EDEM
Escuela de Empresarios

GFT

Data Strategy

Master Data Analytics para para la empresa

Pedro Nieto

EDEM
Escuela de Empresarios

Agenda

1. Data Strategy
2. Roles y Responsabilidades
    1. Data Consumer
    2. Platform Owner
    3. Data Steward
    4. Content Owner
3. Physical Model
    1. SQL
    2. Views
4. Comités de funcionamiento
5. Glosario de Términos
6. Capa Semántica / Ontología
7. Trazabilidad
8. Flujos de Trabajo

Data Strategy

From our experience there is not such a thing as a magic receipe for establishing a succesful Data Strategy.

Combination of different aspects on an iterative mode, can evolve your organization into a data oriented approach.

Small changes driven by a clear mindset can make big changes in your organization.

Data Pillars

We need to analyse what do we expect from a Data Strategy.

These will be the requirements that every policy and principle needs to aim, these requirements needs to be on a very high common sense model.

Data Governance Framework

“Data can be found quickly”

“Data needs to be easily understood”

“Data is factually correct”

“Data is always complete”

“Data is consistent, at any point or any time”

“Data needs to be traced”

Why a Data Strategy?

Towards data-centric architectures

Analysis and reporting issues are more often related to data governance issues, not technology issues.

As the organizations move into a more **data-centric** world, data governance becomes more critical for ensuring the **data is consistent, reliable and usable for analysis.**

Regulatory pressure

Financial institutions are subject to an ever-growing set of regulations, putting immense pressure on staff to comply with each requirement. **Non-compliance is not an option**.

As the need to produce **higher-quality, real-time information** accelerates and compliance draws more attention from internal and external stakeholders, finance and tax teams face growing **complexity**, along with mounting pressure on their stretched resources to get compliance right.

Data Strategy

A data strategy helps by **ensuring that data is managed and used like an asset**.

It provides a common set of goals and objectives across projects to ensure data is used both effectively and efficiently.

Cultural mind-set change

Education is the first flag we need to raise in this area, the way we teach our people to approach data as an asset instead of a tool will set up the basics to build an strategy capable of dealing with this ecosystem.

Offering

- After several successful engagements GFT has consolidated best practices and lessons learned, making this experience available to its clients when handling data change in large organisations
- Some key aspects are:

Define

- Evaluate current situation
- Define vision
- Introduce Data Principles as corporate policy
- Introduce the rigor and support around data into the Project Delivery Model
- Set up a controlled change framework

Align

- Align structure and strategy
- Ensure stakeholder participation
- Build motivation and ask for feedback
- Engage with in-flight change programs to determine need for support, alignment and explore synergies
- Train and educate people

Manage

- Manage priorities
- Manage emotional response
- Monitor progress
- Drive ongoing communication and campaigns

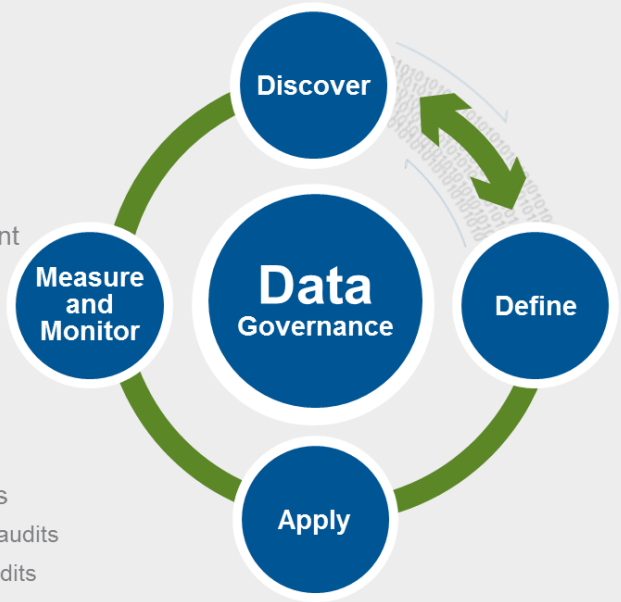
“Data is part of your business, not a tool for it”

Discover

- Data discovery
- Data profiling
- Data inventories
- Process inventories
- CRUD analysis
- Capabilities assessment

Measure and Monitor

- Proactive monitoring
- Operational dashboards
  - Reactive operational DQ audits
  - Dashboard monitoring/audits
- Data lineage analysis
- Program performance
- Business value/ROI



Define

- Business glossary creation
- Data classifications
- Data relationships
- Reference data
- Business rules
- Data governance policies
- Other dependent policies
- Key Performance Indicators

Apply

- Automated rules
- Manual rules
- End to end workflows
- Business/IT collaboration

Business and IT shared objectives

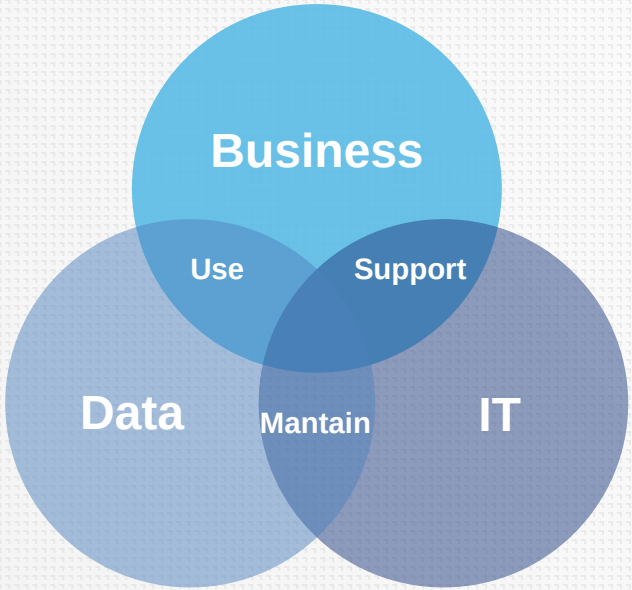
Data is the shared area between IT and Business, and must be seen as a cooperative point of contact between them.

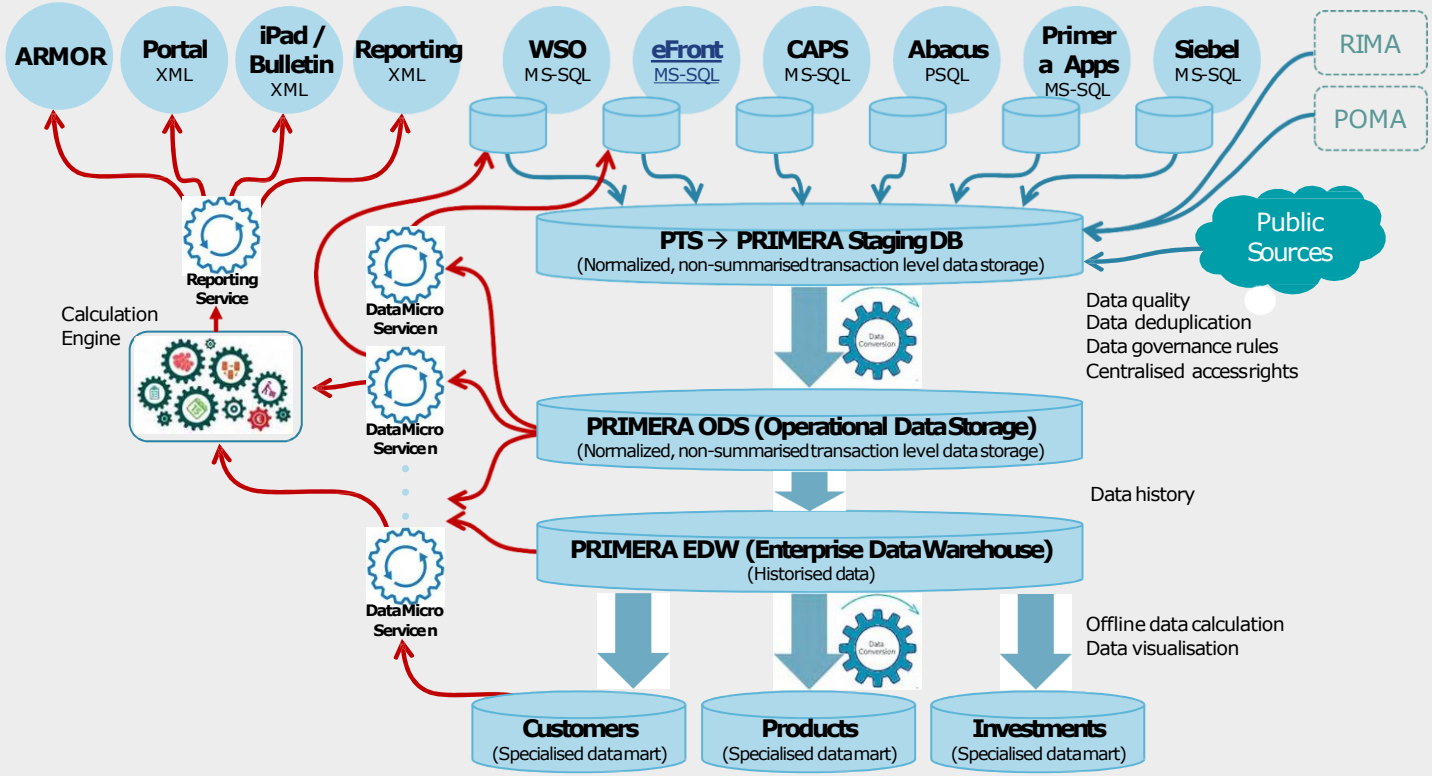
Data strategy must draw the guidelines to establish this relationship and set common goals that both teams should accomplish.

Shared Asset: No one owns the data, data is an asset

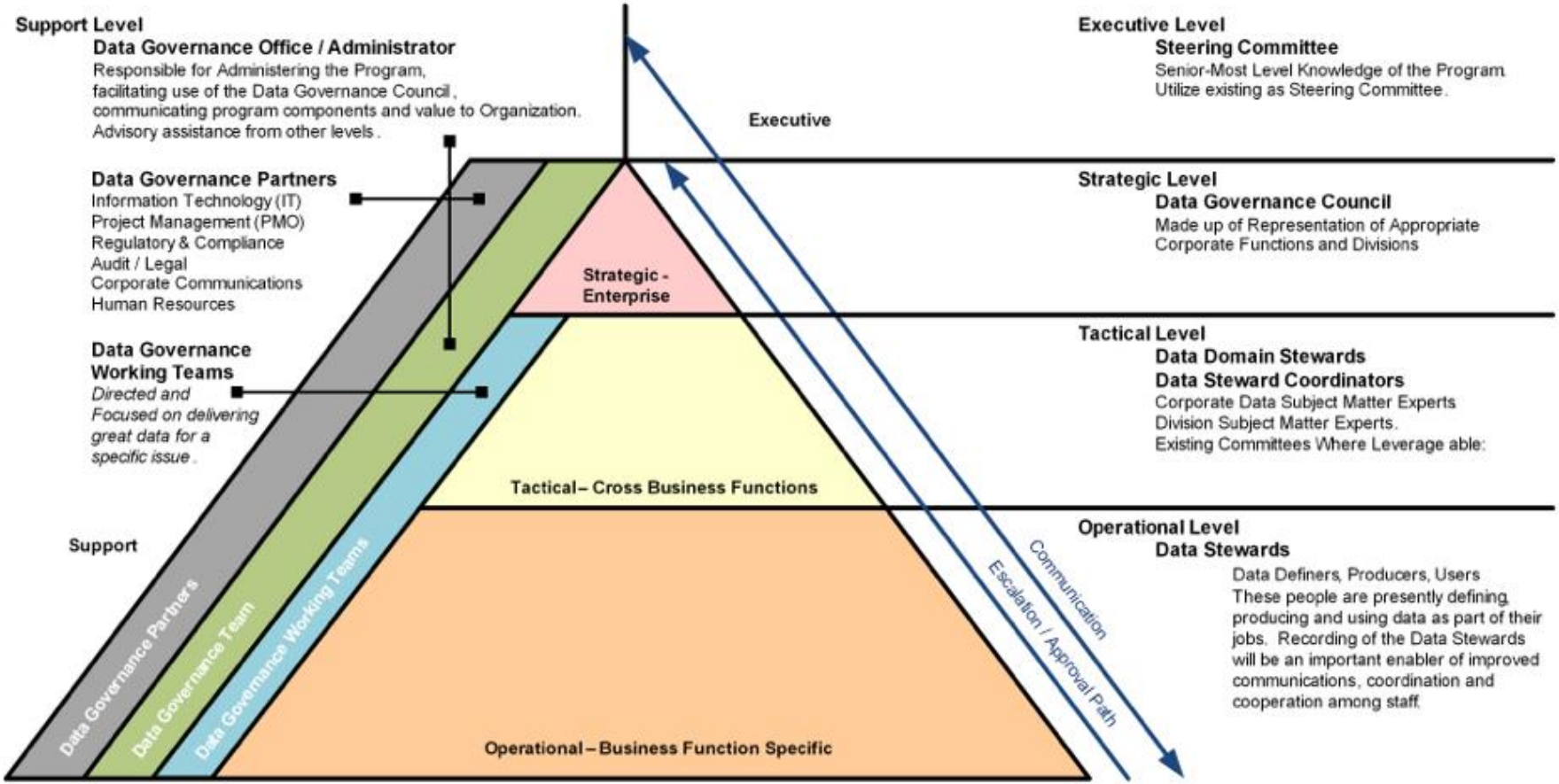
Shared Responsability: Everyone is in charge of mantaining and enriching data

Data as a Service: Data must fit for purpose




ARMOR
Portal
XML
iPad / Bulletin
XML
Reporting
XML
WSO
MS-SQL
eFront
MS-SQL
CAPS
MS-SQL
Abacus
PSQL
Primer a Apps
MS-SQL
Siebel
MS-SQL
RIMA
POMA
Public Sources
Reporting Service
Calculation Engine
DataMicro Service n
DataMicro Service n
DataMicro Service n
PTS → PRIMERA Staging DB
(Normalized, non-summarised transaction level data storage)
Data quality
Data deduplication
Data governance rules
Centralised accessrights
PRIMERA ODS (Operational Data Storage)
(Normalized, non-summarised transaction level data storage)
Data history
PRIMERA EDW (Enterprise Data Warehouse)
(Historised data)
Offline data calculation
Data visualisation
Customers
(Specialised data mart)
Products
(Specialised data mart)
Investments
(Specialised data mart)

Roles y Responsabilidades


Support Level
Data Governance Office / Administrator
Responsible for Administering the Program, facilitating use of the Data Governance Council, communicating program components and value to Organization. Advisory assistance from other levels.
Data Governance Partners
Information Technology (IT)
Project Management (PMO)
Regulatory & Compliance
Audit / Legal
Corporate Communications
Human Resources
Data Governance Working Teams
Directed and Focused on delivering great data for a specific issue.
Support
Data Governance Partners
Data Governance Team
Data Governance Working Teams
Executive
Strategic - Enterprise
Tactical – Cross Business Functions
Operational – Business Function Specific
Communication
Escalation / Approval Path
Executive Level
Steering Committee
Senior-Most Level Knowledge of the Program. Utilize existing as Steering Committee.
Strategic Level
Data Governance Council
Made up of Representation of Appropriate Corporate Functions and Divisions
Tactical Level
Data Domain Stewards
Data Steward Coordinators
Corporate Data Subject Matter Experts
Division Subject Matter Experts.
Existing Committees Where Leverage able:
Operational Level
Data Stewards
Data Definers, Producers, Users
These people are presently defining, producing and using data as part of their jobs. Recording of the Data Stewards will be an important enabler of improved communications, coordination and cooperation among staff.

Framework Definition

Roles & Responsabilities

Data Consumer

- Define data requirements based on Glossary
- Use approved data sources by owner
- Support data lineage policies
- Report data quality issues
- Report data modifications to the owner

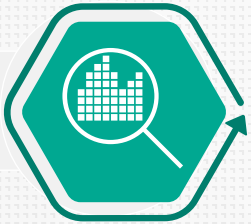
Content Owner

- Creating, maintaining and distributing data.
- Identify Stewards
- Manage and Fulfill Consumers Requirements
- Maintain Glossary and Lineage
- Ensure data quality
- Handle & Track quality issues

Platform Owner

- Control Data Access to authorized data consumers
- Capture technical requirements for the data platform
- Ensure integrity of the data platform
- Define and Maintain physical data model
- Establish archiving policies

Data Steward



- Day-to-Day operational Management on behalf of Owner
- Collaborate with Owners to ensure data policies
- Investigate data quality issues
- Escalate data quality issues
- Identify gaps in data definitions

Data Consumer

Data Consumer: Data final user which is in charge of requesting what the business need to accomplish their functions. Consumers can be seen as units, applications, people…

Data Consumer

Specific Accountabilities:

- Notify data usage
- Do not resell data out of contracted
- Maintain control of data and raise all issues generated by data
- Do not modify data
- Report data breaches

Content Owner / Data Owner

Data Owner: The Data Owner is accountable for the data within a specific Data Domain. They are responsible to ensure that information within their Domain is governed across systems and lines of business. Data Owners usually are part of the Steering Committee, either as voting or non-voting members.

Content Owner

Specific Accountabilities:

- Approve data Glossaries and other data definitions
- Ensure the accuracy of information as used across the Enterprise
- Direct Data Quality activities
- Review and Approve Master Data Management approach, outcomes, and activities
- Work with other Data Owners to resolve data issues and dissonance across business units
- Second level review for issues identified by Data Stewards
- Provide input to the Steering Committee on software solutions, policies or Regulatory Requirements that impact their data domain

Content Owner / Data Owner

Data Steward: The Data Steward has the accountability for the day-to-day management of data. They are the Subject Matter Experts who understand and communicate the meaning and use of information. They are responsible to work with the other Stewards across the organization as the governing body for most data decisions and issue resolutions. They will represent the Data Owner in most discussions. Data Stewards utilize the Data Owners and the Governance Steering Committee as "Appellate" organizations when the Stewards Council cannot resolve a data issue.

Data Steward

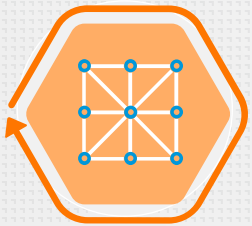
Specific Responsibilities:

- Serve as a subject matter expert (SME) for your data domain
- Identify and work with other data steward to resolve data issues
- Act as a member of the Data Steward council
- Propose, discuss, and vote on data policies and committee activities
- Report activities and decision of the Stewards to the Data Owner and the other Stakeholders within a data domain
- Ensure that Stakeholders' interests are represented at the Steward's Council.
- Work cross functionally across lines of business to ensure their domain's data is managed and understood

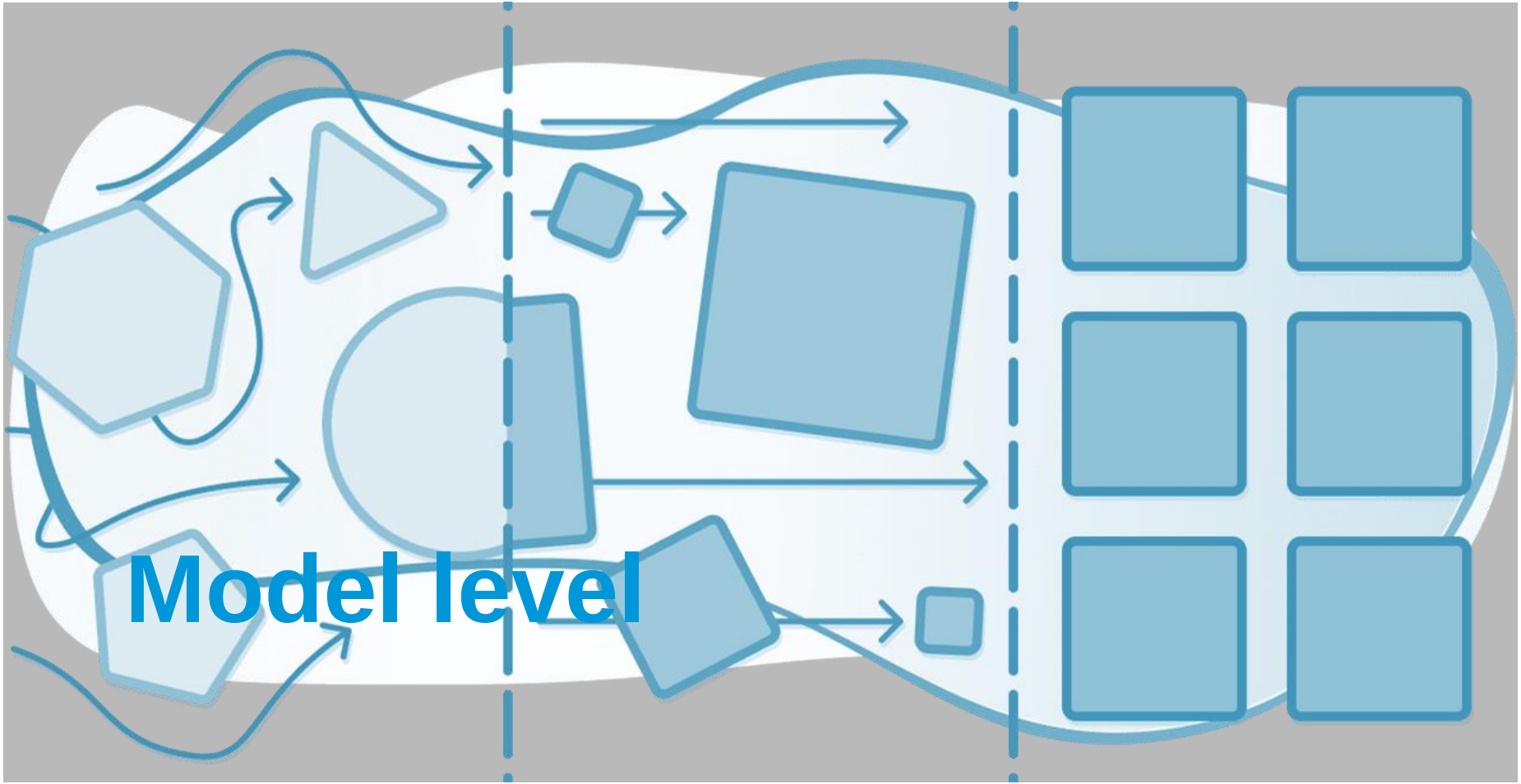
Platform Owner / Technical Owner

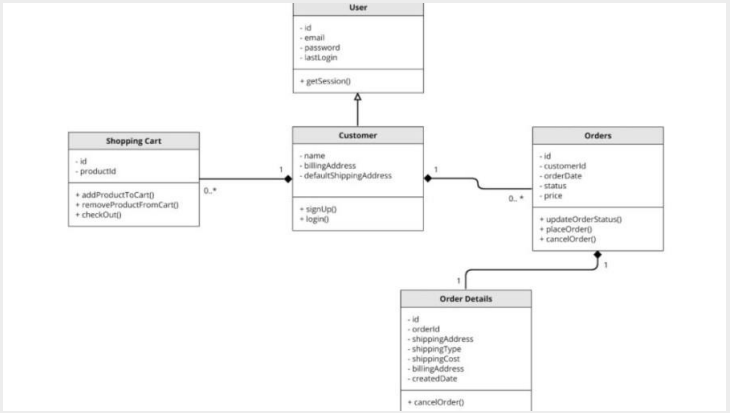
Platform Owner: Platform owner is accountable for all needs regarding technology in the organization and the impact that every business requirement needs to be addressed. The stability of the platform and the image of the organization may relay on a correct evaluation of the non-technical needs.

Platform Owner



- Technical handling of data to meet data classification requirements.
- Securing IT infrastructure on behalf of the business units that own or have responsibility for data.
- Assuring that sensitive data, regardless of format, is protected at all times by only using approved equipment, networks, and other controls.
- Ensuring that standard project methodology is followed and that policies, procedures and metrics are in place for maintaining/improving data quality and the creation, capture, and maintenance of metadata.
- Providing technical support for ensuring data quality.
- Providing technical support for data governance and data cleansing efforts where required.
- Ensuring that metadata critical to data governance is included in the metadata resource and is accessible.


Model level



UNIFIED
MODELING
LANGUAGE

amazon

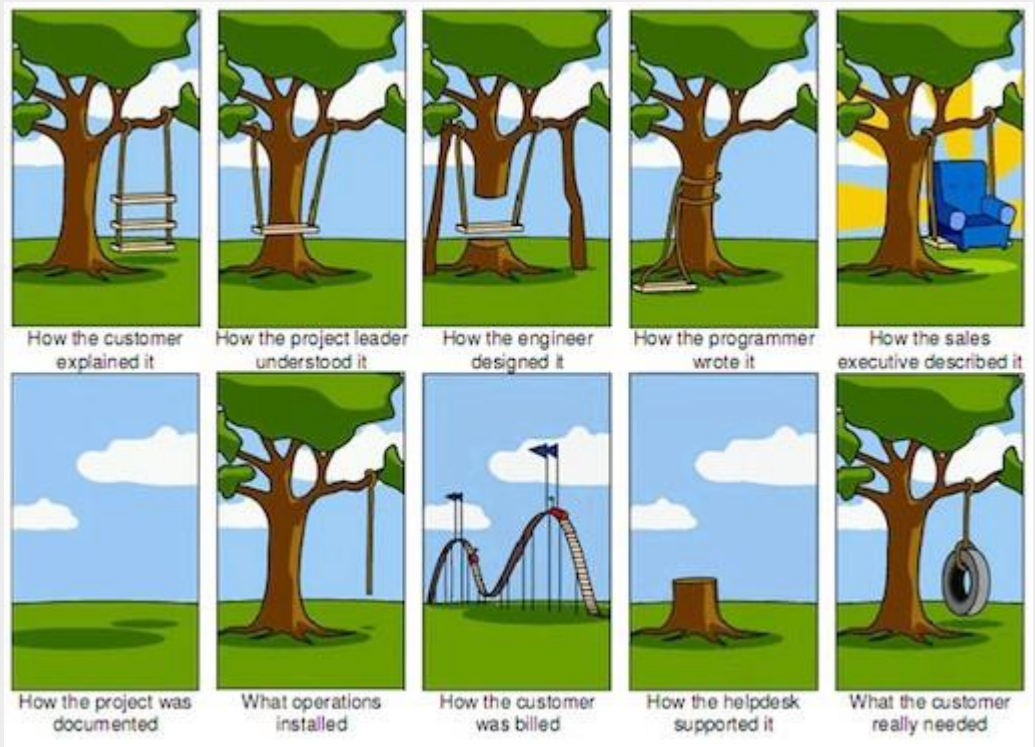
https://app.diagrams.net/

https://www.uml.org/

LOGICAL DATA MODEL VERSUS PHYSICAL DATA MODEL

LOGICAL DATA MODEL	PHYSICAL DATA MODEL
Model that describes the data as much as possible, without regard to how they will be physical implemented in the database	Model that represents how the actual database is built
Defines the data elements and their relationships	Allows developing the actual database
Data Architects and business analyts create logical data model	Database Administrators and developers create physical data model
The objective of logical data model is to develop a technical map of rules and data structures	The objective of physical model is to implement the actual database
Simpler than the physical data model	Complex than the logical data model

Visit www.PEDIAA.com



Generemos los modelos

Físico

Lógico

Conceptual

Physical

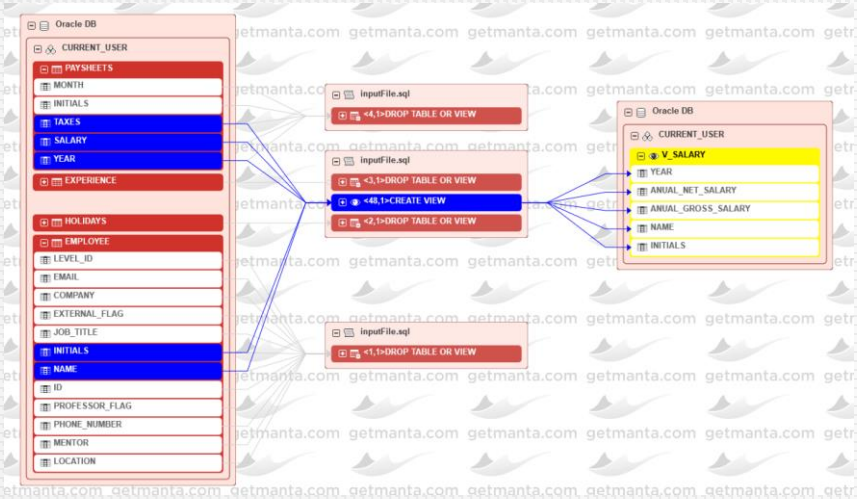
A **physical data model** defines all of the logical **database** components and services that are required to build a **database** or can be the layout of an existing **database**. A **physical data model** consists of the table's structure, column names and values, foreign and primary keys and the relationships among the tables.

Physical Model

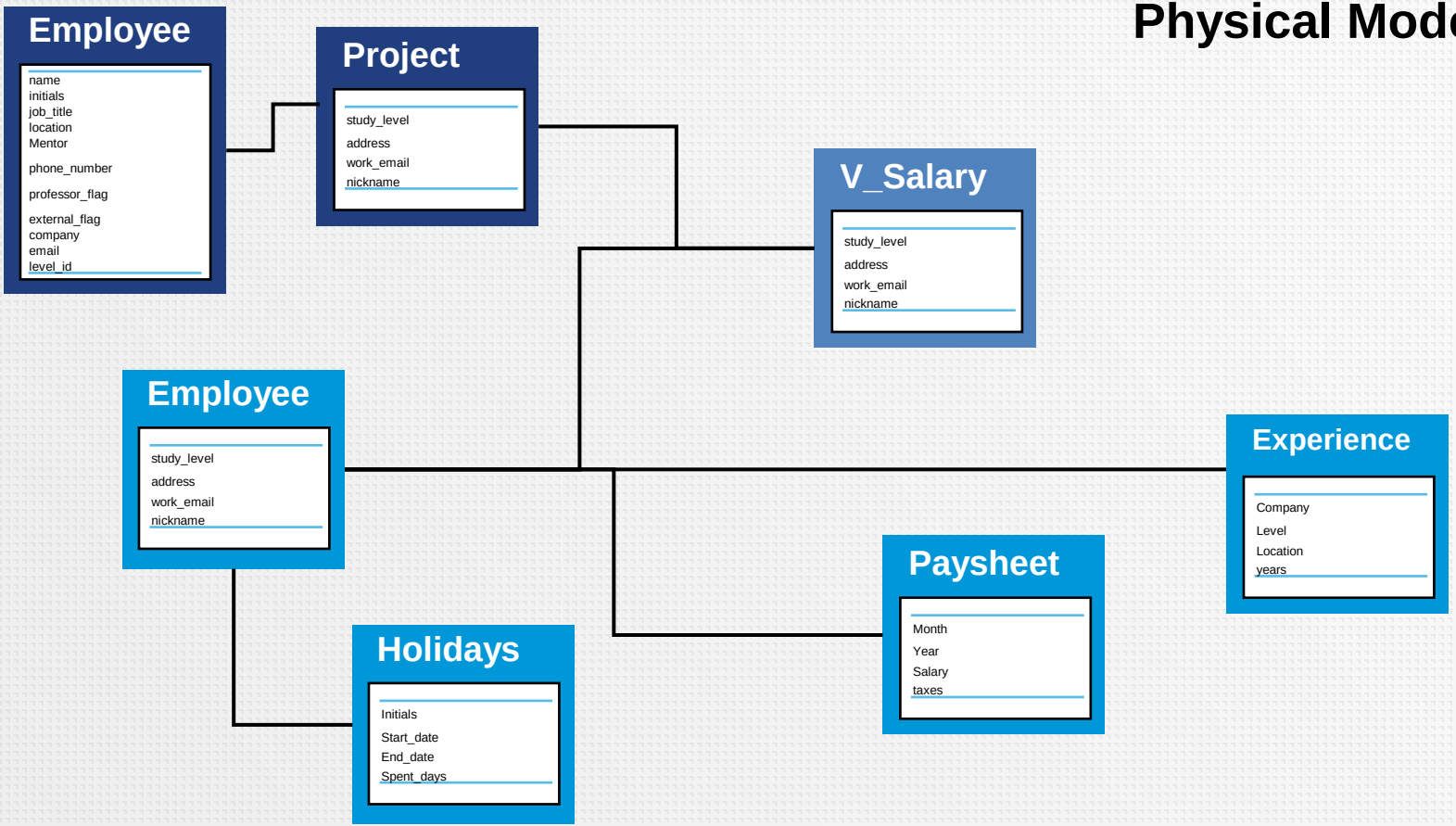
Metadata, or information about data, gives you the ability to understand lineage, quality, and lifecycle, and provides crucial visibility into today's data-rich environments.

First step will be to automatically interrogate our systems to build a human readable interpretation of our systems. Using this approach we will be capable of finding:

- **Systems Involved:** Candidate to Data Providers
- **Backbone tables:** Candidates to find a Entities and Attributes
- **Physical Lineage:** Data Consumers will be highlighted following data flows



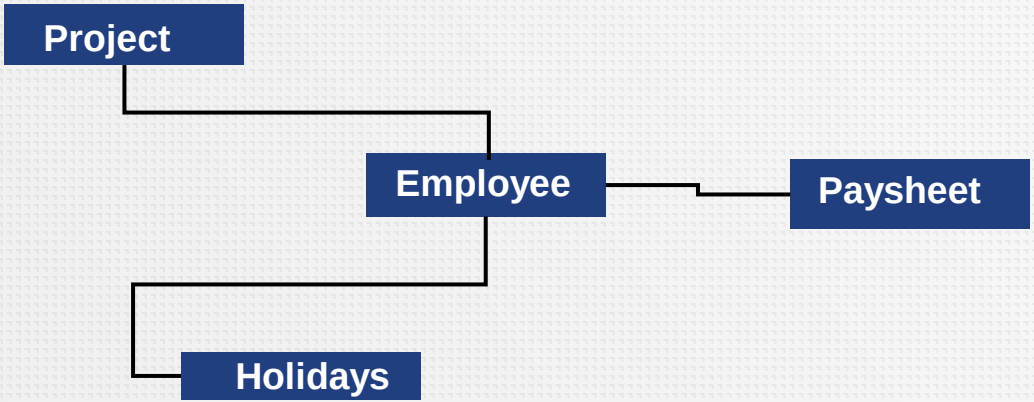
Physical Model



Logical

A **logical data model** describes the **data** in as much detail as possible, without regard to how they will be physical implemented in the **database**. Features of a **logical data model** include: Includes all entities and relationships among them. All attributes for each entity are specified.

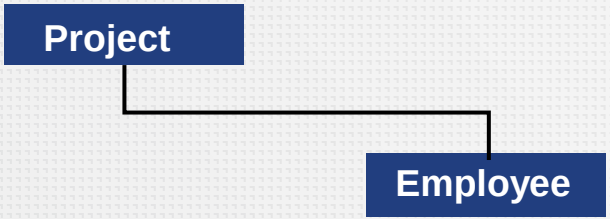
Entity Extraction



Conceptual

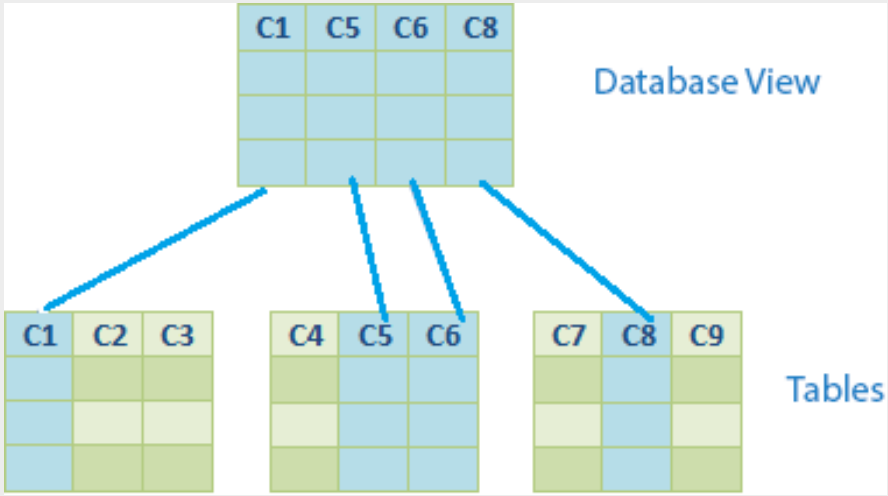
Conceptual Data Model. The **conceptual data model** is a structured business view of the **data** required to support business processes, record business events, and track related performance measures. This **model** focuses on identifying the **data** used in the business but not its processing flow or physical characteristics.

Entity Extraction



Vistas

Una vista es una alternativa para mostrar datos de varias tablas. Una vista es como una tabla virtual que almacena una consulta. Los datos accesibles a través de la vista no están almacenados en la base de datos como un objeto.



•**Ocultar información:** Con una vista se permite el acceso a algunos datos, manteniendo oculto el resto de la información que no se incluye en la vista. El usuario solo puede consultar la vista.
•**Simplificar la administración de los permisos de usuario:** Permite dar al usuario permisos para que solamente pueda acceder a los datos a través de vistas, en lugar de concederle permisos para acceder a ciertos campos, así se protegen las tablas base de cambios en su estructura.

Example

- SQL / View

- 1) Crear una tabla Sales account con nombre SLS_ACCT con esta estructura:
  - ID
  - VND
  - FVenta
  - AMT

- 2) Generar una vista de Sales lógica con los campos correctos

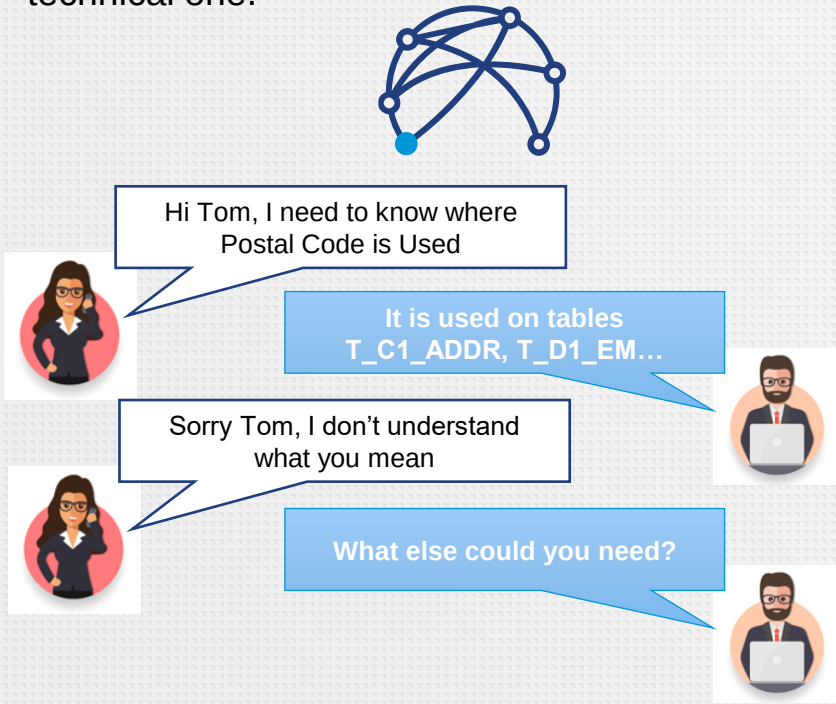
Example

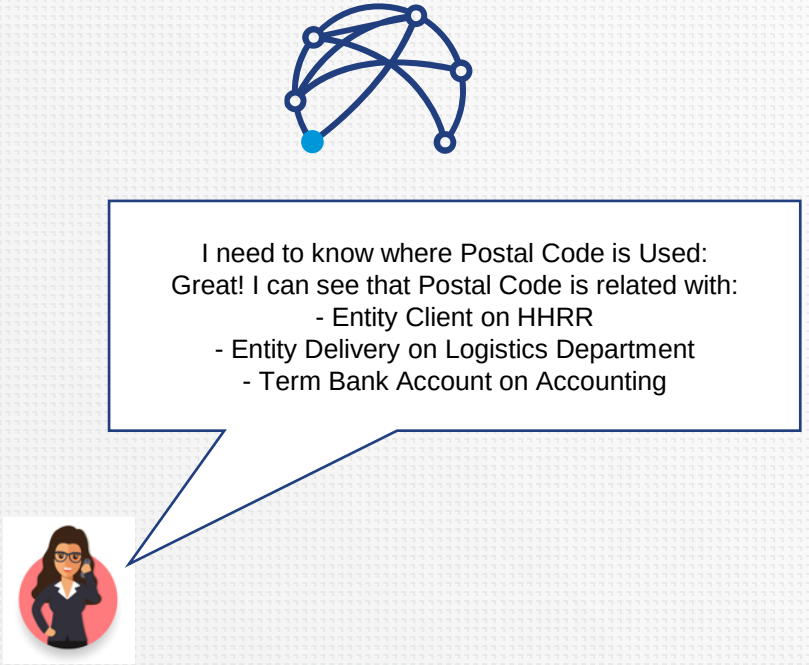
```
CREATE TABLE SLS_ACCT(
   id serial PRIMARY KEY,
   vnd VARCHAR (50) NOT NULL,
   amt double NOT NULL,
   fventa TIMESTAMP NOT NULL,
);
```

```
Create view SALES  AS SELECT ID AS Identifier,VND as Vendor,AMT as
Amount,FVENTA as Fecha_Venta FROM SLS_ACCT;
```

Metadata Lineage – Business Lineage

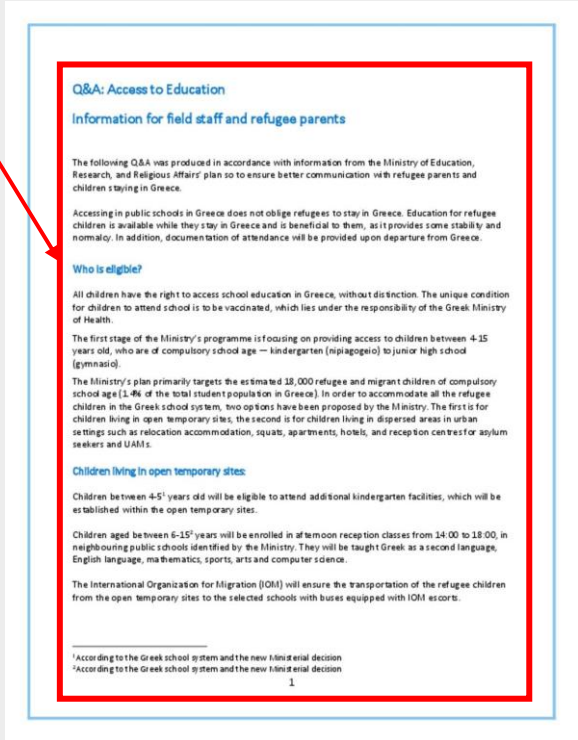
- Sometimes communication barriers becomes an invisible problem in IT Process. In order to minimize this, business should be able to interrogate systems using their language instead of technical one.





What is metadata?

Data



Q&A: Access to Education

Information for field staff and refugee parents

The following Q&A was produced in accordance with information from the Ministry of Education, Research, and Religious Affairs' plan so to ensure better communication with refugee parents and children staying in Greece.

Accessing in public schools in Greece does not oblige refugees to stay in Greece. Education for refugee children is available while they stay in Greece and is beneficial to them, as it provides some stability and normalcy. In addition, documentation of attendance will be provided upon departure from Greece.

Who is eligible?

All children have the right to access school education in Greece, without distinction. The unique condition for children to attend school is to be vaccinated, which lies under the responsibility of the Greek Ministry of Health.

The first stage of the Ministry's programme is focusing on providing access to children between 4-15 years old, who are of compulsory school age – kindergarten (nipiagogeio) to junior high school (gymnasio).

The Ministry's plan primarily targets the estimated 18,000 refugee and migrant children of compulsory school age (1.4% of the total student population in Greece). In order to accommodate all the refugee children in the Greek school system, two options have been proposed by the Ministry. The first is for children living in open temporary sites, the second is for children living in dispersed areas in urban settings such as relocation accommodation, squats, apartments, hotels, and reception centres for asylum seekers and UAMs.

Children living in open temporary sites:

Children between 4-5[1] years old will be eligible to attend additional kindergarten facilities, which will be established within the open temporary sites.

Children aged between 6-15[2] years will be enrolled in afternoon reception classes from 14:00 to 18:00, in neighbouring public schools identified by the Ministry. They will be taught Greek as a second language, English language, mathematics, sports, arts and computer science.

The International Organization for Migration (IOM) will ensure the transportation of the refugee children from the open temporary sites to the selected schools with buses equipped with IOM escorts.

[1] According to the Greek school system and the new Ministerial decision
[2] According to the Greek school system and the new Ministerial decision

1

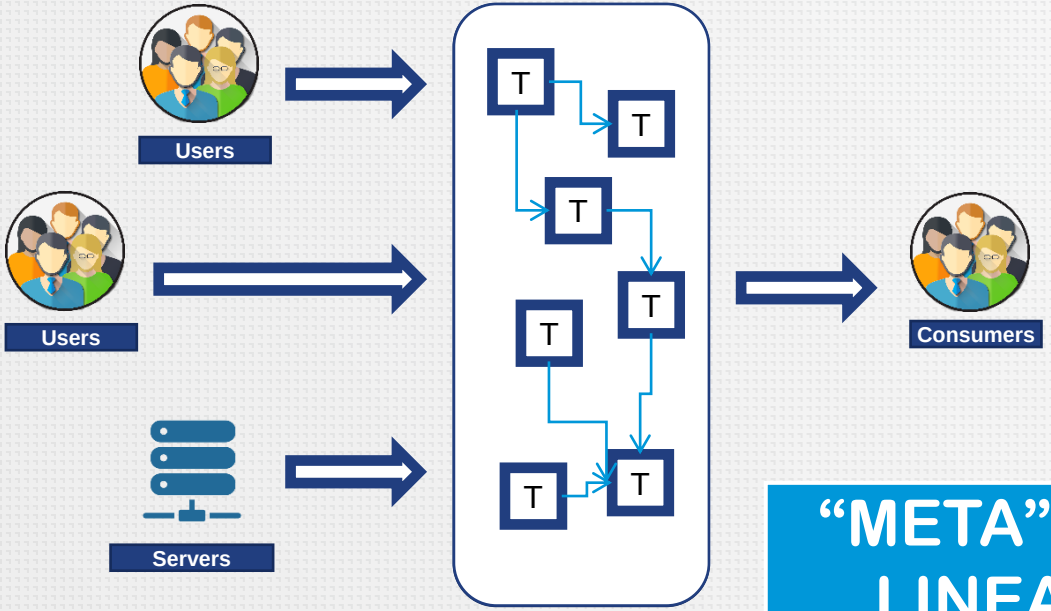
Metadata describes other data. It provides information about a certain item's content.

Metadata:

Fecha Publicación: 01/10/2019
Autor: Pedro Nieto
Tamaño: 3Mb
Formato: docx

Metadata Lineage

Metadata lineage is **defined** as a **data** life cycle that includes the **data's** origins and where it moves over time. It describes what happens to **data** as it goes through diverse processes. It helps provide visibility into the analytics pipeline and simplifies tracing errors back to their sources.



On a normal scenario, there are always Data producers and data consumers but what happens when the Consumer starts wondering…

- How does data get Calculated?
- Which system provided the data I am watching?
- My data is wrong, is it all wrong or at which stage it got corrupted?
- …

“META”DATA LINEAGE

Gartner®

Figure 1. Magic Quadrant for Metadata Management Solutions

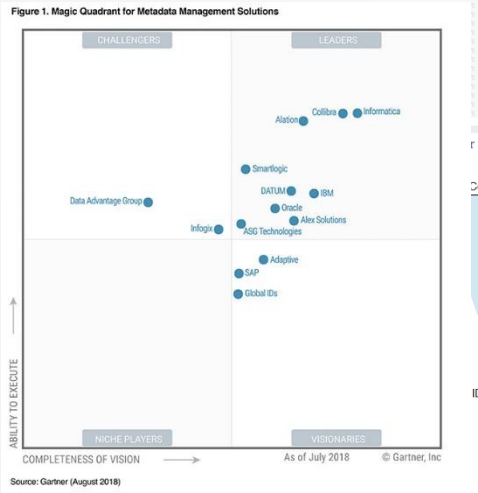
Source: Gartner (October 2019)

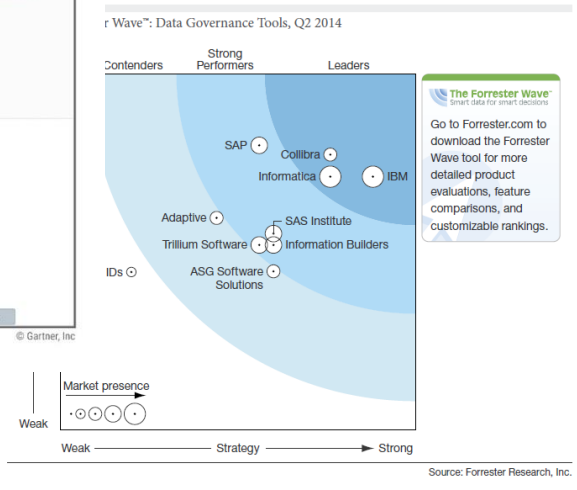
Which tool to use?

There are plenty of options in the market to satisfy your requirements. But there are three big requirements a tool must cover:

Data Catalogue & Lineage
Data Governance
Data Glossary

In GFT we evaluate market leaders and establish our opinion based on POCs and experiences.





Tooling Election

Apache Atlas

Waterline Data

IBM Infosphere

Informatica

TRILLIUM SOFTWARE

pentaho

Waterline Data

Data Catalogue

Used to document and manage metadata, and to perform analysis using the metadata. Organizations can also use repositories to publish information about reusable assets, enabling users to browse metadata during life cycle activities such as design, testing and release management.

Data Governance

Data governance (DG) is the overall management of the availability, usability, integrity and security of data used in an enterprise. A sound data governance program includes a governing body or council, a defined set of procedures and a plan to execute those procedures.

collibra

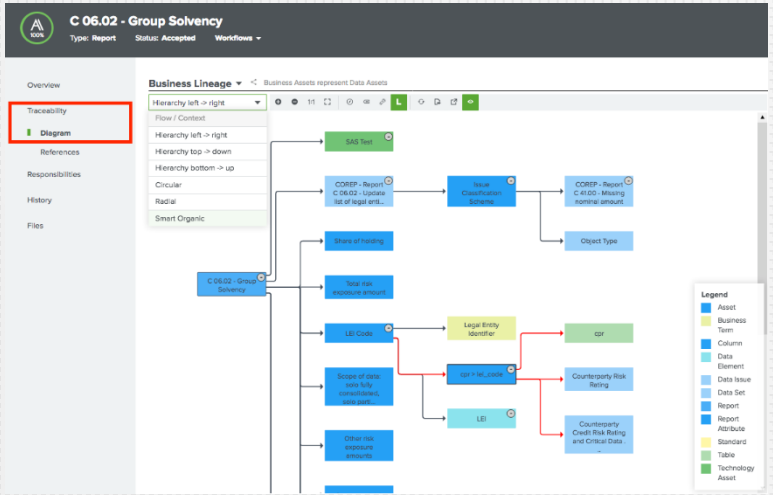
Informatica

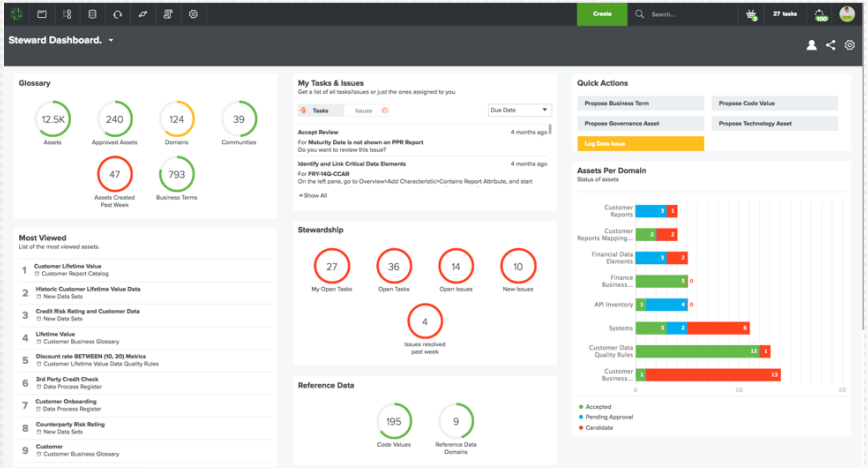
Alation

Data Quality

Data quality refers to the condition of a set of values of qualitative or quantitative variables. There are many definitions of data quality but data is generally considered high quality if it is "fit for [its] intended uses in operations, decision making and planning".

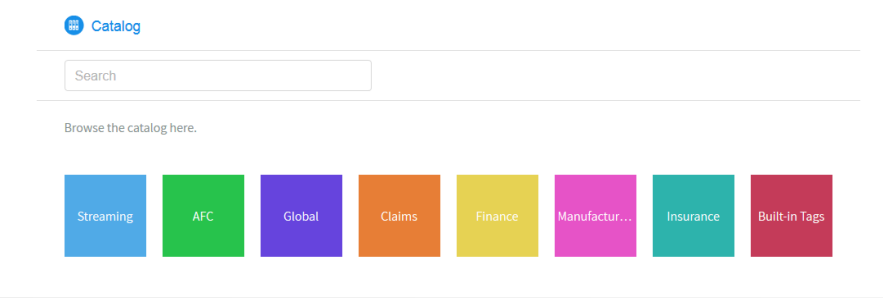
Data Governance

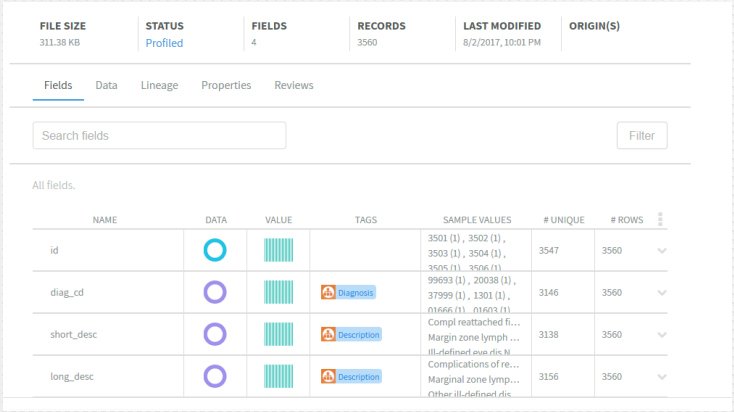


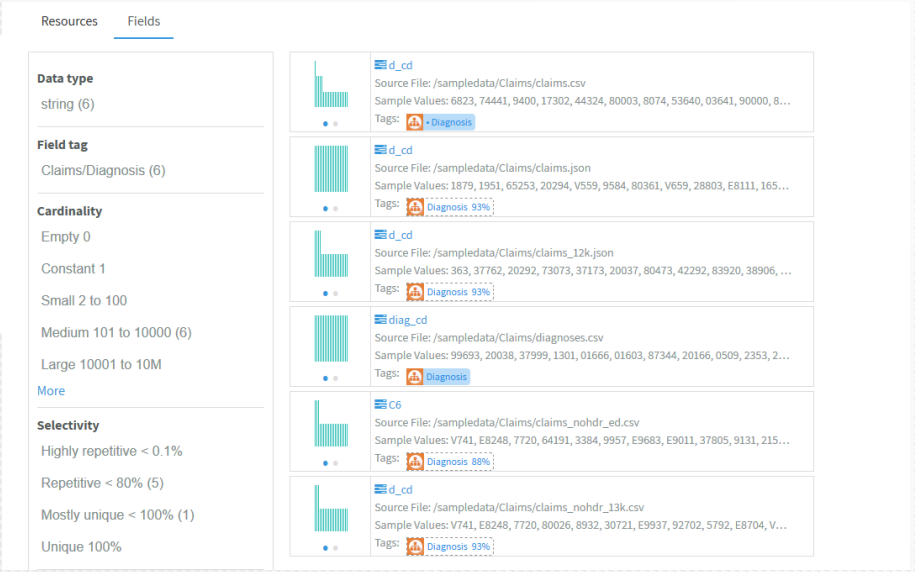


collibra™

Data Catalog

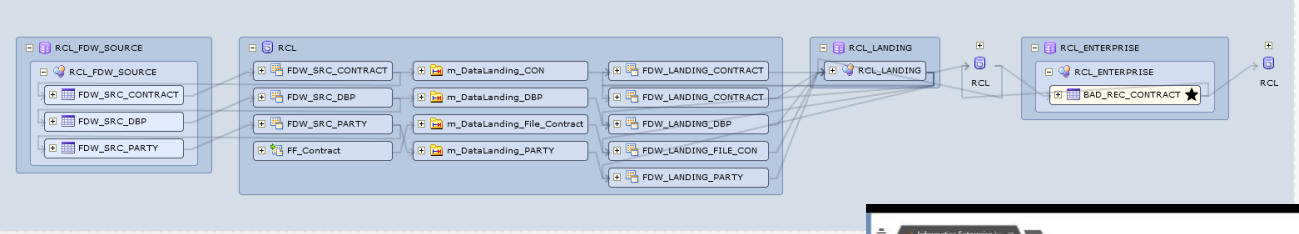






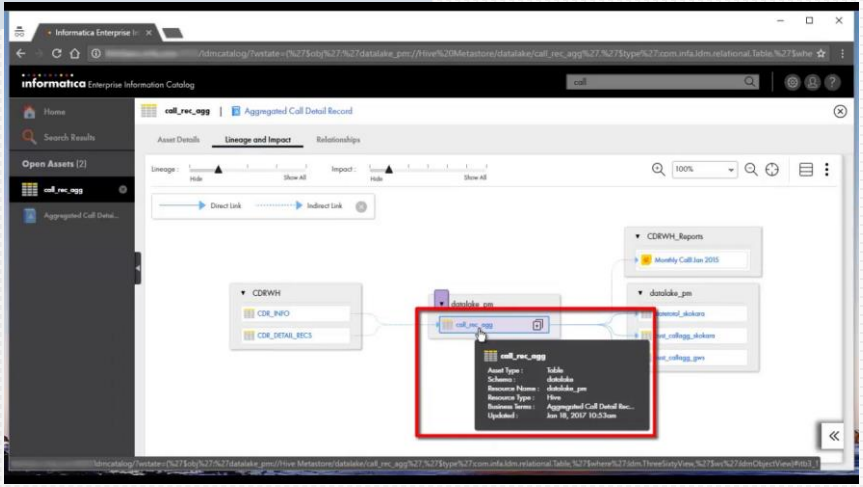
Waterline Data

Data Lineage



MANTA

Informatica




Business Glossary

Ejemplo

Haz un dibujo que represente lo siguiente:

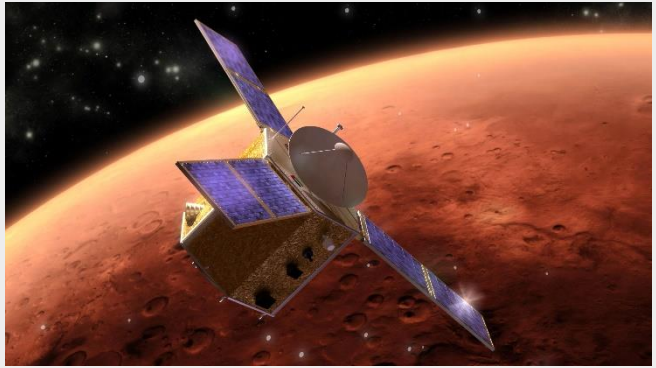
Gemelos

Define con palabras lo siguiente:

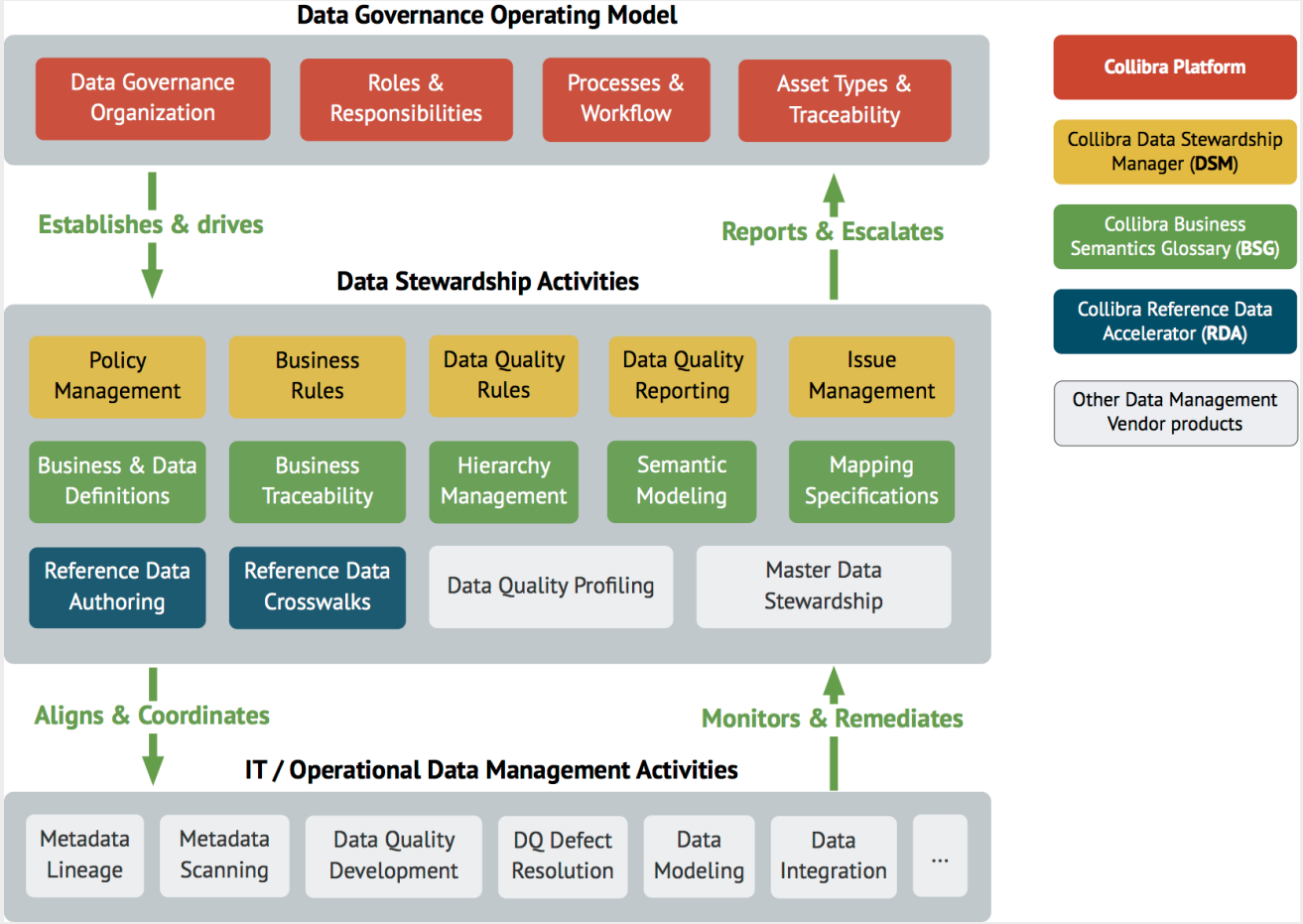
Pronto

Business Glossary

Business Glossary enables data stewards to build and manage a common **business vocabulary** and make it available across an organization. This **vocabulary** furnishes clear meaning and **business** context and can be linked to the underlying technical metadata to provide a direct association between **business** terms and objects.



Comités y Funcionamiento


Data Governance Operating Model
Data Governance Organization
Roles & Responsibilities
Processes & Workflow
Asset Types & Traceability
Establishes & drives
Reports & Escalates
Data Stewardship Activities
Policy Management
Business Rules
Data Quality Rules
Data Quality Reporting
Issue Management
Business & Data Definitions
Business Traceability
Hierarchy Management
Semantic Modeling
Mapping Specifications
Reference Data Authoring
Reference Data Crosswalks
Data Quality Profiling
Master Data Stewardship
Aligns & Coordinates
Monitors & Remediates
IT / Operational Data Management Activities
Metadata Lineage
Metadata Scanning
Data Quality Development
DQ Defect Resolution
Data Modeling
Data Integration
...
Collibra Platform
Collibra Data Stewardship Manager (DSM)
Collibra Business Semantics Glossary (BSG)
Collibra Reference Data Accelerator (RDA)
Other Data Management Vendor products

Real Scenarios

Scenario 1

New Entity

In case a new entity needs to be added, entity analysis process needs to be triggered. From a governance perspective no data should be sent out without the proper consumer approval. This process should be escalated to DGC to lead these actions.

Scenario 2

New Attribute

In case a new attribute is proposed in an existing entity, owner should validate and provide a definition for it. As part of this validation Owner should involve different stakeholders for reviewing. Once attribute is defined, it should be brought to DGC to new data.

Scenario 3

Quality Issue

Once a data steward has identified a data quality issue, this needs to be escalated to Content Owner. Once owner has evaluated the issue, he will need to set up the fix aligned with Platform owner in the correct source of the data and define metrics to avoid this error in the future.

Scenario 4

New Report

Reports must be requested at entity level. Relaying on data lineage business users can interrogate the system from an abstract perspective. Content Owner and Platform should handle the request.

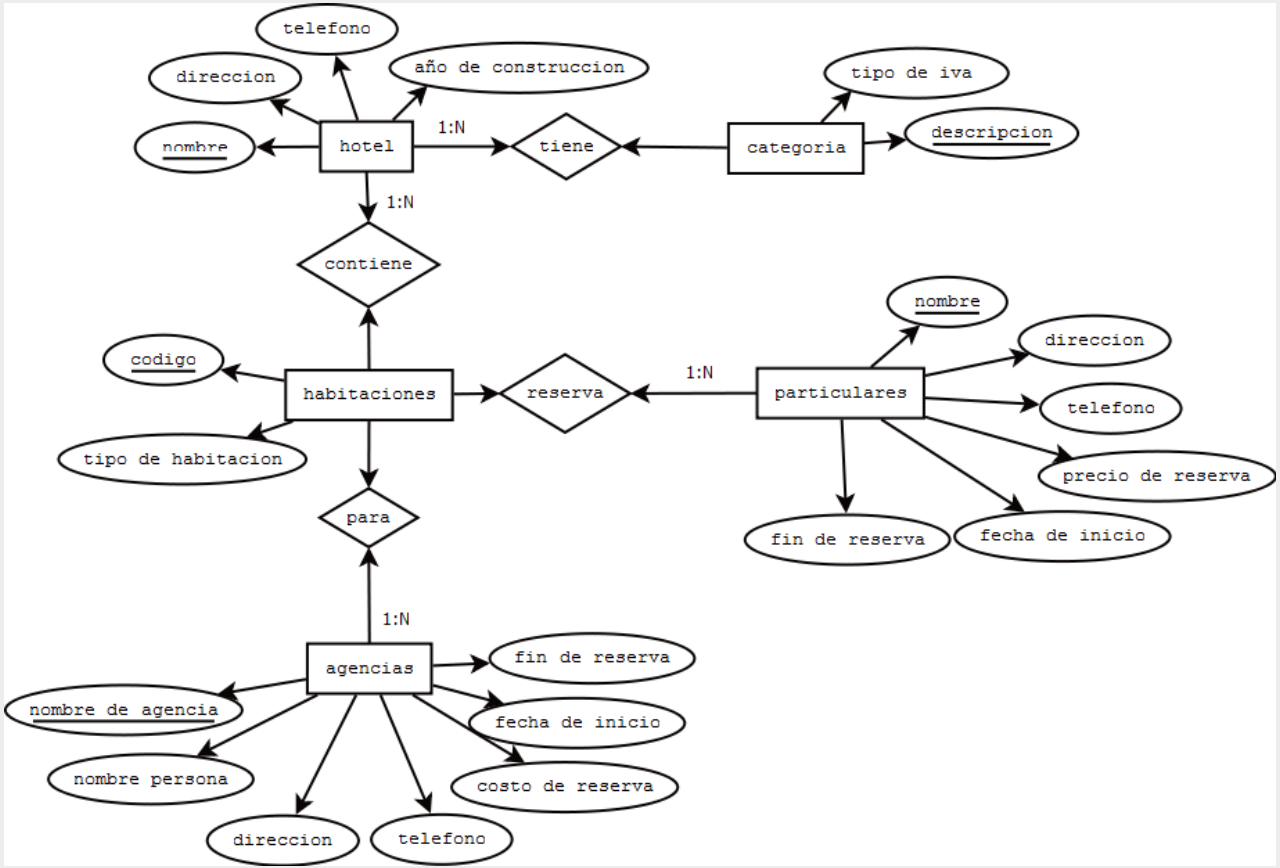

Learn by DOING.


HOLA KIERO Q JUGUEMOS
UN JUEGO
Imagen creada en GeneradorMemes.com

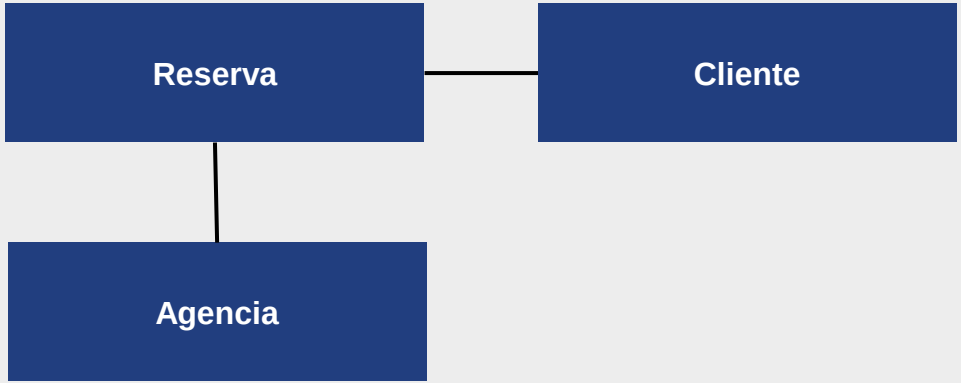
- Elegir 2 CDOs
  - Darles un problema a cada equipo y realizar el modelo lógico
  - Generar los dominios que van a tener (min 3):
    - Cada dominio debe definir sus Roles y comités necesarios
    - Cada dominio tiene que definir sus campos para hacer un glosario
    - Generar un modelo físico en postgres de su dominio y de las relaciones que tiene así como los atributos y sus tipos

Equipo 1

"Cada hotel (del que interesa almacenar su nombre, dirección, teléfono, año de construcción, etc.) se encuentra clasificado obligatoriamente en una categoría (por ejemplo, tres estrellas) pudiendo bajar o aumentar de categoría.
Cada categoría tiene asociada diversas informaciones, como, por ejemplo, el tipo de IVA que le corresponde y la descripción.
Los hoteles tiene diferentes clases de habitaciones (suites, dobles, individuales, etc.), que se numeran de forma que se pueda identificar fácilmente la planta en la que se encuentran. Así pues, de cada habitación se desea guardar el código y el tipo de habitación.
Los particulares pueden realizar reservas de las habitaciones de los hoteles. En la reserva de los particulares figurarán el nombre, la dirección y el teléfono.
Las agencias de viaje también pueden realizar reservas de las habitaciones. En caso de que la reserva la realiza una agencia de viajes, se necesitarán los mismos datos que para los particulares, además del nombre de la persona para quien la agencia de viajes está realizando la reserva.
En los dos casos anteriores también se debe almacenar el precio de la reserva, la fecha de inicio y la fecha de fin de la reserva".

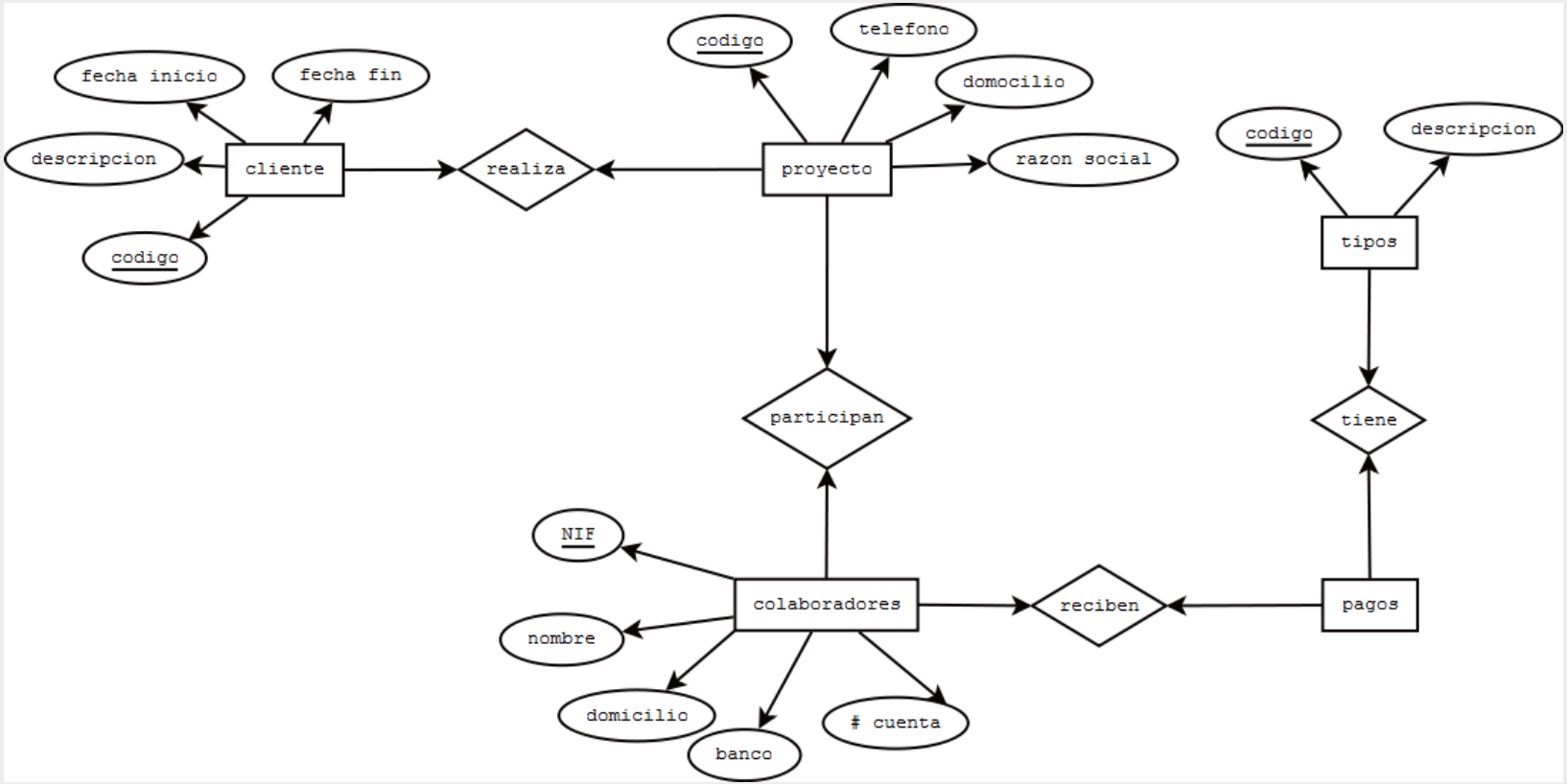

telefono
direccion
año de construccion
tipo de iva
nombre
hotel
1:N
tiene
categoria
descripcion
1:N
contiene
nombre
codigo
habitaciones
reserva
1:N
particulares
direccion
telefono
tipo de habitacion
precio de reserva
para
fin de reserva
fecha de inicio
1:N
agencias
fin de reserva
nombre de agencia
fecha de inicio
nombre persona
costo de reserva
direccion
telefono

Modelo Conceptual

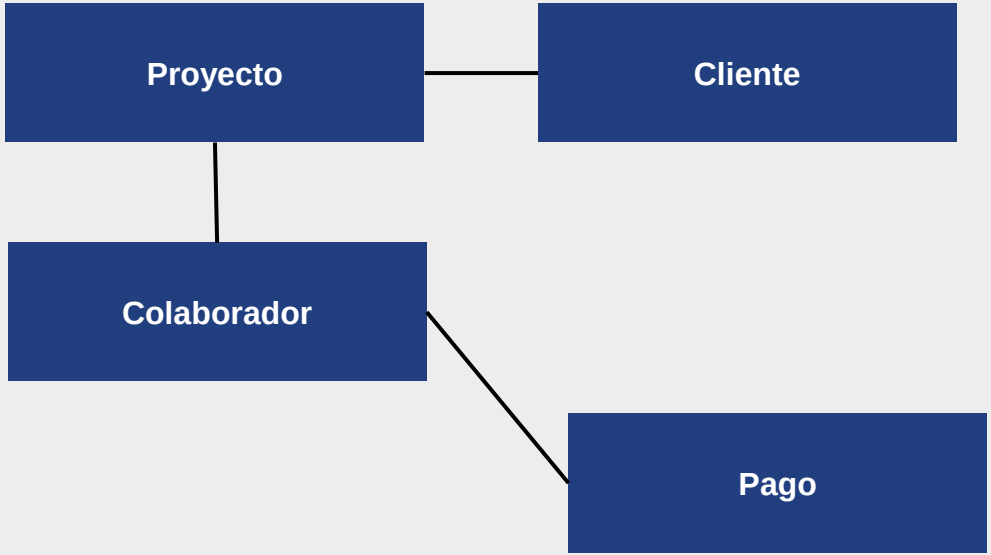


Equipo 2

"De cada uno de los proyectos realizados interesa almacenar el código, descripción, cuantía del proyecto, fecha de inicio y fecha de fin. Los proyectos son realizados por clientes de los que se desea guardar el código, teléfono, domicilio y razón social. Un cliente puede realizar varios proyectos, pero un solo proyecto es realizado por un único cliente.
En los proyectos participan colaboradores de los que se dispone la siguiente información: nif, nombre, domicilio, teléfono, banco y número de cuenta. Un colaborador puede participar en varios proyectos. Los proyectos son realizados por uno o más colaboradores. Los colaboradores de los proyectos reciben pagos. De los pagos realizados se quiere guardar el número de pago, concepto, cantidad y fecha de pago. También interesa almacenar los diferentes tipos de pagos que puede realizar la empresa. De cada uno de los tipos de pagos se desea guardar el código y descripción. Un tipo de pago puede pertenecer a varios pagos".


codigo
telefono
fecha inicio
fecha fin
domocilio
descripcion
cliente
realiza
proyecto
razon social
codigo
codigo
descripcion
tipos
participan
tiene
NIF
colaboradores
reciben
pagos
nombre
domicilio
cuenta
banco

Modelo Conceptual



Retos Equipo 1

- Necesitamos saber el numero de planta de la habitación
- Necesito una consulta con las habitaciones reservadas por cada particular
- Necesitamos un segundo teléfono para los clientes
- Necesitamos almacenar el dueño del hotel y sus datos
- GDPR ha llegado hay que reportar todos los datos sensibles
- ¿Qué quiere decir Fin de la reserva, fecha de pago, fecha de factura o fecha de salida?
- Genera una vista de habitaciones reservadas
- Necesitamos almacenar el coste de cada tipo de habitación
- Necesitamos almacenar el nombre de la cadena hotelera y la sede de la misma

Retos Equipo 2

- Necesitamos saber país del domicilio del cliente
- Necesitamos una consulta que muestre para colaborador los proyectos implicados
- Necesitamos un teléfono para los colaboradores
- Necesitamos almacenar el banco donde se hace el pago y el nombre del director
- GDPR ha llegado hay que reportar todos los datos sensibles
- ¿Qué quiere decir código de cliente, razón social o cuenta ?
- Genera una vista de proyecto que tengan colaboradores
- Necesitamos almacenar el iban del numero de cuenta y la fecha de creación de la cuenta
- Necesitamos almacenar la cantidad máxima autorizada por tipo de pago