

# Data Management:

## Gestión holística de los datos empresariales



**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.

# Data Management:

## Gestión holística de los datos empresariales

### Agenda:

#### Día 1.

1. Objetivos
2. Introducción al Data Management.
3. Gobierno de datos.
4. Metadatos.
5. Ejercicios Modelado, utilizando Oracle Data Modeler
6. Calidad de datos.
7. Ejercicios Calidad de datos, utilizando Talend y Trifacta

#### Día 2.

10. Arquitectura de Datos
11. Seguridad
12. Datos Maestros y De Referencia
13. Integración de Datos
14. Ejercicios de Integración de datos con ODI y Talend





# Data Management:

## Gestión holística de los datos empresariales

### Objetivo Teórico:

- Comprender los conceptos principales relacionados con las funciones de la gestión de datos desde la perspectiva de personas (organización), procesos (actividades) y tecnología.



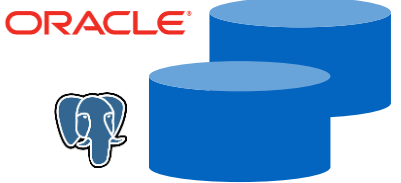
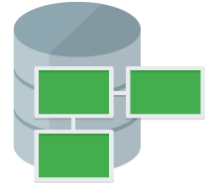





# Data Management:

## Gestión holística de los datos empresariales

### Objetivo Práctico:

- Conocer las tecnologías que permitan la gestión eficiente del dato para dar respuesta a las preguntas ¿de dónde vienen los datos? ¿tienen la calidad necesaria? ¿existe la misma visión en toda la organización?

Datos	¿De dónde vienen?	¿Tienen la calidad necesaria?	¿Existe la misma visión?
<b>Base de Datos</b> 	<b>Gobierno, Metadatos, Arquitectura y Modelado</b> <b>Data Modeler</b> 	<b>Calidad</b> <b>Talend Data Preparation</b>  <b>Trifacta Wrangler</b> 	<b>Integración</b> <b>Oracle Data Integrator</b>  <b>Talend Data Integration</b> 



# Data Management:

Gestión holística de los datos empresariales

## TECNOLOGÍAS:

- Bases de datos: Oracle Database 12c y PostgreSQL
- Modelado: SQL Developer con Data Modeler
- Calidad de Datos: Talend Preparation y Trifacta Wrangler.
- Integración de Datos: Oracle Data Integrator y Talend Data Integration



# Data Management:

## Gestión holística de los datos empresariales

# INTRODUCCIÓN



**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.



# Data Management:

## Gestión holística de los datos empresariales

### Incremento Exponencial del volumen de datos

**Se han creado más datos en los últimos dos años, que en toda la historia anterior de la raza humana.**

**Forbes**

**2020, se crearan aproximadamente 1,7 Megabytes de nueva información por segundo para cada ser humano en el planeta**

**El universo digital acumulado de datos crecerá de 4,4 zettabytes hoy, a alrededor de 44 zettabytes, o 44 billones de gigabytes en el 2020**

**300 horas de video son subidas a Youtube cada minuto**



**Los usuarios de Facebook envían un promedio de 31,25 millones de mensajes y ven 2,77 millones de videos cada minuto**



**El 40% de los usuarios corporativos indican No Encontrar la información que requieren sus Funciones**



**Menos del 0,5 del total de los datos son analizados y explotados**





# Data Management:

## Gestión holística de los datos empresariales

### Nuevas áreas de conocimiento

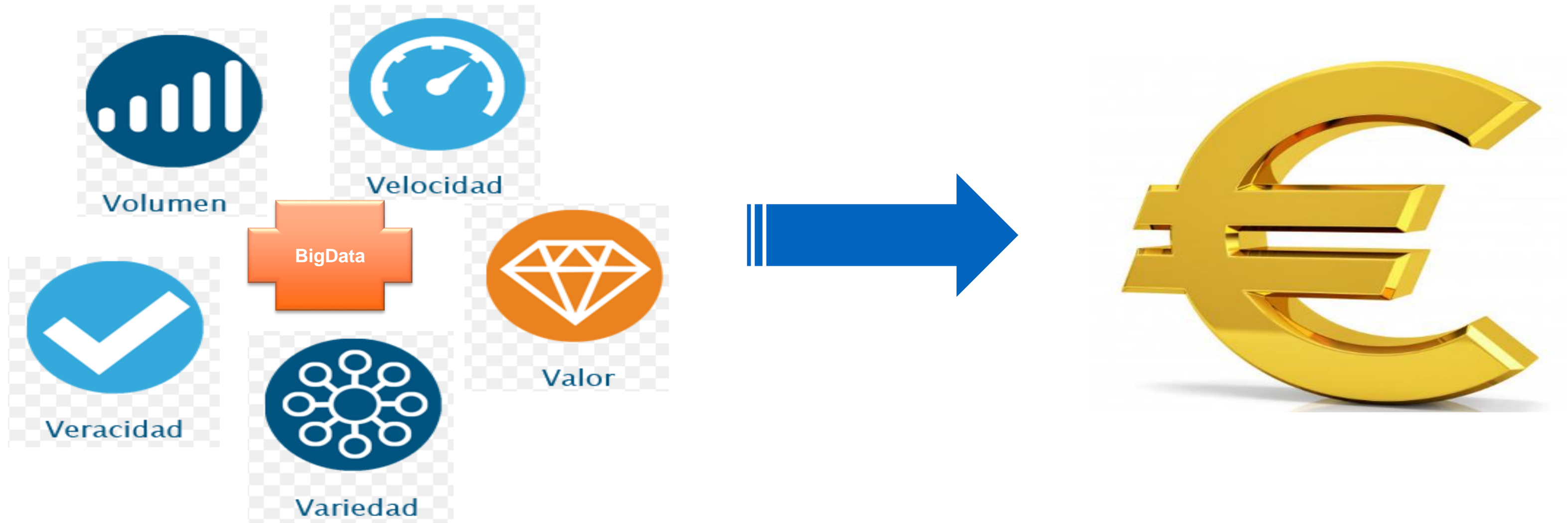




# Data Management:

## Gestión holística de los datos empresariales

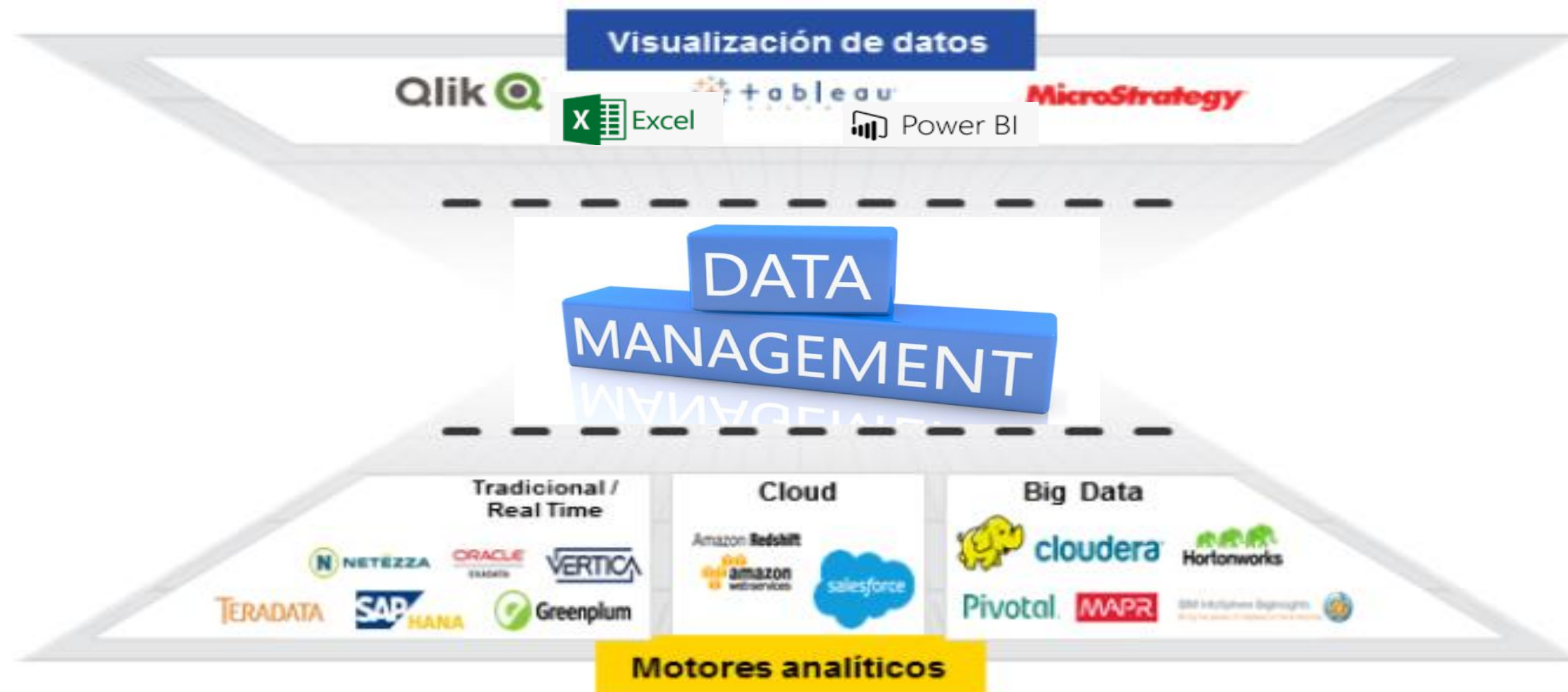
La realidad de los datos



# Data Management:

## Gestión holística de los datos empresariales

Generar VALOR de los Datos a través de múltiples tecnologías



# Data Management:

## Gestión holística de los datos empresariales

¿Quién se encarga?

Smarter With Gartner

Subscribe



Gartner

### 3 Top Take-Aways From the Gartner Chief Data Officer Survey

Data & Analytics

January 29, 2018 | Contributor: Christy Pettey

**CDOs are accountable and impactful change agents, leading the data-driven transformation of their organizations.**

Today's chief data officers (CDOs) have one of the toughest seats at the executive table. Within this challenging role, they unlock data-driven innovation as well as integrate disparate data and analytics capabilities into a strategic discipline, all while delivering a steady drumbeat of quick-win, high-value projects. [A successful CDO requires the skills of a seasoned high-wire performer](#) who is consistently both stable and agile.





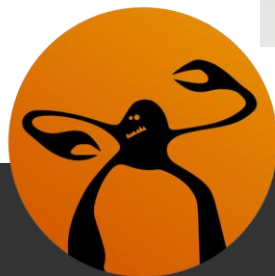
# Data Management:

## Gestión holística de los datos empresariales

¿A qué se aplica?

### Gartner Data & Analytics Summit 2019

- Advanced **Analytics**/Machine Learning
- Artificial Intelligence
- Blockchain
- Building Data and Analytics Skills
- **CDO and Leadership**
- Cloud Computing
- Compliance/GDPR
- Cost Optimization
- **Data and Analytics Architecture**
- **Data and Analytics Governance**
- **Data and Analytics Technology**
- Data Integration
- Data Lakes
- **Data Quality**
- **Data Security and Privacy**
- Data Warehousing
- Digital Business
- Domain Analytics (HR, Supply Chain, Finance etc.)
- Internet of Things
- Leadership
- Master Data Management
- Overall Data and Analytics Strategy
- Roles and Organizational Structure
- Self-service **BI**
- Streamline your BI/Analytics Tools Portfolio



# Data Management:

## Gestión holística de los datos empresariales

### ¿Qué es la gestión de datos?



- La gestión de datos se dedica a definir métodos efectivos para que la información se encuentre en el **LUGAR** donde lo necesitamos, en el **MOMENTO** en el que lo necesitamos y con las **CARACTERISTICAS** que necesitamos.
- El Gobierno de dato es un responsabilidad conjunta del Negocio y IT
- La gobernabilidad de los datos requiere mejora y desarrollo continuos
- Utilizando marcos de mejores practicas



# Data Management:

## Gestión holística de los datos empresariales



- El framework propuesto por DAMA especifica 11 funciones a tener en cuenta para la realización de la gestión del dato integral y de su gobierno.
- Cada una de las dimensiones propuestas en el framework DAMA detalla las diferentes actividades que deberían ser llevadas a cabo para cubrir todos los aspectos relacionados con ese ámbito de la gestión del gobierno del dato.





# Data Management:

## Gestión holística de los datos empresariales

De DAMA



A



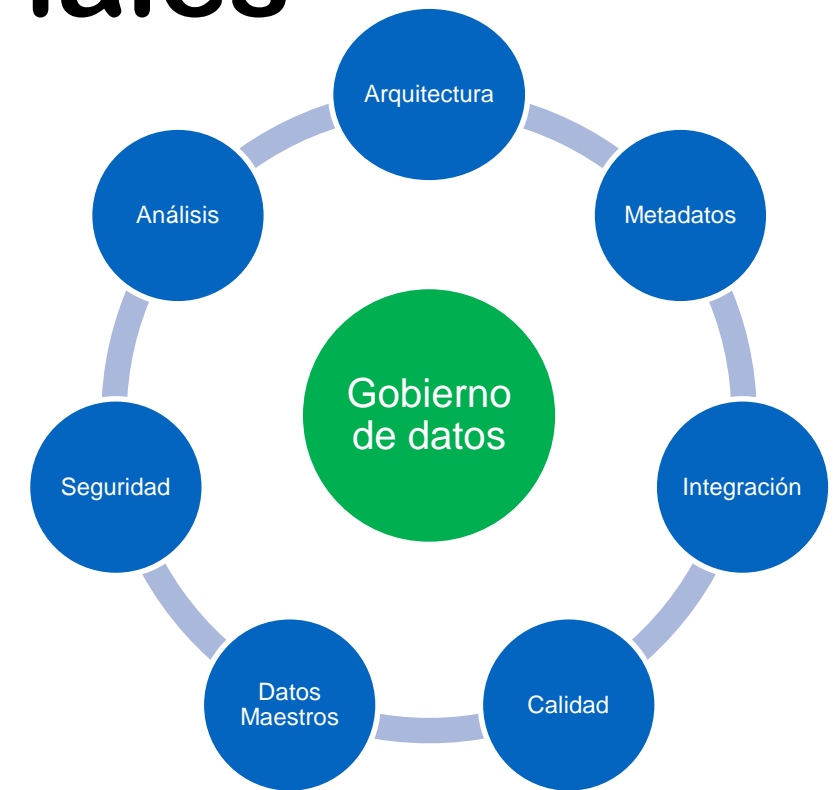
Enfoque Reducido



# Data Management:

Gestión holística de los datos empresariales

## GOBIERNO DE DATOS



**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.

# Gobierno de Datos

El **Gobierno de Datos** se refiere a la administración general de la disponibilidad, usabilidad, integridad y seguridad de los datos utilizados dentro de una organización. La gobernanza incluye un **órgano de gobierno** (Consejo), un conjunto de **estándares y políticas** y un **plan** para ejecutar esos estándares y políticas



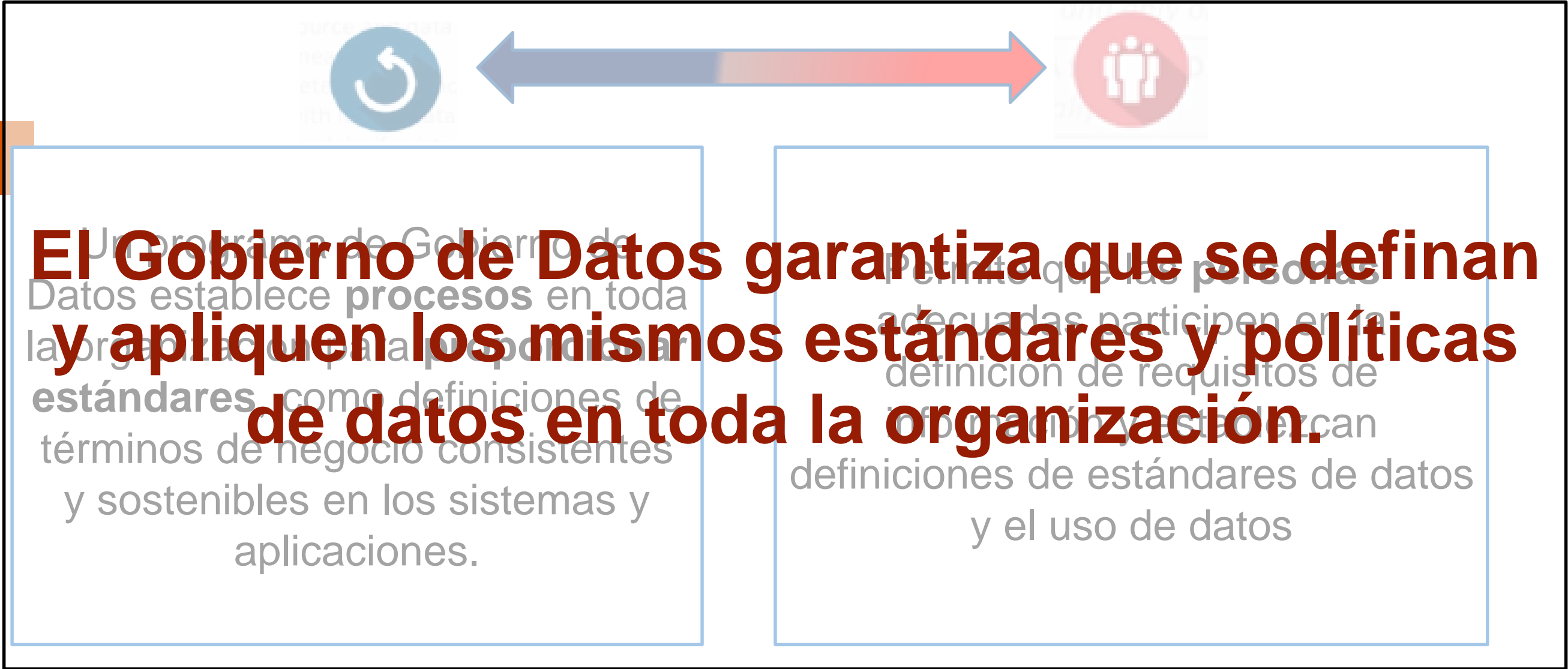
Un programa de Gobierno de Datos establece **procesos** en toda la organización para **proporcionar estándares**, como definiciones de términos y reglas de negocio consistentes, en los sistemas y aplicaciones.

Determina las **personas** adecuadas para que participen en la definición de requisitos de información y establezcan definiciones de estándares de datos y el uso de datos



# Gobierno de Datos

El **Gobierno de Datos** se refiere a la administración general de la disponibilidad, usabilidad, integridad y seguridad de los datos utilizados dentro de una organización. La gobernanza incluye un **órgano de gobierno** (Consejo), un conjunto de **estándares y políticas** y un **plan** para ejecutar esas estándares y políticas



The diagram illustrates the components of Data Governance. At the top, a large double-headed arrow connects two circular icons: a blue circle with a white circular arrow (representing processes) on the left and a red circle with a white group of people icon (representing standards and policies) on the right. Below this, two light blue rectangular boxes are positioned side-by-side. The left box contains text about establishing processes and standards, while the right box contains text about ensuring participation in defining requirements and standards. The main title 'El Gobierno de Datos garantiza que se definan y apliquen los mismos estándares y políticas de datos en toda la organización.' is centered across both boxes in a large, bold, dark red font.

**El Gobierno de Datos garantiza que se definan y apliquen los mismos estándares y políticas de datos en toda la organización.**

Un programa de Gobierno de Datos establece **procesos** en toda la organización para proporcionar **estándares**, como definiciones de términos de negocio consistentes y sostenibles en los sistemas y aplicaciones.

Permite que las personas adecuadas participen en la definición de requisitos de información y establezcan definiciones de estándares de datos y el uso de datos



# Principios del Gobierno de Datos

El **Gobierno de Datos** es la función es menos técnica que cualquier otra función del Data Management . Se refiere a la organización, la estandarización y las políticas. Está estrechamente relacionada con la **Calidad de Datos** y la gestión de **Metadatos**.



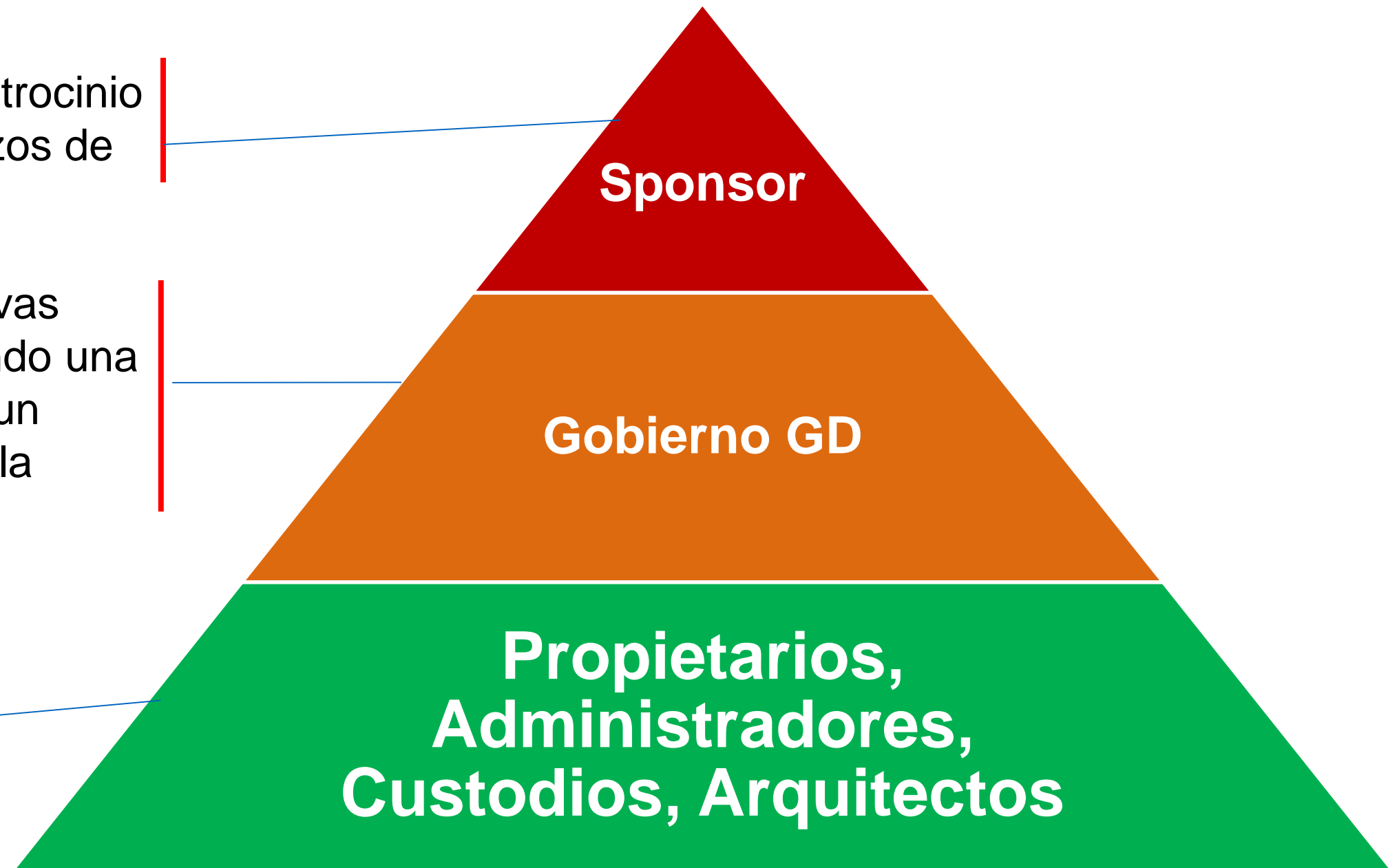
# Organización del Gobierno de Datos

La Organización del Gobierno de Datos implica una combinación de múltiples niveles de roles de negocios y tecnología.

Proporciona liderazgo y patrocinio general a todos los esfuerzos de GD.

Proporciona coherencia y coordinación de las iniciativas multi-funcional , manteniendo una perspectiva empresarial y un enfoque estratégico sobre la calidad de los datos.

Propone y aplica los estándares, políticas y procedimientos.





# Organización del Gobierno de Datos

La Organización del Gobierno de Datos implica una combinación de múltiples niveles de roles de negocios y tecnología.

Proporciona liderazgo y patrocinio general a todos los esfuerzos de GD.

## PATROCINIO

Sponsor

Proporciona coherencia y coordinación de las iniciativas multi-funcional , manteniendo una perspectiva empresarial y un enfoque estratégico sobre la calidad de los datos.

## DIRECCIÓN

Gobierno GD

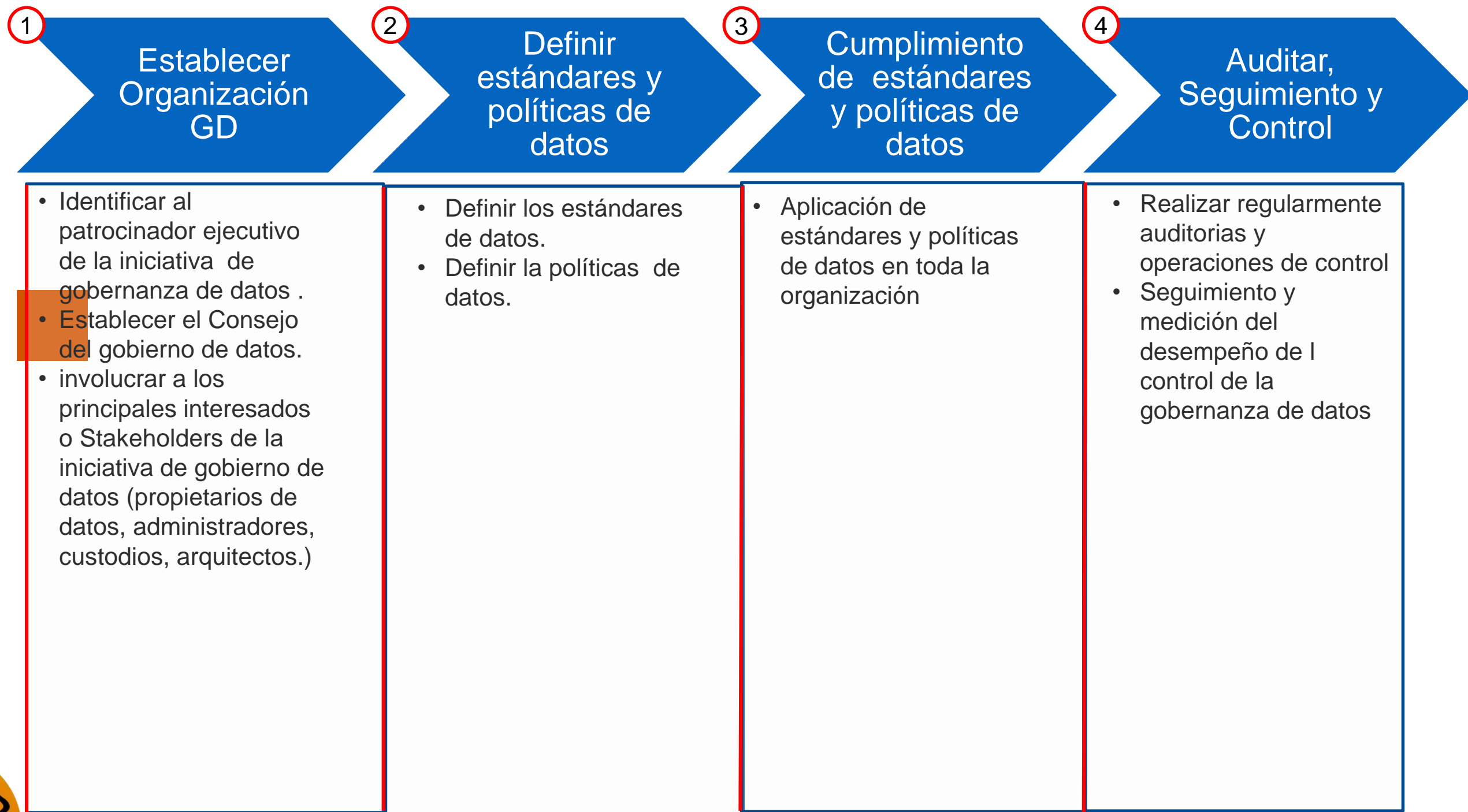
Propone y aplica los estándares, políticas y procedimientos.

## EJECUCIÓN

Propietarios,  
Administradores,  
Custodios, Arquitectos



# Proceso de Gobierno de Datos

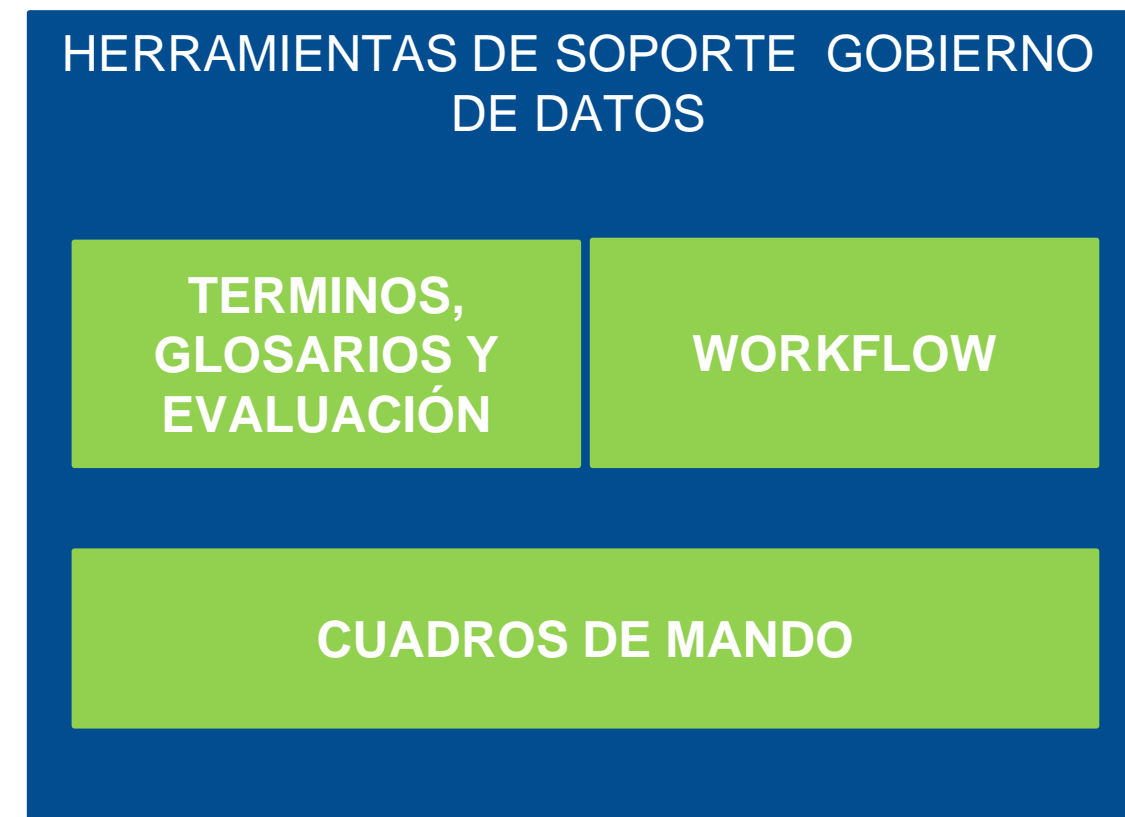


# Proceso de Gobierno de Datos



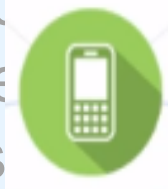
# Herramientas Tecnológicas para Gobierno de Datos

- Las herramientas de soporte de Metadatos, y de soporte de calidad de datos, se utilizan a menudo para la implementación del Gobierno de datos. Existen muy pocas herramientas en el mercado que se consideran específicas de soporte de Gobierno de datos.
- Una funcionalidad importante para el gobierno de los datos es el diseño de Workflows o flujos de trabajo que se puede usar para implementar estándares y políticas de datos.
- Otra funcionalidad importante es el diseño de cuadros de mando para el seguimiento y control de las actividades de Gobierno de datos.



# Herramientas Tecnológicas para Gobierno de Datos

- Las herramientas de soporte de Metadatos, y de soporte de calidad de datos, se utilizan a menudo para la implementación del Gobierno de datos. Existen muy pocas herramientas en el mercado que se consideran esenciales para el soporte de Gobierno de datos.
- Una funcionalidad importante para el gobierno de los datos es el diseño de Workflows o flujos de trabajo que se puede usar para implementar estándares y políticas de datos.
- Otra funcionalidad importante es el diseño de cuadros de mando para el seguimiento y control de las actividades de Gobierno de datos.



## TECNOLOGÍA

### HERRAMIENTAS DE SOPORTE GOBIERNO DE DATOS


TERMINOS,  
GLOSARIOS Y  
EVALUACIÓN

WORKFLOW

CUADROS DE MANDO




# Resumen – Puntos Clave




El **Gobierno de Datos** se trata de estándares y políticas de datos que gestionan la **disponibilidad, usabilidad, integridad y seguridad** de los datos empleados en una organización.




El Gobierno de Datos es la función de gestión de datos menos técnica.



Los componentes clave de la organización de gobierno de datos son: **Patrocinador, Consejo y Stakeholders** (propietarios, administradores, custodios y arquitectos de datos).



El **Proceso Gobierno de Datos** incluye las actividades de: Establecer la Organización de GD, Definir, Aplicar estándares y políticas y Auditar, Seguimiento y Control de las actividades de GD.



Las **Herramientas de gestión de metadatos** y calidad de datos se utilizan a menudo para implementar el gobierno de datos. Hay muy pocas herramientas específicas de gobierno de datos en el mercado.

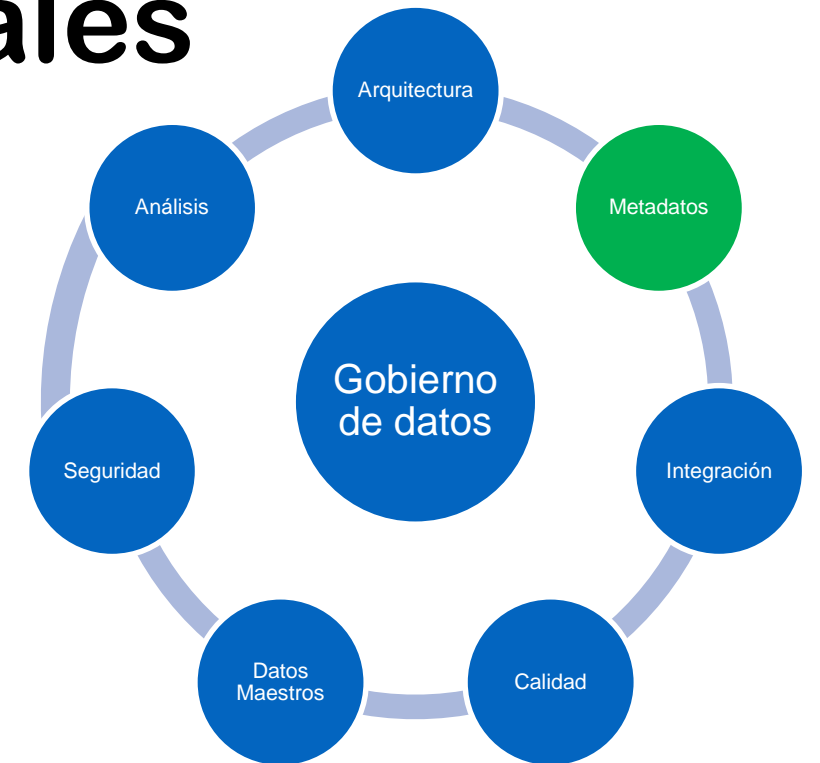




# Data Management:

Gestión holística de los datos empresariales

## GESTIÓN DE METADATOS



**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.

# Conceptos Básicos – ¿Qué es elemento de dato?

ED

Un **elemento de dato** (ED) es una unidad atómica de datos compuesta por un conjunto de atributos (definición, identificación, representación y valores permitidos)

¿Cuál es el criterio para que un elemento de datos se vuelva crítico?

1. Datos de negocio que se consideran críticos para los responsables de la organización.
2. Datos de procesos críticos del negocio y sus componentes, por ejemplo, contactar con clientes.
3. Datos clave utilizados en informes de alto nivel del negocio.
4. Identificadores únicos de cosas importantes de la empresa, por ejemplo, ID de cliente

EDC

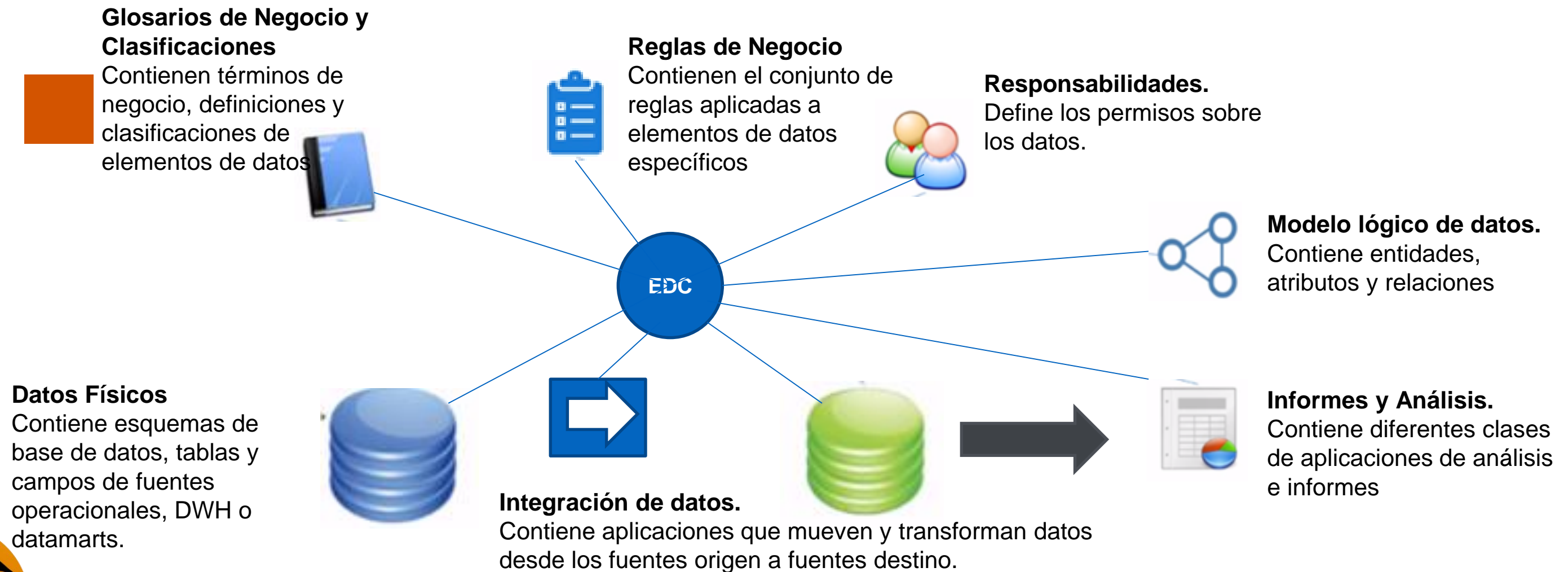
El **elemento dato crítico** (EDC) es el elemento de datos que es crítico para el éxito en un área de negocio o proceso de negocio específico.

**Los elementos de datos (ED) y elementos de datos críticos (EDC) son conceptos fundamentales en la gestión de datos.**



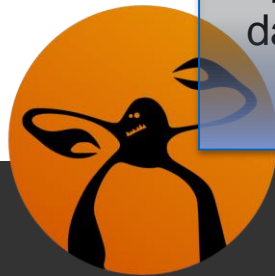
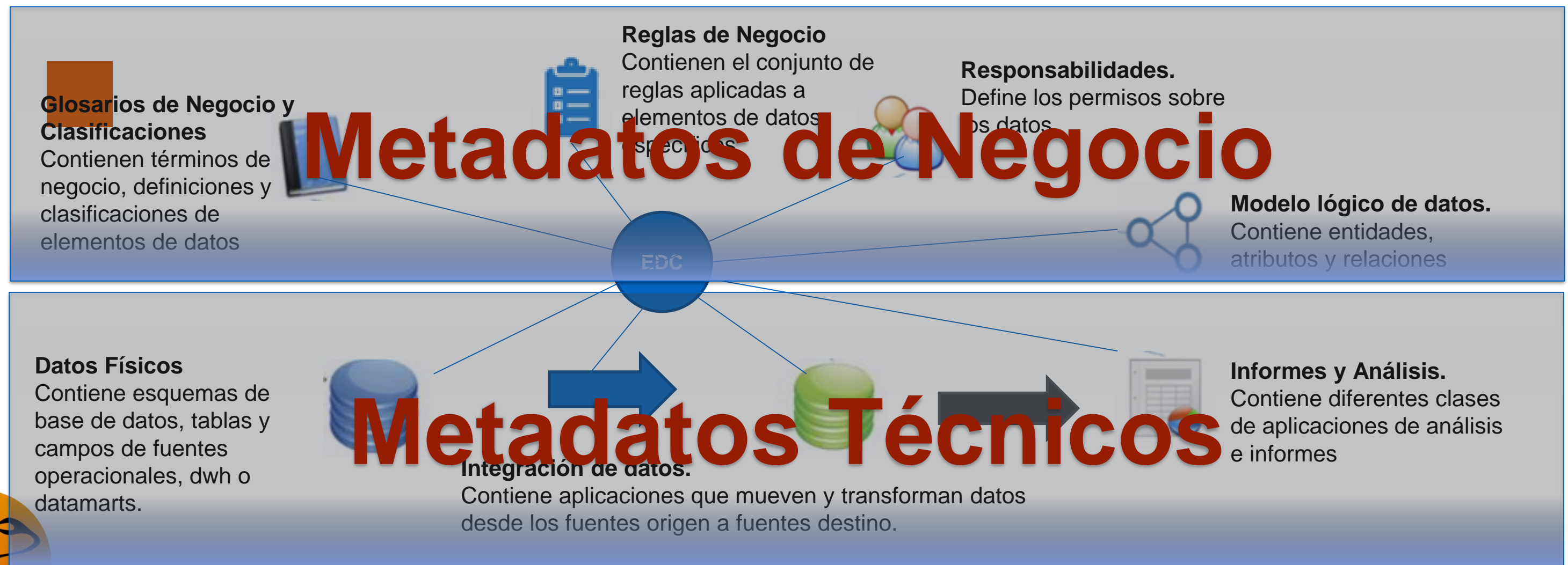
# Gestión de Metadatos

La **gestión de metadatos** implica la gestión de datos sobre otros datos, se entiende por “otros datos” a modelos de datos y estructuras, no al contenido. Incluye la gestión de información sobre estructuras de datos de diferentes modelos y sus asociaciones (e.j, términos de negocio en glosarios y atributos en los modelos lógicos de datos o tablas y columnas en la base de datos, así como sus asociaciones)



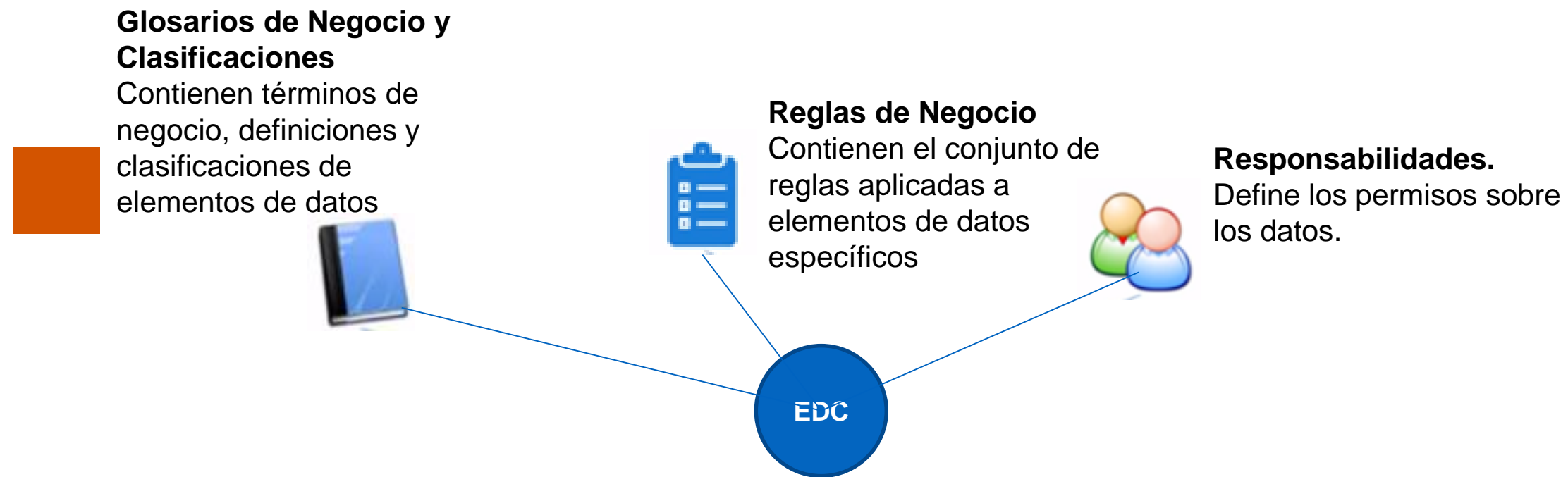
# Gestión de Metadatos

La **gestión de metadatos** implica la gestión de datos sobre otros datos, se entiende por “otros datos” a modelos de datos y estructuras, no al contenido. Incluye la gestión de información sobre estructuras de datos de diferentes modelos y sus asociaciones (e.j, términos de negocio en glosarios y atributos en los modelos lógicos de datos o tablas y columnas en la base de datos, así como sus asociaciones,)

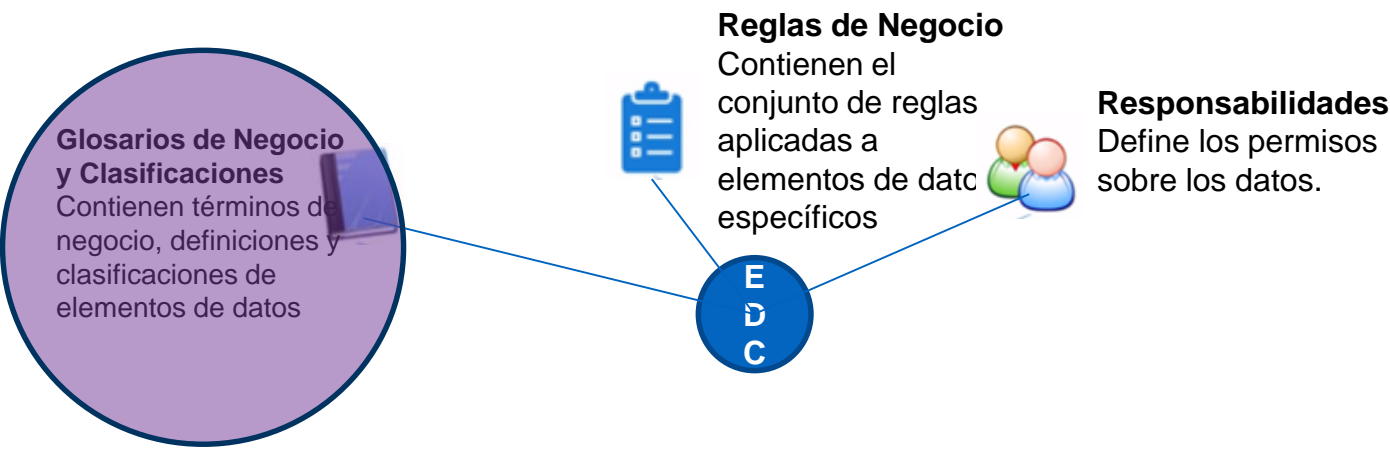


# Metadatos de Negocio

Los **metadatos de negocio** describen el elemento de datos desde la perspectiva de uso del negocio, incluyen información como glosario de negocios con términos y definiciones, sinónimos, siglas, reglas de negocios y responsabilidades.



# Ejemplo: Nombre Cliente



Glosario de Negocio y Clasificaciones

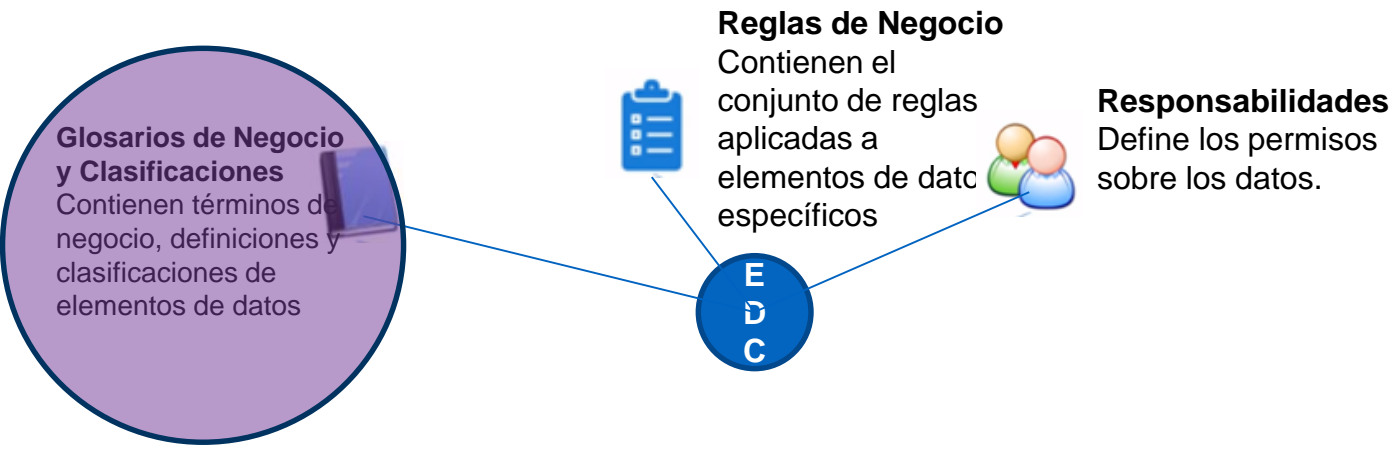


TÉRMINO	Nombre Cliente	
DEFINICIÓN	representa una palabra o una combinación de palabras por las que se llama o se conoce a un cliente (persona o entidad).	
SINÓNIMOS	Nombre Consumidor	
CLASIFICACIÓN	Nombre Legal	El nombre legal de una empresa es el nombre de la persona o entidad dueño de la empresa, es el nombre que identifica a un cliente para fines legales, administrativos y otros propósitos oficiales
	Nombre Comercial	Generalmente es el nombre para usos comerciales como publicidad y ventas, es diferente del nombre legal que documentos oficiales de organización





# Ejemplo: Nombre Cliente



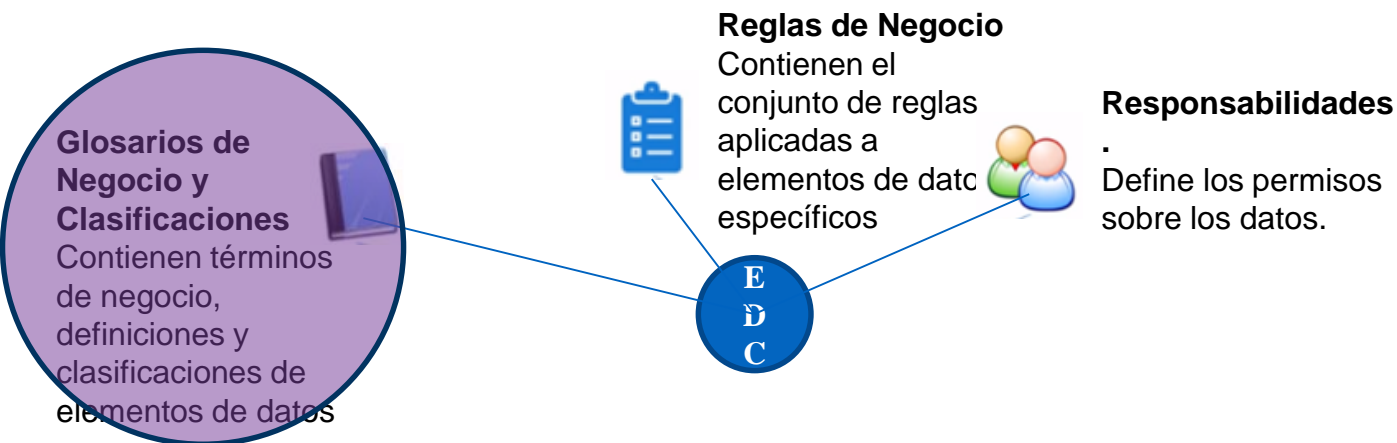
## Glosario de Negocio y Clasificaciones



TÉRMINO	Nombre Cliente	
DEFINICIÓN	representa una palabra o una combinación de palabras por las que se llama o se conoce a un cliente (persona o entidad).	
SINÓNIMOS	Nombre Consumidor	
CLASIFICACIÓN	Nombre Legal	El nombre legal de una empresa es el nombre de la persona o entidad dueño de la empresa, es el nombre que identifica a un cliente para fines legales, administrativos y otros propósitos oficiales
	Nombre Comercial	Generalmente es el nombre para usos comerciales como publicidad y ventas, es diferente del nombre legal que documentos oficiales de organización



# Ejemplo: Nombre Cliente



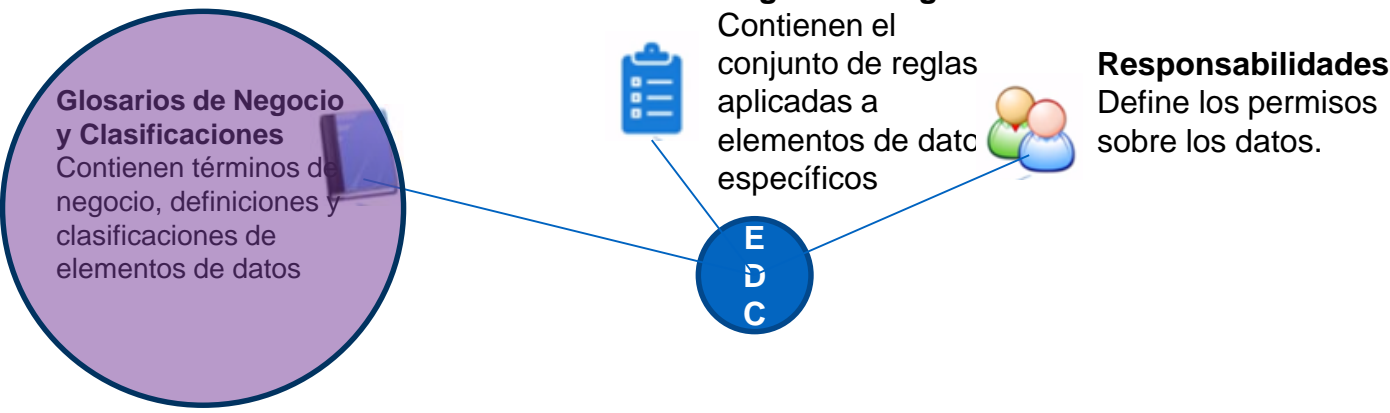
## Glosario de Negocio y Clasificaciones



TÉRMINO	Nombre Cliente	
DEFINICIÓN	representa una palabra o una combinación de palabras por las que se llama o se conoce a un cliente (persona o entidad).	
SINÓNIMOS	Nombre Consumidor	
CLASIFICACIÓN	Nombre Legal	El nombre legal de una empresa es el nombre de la persona o entidad dueño de la empresa, es el nombre que identifica a un cliente para fines legales, administrativos y otros propósitos oficiales
	Nombre Comercial	Generalmente es el nombre para usos comerciales como publicidad y ventas, es diferente del nombre legal que documentos oficiales de organización



# Ejemplo: Nombre Cliente



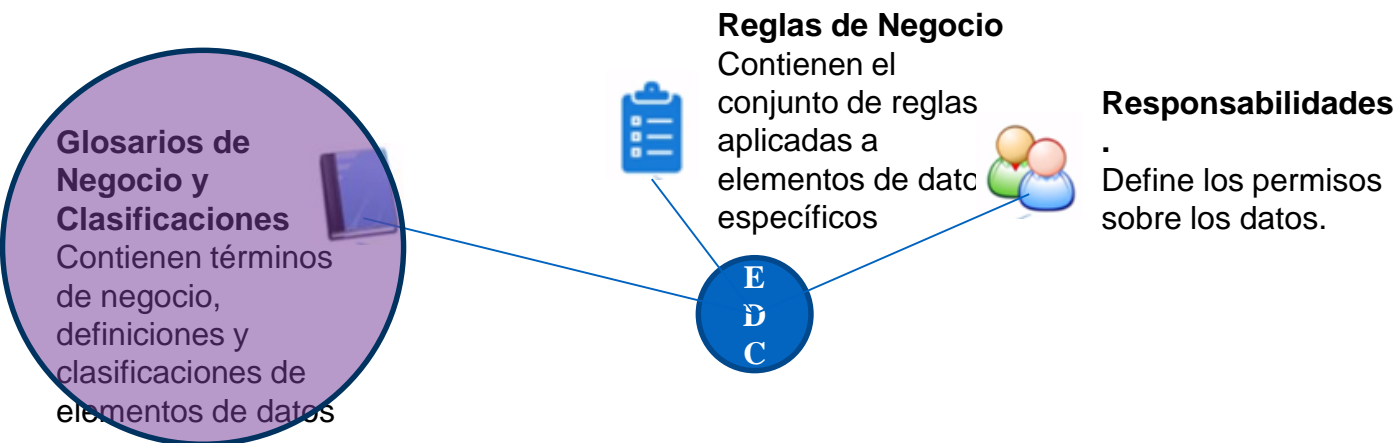
Glosario de Negocio y Clasificaciones



TÉRMINO	Nombre Cliente	
DEFINICIÓN	representa una palabra o una combinación de palabras por las que se llama o se conoce a un cliente (persona o entidad).	
SINÓNIMOS	Nombre Consumidor	
CLASIFICACIÓN	Nombre Legal	El nombre legal de una empresa es el nombre de la persona o entidad dueño de la empresa, es el nombre que identifica a un cliente para fines legales, administrativos y otros propósitos oficiales
	Nombre Comercial	Generalmente es el nombre para usos comerciales como publicidad y ventas, es diferente del nombre legal que documentos oficiales de organización



# Ejemplo: Nombre Cliente



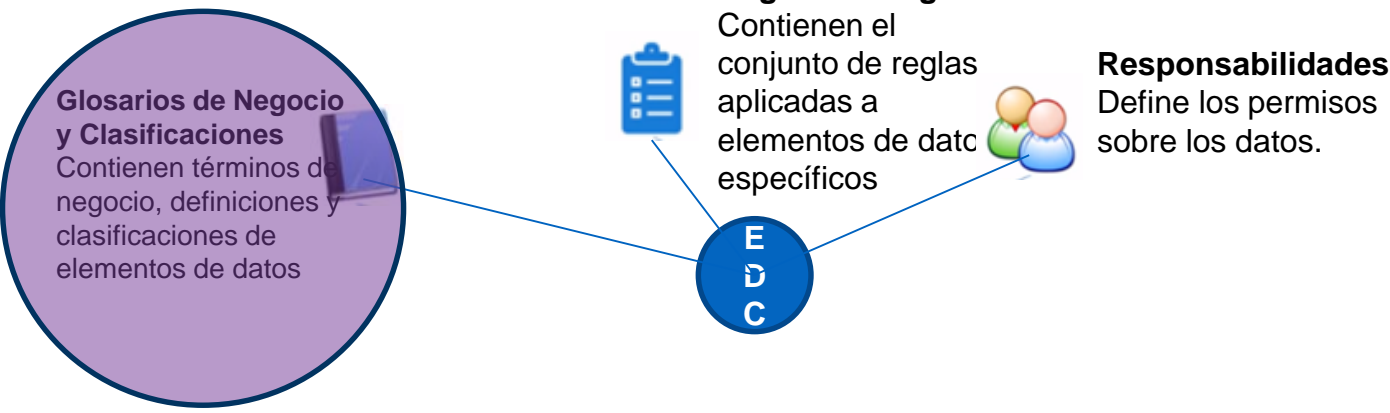
## Glosario de Negocio y Clasificaciones



TÉRMINO	Nombre Cliente	
DEFINICIÓN	representa una palabra o una combinación de palabras por las que se llama o se conoce a un cliente (persona o entidad).	
SINÓNIMOS	Nombre Consumidor	
CLASIFICACIÓN	Nombre Legal	El nombre legal de una empresa es el nombre de la persona o entidad dueño de la empresa, es el nombre que identifica a un cliente para fines legales, administrativos y otros propósitos oficiales
	Nombre Comercial	Generalmente es el nombre para usos comerciales como publicidad y ventas, es diferente del nombre legal que documentos oficiales de organización



# Ejemplo: Nombre Cliente



Glosario de Negocio y Clasificaciones

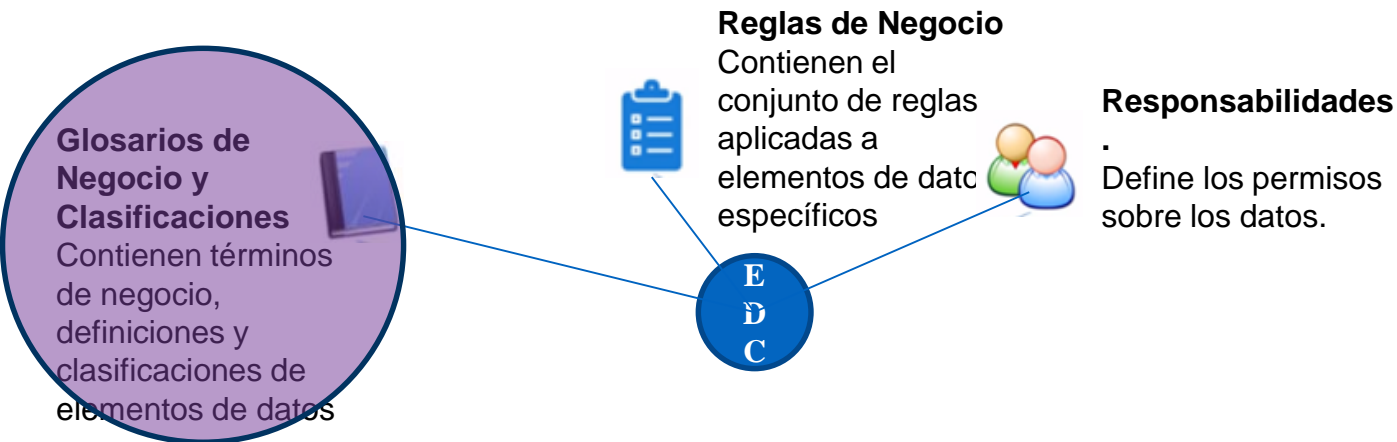


TÉRMINO	Nombre Cliente	
DEFINICIÓN	representa una palabra o una combinación de palabras por las que se llama o se conoce a un cliente (persona o entidad).	
SINÓNIMOS	Consumidor	
CLASIFICACIÓN	Nombre Legal	El nombre legal de una empresa es el nombre de la persona o entidad dueño de la empresa, es el nombre que identifica a un cliente para fines legales, administrativos y otros propósitos oficiales
	Nombre Comercial	Generalmente es el nombre para usos comerciales como publicidad y ventas, es diferente del nombre legal que documentos oficiales de organización





# Ejemplo: Nombre Cliente



## Glosario de Negocio y Clasificaciones



TÉRMINO	Nombre Cliente	
DEFINICIÓN	representa una palabra o una combinación de palabras por las que se llama o se conoce a un cliente (persona o entidad).	
SINÓNIMOS	Consumidor	
CLASIFICACIÓN	Nombre Legal	El nombre legal de una empresa es el nombre de la persona o entidad dueño de la empresa, es el nombre que identifica a un cliente para fines legales, administrativos y otros propósitos oficiales
	Nombre Comercial	Generalmente es el nombre para usos comerciales como publicidad y ventas, es diferente del nombre legal que documentos oficiales de organización



# Ejemplo: Nombre Cliente

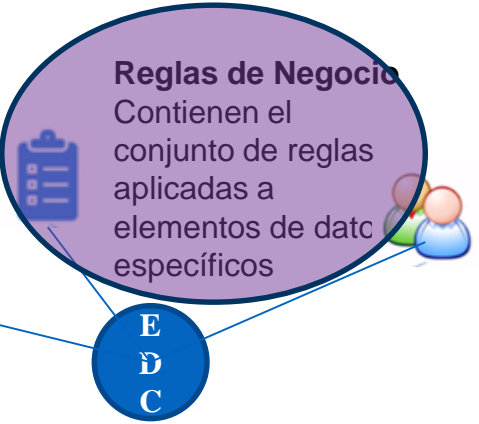


## Reglas de Negocio



REGLA	DESCRIPCIÓN
Nombre Legal <b>Acordada</b>	Para las compañías de responsabilidad limitada, SL y corporaciones, el nombre legal de la empresa es con el que se registró en el Registro Mercantil. Estos nombres a menudo tendrán al final una siglas como SL (Sociedad Limitada) , SC (Sociedad Civil), SA (Sociedad Anónima)
Nombre Legal <b>Acordada</b>	Si la empresa es una colaboración de varias empresas, el nombre legal es el nombre dado en el acuerdo de sociedad o de los socios,
Nombre Comercial <b>Acordada</b>	El nombre comercial puede no incluir terminaciones legales similares o Corp. Un nombre comercial también puede ser una marca registrada. Un nombre comercial no es en sí mismo una forma de propiedad intelectual.

**Glosarios de Negocio y Clasificaciones**  
Contienen términos de negocio, definiciones y clasificaciones de elementos de datos



**Responsabilidades**  
Define los permisos sobre los datos.



Las reglas de negocios son fundamentales para la evaluación de la calidad de los datos.

# Roles en Metadatos Management



**Propietario del negocio** es el responsable final de la definición, la calidad de los datos y el valor de los datos. Es el responsable de confirmar que los datos se utilizan de manera coherente con la estrategia general de los datos (Gobierno de datos). Son también los responsables de dirigir las actividades y procesos de la gestión de los datos.



**Administrador de datos** o data Stewart es el responsable de la supervisión operativa de los datos y de las interacciones con los expertos de cada área en toda la organización, así como la identificación del enfoque para estandarizar, medir y monitorear la calidad de los datos. Se encarga de confirmar que se han definido estándares de datos y que los procesos y practicas se han implementado.



**Propietario técnico** es un especialista en tecnología, responsable final de que los datos de los sistemas se gestionen y utilicen de acuerdo con los estándares de datos definidos. Que incluyen el cumplimiento de las reglas de negocio, las políticas y procedimientos.



**Custodio de datos** es un especialista en tecnología que es responsable de la seguridad y la administración de los datos en los sistemas. Se encarga de confirmar que los estándares de datos definidos se aplican.




Rol de  
Negocio




Rol Técnico



# Roles en Metadatos Management



**Propietario del negocio** es el responsable final de la definición, la calidad de los datos y el valor de los datos . Es el responsable de confirmar que los datos se utilizan de manera coherente con la estrategia general de los datos (Gobierno de datos). Son también los responsables de dirigir las actividades y procesos de la gestión de los datos.



**Administrador de datos** es responsable de la supervisión operativa de los datos y de las interacciones con los expertos de cada área en toda la organización, así como la identificación del enfoque para estandarizar, medir y monitorear la calidad de los datos. Se encarga de confirmar que se han definido estándares de datos y que los procesos y practicas se han implementado.

## Roles de Negocio



**Propietario técnico** es un especialista en tecnología, responsable final de que los datos de los sistemas se gestionen y utilicen de acuerdo con los estándares de datos definidos. Que incluyen el cumplimiento de las reglas de negocio, las políticas y procedimientos.



**custodio de datos** es un especialista en tecnología que es responsable de la seguridad y la administración de los datos en los sistemas. Se encarga de confirmar que los estándares de datos definidos se aplican.

## Roles Técnicos



Rol de  
Negocio



Rol Técnico



# Ejemplo: Nombre Cliente

**Glosarios de Negocio y Clasificaciones**  
Contienen términos de negocio, definiciones y clasificaciones de elementos de datos

**Reglas de Negocio**  
Contienen el conjunto de reglas aplicadas a elementos de datos específicos

**Responsabilidades**  
Define los permisos sobre los datos.



## Responsabilidades – Roles Clave

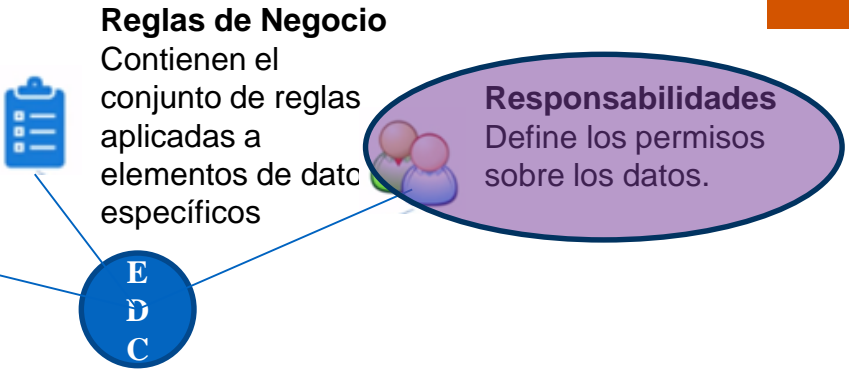
Rol	NOMBRE	ÁREA	EMAIL	TELÉFONO
<b>PROPIETARIO DEL NEGOCIO</b> <i>(Uno y sólo uno)</i>	Juan L.	Finanzas	<i>Alguna dirección de email</i>	<i>Algún número de teléfono</i>
<b>ADMINISTRADOR DEL DATO</b> <i>(Normalmente uno)</i>	Ana M.	Finanzas	<i>Alguna dirección de email</i>	<i>Algún número de teléfono</i>
<b>PROPIETARIO TÉCNICO</b> <i>(Pueden ser varios)</i>	Carlos S.	CDO - SAP	<i>Alguna dirección de email</i>	<i>Algún número de teléfono</i>
<b>CUSTODIO DEL DATO</b> <i>(Pueden ser varios)</i>	Marta R.	CDO - SAP	<i>Alguna dirección de email</i>	<i>Algún número de teléfono</i>





# Ejemplo: Nombre Cliente

**Glosarios de Negocio y Clasificaciones**  
Contienen términos de negocio, definiciones y clasificaciones de elementos de datos



## Responsabilidades – Roles Clave

Rol	NOMBRE	ÁREA	EMAIL	TELÉFONO
PROPIETARIO DEL NEGOCIO (Uno y sólo uno)	Juan L.	Finanzas	Alguna dirección de email	Algún número de teléfono
ADMINISTRADOR DEL DATO (Normalmente uno)	Ana M.	Finanzas	Alguna dirección de email	Algún número de teléfono
PROPIETARIO TÉCNICO (Pueden ser varios)	Carlos S.	CDO - SAP	Alguna dirección de email	Algún número de teléfono
CUSTODIO DEL DATO (Pueden ser varios)	Marta R.	CDO - SAP	Alguna dirección de email	Algún número de teléfono

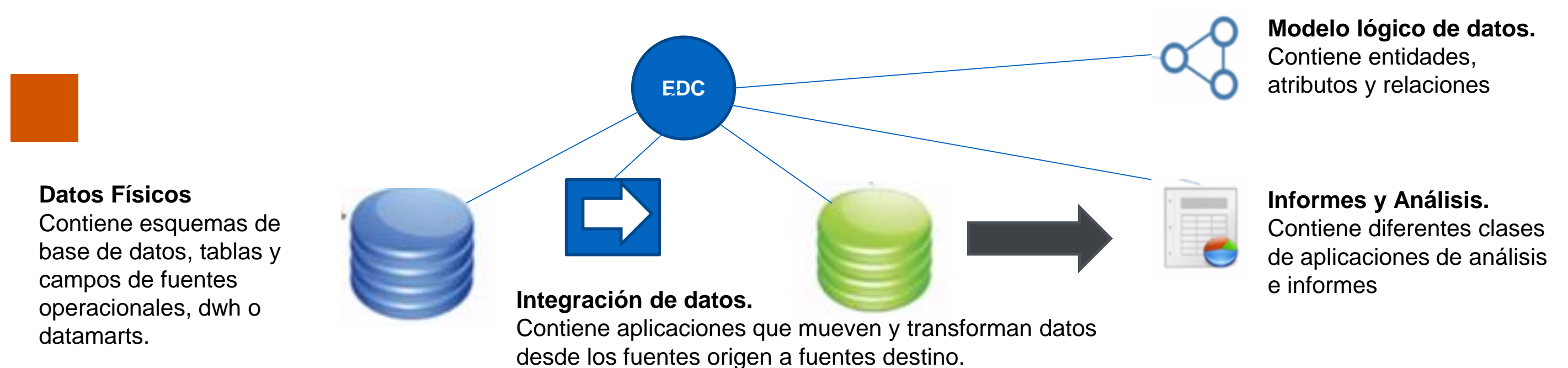
**PERSONAS**



La definición de Roles y sus responsabilidades de los metadatos cubren la **perspectiva de personas** dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management

# Metadatos Técnicos

Los **metadatos técnicos**, describen los elementos de datos desde la perspectiva de tecnología e incluye información como modelos lógicos de datos, sistemas de origen y destino, estructuras de tablas y campos, así como dependencias entre sistemas.



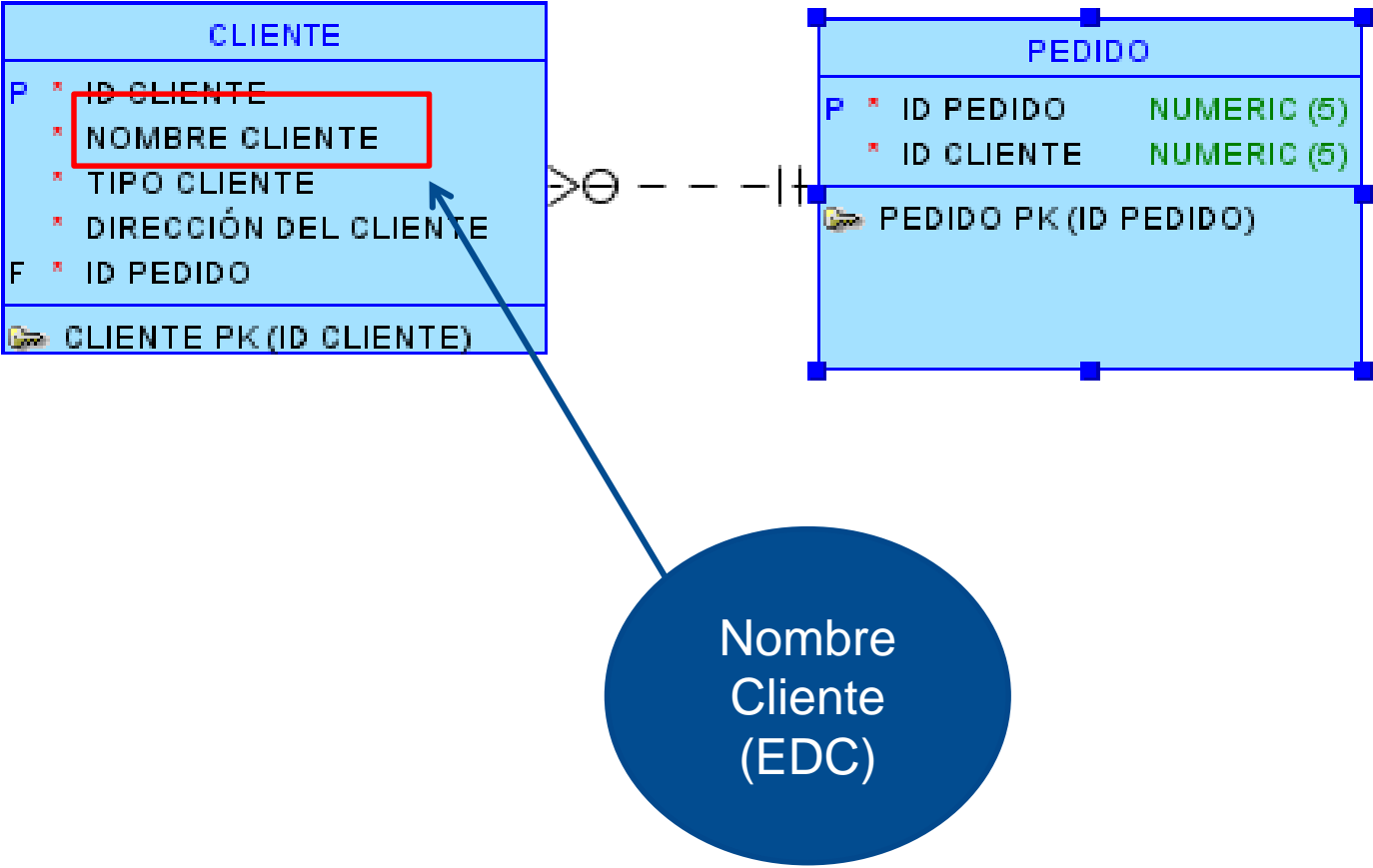
*Los **metadatos operativos** incluyen información sobre la ejecución de las aplicaciones : frecuencia, cuentas de registros y componente por componente de los análisis y otras estadísticas para fines de auditoría.*



# Ejemplo: Nombre Cliente

## Metadatos del Modelo Lógico de Datos

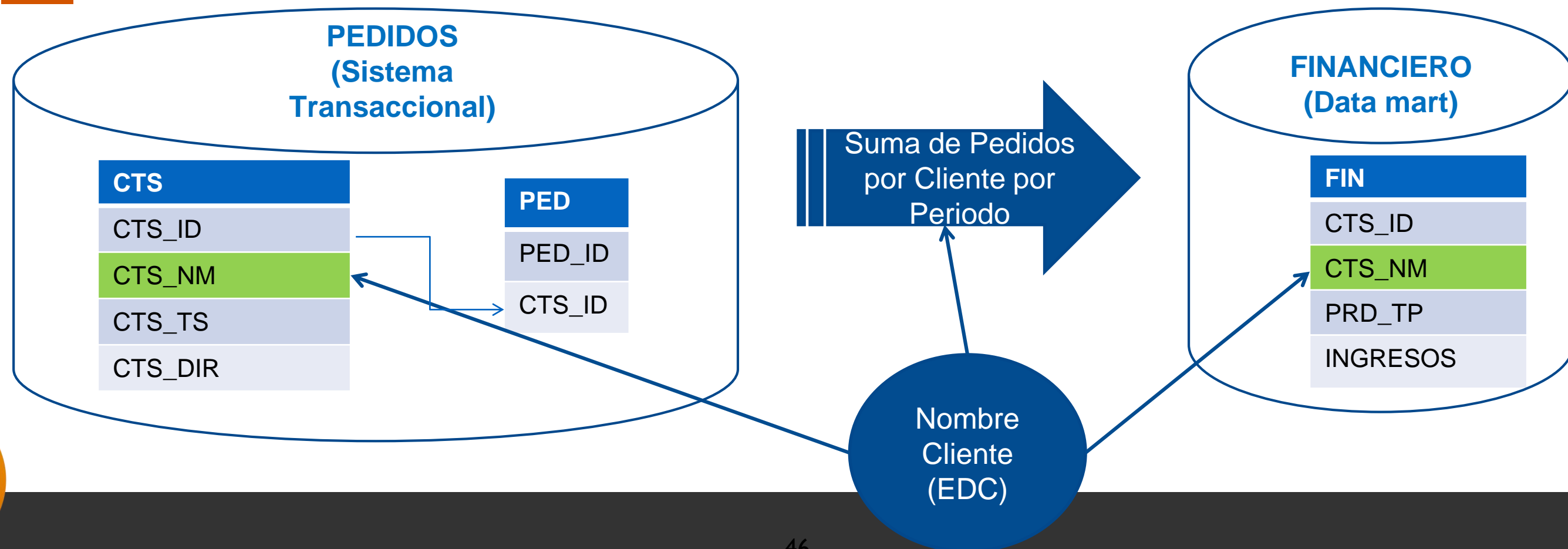
El modelo lógico de datos se refiere a la información sobre las entidades, los atributos y sus relaciones, también es importante establecer una asociación entre el elemento crítico de datos (EDC) y todas las representaciones del mismo en el modelo lógico de datos, de esa manera podrá alinear el EDC con la tecnología.



# Ejemplo: Nombre Cliente

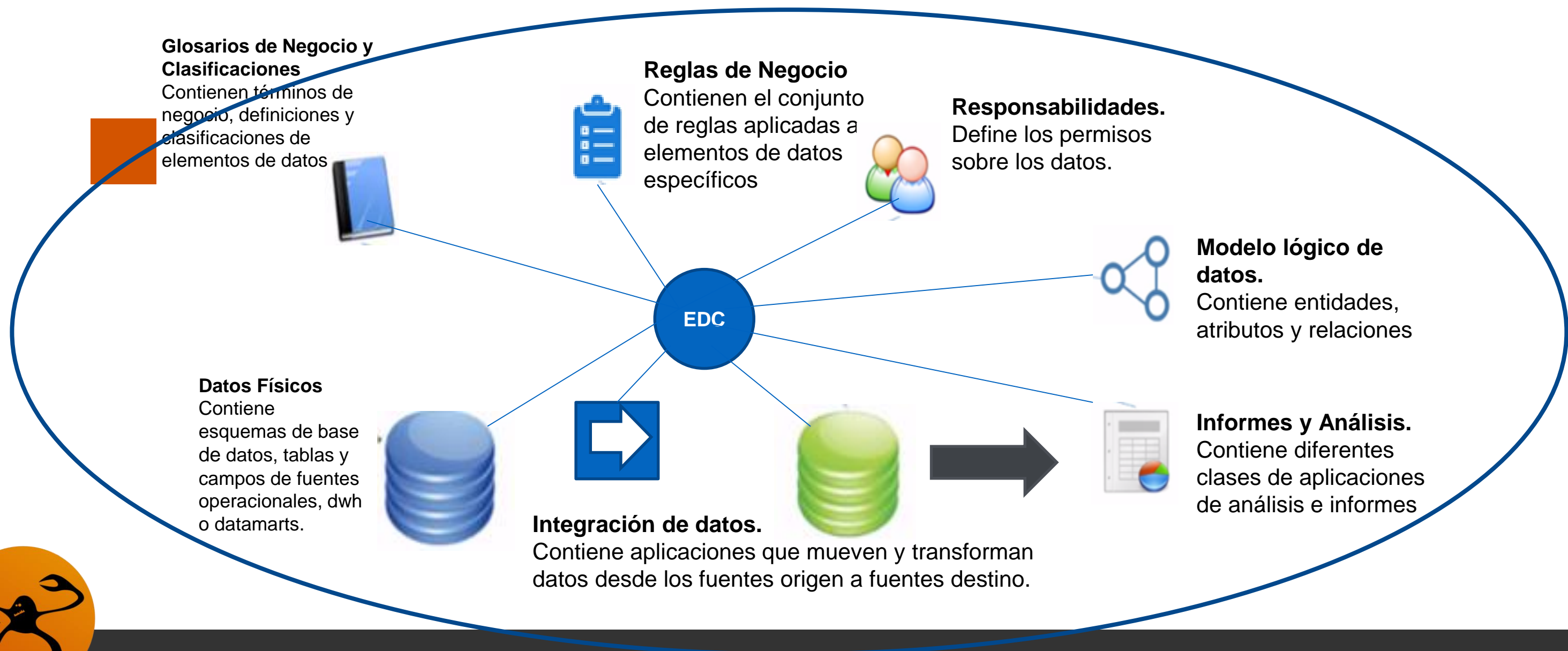
**Metadatos del Modelo Físico de Datos** Los metadatos del modelo físico incluyen información sobre sistemas, esquemas de base de datos, tablas, columnas, restricciones, transformaciones, etc. También es importante establecer los metadatos de negocio y todas las representaciones del EDC en modelos físicos de datos, así como recopilar información sobre el linaje de datos.

El **Linaje de datos**, se refiere a la trazabilidad de EDC's, desde los informes del usuario final en sentido ascendente hasta la fuente final, (desde agregaciones en fuentes como el datawarehouse y datamarts, los almacenes operativos de datos (ODS's), áreas de staging y sistemas transaccionales.



# Estándar EDC – Vista 360° EDC

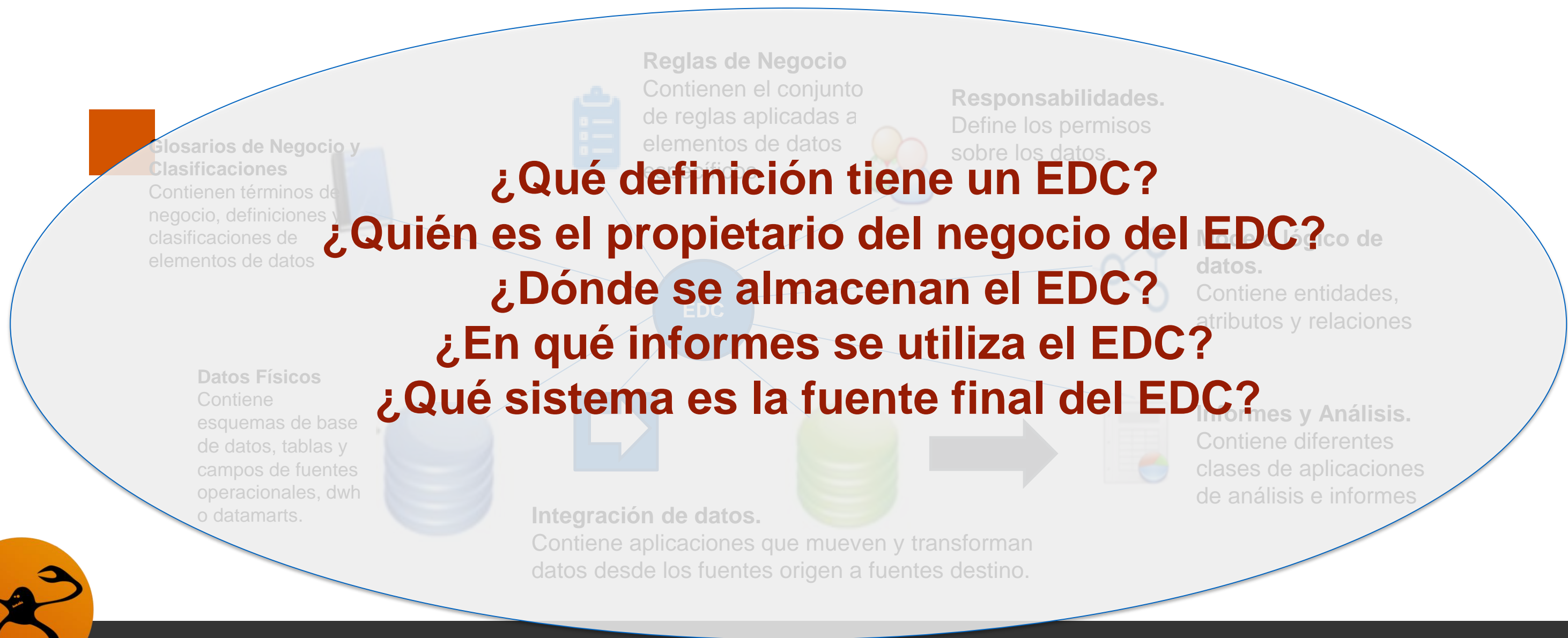
Los metadatos técnicos y de negocios juntos representan **el estándar EDC**, significa que el elemento de datos críticos se describe completamente desde la perspectiva técnica y de negocios. El estándar EDC también se denomina vista 360° del EDC. Al obtener una **vista de 360° del EDC**, se proporcionan las condiciones optimas para gestión de los datos en toda la organización que es el objetivo final de la gestión de datos.



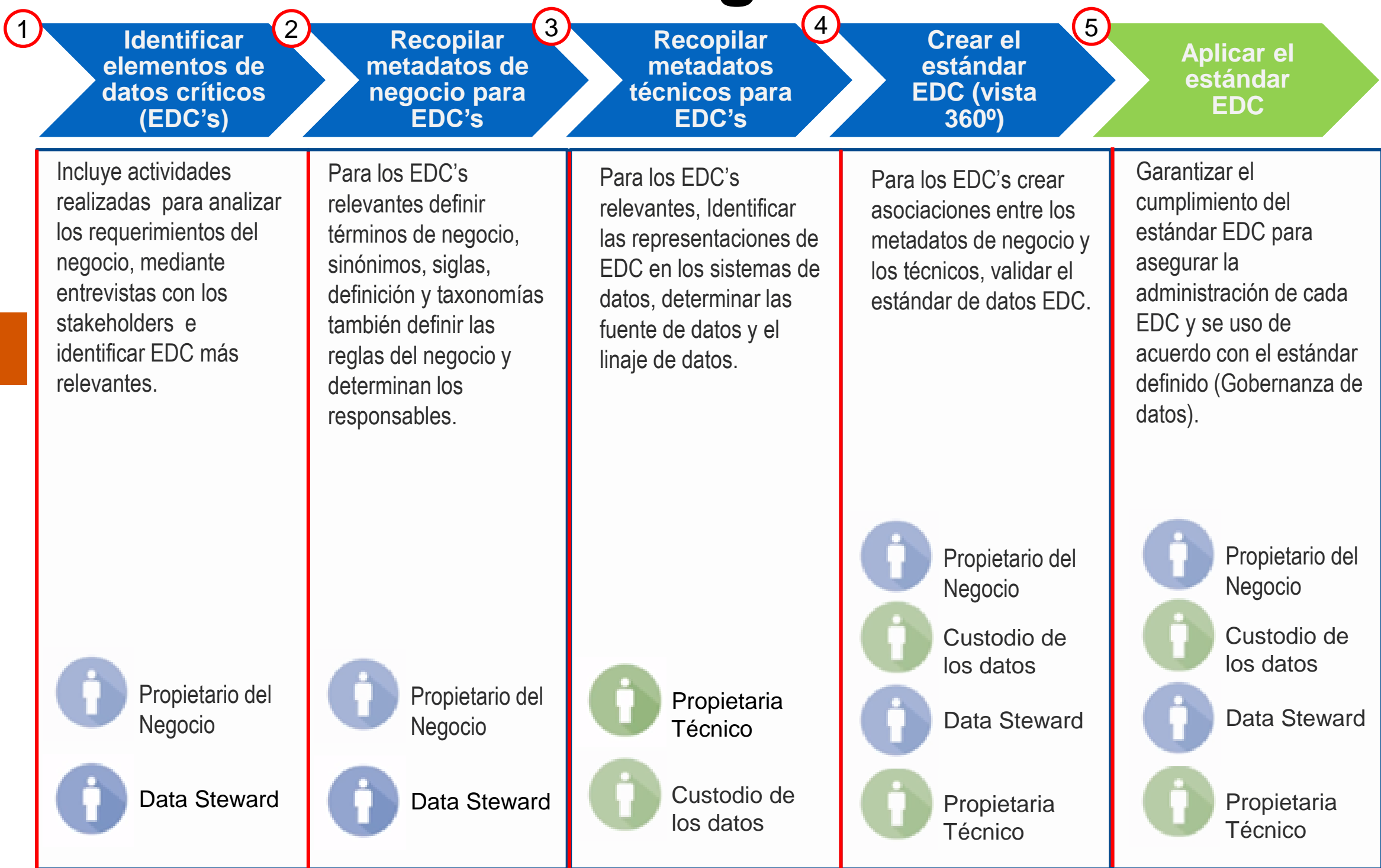


# Estándar EDC – Vista 360° EDC

Los metadatos técnicos y de negocios juntos representan **el estándar EDC**, significa que el elemento de datos críticos se describe completamente desde la perspectiva técnica y de negocios. El estándar EDC también se denomina vista 360° del EDC. Al obtener **una vista de 360° del EDC**, se proporcionan las condiciones optimas para gestión de los datos en toda la organización que es el objetivo final de la gestión de datos.



# Proceso Metadatos Managament



# Proceso Metadatos Managament



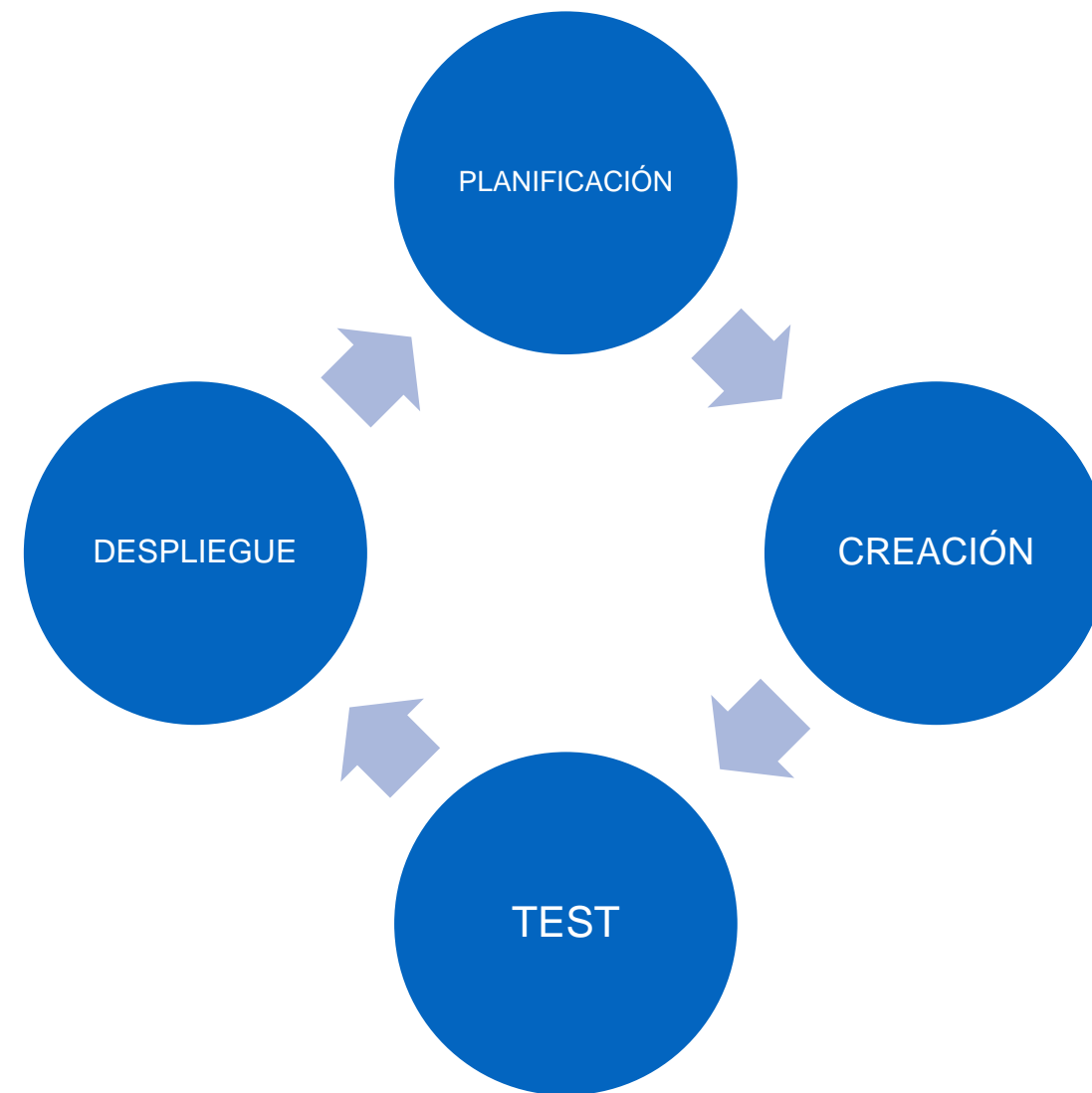
## PROCESOS

Las actividades del proceso del metadatos cubren la perspectiva de procesos dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management



# El ciclo de vida del desarrollo de sistemas (SDLC)

El ciclo de vida del desarrollo del sistema (SDLC) se refiere al proceso de planificación, creación, pruebas y despliegue en un sistema de información.



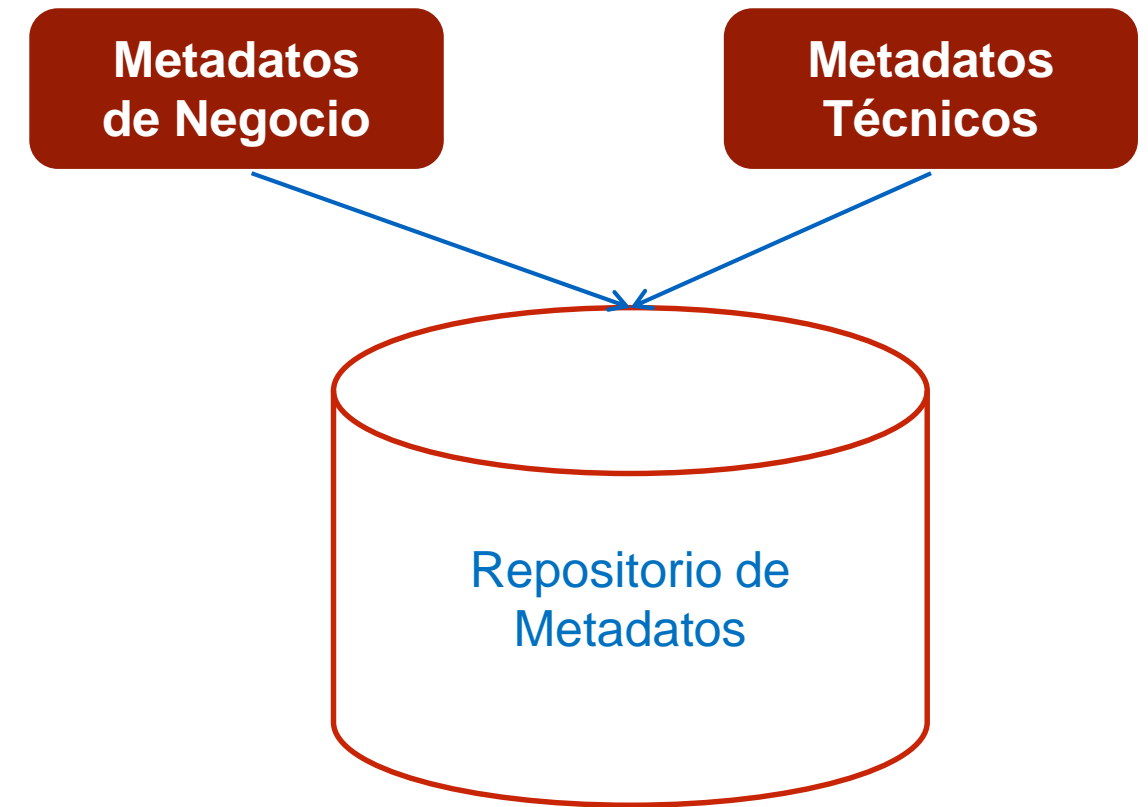
*Es importante que se definan los metadatos derivados del ciclo de vida del desarrollo de los sistemas, y que formen parte del proceso de la función de metadatos.*



# Herramientas Tecnológicas para Metadatos

## Requerimientos de las Herramientas Tecnológicas para la Gestión de Metadatos:


- Capacidad para establecer y desplegar un repositorio de metadatos centralizado.
- Capacidad para definir términos coherentes, definiciones válidas, valores válidos, dominios de datos a través de una jerarquía de glosarios
- Capacidad para extraer metadatos desde las sistemas de gestión de bases de datos más populares (Oracle, SQLServer, NoSQL), desde herramientas de modelado de datos, herramientas de inteligencia empresarial y ETL.
- Capacidad para crear y definir asociaciones entre artefactos del repositorio de metadatos
- Capacidad para administrar versiones de metadatos desde múltiples fuentes y tecnologías.



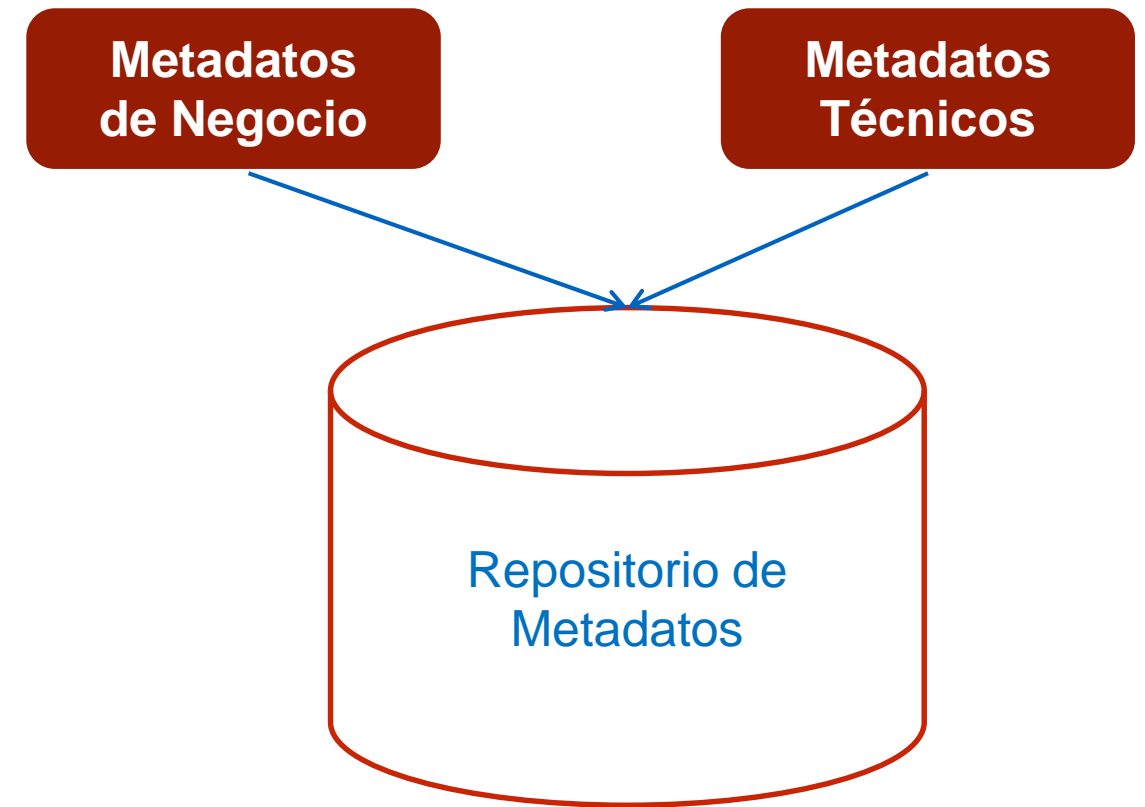


# Herramientas Tecnológicas para Metadatos

Requerimientos de las Herramientas Tecnológicas para la Gestión de Metadatos:


- Capacidad para establecer y desplegar un repositorio de metadatos centralizado.
- Capacidad para definir términos diferentes, definiciones válidas, valores válidos, dominios de  a través de una jerarquía de glosarios
- Capacidad para extraer metadatos desde los sistemas de gestión de bases de datos más populares (Oracle, SQLServer, NoSQL), desde herramientas de modelado de datos, herramientas de inteligencia empresarial y ETL.
- Capacidad para crear y definir asociaciones entre artefactos del repositorio de metadatos
- Capacidad para administrar versiones de metadatos desde múltiples fuentes y tecnologías.

## TECNOLOGÍA





*Las Herramientas Tecnológicas que dan soporte a la gestión de metadatos hacen referencia a la **perspectiva de tecnología***


# Resumen – Puntos Clave




Los **Elementos de Datos Críticos** (EDC's) son componentes fundamentales de la Metadata Management



Los **Metadatos de Negocios** describen los EDC 's desde la perspectiva del negocio e incluyen términos, definición, sinónimos, siglas, reglas de negocios, clasificaciones y responsabilidades.




Los **Roles** de metadatos definen los permisos y responsabilidades en el proceso de gestión de metadatos, incluyen al propietario de negocio, el administrador de datos, propietario técnico y custodio de los datos.




Los **Metadatos Técnicos** describen los EDC's desde la perspectiva de la tecnología e incluyen modelos lógicos de datos, modelos físicos de datos, integración de datos y linaje de datos.



El **Proceso de gestión de metadatos** formaliza las actividades relativas a los metadatos.



Los **Estándares de datos o Vistas 360º** proporcionan una vista completa de los EDC's, incluidos los metadatos de negocio y técnicos.



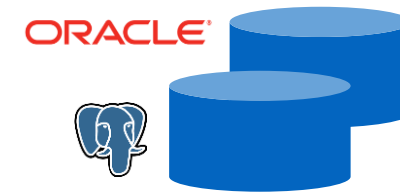
Las **Herramientas Tecnológicas** para gestión de metadatos brindan soporte tecnológico al proceso de gestión de metadatos para garantizar que los EDC's se administren y utilicen adecuadamente.



# Data Management:

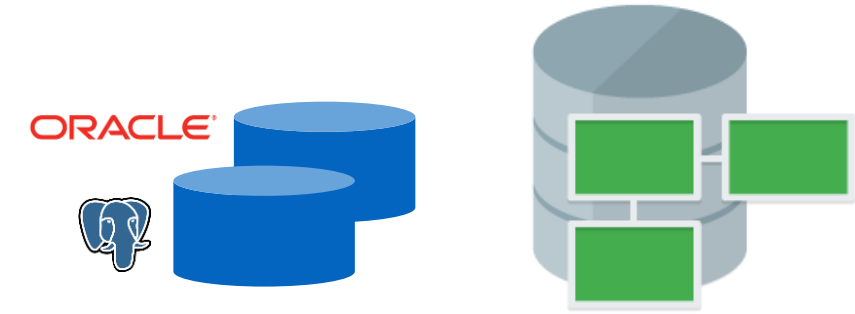
## Gestión holística de los datos empresariales

# Ejercicios Modelado



**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.

# Ejercicios Modelado



## PARTE I . CREAR OBJETOS en BASES DE DATOS POSTGRES

- Ejecutar script BBDD\_Recetas.sql que se entrega con el material de modelado, en PostgreSQL

## PARTE II . CONEXIONES JDBC

- Creación de conexiones JDBC a las bases de datos en SQLDeveloper
- Creación del esquema “lab” y realizar la conexión JDBC

## PARTE III . DATA MODELER

- Creación del modelo lógico y físico de un sistema de Roles y Privilegios de Usuarios.
- Creación de modelo físico utilizando un script de DDL's.
- Ingeniería Inversa de PostgreSQL para obtener el modelo E/R.



# Ejercicios Modelado.

## PARTE I. PostgreSQL. Crear BD

Modelo E/R

El Servicio de Salud dispone de una base de datos para registrar toda la información acerca de las visitas médicas de pacientes, los médicos que trabajan y las recetas que estos dispensan.

Las tablas son las siguientes:

### MÉDICO:



Información acerca de los médicos.

### PACIENTE:

Información acerca de los pacientes.

### VISITA:

Información acerca de las visitas que los pacientes realizan a los médicos.

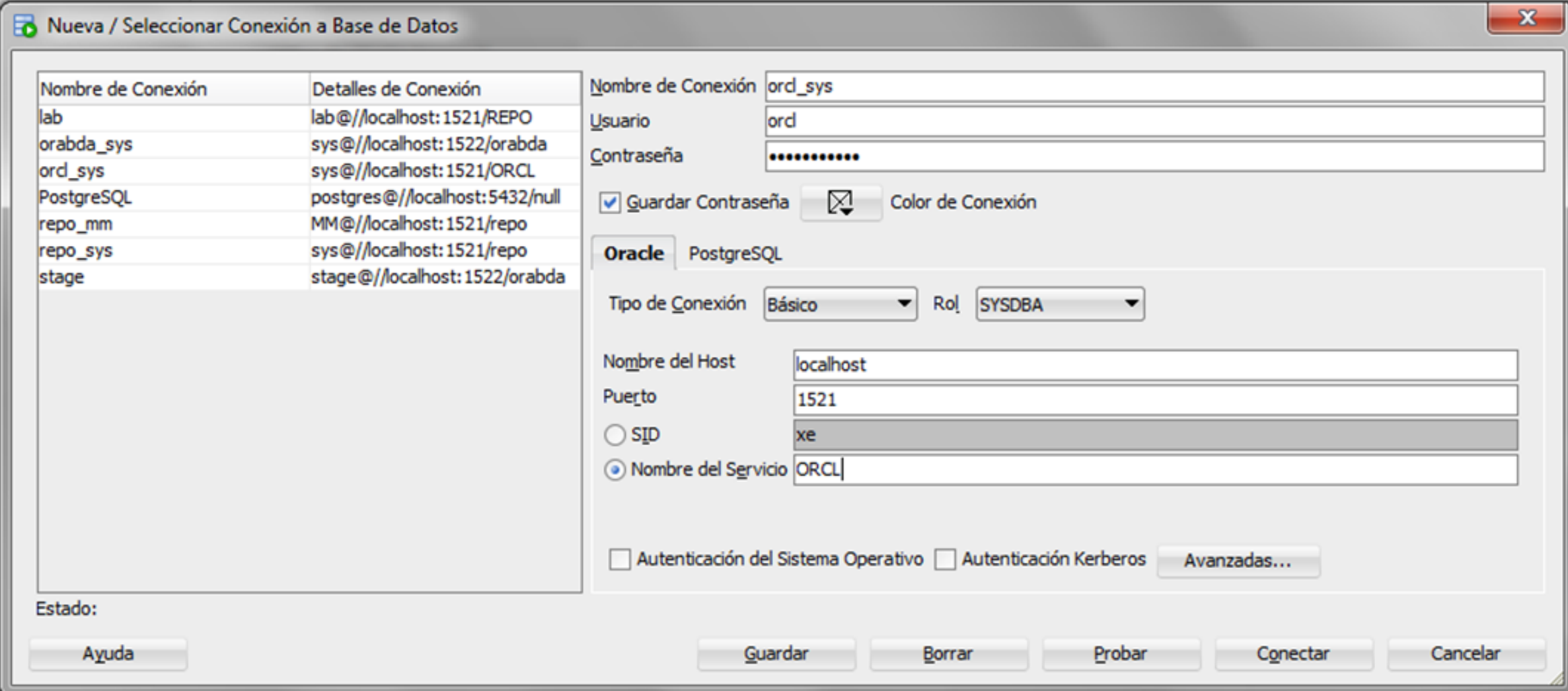
### RECETA.

Información acerca de las recetas que se han dispensado los médicos en las visitas que han tenido con los pacientes.



# PARTE II. SQL Developer. Conexiones JDBC Oracle

Rol: SYSDBA



Nombre de Conexión	Detalles de Conexión
lab	lab@//localhost:1521/REPO
orabda_sys	sys@//localhost:1522/orabda
ord_sys	sys@//localhost:1521/ORCL
PostgreSQL	postgres@//localhost:5432/null
repo_mm	MM@//localhost:1521/repo
repo_sys	sys@//localhost:1521/repo
stage	stage@//localhost:1522/orabda

Nombre de Conexión: ord\_sys

Usuario: ord

Contraseña: .....

☒ Guardar Contraseña ☐ Color de Conexión

**Oracle** PostgreSQL

Tipo de Conexión: Básico Rol: SYSDBA

Nombre del Host: localhost

Puerto: 1521

☐ SID xe

☒ Nombre del Servicio ORCL

☐ Autenticación del Sistema Operativo ☐ Autenticación Kerberos Avanzadas...

Estado:

Ayuda Guardar Borrar Probar Conectar Cancelar





## PARTE II. SQL Developer. Crear Nuevo Esquema

-- Con usuario sys y desde ventana SQL en base de datos Oracle12c creada



```
CREATE USER lab IDENTIFIED BY lab;  
GRANT CONNECT, RESOURCE, DBA TO lab;  
GRANT UNLIMITED TABLESPACE TO TEMP;
```

-- DROP USER lab CASCADE; (Para borrar el usuario creado)



# PARTE II. SQL Developer. Conexiones JDBC Terceros

Pasos para crear conexión JDBC de Terceros como PostgreSQL

1.- El driver JDBC para PostgreSQL se obtendrá desde esta URL:  
<http://jdbc.postgresql.org/download.html>

(La URL se proporciona en el material de modelad, URL\_DriverJBC.txt)



2.- Mirar versión Java, Programas; Java

Descargaremos la [PostgreSQL JDBC 4.2 Driver, 42.2.5](#)

3.- Una vez descargado el driver JDBC para PostgreSQL, lo dejaremos en el path de drivers jdbc de la aplicación SQL Developer. \$SQLDEVELOPER\_HOME/jdbc/lib

4.- Añadir el controlador en SQL Developer. Desde Herramientas; Preferencias.



# PARTE III. Data Modeler.

## Creación Modelo Lógico y Físico

### Sistema de gestión de roles y privilegios

#### Usuarios

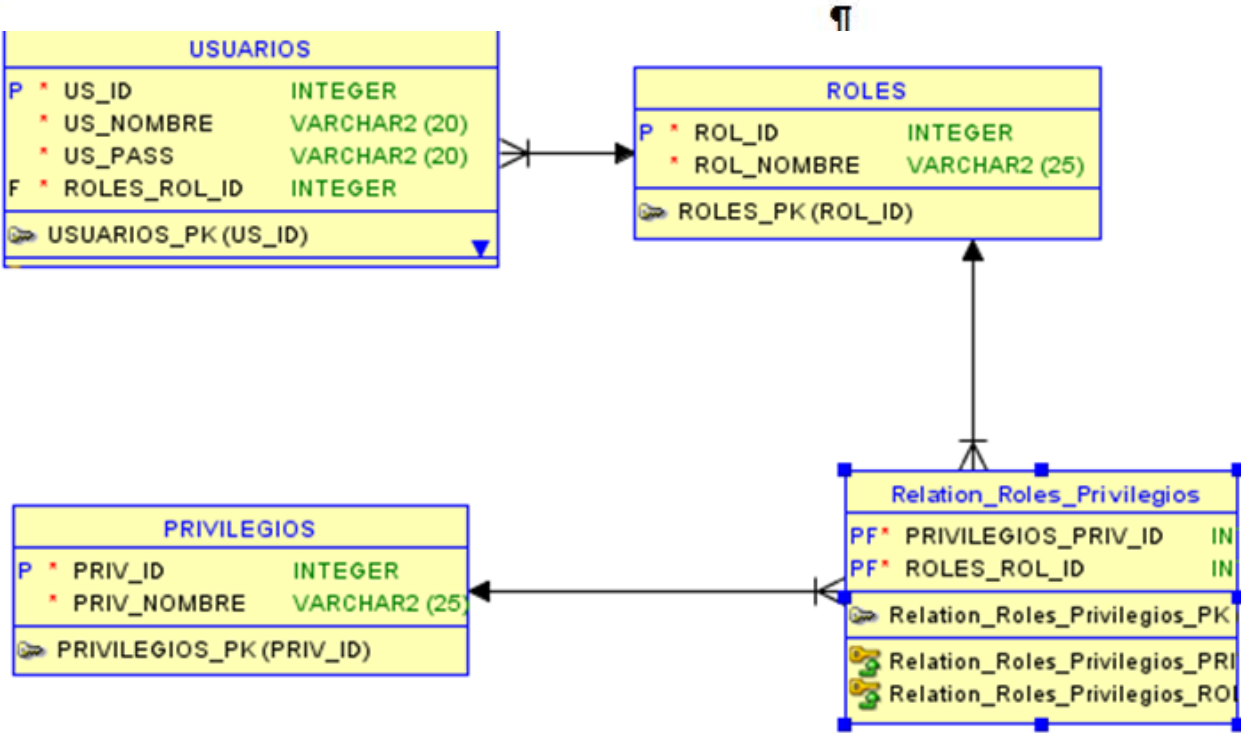
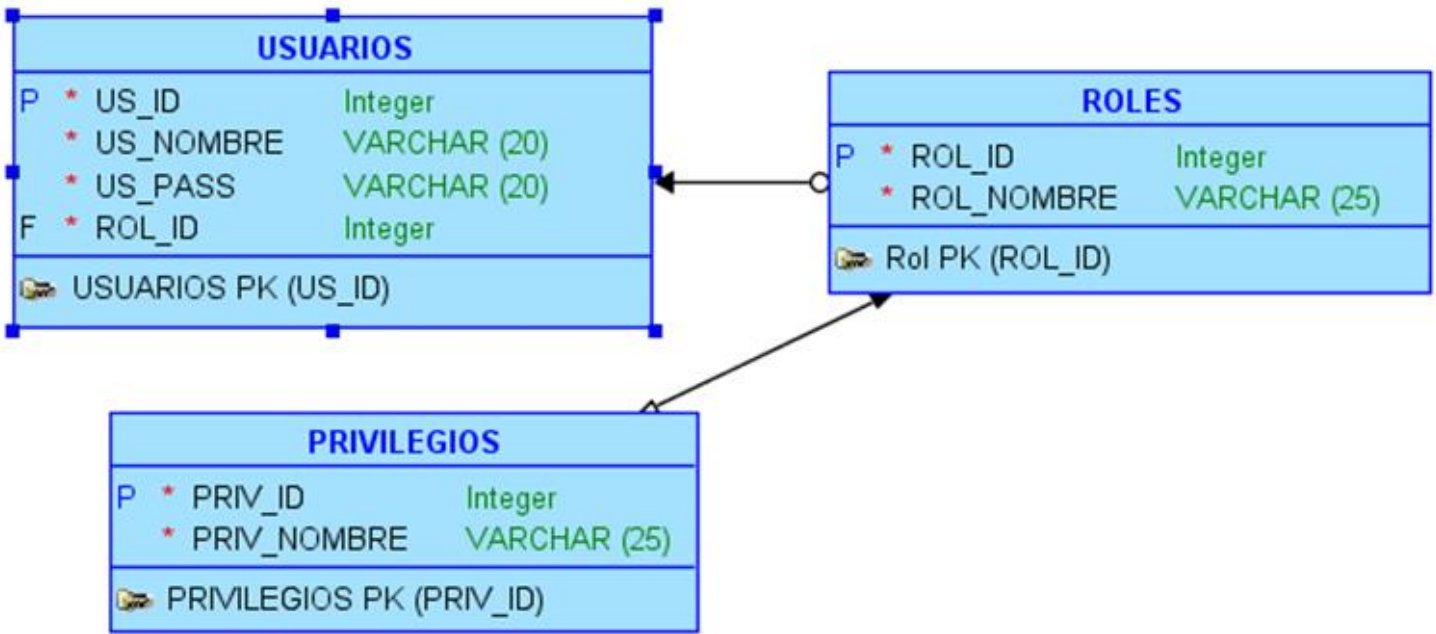
US\_ID (INTEGER) Identificador del usuario. Clave primar  
US\_NOMBRE (VARCHAR 20) Nombre del usuario. Cam  
US\_PASS (VARCHAR 20) Password. Campo obligatorio

#### Roles

ROL\_ID (INTEGER) Identificador del rol. Clave primaria  
ROL\_NOMBRE (VARCHAR 25) Nombre del rol. Campo obligatorio

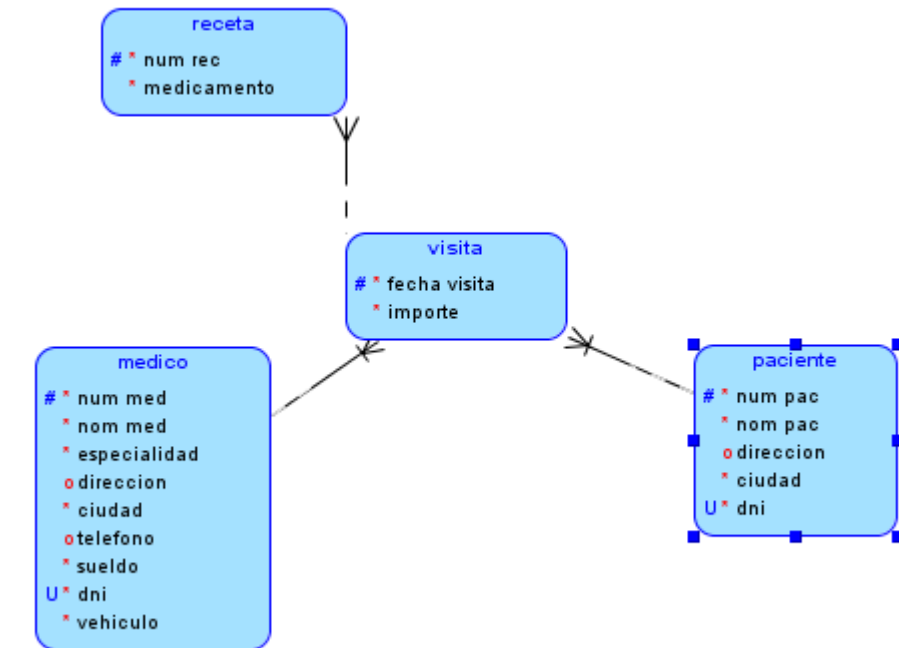
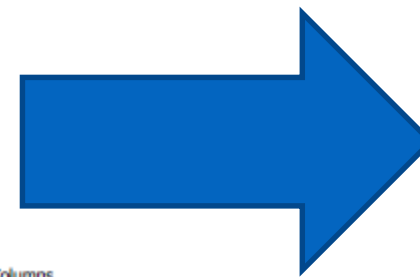
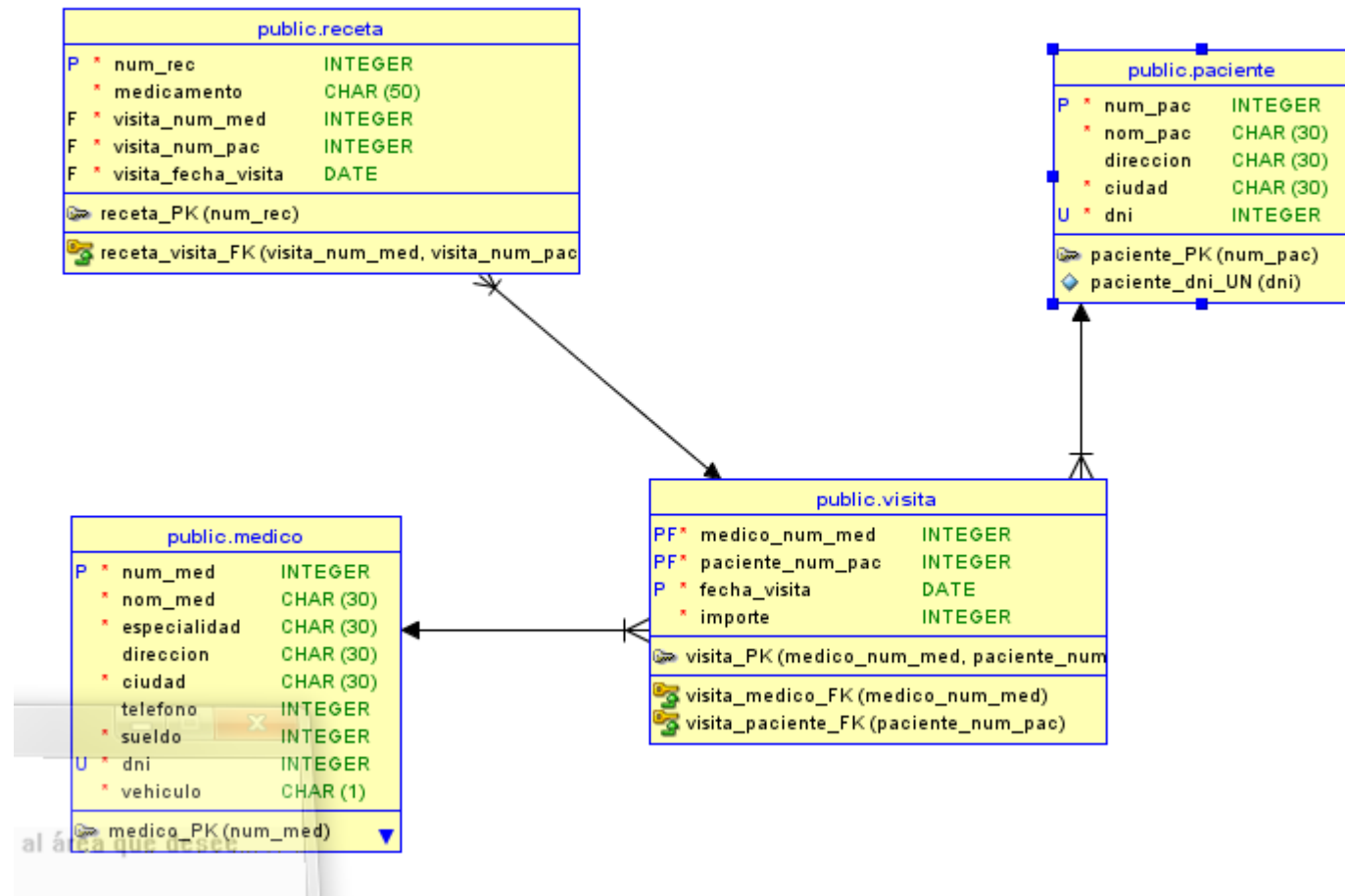
#### Privilegios.

PRIV\_ID (INTEGER) Identificador del privilegio. Clave primaria  
PRIV\_NOMBRE (VARCHAR 25) Nombre del privilegio. Campo obligatorio



# PARTE III. Data Modeler.

## Ingeniería Inversa con PostgreSQL



Columns										
No	Column Name	PK	FK	M	Data Type	DT Kind	Domain Name	Formula (Default Value)	Security	Abbreviation
1	num_med	P		Y	Integer	LT				
2	nom_med			Y	CHAR (30)	LT				
3	especialidad			Y	CHAR (30)	LT				
4	direccion				CHAR (30)	LT				
5	ciudad			Y	CHAR (30)	LT				
6	telefono				Integer (10)	LT				
7	sueldo			Y	Integer (10)	LT				
8	dni			Y	Integer	LT				
9	vehiculo			Y	CHAR (1)	LT				

Columns Comments			
No	Column Name	Description	Notes
9	vehiculo	1.Tiene Vehiculo 0.No tiene vehiculo	



# CONCLUSIONES.

Un modelo lógico es una simplificación de la realidad que nos apoya en cuatro cosas:

1. Nos ayuda a visualizar un sistema como deseamos que sea diseñado. Un modelo ayuda al equipo de trabajo a comunicar la visión del sistema que se está construyendo. Es muy difícil compartir una misma visión si solo se cuenta con especificaciones.
2. Nos permite especificar la estructura y conducta del sistema. Un modelo permite documentar la estructura y conducta de un sistema antes de que sea codificado.
3. Nos da una plantilla que guía el proceso de construcción. Un modelo es una herramienta invaluable durante la construcción, sirve como guía para el programador. ¿Alguna vez ha tenido la experiencia de que un programador codifique la funcionalidad incorrecta debido a que confundió la descripción textual de un requerimiento? El modelado ayuda a aliviar esta situación.
4. Documenta las decisiones que hemos tomado. Los modelos son herramientas que apoyan al proyecto a largo plazo, ya que documentan las decisiones de diseño tomadas y ya no solo se depende de la memoria.



# Data Management:

Gestión holística de los datos empresariales

## CALIDAD DE DATOS

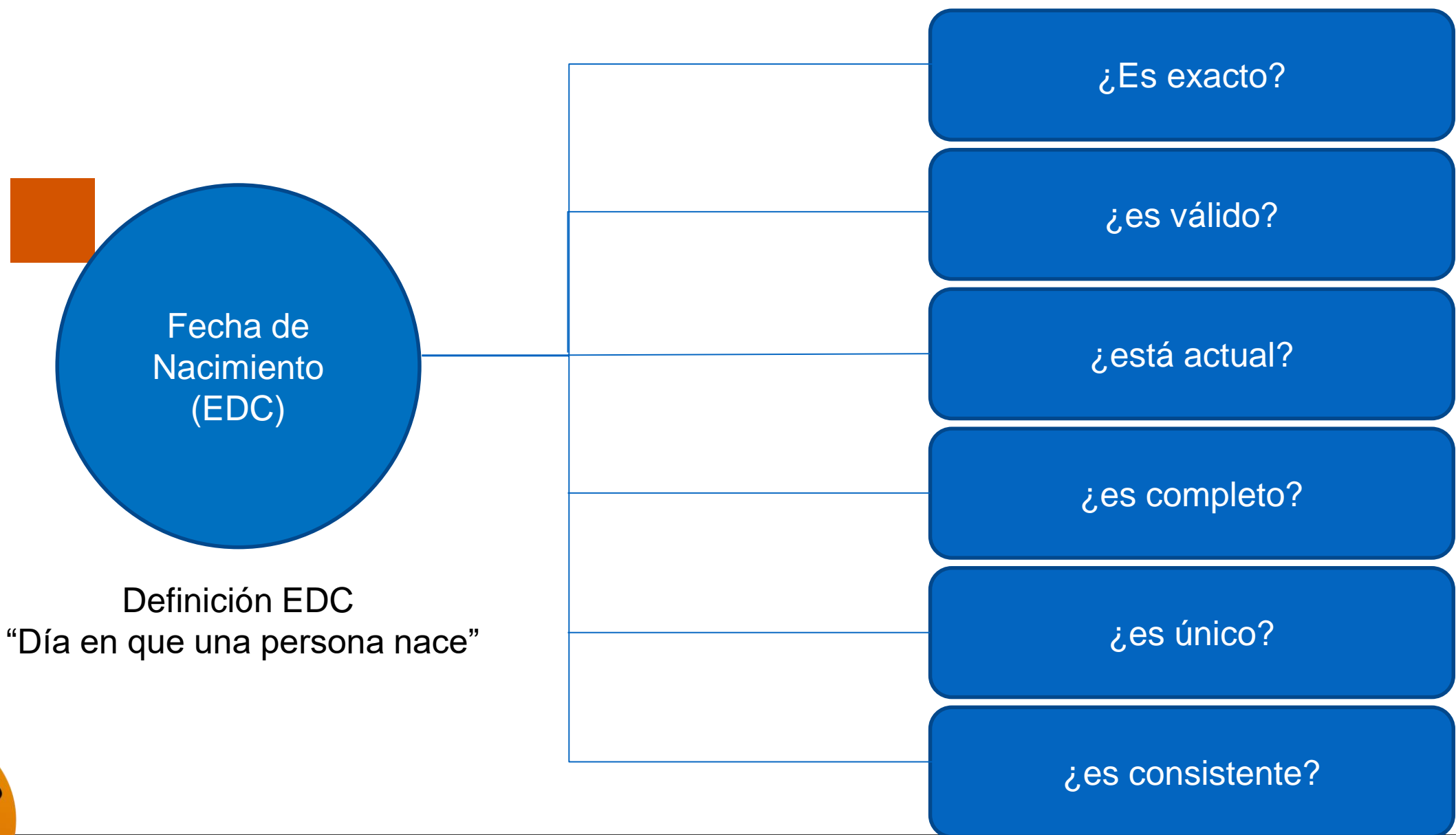


**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.



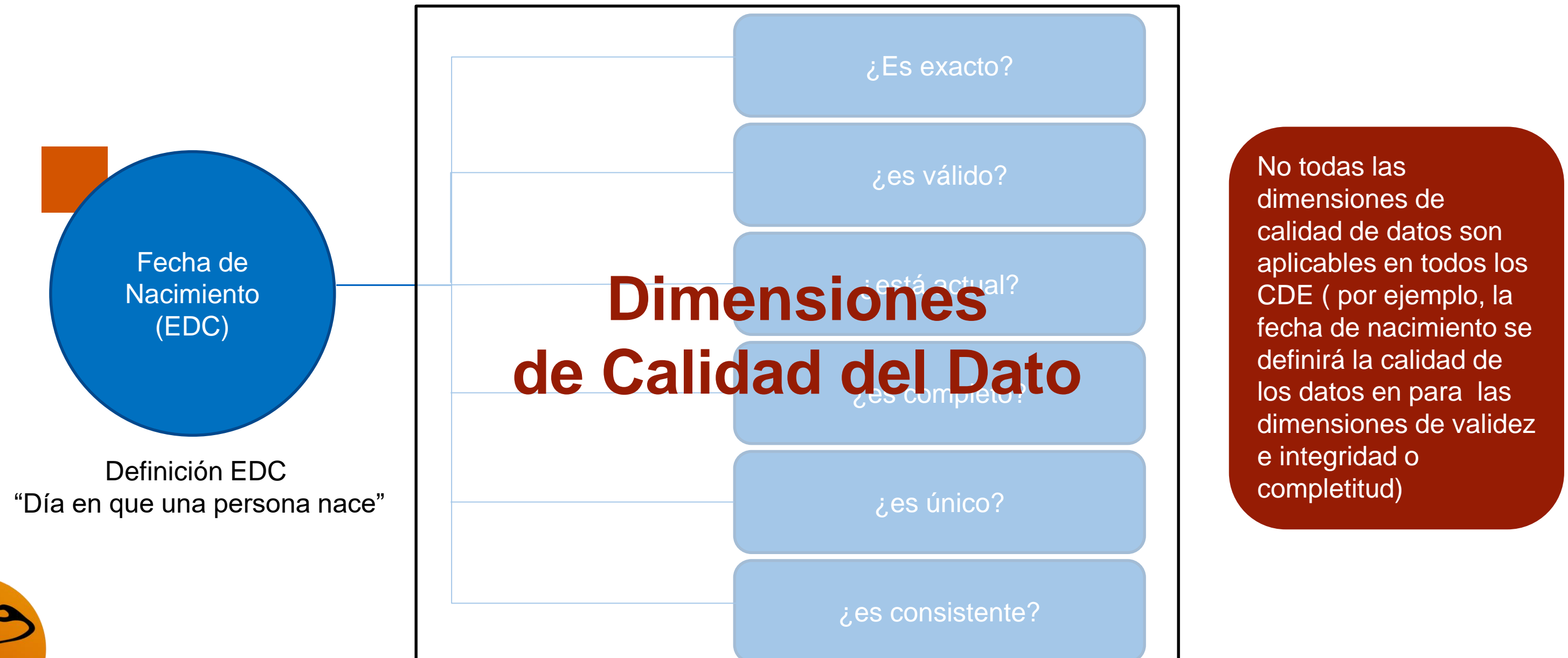
# Gestión de la Calidad de datos

La **calidad de los datos** hace referencia a la metodología, políticas y los procesos mediante los cuales una organización garantiza la exactitud, la consistencia, conformidad, la integridad, la unicidad y la consistencia de sus datos en los sistemas y los flujos de datos.



# Gestión de la Calidad de datos

La **calidad de los datos** hace referencia a la metodología, políticas y los procesos mediante los cuales una organización garantiza la exactitud, la consistencia, conformidad, la integridad, la unicidad y la consistencia de sus datos en los sistemas y los flujos de datos.



# Dimensiones de Calidad de Datos

Las **dimensiones de la calidad de los datos** se refiere a los aspectos o características de la información que se pueden evaluar y utilizar para determinar la calidad de los datos.

## 6 Dimensiones Clave en la calidad de los datos

EXACTITUD	VALIDEZ	ACTUALES	COMPLETITUD	UNICIDAD	CONSISTENCIA
Significa que los datos representan con precisión el mundo real	Los datos se ajustan a la sintaxis de su definición (formatos, tipos y rangos)	Los datos representan la realidad desde el punto de vista del tiempo.	Los datos completos en términos de importancia para el negocio.	Los datos se identifican correctamente y se registran solo una vez.	Los datos se representa de la misma manera en todo el conjunto de datos.

Ejemplo: Errores de ortografía en nombres de productos, personas o direcciones

Ejemplo: Valores incorrectos para el género y tipos de clientes.

Ejemplo: Cambio de dirección de un cliente que entra en vigor el 1Julio y se introduce al sistema el 15deJulio

Ejemplo: Dirección de cliente a la que le falta el código postal

Ejemplo: cliente único que se registra dos veces en la base de datos con diferentes identificadores.

Ejemplo: Borrado del num de cuenta de cliente pero hay un pedido asociado a esa cuenta.



# Reglas de Calidad de datos

Las reglas de **calidad de los datos** hacen referencia a las reglas de negocio destinadas a garantizar la calidad de los datos en términos en términos de exactitud, validez, vigencia, integridad, unicidad y consistencia..



Definición EDC

“Día en que una persona nace”

**Validez:** La validez de la fecha de nacimiento, debe estar comprendida en el rango desde el año 1900 hasta la fecha actual.

**Compleitud:** La fecha de nacimiento debe ser insertada para cada individuo, no se permiten campos vacíos.



# Reglas de Calidad de datos

The screenshot shows the 'Propiedades de Entidad - CLIENTE' window. The 'Atributos' tab is active, displaying a list of attributes. The attribute '6 FECHA NACIMIEN...' is selected, and its properties are shown on the right. The 'Tipo de Origen' is set to 'Date'. The 'Obligatorio' checkbox is checked. A 'Restricción' dialog is open, showing a list of database versions. The restriction 'Between 1990 and sysdate' is selected for the 'Oracle Database 12c' version.

Nombre	Tipo de Dato
1 ID CLIENTE	NUMERIC (5)
2 NOMBRE CLIENTE	VARCHAR (50)
3 TIPO CLIENTE	CHAR (1)
4 DIRECCIÓN DEL ...	VARCHAR (100)
5 ID PEDIDO	NUMERIC
6 FECHA NACIMIEN...	Date

Propiedades de Atributo

Nombre: FECHA NACIMIENTO

Tipo de Dato: ☐ Dominio ☒ Lógico ☐ Distinto

☐ Estructurado ☐ Recopilación

Tipo de Origen: Date

Preferencia: ☐

☐ UID Primario ☐ UID de Relación ☒ Obligatorio ☐ Anticuoado

Comentarios: Comentarios en RDBMS: Notas:

Restricción

Tipo	Restricción
Generic Constraint	
Oracle Database 12cR2	
Oracle Database 12c	Between 1990 and sysdate
Oracle Database 11g	
Oracle Database 10g	
Oracle9i	
SQL Server 2012	
SQL Server 2008	
SQL Server 2005	
SQL Server 2000	
DB2/390 8	
DB2/390 7	
DB2/UDB 9	
DB2/UDB 8.1	
DB2/UDB 7.1	

Aceptar Aplicar Cancelar

**Complejidad:** La fecha de nacimiento debe ser insertada para cada individuo, no se permiten campos vacíos.

**Validez:** La validez de la fecha de nacimiento, debe estar comprendida en el rango desde el año 1900 hasta la fecha actual.



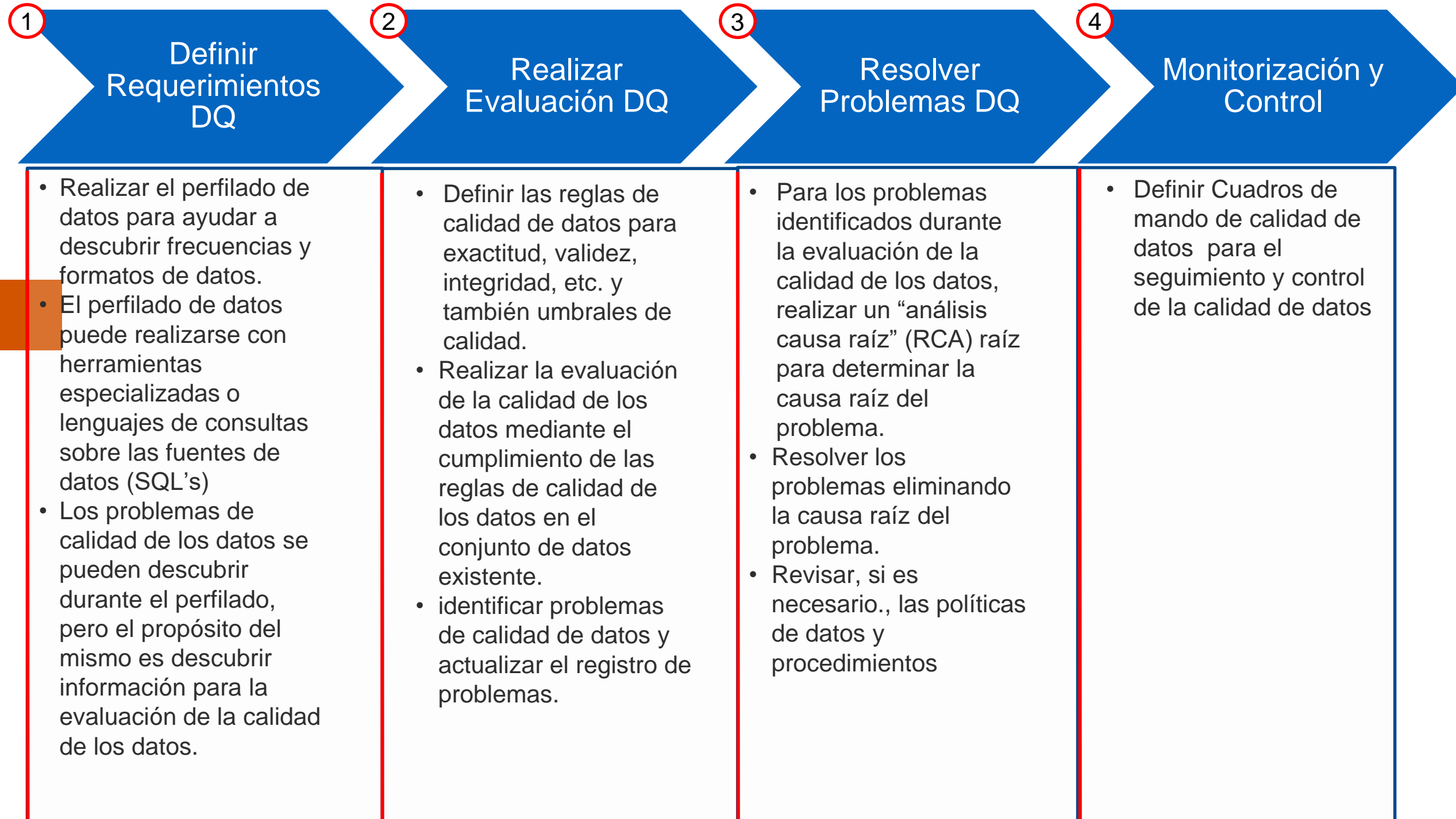
# Reglas de Calidad de datos

Las reglas de **calidad de los datos** hacen referencia a las reglas de negocio destinadas a garantizar la calidad de los datos en términos en términos de exactitud, validez, vigencia, integridad, unicidad y consistencia..

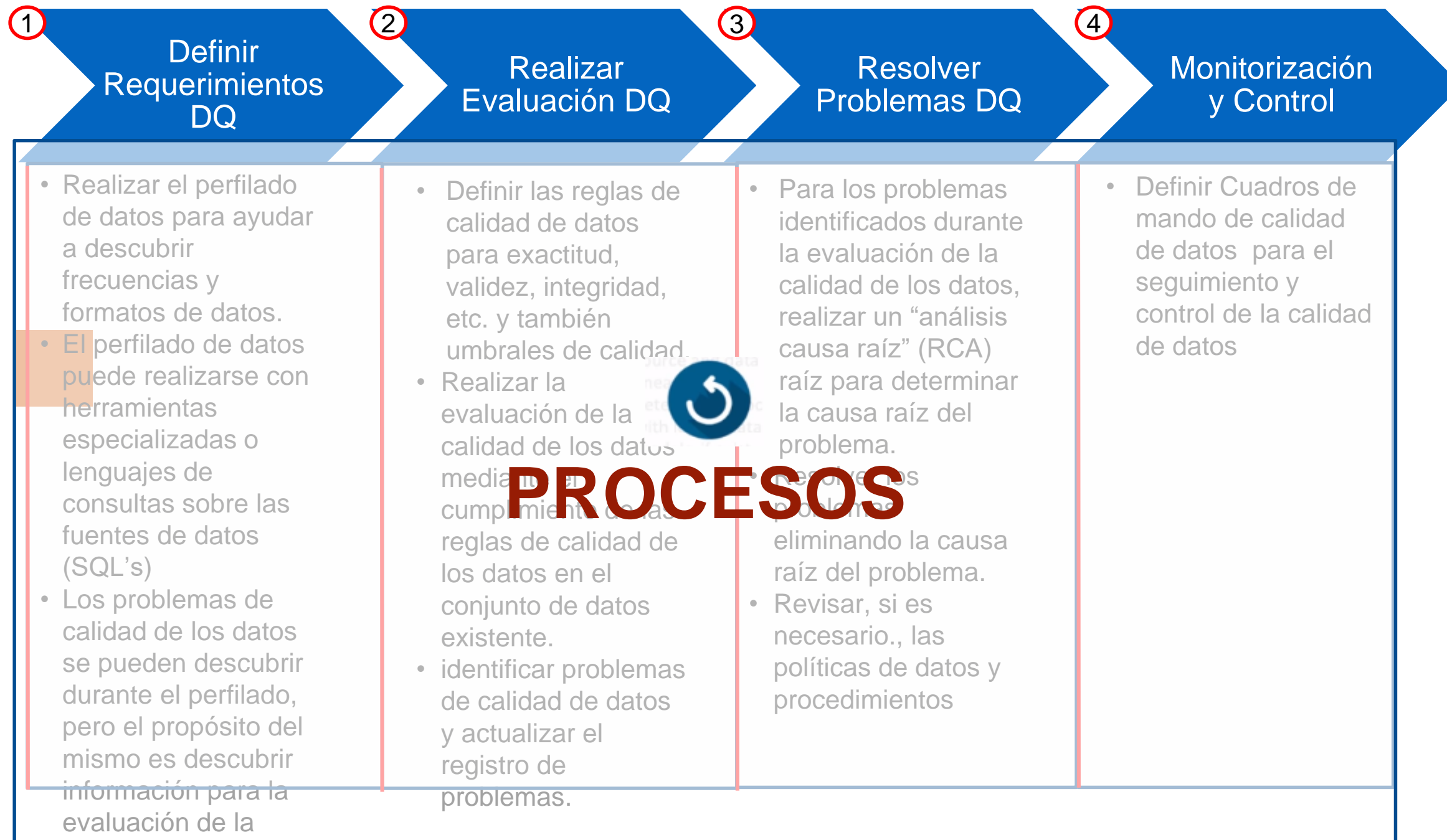




# Proceso de Calidad de datos



# Proceso de Calidad de datos



*Las actividades del proceso del calidad del dato cubren la perspectiva de procesos dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management*



# Roles de Calidad de datos



El **Analista de calidad de datos** representa el rol clave de la calidad del dato y es responsable de realizar las actividades asociadas con proceso de calidad de datos. Si bien es el único rol específico de la calidad de los datos, trabajará estrechamente con el propietario del negocio, los administradores de datos, los propietarios técnicos y custodios de los datos. Esto incluye, entre otras, la definición de las reglas de calidad de los datos, el análisis de los resultados, la elaboración de perfilados, la evaluación, la investigación de las causas de los problemas de calidad de los datos, etc.



Analista DQ



Propietario del negocio



Administrador de datos  
(Data Steward)



Propietario Técnico



Custodios de los datos



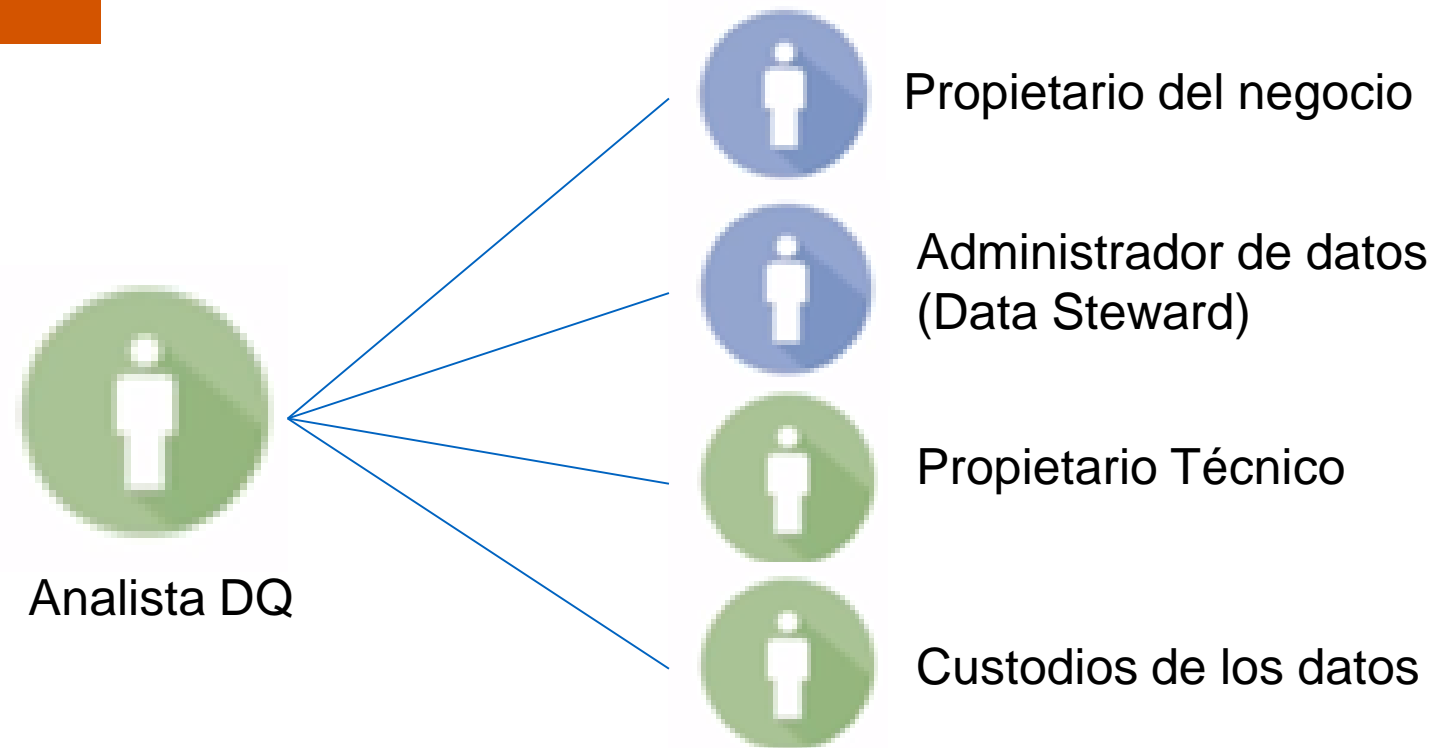
Rol de  
Negocio



Rol Técnico



# Roles de Calidad de datos



La definición del Rol en la Calidad de Datos y sus responsabilidades cubren la perspectiva de personas dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management



# Ejemplo de Calidad de Datos

## Definición de un Caso de Uso

Realizar la evaluación de la calidad de datos para los datos de los Empleados, concretamente del elemento crítico de datos, Fecha de Nacimiento. El umbral para la calidad de los datos es del 99%, para cada dimensión de calidad de los datos



Definición EDC

“Día en que una persona nace”

**Validez:** La validez de la fecha de nacimiento, debe estar comprendida en el rango desde el año 1900 hasta la fecha actual.

**Compleitud:** La fecha de nacimiento debe ser insertada para cada individuo, no se permiten campos vacíos.



# Ejemplo de Calidad de Datos

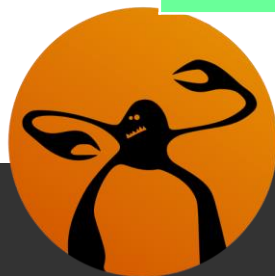
## Perfilado de datos – Profiling

Datos de Empleados

ID	Nombre Completo	Fecha Nacimiento
1	Susana López	10/01/1968
2	Alejandro Dumas	12/03/1990
3	Manuel González	07/24/1973
4	Pedro Santisteban	02/30/1968
5	Dolores Detripa	
6	Sergio García	12/12/1981
7	Ana Sanjuan	09/08/1983
8	Marisa Paredes	1/22/1991

Datos de Empleados - Resultados del Perfilado de datos

Categoría	Resultado
Número de Registros	8
Número de Valores Únicos	7
Número de Nulos	1





# Ejemplo de Calidad de Datos

## Evaluación DQ

### Reglas DQ

- Validez: La fecha de nacimiento será válida si está comprendida en el rango desde el año 1900 hasta la fecha actual.
- Completitud: La fecha de nacimiento debe ser insertada para cada individuo.

### Datos de Empleados

ID	Nombre Completo	Fecha Nacimiento
1	Susana López	10/01/1968
2	Alejandro Dumas	12/03/1990
3	Manuel González	07/24/1973
4	Pedro Santisteban	02/30/1868
5	Dolores Detripa	
6	Sergio García	12/12/1981
7	Ana Sanjuan	09/08/1983
8	Marisa Paredes	1/22/1991

### Datos de Empleados – Resultados de la Evaluación

Dimensión	Regla	Válidos	Fallos	Rtdo
Validez	Fecha Nacimiento válida en el rango entre 1900 -Fecha actual	6	2	75%
Completitud	Fecha Nacimiento insertada para cada individuo	7	1	88%



# Ejemplo de Calidad de Datos

## Evaluación DQ

### Reglas DQ

- Validez: La fecha de nacimiento será válida si está comprendida en el rango desde el año 1900 hasta la fecha actual.
- Completitud: La fecha de nacimiento debe ser insertada para cada individuo.

### Datos de Empleados

ID	Nombre Completo	Fecha Nacimiento
1	Susana López	10/01/1968
2	Alejandro Dumas	12/03/1990
3	Manuel González	07/24/1973
4	Pedro Santisteban	02/30/1868
5	Dolores Detripa	
6	Sergio García	12/12/1981
7	Ana Sanjuan	09/08/1983
8	Marisa Paredes	1/22/1991

### Datos de Empleados – Resultados de la Evaluación

Dimensión	Regla	Válidos	Fallos	Rtdo
Validez	Fecha Nacimiento válida en el rango desde 1900 hasta la fecha actual	6	2	75%
Completitud	Fecha Nacimiento insertada para cada individuo	7	1	88%

**Scorecard DQ**



# Ejemplo de Calidad de Datos

## Resolución de Problemas de DQ

### Causa Raíz de Problemas DQ

- Validez: los datos introducidos en este campo no se validan con el formato de fecha válido
- Completitud: la fecha de nacimiento no se establece como obligatorio en la base de datos.

### Datos de Empleados

ID	Nombre Completo	Fecha Nacimiento
1	Susana López	10/01/1968
2	Alejandro Dumas	12/03/1990
3	Manuel González	07/24/1973
4	Pedro Santisteban	02/30/1968
5	Dolores Detripa	4/14/1971
6	Sergio García	12/12/1981
7	Ana Sanjuan	09/08/1983
8	Marisa Paredes	1/22/1991

### Datos de Empleados – Resultados de la Evaluación

Dimensión	Regla	Válidos	Fallos	Rtdo
Validez	Fecha Nacimiento válida en el rango entre 1900 -Fecha actual	8	0	100%
Completitud	Fecha Nacimiento insertada para cada individuo	8	0	100%

### Medidas Preventivas

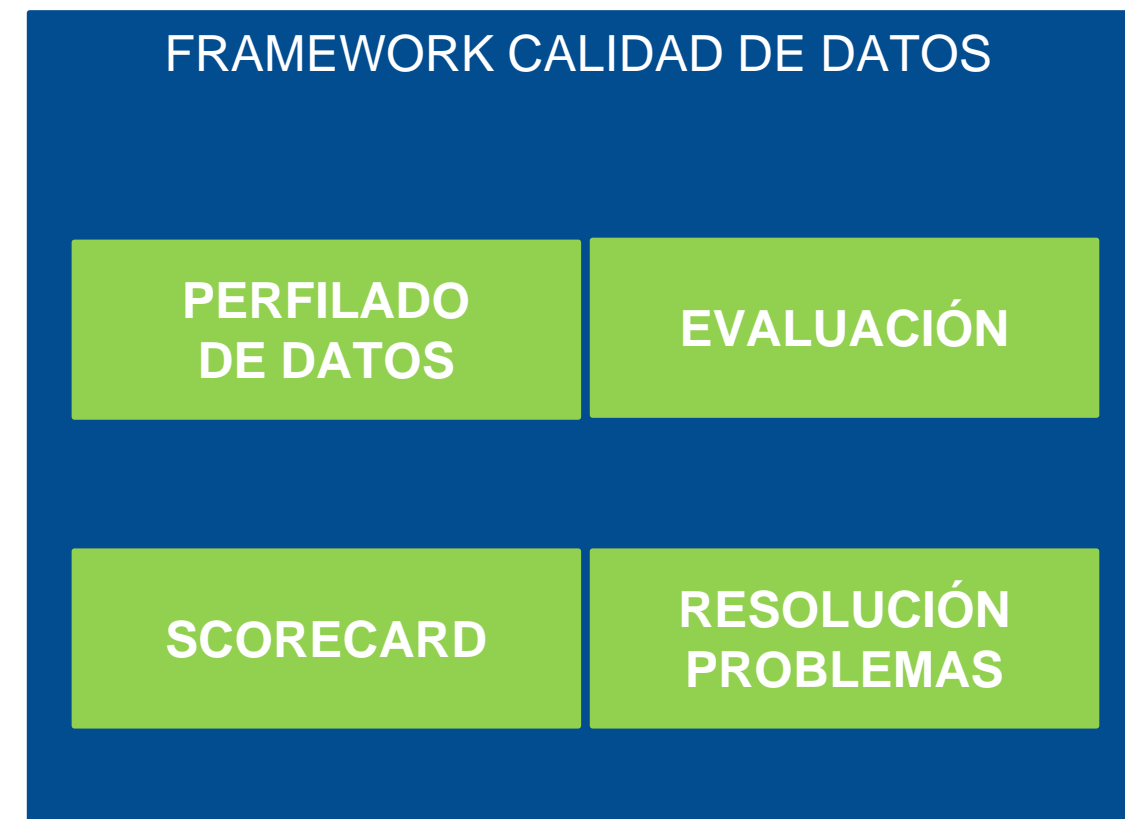
1. Implementar una validación en el campo para controlar que la fecha de nacimiento está entre el rango de fechas definido en la Regla DQ.
2. Establecer el campo de nacimiento de datos como obligatorio en la base de datos.



# Herramientas Tecnológicas para Calidad de datos

## Requerimientos de las Herramientas Tecnológicas para la Gestión de Calidad de Datos:

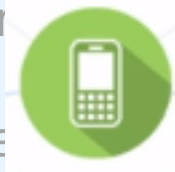
- Capacidad para realizar perfilado de datos, incluyendo análisis estadísticos de datasets.
- Capacidad para definir y ejecutar reglas de calidad de datos para control de calidad de datos críticos
- Capacidad para almacenar evaluaciones y resultados de calidad de datos.
- Capacidad para llevar a cabo el proceso de resolución y descubrimiento de problemas.
- Capacidad para crear y visualizar Scorecard de calidad de datos



# Herramientas de Soporte para Calidad de datos

## Requerimientos de las Herramientas Tecnológicas para la Gestión de Calidad de Datos:

- Capacidad para realizar perfiles de datos, incluido el análisis estadístico de conjuntos de datos.
- Capacidad para definir y ejecutar reglas de calidad de datos para elementos de datos críticos que son temas de control de calidad de datos.
- Capacidad para almacenar datos de calidad de datos y resultados de evaluación.
- Capacidad para llevar a cabo el proceso de resolución de problemas y descubrir patrones de problemas.
- Capacidad para crear y visualizar cuadros de mando de calidad de datos



## TECNOLOGÍA

### FRAMEWORK CALIDAD DE DATOS

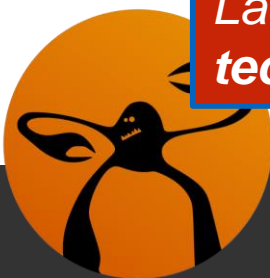
PERFILADO  
DE DATOS

EVALUACIÓN

SCORECARD

RESOLUCIÓN  
PROBLEMAS

*Las Herramientas Tecnológicas que dan soporte a la gestión de calidad de datos hacen referencia a la **perspectiva de tecnología** dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management*



# Resumen – Puntos Clave

La **Calidad de los Datos** es importante en los **elementos de datos críticos** (EDC's).

La calidad de los datos se refiere a diferentes aspectos o características de los EDC's. A aspectos o características se les denomina **Dimensiones de Calidad de Datos**.

Hay 6 Dimensiones Clave de la calidad de los datos: exactitud, validez, actualidad, completitud, unicidad y consistencia

El **Proceso de Calidad de Datos** consta de 4 actividades: Definir Requerimientos DQ. Realizar la evaluación DQ. Resolver Problemas DQ y Monitorizar y Control. Todas son realizadas por Analistas de Calidad de Datos en cooperación con el Propietario del negocio, Data Steward, el Propietario Técnico y el Custodio de los datos.

El **Perfilado de Datos** (Profiling) consiste en obtener información estadística sobre los datos.

La **Evaluación de la Calidad Datos** se refiere a evaluar la calidad de los datos e identificar los problemas de calidad de los datos. Se basa en Reglas y Umbrales. Las reglas de calidad se asocian con las dimensiones de la calidad de los datos. Los **ScoreCards DQ** visualizan los resultados de la evaluación.

Los **Análisis de Causa Raíz** sirven para encontrar la causa raíz de un problema de calidad de datos en particular.

Las **Herramientas de calidad de datos** proporcionan apoyo tecnológico al proceso de calidad de datos para garantizar la calidad de los datos de acuerdo con los requerimientos del negocio.



# Data Management:

## Gestión holística de los datos empresariales

### Ejercicios Calidad



TRIFACTA



**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.

# Ejercicios Calidad



## PARTE II . TRIFACTA WRANGLER ONLINE

- Preparación de datos utilizando el dataset `potenciales_clientes.xls` entregado como materia
- Practicar la limpieza sobre un dataset de datos abiertos utilizado anteriormente pero ahora con Trifacta Wrangler.



# Ejercicios Calidad



## PARTE I . TALEND PREPARATION

- Preparación de datos utilizando el dataset `potenciales_clientes.xls` entregado como material
- Realizar ejercicio de limpieza sobre un dataset de datos abiertos.

Utilizar el fichero de datos abiertos del Gobierno de Canarias

El dataset contiene observaciones de cetáceos por parte de empresas autorizadas en las costas de las islas. Son datos recogidos por diferentes empresas y por esta razón se registran datos con errores de calidad de datos. Se puede descargar en formato xls o csv.



# CONCLUSIONES.

- La tarea de depuración de datos (Data Cleaning) es una de las que más tiempo ocupa a los Data Scientist. Es una tarea que se puede realizar sin necesidad de tener conocimientos de programación utilizando herramientas más especializadas como Talend Preparation o Trifacta Wrangler.

