérico: Un metamodelo es genérico cuando busca almacenar metadatos por área temática de metadatos en lugar de ser específico de la aplicación. Al ser genéricos, los metamodelos pueden manejar los cambios en la infraestructura de la empresa, como la adopción de nuevos sistemas de bases de datos o cambios en las normas de la empresa.

Por ejemplo, se puede tener un metamodelo genérico que incluye un atributo llamado "FECHA TRANSACCIÓN" para registrar las fechas en que se realizan las transacciones en cualquier parte de la empresa.

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

Por otro lado, un metamodelo específico de la aplicación podría tener un atributo llamado "FECHA VENTA" en el sistema de punto de venta, y otro atributo llamado "FECHA COMPRA" en el sistema de gestión de inventario. Este enfoque se vuelve problemático si la empresa desea obtener una visión unificada de todas las transacciones en toda la empresa. La necesidad de tener que reconciliar "FECHA VENTA" y "FECHA COMPRA" en un marco de tiempo común puede ser un desafío innecesario.

En este caso, un metamodelo genérico con un atributo "FECHA TRANSACCIÓN" puede ser mucho más útil para proporcionar una visión coherente y unificada de las transacciones en toda la empresa, independientemente del departamento o la aplicación de origen.

2. Integrado: Los metamodelos proporcionan una vista integrada de las principales áreas temáticas de los metadatos de la empresa. Son capaces de vincular diferentes tipos de metadatos, como los metadatos técnicos y de negocio, en un solo marco. Esto facilita la toma de decisiones basada en datos y permite una comprensión más completa y coherente de los datos.

Un claro ejemplo de esto sucede en las organizaciones que administran un CRM y ERP. Cada uno de estos sistemas tiene sus propios metadatos asociados. Por ejemplo, el sistema CRM puede tener metadatos que describen los atributos del cliente, como el nombre del cliente, la ubicación, las preferencias de compra, entre otros, mientras que el sistema ERP puede tener metadatos que describen detalles del producto, como el nombre del producto, el precio, la disponibilidad de inventario, entre otros.

Si estos metadatos se almacenan en repositorios separados y no están integrados, la organización tendría dificultades para obtener una visión unificada de cómo los clientes interactúan con sus productos. Diferente, si existe un metamodelo integrado ya que permitiría a la organización vincular fácilmente los metadatos del sistema CRM con los metadatos del sistema ERP. De esta manera, podrían relacionar fácilmente los atributos del cliente con los detalles del producto, proporcionando una visión completa y unificada que apoye el análisis de datos integral.

3. Actual: Un metamodelo efectivo contiene metadatos que se relacionan tanto con el entorno actual como con el futuro o planificado. Deben ser capaces de acomodar y reflejar los cambios en los datos y las operaciones de la empresa, lo cual es crucial para apoyar la toma de decisiones estratégicas y la planificación.

Un ejemplo, es si una organización está considerando actualizar su sistema de registro médico electrónico (EMR), el cual almacena una variedad de metadatos, como detalles de pacientes, historiales médicos, tipos de procedimientos, medicamentos recetados, entre otros.

Un metamodelo actualizado ayuda en este caso, permitiendo a la organización comparar los metadatos actuales con los del nuevo sistema. Por ejemplo, si el nuevo sistema EMR tiene campos

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

para metadatos que el sistema actual no tiene, estos pueden ser identificados por adelantado. De la misma manera, si hay campos de metadatos en el sistema actual que no se encuentran en el nuevo sistema, estos también pueden ser identificados.

Esto permite a la organización a prepararse para la migración y a planificar cualquier adaptación necesaria en los procesos de trabajo.

4. Histórico: Los metamodelos deben ser capaces de mantener un historial de los metadatos. Esto permite a las empresas entender cómo han evolucionado sus datos y operaciones a lo largo del tiempo. Es especialmente útil para el análisis de tendencias y la toma de decisiones basada en datos históricos.

Para ilustrar la característica histórica de un metamodelo, un claro ejemplo es un banco que ha estado en operación durante varias décadas. A lo largo de esos años, el banco ha introducido y eliminado diversos productos financieros (como cuentas de ahorro, tarjetas de crédito, hipotecas, préstamos para estudiantes, etc.) y ha hecho cambios en la estructura y condiciones de estos productos basándose en factores como cambios regulatorios, condiciones del mercado y estrategias internas.

Un metamodelo histórico almacenaría metadatos que representan todas las versiones de estos productos financieros a lo largo del tiempo, incluyendo detalles como los términos y condiciones, las tasas de interés y las características del producto. Esto permitiría al banco entender cómo ha evolucionado su oferta de productos y analizar tendencias a lo largo del tiempo.

Por ejemplo, si el banco está planeando introducir un nuevo producto, puede mirar los metadatos históricos de productos similares para predecir posibles resultados. O, si se produce un problema con un producto (como un alto índice de impago en una determinada hipoteca), el banco podría revisar los metadatos históricos para entender cómo han cambiado las condiciones de ese producto a lo largo del tiempo y si esos cambios pueden haber contribuido al problema.

En un artículo académico de Kyung-Seok Ryu, Joo-Seok Park, y Jae-Hong Park (2021), se define la arquitectura de metadatos en el contexto de la gestión de la calidad de los datos. Sugieren que los metadatos deben categorizarse en metadatos lógicos, físicos y de mapeo para mejorar la calidad de los datos en toda la organización. Este enfoque de clasificación de metadatos está en línea con las características esenciales de un metamodelo eficaz.

Los metadatos lógicos, que incluyen elementos de datos y modelos de datos, alinean con la idea de un metamodelo genérico e integrado, que busca almacenar metadatos por área temática de metadatos en lugar de ser específico de la aplicación y proporciona una visión integrada de las principales áreas temáticas de los metadatos de la empresa.

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

Los metadatos físicos, que abarcan aspectos como el nombre de la tabla y el esquema, son esenciales para mantener un metamodelo actual, que refleja el entorno actual de la infraestructura de datos de la empresa.

Finalmente, la información de mapeo de metadatos, que incluye reglas de transformación, es crucial para mantener un metamodelo histórico, que permite a las empresas entender cómo han evolucionado sus datos y operaciones a lo largo del tiempo.

Por lo tanto, la clasificación de los metadatos en lógicos, físicos y de mapeo puede verse como un enfoque pragmático para implementar un metamodelo efectivo, que puede apoyar la gestión de la calidad de los datos en toda la organización.

Category	Definition	Example
Standard data information	Logical information of common data factors (business object) which are defined at the enterprise level	Information of common data for the enterprise data model: entity information, attribute information, domain information, etc.
Physical database information	System catalogue information of the physical database systems which are generated at each base systems	Physical DBMS management information: database names, table names, view names, column name, etc.
Mapping information of standard data and physical database	Information which manages the relationship between standard data and physical database	Mediator information: source and target mapping information, transformation rules, etc.

Ilustración 3 Clasificación de los metadatos en función de la estructura de las capas

Además, el artículo propone que los factores de gestión de la estructura de los datos se deben derivar de acuerdo con cada proceso de madurez. Las empresas pueden evaluar su nivel de gestión de calidad de datos y mejorar la calidad de los datos a un nivel más avanzado a través de la definición de procesos de madurez.

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

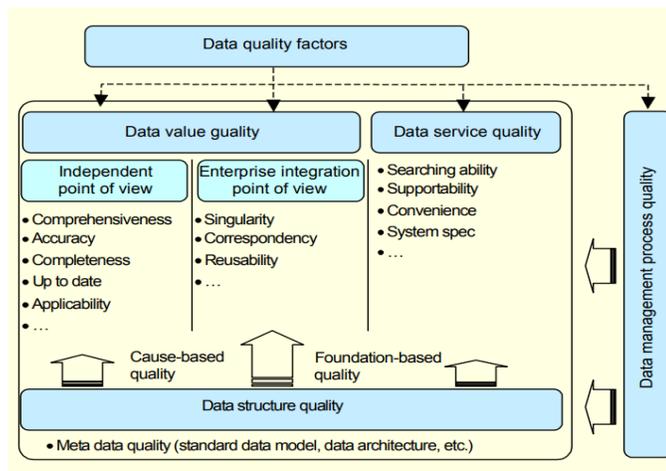


Ilustración 4 Modelo ampliado de arquitectura de calidad de datos

12. CATEGORÍA DE LOS METADATOS

Un catálogo de datos es una herramienta que ayuda a las organizaciones a gestionar sus datos. Esta recopila y organiza información sobre los datos de una organización, lo que facilita a los usuarios encontrar y comprender los datos que necesitan.

El catálogo de datos se ocupa principalmente de tres tipos de metadatos: técnicos, de proceso y empresariales.

Metadatos técnicos (también conocidos como metadatos estructurales): describen la estructura de los datos. Incluyendo detalles como tablas, columnas, filas, índices y conexiones. Proporcionan información sobre cómo trabajar con los datos, y si es necesario, cómo transformarlos para su análisis o integración.

Los metadatos técnicos pueden abarcar diferentes tablas como por ejemplo "Equipos", "Localizaciones", "Componentes Físicos", "Programas", entre otros. En la tabla "Equipos", por ejemplo, se podrían encontrar campos como el identificador único del equipo, la dirección IP, su localización, los componentes físicos que integra y los programas que tiene instalados.

Metadatos de procesos (también llamados metadatos administrativos): describen el valor de los datos para la organización. Incluyen la creación y manipulación de los datos, responden al cuándo, cómo y quién accedió, usó, actualizó o modificó los datos. Asimismo, indican quién tiene permiso para acceder y utilizar los datos.

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

Los metadatos de procesos podrían incluir registros incluyendo la identidad de la persona que lo ha consultado o realizado cambios, y cuándo tuvo lugar esta acción. Adicionalmente, los metadatos de procesos demarcan los permisos de acceso y modificación de los datos del servidor, identificando claramente quién tiene la autoridad para manipular dicha información.

Metadatos empresariales (a veces denominados metadatos externos): se refieren a la información comercial relevante de los datos, incluyendo su valor para la organización, su aplicabilidad para diversos propósitos y su cumplimiento con las normativas. Estos metadatos facilitan la comunicación entre los profesionales de los datos y los usuarios empresariales.

Un ejemplo podría ser que los metadatos reflejen la importancia de cada producto/servicio en términos de ventas o beneficios. Un producto que se vende frecuentemente o aporta un alto margen de beneficio puede tener un alto valor comercial para la empresa.

13. GESTIÓN DE METADATOS

La gestión de metadatos es un proceso continuo que se puede dividir en varios pasos coherentes para una comprensión más fácil. A continuación, se presenta una descripción detallada de cada uno de estos pasos:

1. Identificación de los metadatos: El primer paso en la gestión de metadatos es la identificación. debes determinar qué metadatos necesitas para satisfacer tus necesidades de negocios y técnicas. Estos pueden variar desde metadatos descriptivos, como los nombres de las columnas de la base de datos, hasta metadatos operativos, como la fecha y hora de la última actualización de un conjunto de datos.

En esta etapa inicial, una organización necesita identificar cuáles metadatos son necesarios para su operación. Por ejemplo, una organización puede necesitar metadatos descriptivos, como "nombre del producto", "precio", "descripción del producto", y metadatos de uso, como "veces visto" o "veces comprado".

2. Recopilación de metadatos: En la fase de recolección de metadatos, se recolectan de forma activa todos los metadatos identificados en la etapa uno. Esta recopilación de metadatos implica reunir datos descriptivos y técnicos de todas las fuentes relevantes.

Es posible realizar esta recopilación de manera manual, a través de un esfuerzo intensivo, sin embargo, dada la cantidad y complejidad de los metadatos, este enfoque manual puede resultar en errores humanos y ser extremadamente lento y costoso.

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

Por tanto, la mayoría de las organizaciones optan por utilizar herramientas de recopilación de metadatos automatizadas para llevar a cabo esta tarea de manera más eficiente y precisa. Estas herramientas, que incluyen software especializado como Apache Atlas, Informática Metadata Manager, IBM InfoSphere, entre otros, que pueden interactuar con una variedad de sistemas de datos y extraer automáticamente los metadatos requeridos.

Estas herramientas de recopilación de metadatos suelen ofrecer varias funcionalidades clave. Por ejemplo, pueden interactuar con bases de datos SQL y NoSQL, sistemas de archivos, servicios en la nube y otras fuentes de datos para extraer metadatos. Además, pueden identificar y extraer metadatos de estructuras de datos complejas, como tablas de bases de datos y esquemas XML. También pueden extraer metadatos de transformaciones de datos, como scripts de Extracción, Transformación y Carga (ETL) y procesos de limpieza de datos. Una vez que los metadatos han sido extraídos, estas herramientas pueden también normalizar los metadatos a un formato estándar, facilitando su almacenamiento y análisis posterior.

3. Almacenamiento de metadatos: Una vez que se han recolectado y normalizado los metadatos, el siguiente paso es encontrar un lugar adecuado para almacenarlos. Para esto se utiliza un repositorio de metadatos que es una base de datos especialmente diseñada para contener metadatos. Existen dos variantes de estos repositorios: los centralizados y los federados.

En un repositorio de metadatos centralizado, se estaría eligiendo un sistema en el que todos los metadatos de la organización se guardan en un solo lugar. Dependiendo de las necesidades concretas de la organización, este lugar puede ser una base de datos SQL, una base de datos NoSQL, o incluso un sistema de almacenamiento en la nube. La ventaja de este sistema es su simplicidad: tener todos los metadatos en un solo lugar puede facilitar y agilizar su acceso y análisis.

Por otro lado, en organizaciones más grandes o con múltiples departamentos, tener un repositorio centralizado podría no ser la opción más eficiente. Para estos casos, existe el repositorio de metadatos federado. Este tipo de repositorio guarda metadatos en distintas ubicaciones, permitiendo que cada departamento administre su propia parte. Aunque esto pueda ser más complejo desde el punto de vista técnico, estos repositorios están diseñados para proporcionar un punto de acceso unificado. Esto significa que, sin importar en qué lugar se encuentren almacenados, se pueden consultar y acceder a los metadatos como si estuvieran en un solo lugar.

4. Uso de metadatos: Esta fase implica poner en práctica los metadatos recolectados para mejorar las operaciones y tomar decisiones informadas. Los metadatos pueden emplearse para realizar análisis de impacto, rastrear el linaje de los datos y obtener conocimientos valiosos sobre diferentes aspectos de la empresa.

El linaje de datos es un componente crítico en la gestión de metadatos, ya que representa la vida de un dato específico: dónde se originó, qué transformaciones ha sufrido, dónde se ha movido y cómo se



MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

ha utilizado a lo largo del tiempo. Esto proporciona una visión completa del dato, permitiendo un mejor entendimiento de su contexto y confiabilidad.

Las herramientas ETL capturan metadatos durante cada etapa del proceso ETL, lo que permite rastrear el movimiento y transformación de los datos desde su origen hasta su destino final. Esto es especialmente útil en entornos complejos, donde los datos pueden moverse y transformarse a través de varios sistemas y procesos.

Además, el linaje de datos está el gobierno de datos, términos interconectados que, juntos, forman un marco sólido para garantizar la calidad, seguridad y trazabilidad de los datos a lo largo del tiempo. El gobierno de datos define una estructura organizacional para la gestión de los activos de datos, estableciendo propietarios de datos, términos de negocio, reglas y políticas a lo largo del ciclo de vida de los datos. Las soluciones de linaje de datos ayudan a los equipos de gobierno de datos a garantizar que los datos cumplan con estos estándares, brindando visibilidad sobre los cambios de los datos a lo largo de su ciclo de vida.

5. Mantenimiento de metadatos: El mantenimiento de metadatos es un proceso continuo que garantiza que los metadatos sean precisos, relevantes y actualizados. Esto es importante para garantizar que los metadatos sean útiles para los usuarios y que los datos subyacentes sean accesibles y comprensibles.

El mantenimiento de metadatos puede incluir las siguientes actividades:

- Actualización de los metadatos para reflejar los cambios en los datos subyacentes.
- Corregir cualquier inexactitud o inconsistencia en los metadatos.
- Eliminar los metadatos obsoletos o innecesarios.
- Hay que asegurar que los metadatos están organizados y fácilmente accesibles.
- Proteger los metadatos de acceso no autorizado o modificación.

El mantenimiento de metadatos es una tarea importante que debe ser realizada por un equipo de profesionales con experiencia en gestión de datos.

6. Seguridad de los metadatos: La seguridad de los metadatos es un aspecto muy importante que debe ser considerado al gestionarlos. Los metadatos pueden contener información confidencial, como nombres de usuarios, contraseñas, números de tarjetas de crédito, etc., por lo que es importante protegerlos de accesos no autorizados.

Algunas medidas de seguridad que se pueden tomar para proteger los metadatos incluyen:

- Restricción del acceso a los metadatos a personas autorizadas.
- Encriptación de los metadatos almacenados.
- Monitorización constante del acceso a los metadatos para detectar cualquier actividad sospechosa.

MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

- Implementación de políticas y procedimientos de seguridad para el manejo de metadatos.

La seguridad de los metadatos es una responsabilidad compartida entre los dueños de los datos, los administradores de datos y los usuarios de datos. Todos los involucrados deben ser conscientes de la importancia de proteger los metadatos y deben tomar las medidas necesarias para hacerlo.



MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	METADATOS	 Sistema Integrado de Gestión
	Proceso: Gestión Estratégica de Tecnologías de la Información	
Versión: 1	Vigencia: 25/04/2024	Código: G-E-GET-45

14. BIBLIOGRAFÍA

Godinez, M., Hechler, E., Koenig, K., Lockwood, S., Oberhofer, M., Schroeck, M. (2010). The Art of Enterprise Information Architecture. IBM Press.

Graham, A. (2012). The Enterprise Data Model: A framework for enterprise data architecture. 2ª edición. Koios Associates Ltd.

NISO. (2019). Metadata Demystified. Recuperado de https://www.anzsi.org/wp-content/uploads/2019/05/Metadata_Demystified-1.pdf

Sciarra, F. (2005). What is LOM. Recuperado de <http://www.dia.uniroma3.it/~sciarro/e-learning/WhatIsLOMScreen.pdf>

Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (3rd ed.). Wiley.

Ryu, K. S., Park, J. S., & Park, J. H. (2006). A data quality management maturity model. ETRI journal, 28(2), 191-204.

<https://www.ewsolutions.com/metadata-management-fundamentals/>

<https://www.ibm.com/es-es/topics/data-catalog>

<https://www.zencos.com/blog/analytics-maturity-model-5-step-guide/>

