

data Bizkaia

Gobierno de Datos en la Era Digital

Desafíos y Oportunidades



Gestión de Metadatos: tipos y fuentes



¿Qué es un Metadato?

“Datos sobre datos”

Describen los datos en sí.

Por ejemplo:

- BBDD
- Reglas
- Restricciones y permisos de acceso
- Procesos técnicos o de negocio
- Aplicaciones, Infraestructura tecnológica
- Relaciones y conexiones
- Responsable, autoría, contacto...



En definitiva, ayudan a las organizaciones a comprender sus datos, sus sistemas y sus flujos de trabajo.

Importancia

Sin metadatos fiables, una organización **no sabe qué datos tiene**, qué representan, dónde se originan, cómo se mueven por los sistemas, quién tiene acceso a ellos o qué significa que los datos sean de alta calidad.

Sin metadatos, una organización no puede gestionar sus datos como un activo.

”

Sin metadatos, una organización puede no ser capaz de gestionar sus datos en absoluto.

-DAMA, DMBOK



Vídeo

Link: <https://www.youtube.com/watch?v=W9df022FHLE>





¿Qué es?



¿Ahora?





DATA

Filename: Tadzik.jpg
Author: Piotr Kononow
Date: August 15, 2016 6:40:10PM
File: 5,312 × 2,988 JPEG
15.9 megapixels
3,393,448 bytes
(3.2 megabytes)
Camera: Samsung SM-G920F
4.3 mm
Lens: Max aperture f/1.9
(shot wide open)
Auto exposure
Program AE
Exposure: 1/402 sec
f/1.9
ISO 40
Flash: none



METADATA

Tipos de **Metadatos**

Dentro del ámbito de TI

Metadatos de Negocio

Se centran en gran medida en el **contenido y la condición** de los datos e incluyen detalles relacionados con el **gobierno** de los datos.

Ejemplos:

- Definiciones y descripciones de conjuntos de datos.
- Origen y linaje
- Información de contacto
- Responsable

Metadatos Técnicos

Proporcionan información sobre los **detalles técnicos** de los datos, los **sistemas** que los almacenan y los **procesos** que los mueven dentro de los sistemas y entre ellos.

Ejemplos:

- Reglas CRUD
- Nombres de programas y aplicaciones
- Backup y copias de seguridad
- Nombres de columnas y tablas

Metadatos Operacionales

Describen los detalles del **procesamiento y el acceso** a los datos

Ejemplos:

- Registro de errores
- Resultado de auditorías
- Fecha y hora de creación, actualización...
- Historial de cambios

Tipos de **Metadatos**

Fuera del ámbito de TI

Metadatos Descriptivos

Describen un recurso y permiten localizarlos (título, autor/a, tema...)

Metadatos Estructurales

Describen las relaciones entre los recursos, componentes (nº capítulos, nº páginas).

Metadatos Administrativos

Para la gestión de los recursos en su ciclo de vida: (nº versión, fecha publicación...)



Fuentes de Metadatos



Business glosary o **Glosario de negocio**

- Centrado en el **lenguaje de los negocios**
- **Fácil de entender** por cualquiera.
- Los términos empresariales no están destinados a definir datos, metadatos, transformaciones o ubicaciones, sino más bien a **definir lo que cada término significa** en un sentido empresarial.



¿Qué es una notificación? Y ¿Un expediente?

- Disponer de un glosario de negocios aporta una **comprensión común del vocabulario** utilizado en toda una organización.
- El **ámbito de aplicación** de un glosario de negocios debe ser de **toda la organización**.

Debido a su alcance y a los conocimientos técnicos necesarios, la **responsabilidad** del glosario de negocios recae en la empresa y no en la tecnología.

Business glosary o **Glosario de negocio**

	Plantilla Glosario de Negocio									
Dominio/Área de Negocio	Data Owner									
Identificador	Término	Definición	Referencia	Estado	Origen Definición	Comentario Negocio	Comentario Gobernanza	Sigla	Ejemplo	Identificador de Linaje
7	Comarca	Una comarca es una región geográfica que se diferencia del resto por sus características históricas, sociales, culturales o geográficas. En algunos países, las comarcas son entidades administrativas formadas por varios municipios. La delimitación de una comarca puede obedecer a diferentes razones e intereses. <u>Cada comarca tiene asociada una codificación INE oficial</u>		Candidato	INE				Arratia-Nerviön	
8	Provincia	Entidad local con personalidad jurídica propia, determinada por la agrupación de Municipios y división territorial para el cumplimiento de las actividades del Estado. <u>Cada Provincia tiene asociado un código INE de 2 dígitos comprendido entre el 01 y el 52</u>		Aprobado	INE				Cáceres	
69	Territorio Histórico	Entidad territorial peculiar de la Comunidad Autónoma del País Vasco, coincidente con la provincia de idéntica denominación. Está regido por una diputación foral, la cual ejerce no solo las competencias ordinarias de una diputación provincial, sino también competencias propias del régimen foral como, por ejemplo, en materia fiscal; y por juntas generales, órganos legislativos facultados para aprobar normas o leyes forales (si bien éstas no tienen rango de ley). Cada junta general está compuesta por 51 miembros, elegidos por sufragio universal en cada territorio. Dentro de este epígrafe se encontrarían Álava-Araba, Gipuzkoa y Bizkaia. La codificación oficial asociada sería de origen INE, dentro del compendio de <u>codificaciones de Provincia</u> .		Candidato					Cáceres	
9	Municipio	Entidad compuesta por un territorio claramente definido por un término municipal de límites fijados. Los municipios tienen asociados un código INE oficial de 5 dígitos, de los cuales los dos primeros hacen referencia a la provincia y los 3 últimos al municipio dentro de cada provincia.		Candidato	INE				Alcorcón	
10	Distrito Censal	Agregación de secciones censales completas decidida por cada Ayuntamiento para su término municipal. La división en distritos conduce al resultado siguiente: - C. A. de Euskadi: 343 distritos - Alava: 57 distritos - Gipuzkoa: 119 distritos - Bizkaia: 173 distritos La codificación asociada está compuesta por 10 dígitos de los que los 2 primeros hacen referencia a la provincia, los 3 siguientes al municipio, los 2 siguientes al distrito censal y los 2 últimos de forma unívoca a la sección censal		Candidato	INE				Arganzuela; 0100101	
11	Barrio	División en una ciudad de un área residencial que tiene características afines o simplemente ha sido delimitada por criterios administrativos.		Candidato	INE				Uríbarri	
12	Código Postal	Combinación de números que se asigna a una población y a las distintas zonas dentro de ella para hacer más fácil la clasificación y entrega del correo.		Candidato	CORREOS			CP	28924	
13	Sección Censal	Partición del término municipal caracterizada por: - estar preferentemente definida mediante límites fácilmente identificables, tales como accidentes naturales del terreno, construcciones de carácter permanente y viales. - tener un tamaño entre 1.000 y 3.500 habitantes de derecho, excepto en el caso que el municipio completo tenga una población inferior. Como resultado de la aplicación de estos criterios en el momento de referencia de la		Candidato	INE			SSCC	0100101001	

Catálogo de Datos

El catálogo de datos sirve como un **directorio** único para **localizar información** y además proporciona el mapeo entre el glosario de negocios y los diccionarios de datos.

El catálogo de datos es un activo de toda la empresa que proporciona una única fuente de referencia para la localización de cualquier conjunto de datos necesarios para diversas necesidades como las operacionales, BI, analíticas, de ciencia de los datos, etc.

Normalmente se desarrolla después de la creación satisfactoria tanto del Business Glossary como de un diccionario de datos, pero también puede crearse de forma incremental a medida que los otros dos activos evolucionan con el tiempo.

NETFLIX

Inicio

Series

Películas

Novedades más vistas

Mi lista

Explorar por idiomas

Infantil

6

Películas

Géneros

Acción

Anime

Ciencia ficción

Clásicas

Comedias

Cortos

Deportes

Documentales

Dramas

Españolas

Europeas

Extranjeras

Fantásticas

Halloween

Hollywood

Independientes

Música y musicales

Para toda la familia

Premiadas

Románticas

Terror

Thrillers

Recién llegada

TOP 10

Casa en Llamas

Recién llegada

EL CUARTO PASAJERO

DETONANTES

BAJO TERAPIA

Recién llegada

TOP 10

Solo en Netflix

UNA AVENTURA EN MARRUECOS

LOS HERMANOS MENENDEZ

PROBLEMAS...

SWEET BOBBY

OLIVIA RODRIGO

SOLO UNA NOCHE DE JUEGOS

Novedades en Netflix

Casa en Llamas

NO TE MUEVAS

LOS HOMBRES LOBO

COLLATERAL REALITY

EL EXORCISTA DEL PAPA

OLIVIA RODRIGO



STALLONE
RAMBO
EL REGRESO

PROBA DE FE

MATT D

Diccionario de Datos

- **La fuente de metadatos más técnica** de las 3.
- Debe centrarse en las descripciones y detalles que conlleva el almacenamiento de datos.
- Debería haber **un diccionario para cada fuente** de datos de la empresa.
- Incluye detalles sobre los datos como el **tipo de datos, la longitud permitida, el linaje, las transformaciones**, etc.:
 - Estos metadatos ayudan a entender cómo unir, consultar y reportar los datos, y también explica la granularidad.

Debido a la necesidad de conocimientos técnicos y metadatos, la **responsabilidad** de la propiedad de un diccionario de datos recae en **TI**, frecuentemente con funciones como administración de bases de datos, ingeniería de datos, arquitectura de datos...

Soluciones **informáticas**



Figure 1. Magic Quadrant for Metadata Management Solutions



Source: Gartner (November 2020)

Gestión de **metadatos**

Al igual que los datos,
los metadatos también requieren gestión.

A medida que aumenta la capacidad de las organizaciones para recopilar y almacenar datos, crece la importancia del papel de los metadatos en la gestión de datos.

Para ser data-driven, una organización debe ser metadata-driven.



Gestión de metadatos.

Beneficios.

- Aumentar la **confianza** en los datos proporcionando contexto y permitiendo medir su calidad.
- Aumentar el **valor** de la información estratégica (por ejemplo, los datos maestros) permitiendo múltiples usos.
- Mejorar la **eficiencia operativa** identificando datos y procesos redundantes.
- Evitar el uso de datos obsoletos o incorrectos.
- **Reducir el tiempo de investigación** orientada a los datos.
- **Mejorar la comunicación** entre las consumidoras de datos y las profesionales de TI.
- Crear **análisis de impacto precisos** reduciendo así el riesgo de fracaso del proyecto.
- Mejorar el *time to market*.
- **Reducir los costes de formación** y el impacto de la rotación de personal mediante una documentación exhaustiva del contexto, la historia y el origen de los datos.
- Respalda el **cumplimiento de la normativa**.



Gestión de metadatos. **Valor y Riesgos.**

- Dado que explican los datos y procesos que permiten el funcionamiento de las organizaciones, **los metadatos son fundamentales para el gobierno de los datos.**
 - Son una guía de los datos de una organización → deben gestionarse bien.
- **Mal gestionados** nos conducen a:
 - Datos redundantes
 - Definiciones incoherentes
 - Riesgos asociados al uso indebido de los datos.
 - Fuentes y versiones de metadatos que compiten entre sí y entran en conflicto
 - Dudas sobre la fiabilidad de los datos
 - Falta de confianza en los datos
 - Parches y más parches (if, if, if...)

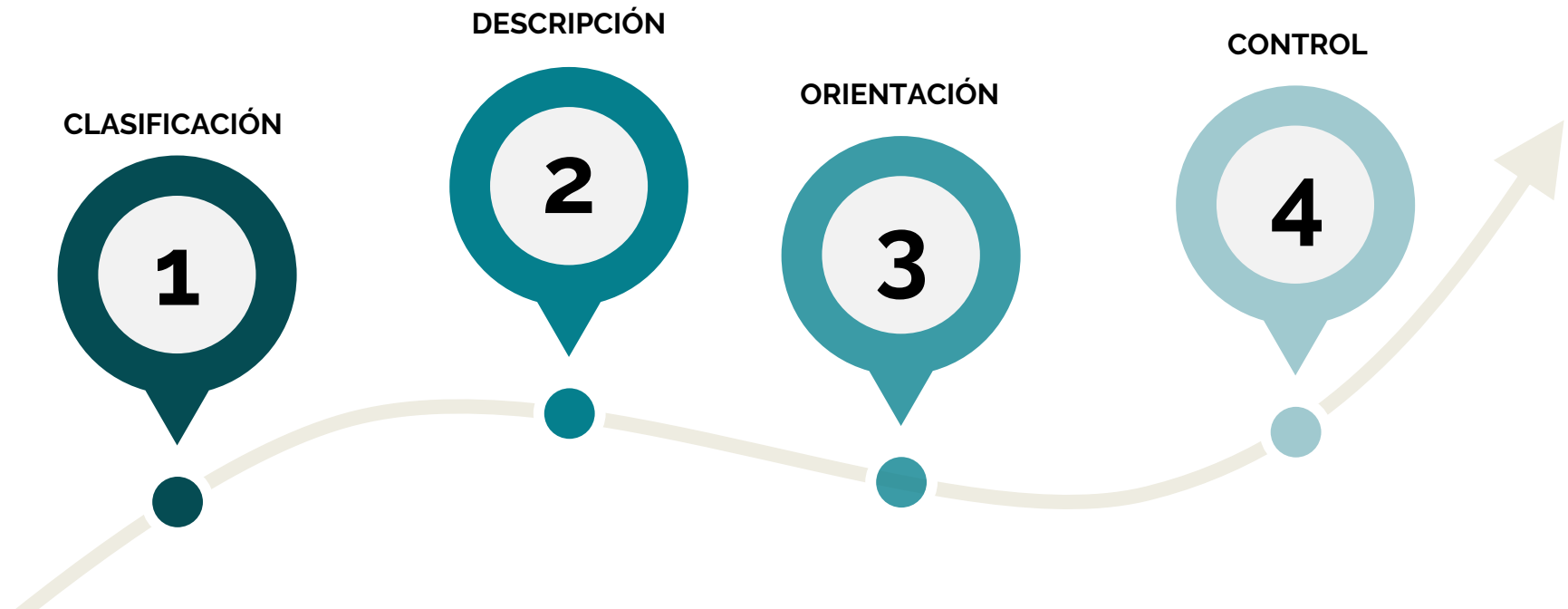


Gestión de metadatos. **Fases.**

Si el metadato es la base del Gobierno de los Datos... entonces es **absolutamente necesario gestionarlo**.

¿Cómo?

Con una **herramienta** de Gestión de Metadatos o con un producto de Gobierno del Dato pero lo que es seguro es que pasaremos por las siguientes **fases**:



1. Clasificación

Los datos pueden tener muchas características por las que se pueden **agrupar o clasificar**. Tener esta categorización **permitirá organizarlos y gestionarlos**. Por ejemplo, por:

- **Tema:** datos financieros, datos de estudiantes, datos fiscales, datos de salud, datos de productos, etc.
- **Uso:** transaccional, analítico, regulatorio, etc.
- **Tiempo:** datos en vivo y actuales, históricos, predictivos, etc.
- **Contenido:** datos geoespaciales, datos de máquinas, datos estructurados vs. No estructurados, etc.
- **Alcance:** datos empresariales, externos, departamentales, maestros, etc.



2. Descripción

Describir los datos **ayuda a entender** tanto sus aspectos lógicos como físicos. Los datos descritos deben incluir:

- **Significado:** Definiciones comerciales, entidades de modelado de datos y atributos.
- **Estructura:** Descripción de los objetos de datos (entidades, tablas, registros, etc.), sus agrupaciones y relaciones lógicas
- **Contenido:** Los tipos de datos como la fecha, la moneda, el texto, el número, etc.
- **Valores:** Qué valores están permitidos, qué datos de referencia están disponibles, qué patrones o rangos de valores deben seguir, qué restricciones deben cumplir, etc.
- **Linaje:**Cuál es la fuente de los datos, cómo se crearon, derivaron y/o calcularon los datos, cómo se transformaron, etc.



3. Orientación

El Metadato puede **servir de guía** a cualquier persona técnica o de negocio para encontrar los datos que necesita, a través de motores de búsqueda u otros procesos. Estos metadatos de orientación pueden estar compuestos por:

- **Palabras clave:** Estos podrían ser cualquier metadato descrito hasta ahora.
- **Taxonomías:** Otro ejemplo más de cómo la clasificación ayuda.
- **Sellos de fecha y hora:** Normalmente se añaden automáticamente a nivel de tabla o fila.
- **Informes asociados, procesos, personas:** Saber dónde surgen los datos, quiénes son las personas usuarias o administradoras de los datos, cómo se capturan y transforman los datos podría servir como un buen punto de partida para encontrar lo que se necesita.
- **Sinónimos, alias, términos relacionados.**



4. Control

Los metadatos pueden proporcionar los conocimientos necesarios para determinar **qué controles deben aplicarse a los datos y qué datos deben ser controlados**.

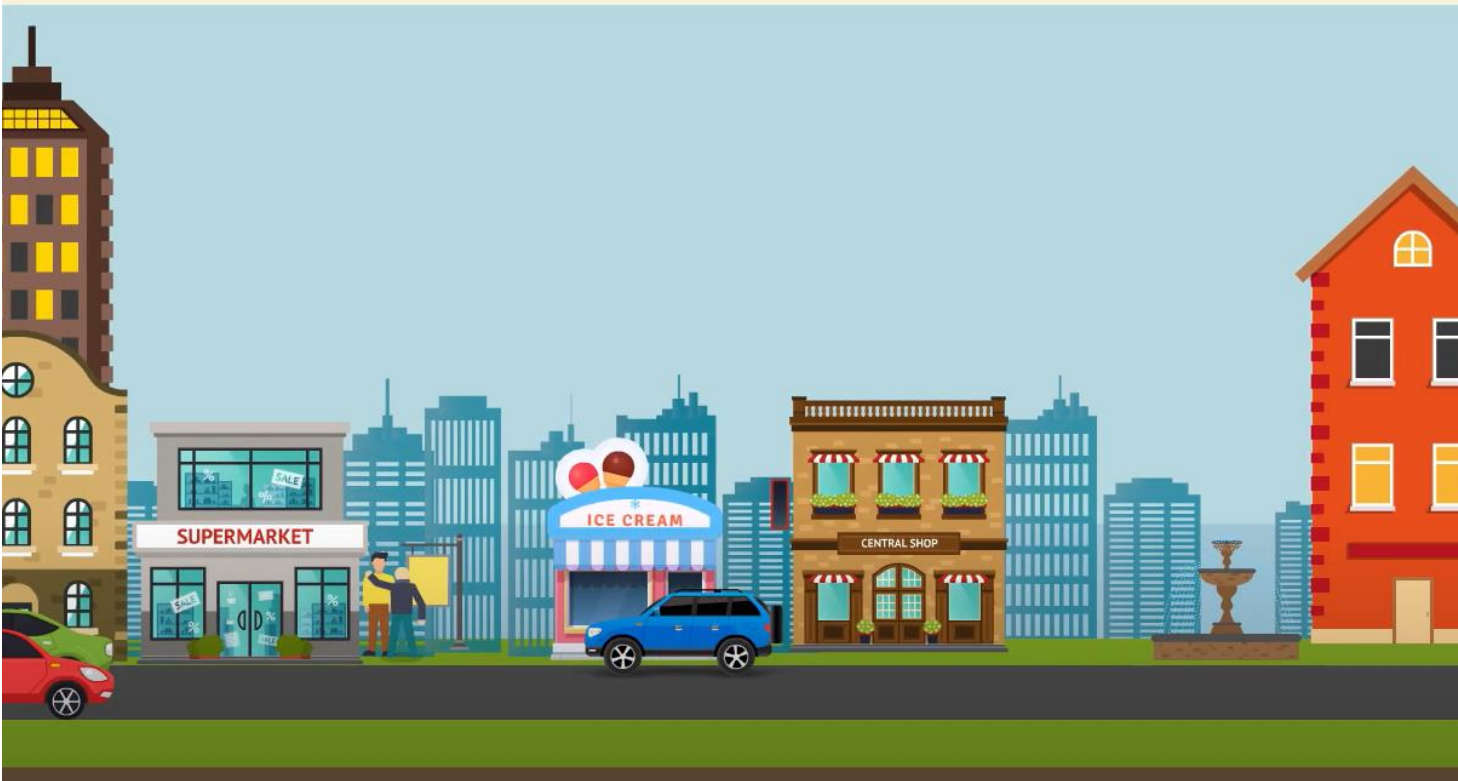
Hace cumplir las limitaciones debidas a:

- Cumplimiento de la normativa y políticas internas.
- Retención y archivo.
- Privacidad y seguridad.
- Niveles de servicio y requisitos comerciales.
- Requisitos técnicos.



Vídeo

Link: <https://youtu.be/rXYwo3hi-j8?si=r-jVX4vhrrsq9TRqm>



¿Qué dicen los decretos sobre los METADATOS?



Transparencia

Proporcionar a la ciudadanía información suficiente para el uso correcto de los datos, a través de metadatos



Interoperabilidad

Los metadatos deben de estar descritos siguiendo las reglas corporativas para facilitar la interoperabilidad y el consumo de datos



Publicación y reutilización

De datos y metadatos generar valor público, potenciar el desarrollo socioeconómico, la creación de empleo y la innovación, tanto pública como privada

Modelo de Gobierno del Dato

¿Qué dicen los decretos sobre los METADATOS?

Nivel ESTRATÉGICO

**CDO**

- Impulsar y promover la política de gestión de metadatos y sus procedimientos

**Responsable de Gobierno del Dato**

- Definir la política de gestión de metadatos y sus procedimientos (Marco Corporativo)
- Velar por el cumplimiento y correcta ejecución de la política de metadatos

Nivel OPERATIVO

**Oficina del Dato**

- Elaborar y mantener actualizados: Diccionario de Datos, Catálogo y Servicios y Glosario de Negocio
- Apoyo y asesoramiento técnico. Eliminar silos.

**Personas administradoras del dato**

- Colaborar en la actualización de: Diccionario de Datos, Catálogo y Servicios y Glosario de Negocio
- Apoyar a los negocios en uso de herramientas

**Apoyo en tecnologías de la información (Lantik)**

- Implementar y mantener una arquitectura de sistemas de información alineada con el Marco Corporativo

DataBarre



