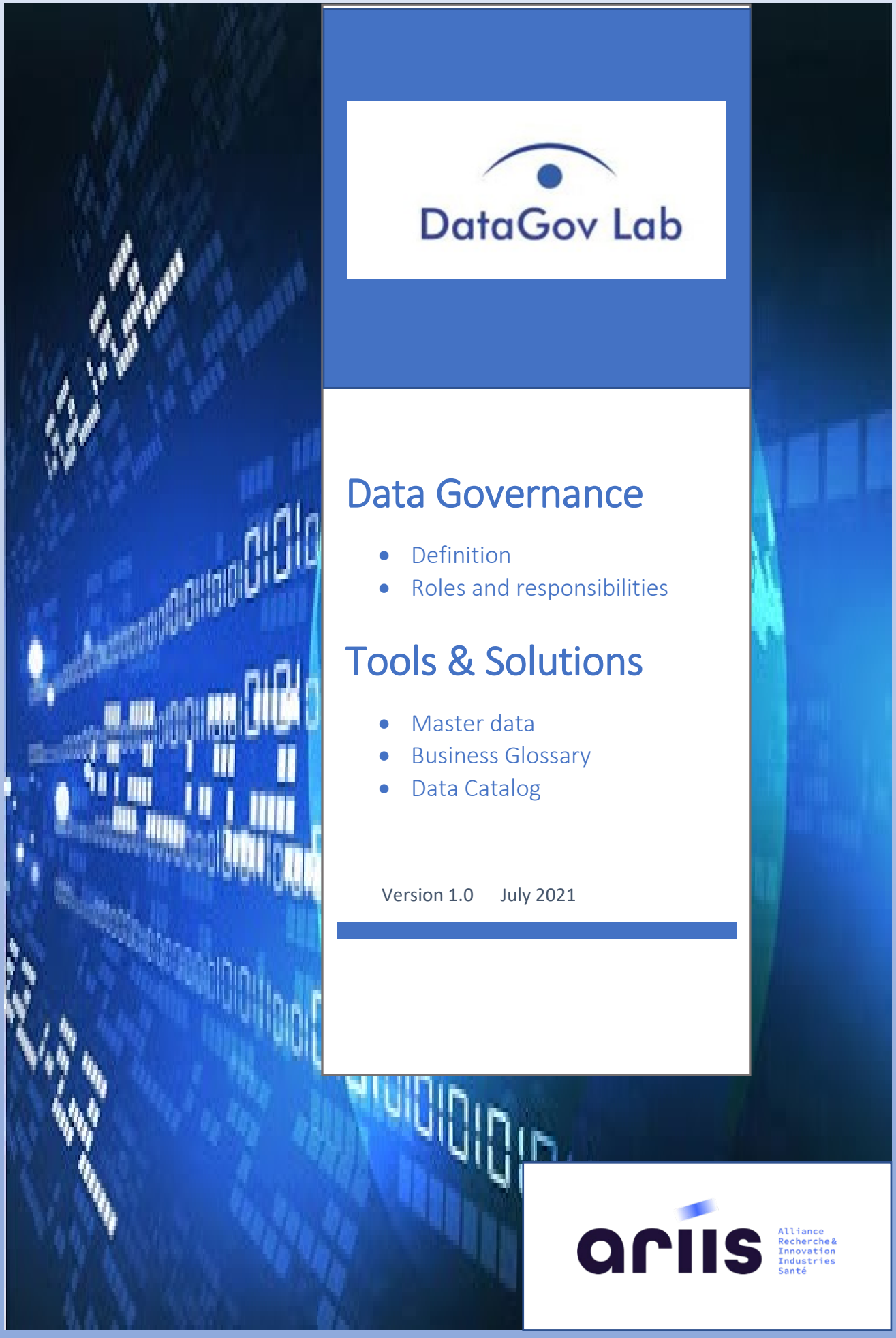


DataGov Lab

Data Governance

- Definition
- Roles and responsibilities

Tools & Solutions

- Master data
- Business Glossary
- Data Catalog

Version 1.0 July 2021

ariis Alliance Recherche& Innovation Industries Santé

DataGov Lab

Contents

1 Introduction 4

1.1 What is the DataGov Lab? 4

1.2 DataGov Lab purpose 4

1.3 Data Gov Lab Road map 5

1.4 DataGov Lab Organization 5

1.4.1 The Board 5

1.4.2 DataGov Lab meeting 6

1.5 DataGov Lab membership 6

1.6 What is ARIIS? 7

2 Data Governance 8

2.1 What is data Governance? 8

2.2 Definition and Mission of “Data Governance” 9

2.3 Data Governance Roles and responsibilities 10

2.3.1 Chief Data Officer (CDO) 10

2.3.2 Data Governance lead (DGL) 10

2.3.3 Data Domain Owner (DDO) 11

2.3.4 Data Owner (DO) 11

2.3.5 Business Data Steward (BDS) 12

2.3.6 Data Producer (DP) 13

2.3.7 Technical Data Steward (TDS) or Data Custodian (DC) 13

3 Master Data 14

3.1 Introduction 14

3.2 What is a Master Data? 15

3.3 Master data -DataGov Lab definition 16

3.4 Benefits of Master Data 16

3.5 Roles and responsibilities 17

3.5.1 Chief Data Officer (CDO) 17

3.5.2 Data Domain Owner (DDO) 17

3.5.3 Data Owner (DO) 17

3.5.4 Data Producer (DP) 17

3.5.5 Business Data Steward (BDS) 17

3.5.6 IT Enterprise Architect 18

3.5.7 Data Governance Lead ... 18
3.6 How does data become master data? ... 19
3.7 Levels to categorize master data ... 20
3.8 Master Data Terminologies and definition ... 21
3.9 Example of Master Data Categories ... 22
3.10 Common examples of Master Data in Pharma (not exhaustive list) ... 23
4 Data Governance tools ... 24
4.1 Introduction ... 24
4.2 Data Governance tools – Takeaway/Summary ... 24
4.3 Business Glossary and Data Catalog: Why these Systems are required? ... 25
5 Business Glossary ... 26
5.1 Introduction ... 26
5.2 What is a Business Glossary? ... 27
5.3 Business Glossary -DataGov Lab definition ... 28
5.4 Benefits of a Business Glossary ... 28
5.5 Scope of the Business Glossary ... 28
5.6 Business Glossary Ownership ... 29
5.7 How define a Business Term? ... 29
5.8 Rules to create a Definition ... 30
5.9 Business Glossary contents ... 30
5.9.1 Mandatory attributes ... 30
5.9.2 "Nice to have" attributes ... 32
5.9.3 Out of scope attributes ... 32
5.9.4 Link between the Business Glossary and Data Catalog ... 32
5.10 Roles and responsibilities ... 33
5.10.1 Requester ... 33
5.10.2 Business Data Steward (BDS) ... 33
5.10.3 Data Owner (DO) ... 33
5.10.4 Data Governance Lead ... 34
5.10.5 Data Domain Owner (DDO) ... 34
5.10.6 Business Glossary Administrator ... 35
5.11 Which tool? ... 35
5.11.1 Business Glossary and MDM tool ... 35
5.12 Business Glossary Template ... 36
5.13 Business Glossary Example ... 37

5.14 The Future/dream tool – what would this look like? 38
5.15 Business Glossary reference 38
6 Data Catalog 39
6.1 Data Catalog introduction 39
6.2 What is a Data Catalog? 41
6.3 Data Catalog -Datagov Lab definition 42
6.4 Benefits of a Data Catalog 42
6.5 Scope of the Data Catalog 42
6.6 Data Catalog Tool Ownership 42
6.7 Users of the Data Catalog 43
6.8 Data Catalog contents 44
6.8.1 Mandatory attributes 44
6.8.2 Nice to have attributes 45
6.9 Data Catalog Functionalities 45
6.9.1 Nice to have functionalities 45
6.10 Roles and responsibilities 45
6.10.1 Business Data Steward (BDS) 45
6.10.2 Technical Data Steward (TDS) 45
6.10.3 Data Owner (DO) 45
6.10.4 Data Governance Lead (DGL) 46
6.10.5 IT Architecture 46
6.11 What should be our Future/dream tool 46
6.12 References 46

1 Introduction

1.1 What is the DataGov Lab?

The DataGov Lab started in 2017 from informal discussions between data governance representatives from different companies. In April 2019, the Datagov Lab was presented to the ARIIS board and became an official "data "group focusing on Data Governance in Life sciences.

Although part of ARIIS, The DataGov Lab is autonomous in terms of memberships and objectives and road map.

The DataGov Lab is a European group of Data Governance experts in life science from Ariis.

ariis.datagovlab@gmail.com

Profil : https://www.linkedin.com/in/ariis-datagovlab/

Group DataGov Lab: https://www.linkedin.com/groups/13671541/

1.2 DataGov Lab purpose

The purpose of the data Gov Lab is to create a community to exchange about data governance, share our data governance processes and methodologies in order to define best practices of Data Governance within Life Sciences.

The DataGov Lab members express their personal views and assumptions without liability for their company.

This document is generic to comply with the confidentiality levels of each company participating in the DataGov Lab, thereby avoiding organization related differences between companies. Obviously, all the definitions, recommendations provide in this book can be adapted to the organization.

1.3 Data Gov Lab Road map

This document is the first output of our road map defined at the start of the DataGov Lab in 2018 and described below. Since the start, we can all confirm that data governance is becoming more and more mature in our company compared to the past few years and we are confident we will have more topics to add to this list.

- What is master data in life sciences?
- How to improve internal data knowledge?

  Tools: Business glossary, data catalog, data flows.

  What can Artificial intelligence offer?
- Roles and responsibilities within data governance?

  Data owner, data stewards, …

  What are the needs for today and how our job will change?
- How do we demonstrate the value of our data?
- Establish data KPIs (Quality, availability, traceability, usage, ...)

1.4 DataGov Lab Organization

1.4.1 The Board

The DataGov Lab Board is composed of the leader and officers. Only one officer per company is accepted in the board.

As per today we have a representant of 6 pharmaceutical industries located in Europe.

The leader(s) of the organization is:

Leader (president): Elisabeth Campain-Teulon

The Leader shall facilitate all meetings and shall have the authority to act on behalf of the organization when matters require immediate action.

The Officers are:

- Pierre Chauvet
- Christophe Fournier
- Quentin Grignet & Maxine Fletcher
- Samuel Mercier
- Charlotte Ratcliffe

The mission of the board is to define the topics as well as the priorities of our meetings.

The officers are in charge to share minutes of the DataGov Lab board meeting within his / her company and to collect feedback from them.

In the case that any member feels an officer is failing to fulfill his/her duties adequately (e.g. lack of attendance in meetings), the board will decide on his/her exclusion.

If the position of leader becomes vacant during the semester it will be the responsibility of the board to assign a new leader.

Decisions at all meetings regarding any action or future actions of the group, or regarding the philosophy of the group, shall be reached by consensus of all present members.

1.4.2 DataGov Lab meeting

Board Team meetings (Board members) will be held at least every two months with the objective of preparing the plenary meetings. During the Pandemic period, the Board Team meeting is held every month (2 hours) as no plenary meeting with the extended team can be organized.

DataGov Lab meetings (in Face to face) will be held at least twice per year with all the members of the organization if the sanitary conditions allow it.

Service providers should not be part of these meetings.

1.5 DataGov Lab membership

This organization will allow membership to all Data Governance experts in Life science except service providers. Future Members are encouraged to attend and participate in all sponsored events.

Any new membership request will be reviewed by the board for acceptation or rejection.

1.6 What is ARIIS?

The aim of ARIIS (Health Industry Alliance for Research and Innovation), founded in 2010 is to federate the healthcare sector around collective events or projects.
ARIIS interacts with all stakeholders of the innovation ecosystem in order to:

- Promote French research abroad,
- Adopt a forward-looking approach to disruptive innovation that may change tomorrow's healthcare,
- Support the digital transformation of the healthcare sector, as regards health data and artificial intelligence.
  ARIIS is defining itself as a center for French research and innovation.

2 Data Governance

2.1 What is data Governance?

Gartner

> ***Data governance is the orchestration of people, processes, and technology to manage the company's critical data assets by using roles, responsibilities, policies, and procedures to ensure the data is accurate, consistent, secure, and aligns with overall company objectives.***

Source: Hallmark; Gartner analysis.

Data Governance Institute Definition:

Short version:

"Data Governance is the exercise of decision-making and authority for data-related matters."

Long version:

"Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods."

Data Governance Professionals Organization (DGPO)

What is Data Governance?

Data Governance is a discipline that provides clear-cut policies; procedures; standards; roles; responsibilities; and accountabilities to ensure that data is well-managed as an enterprise resource.

— from the DGPO Data Governance Glossary

DAMA-DMBOK 2nd Edition

Data Governance and Stewardship

Definition: The exercise of authority, control, and shared decision-making (planning, monitoring, and enforcement) over the management of data assets.

Goals:

1. Enable an organization to manage its data as an asset.
2. Define, approve, communicate, and implement principles, policies, procedures, metrics, tools, and responsibilities for data management.
3. Monitor and guide policy compliance, data usage, and management activities.

Wikipedia Definition:

"Data governance is a control that ensures that the data entry by an operations team member or by an automated process meets precise standards, such as a business rule, a data definition and data integrity constraints in the data model. The data governor uses data quality monitoring against production data to communicate errors in data back to operational team members, or to the technical support team, for corrective action. Data governance is used by organizations to exercise control over processes and methods used by their data stewards and data custodians in order to improve data quality."

Find out more: http://en.wikipedia.org/wiki/Data_governance

2.2 Definition and Mission of "Data Governance"

Data Governance defines and maintains the data asset management framework of the enterprise based on:

- **People** > roles and responsibilities, rules, change management
- **Process** > data life cycle, common language
- **Technologies** > Master Data Management (MDM), Business Glossary, Data Catalog, Indicators

In order to Improve enterprise efficiency

- Better decision making using reliable data (consistent, up to date, secure and accurate
- Contribute to regulatory Compliance (e.g. GDPR, CNIL, IDMP, serialization, ...)
- Generate business value from data (Definition, standard terminology, availability of master data, optimize data flows & systems)

2.3 Data Governance Roles and responsibilities

The main common roles identified in our companies are described below.

Depending of the size of the company, one person should have one role or several roles.

Example: A Data Steward and a Data Producer can be the same person.

2.3.1 Chief Data Officer (CDO)

The main roles of a Chief Data Officer are:

- Ensuring there is funding for a governance organization and that resources are available for data related roles and responsibilities
- Establish the Enterprise Data Strategy in accordance with the Business strategy as regards added value and regulatory constraints related to risk and compliance.
- Onboarding and aligning different functions such as IT, digital, security and risks
- Promote Data Awareness within the enterprise
- Define roles and responsibilities toward the data
- Establish Data Governance standards, policies and procedures
- Oversee the Data Governance activities within the Business and other IT functions (architecture, analytics, MDM, interfaces, etc.).

2.3.2 Data Governance lead (DGL)

The Data Governance leads are generally part of the Business functions supporting one or several data domains

The main roles of a Data Governance Lead is/ are:

- Ensure compliance with data governance policies, standards, roles and responsibilities, and adoption requirements.
- Ensure the integrity, usability, privacy and security of its data assets in conjunction with the relevant stakeholders.
- Collaborate with business to develop & maintain an inventory of the enterprise information maps (Data Domains, End to end Processes and actors, data flows, Data Catalog, Business Glossary)
- Facilitate data governance board meeting with Data owners to establish and share data governance rules and provide support to the data actors.
- Collaborate with stakeholders to define indicators of performance and quality metrics

- Identify new business opportunities with Domain Data Owner and Data Owners pertaining to the use of information assets to achieve efficiency and effectiveness in its domain/enterprise.

2.3.3 Data Domain Owner (DDO)

A Data Domain reflects an interconnected portfolio of data assets that are the key business drivers across functions. e.g. Study, product, supplier, Customer.

The main roles of a Data Domain Owner is/ are:

- Oversee a data domain across Business functions with the goal of maximizing the value derived from the data within his domain.
- Legitimize the roles of data owner within a domain
- Define which information (data assets) are important for the business strategy in terms of either added value (internal to the company) or risk/compliance (external regulations).
- Collaborate with the other DDO, sharing road map, and decisions taken for their domain.

Who are they?

Generally, Data Domain Owners are senior individuals who have enough authority in the hierarchy such as the level of "Director" or "Head of" and have resources at their disposal.
DDO should have a global view on enterprise and a dep domain knowledge within their assigned Data Domain

2.3.4 Data Owner (DO)

Definition: Data Owner has approval authority for decisions about data within their domain. (DAMA book)

The Data Owner should have the same responsibility as a DDO but on a specific scope. The data owner is responsible for ensuring accessibility and data quality (defining validity rules) for data assets.

A Data Owner works in close collaboration with several Data Stewards.

The main roles of a Data Owner is/are:

- Defines, documents and masters data life cycle
- Determines the data quality expectations

- Prioritizes data issues resolution within their domain
- Provides Business definitions which are then approved by the DDO.
- Liaises with consumers of data in their domain to provide additional context on the usage of the data ensuring that metadata & documentation for their domain accurately reflects the real world and is helpful to self-service consumers of the data.
- Defines and builds the business case that will leverage data from their domain, often in collaboration with the Data Owners from other domains

Several Data owners can share responsibilities within a Domain.
The split can be

- Per geographic location (One for Europe and another one for US)
- Per business Unit (CHC, SC)
- Per type of data (e.g. clinical Studies, non-clinical studies, medical Affairs studies,)

However, rules related to life cycle management (e.g. what are the triggers for the creation of a new substance, pharmaceutical product, clinical study etc, what are the mandatory fields before it can be validated in a system, when is it ok to deactivate it, how long do we need to keep it etc..) must be agreed across the different geographic regions and or business units. Otherwise duplicate values will be created

2.3.5 Business Data Steward (BDS)

Definition: Business Data Stewards are business professionals, most often recognized subject matter experts, accountable for a subset of data. They work with stakeholders to define and control data. (DAMA book)

A Data Steward can be attached to a centralized structure working only on reference and transverse data.

The main roles of a Business Data Steward is/are:

- Set up and execute Data Checks to ensure data quality
- Monitor the Data Quality.
- Monitor data for (non) compliance to business information rules as defined by the DO and the DDO.
- Checks that any requests to update data standards (e.g. check that adding a new value within a drop-down list or reference list) meet the business rules and have been approved by the data owner.

Note: The Business Data Steward is not responsible for correcting errors. The errors must be corrected by the role that made them (part of the learning curve and continuous improvement process).

2.3.6 Data Producer (DP)

Definition: Data Producer captures the data in usable form (DAMA book).

The main roles of a Data Producer is/are:

- Responsible for entering data within a form (paper or electronic) or within an IT system.
- Responsible for the data integrity within the process
- Have to respect and adhere to the business information rules.

  Furthermore, this role is empowered to check and escalate any request that violates these business rules to the Data Owner and Data Steward.

2.3.7 Technical Data Steward (TDS) or Data Custodian (DC)

Definition: Technical Data Stewards are IT professionals operating with one of their knowledge areas such as Data Integration specialist, database administrators, Business Intelligence Specialists, Data Quality analysts or metadata administrators. (DAMA book)

The technical Data Steward works closely with the Business Data Steward to implement rules defined by the DDO and DO within the IT System. (e.g. personal data and encryption)

3 Master Data

3.1 Introduction

Although Data Governance covers all type of data, the DataGov Lab group believes that Master Data Management is key to solving many data-related pain points, improving business efficiency and ensuring regulatory compliance, all of which are significant within our companies.

Therefore, the DataGov Lab group decided to dedicate several workshops to this topic, and we are pleased to present the result of our effort is this document.

Thanks to all the participants for their participation to the Face to Face workshops which took place in 2019 before the COVID period

3.2 What is a Master Data?

DAMA Dictionary

Reference Data

Any data used to categorize other data, or for relating data to information beyond the boundaries of the enterprise. See master data.

Master Data

Synonymous with reference data. The data that provides the context for transaction data. It includes the details (definitions and identifiers) of internal and external objects involved in business transactions. Includes data about customers, products, employees, vendors, and controlled domains (code values).

METAWRIGHT

Gartner Definition

Master data is the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies and chart of accounts.

1.3.3 Master Data DAMA-DMBOK2

Master Data is data about the business entities (e.g., employees, customers, products, financial structures, assets, and locations) that provide context for business transactions and analysis. An entity is a real world object (person, organization, place, or thing).

Philip Russom - TDWI

Master data is having "consistent definitions of business entities (e.g., customer or product) and data about them across multiple IT systems and possibly beyond the enterprise to partnering businesses.

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia

Article Talk

Master data

From Wikipedia, the free encyclopedia

Master data represents "data about the business entities that provide context for business transactions".[1] The most commonly found categories of master data are Parties (individuals and organisations, and their roles, such as customers, suppliers, employees), Products, Financial Structures (such as ledgers and cost centres) and Locational Concepts.[1][2]

Master data should be distinguished from Reference Data. While both provide context for business transactions, reference data is concerned with classification and categorisation, while master data is concerned with business entities.

3.3 Master data -DataGov Lab definition

Many definitions of Master Data exist in literature issued from different industry sectors.
Our proposed definition is generic, but works for all, working in pharmaceutical companies.

Master Data is a valuable data, stable and shared across business processes.

Generally, we start to think about master data when data is shared by at least 3 business processes.

Remark: Some data (e.g. Sell- in / sell- out) are not master data according our definition but are critical for companies and should be managed as "master data".

Two other types of data can be identified within a Business process

Transactional Data:

Transactional data is information directly derived as a result of transactions.

e.g. Financial: orders, invoices, payment; Work: plans, activity records

Reference Data:

Reference data is a Data that defines the set of permissible values to be used by other Master data attributes (e.g. Country Iso codes).

3.4 Benefits of Master Data

The main benefits of Master Data are:

Optimization of business efficiency, through:

- Ensuring data availability for key-data-dependent processes
- Interoperability between systems/business functions (sharing of consistent, uniquely defined data)
- Cost reduction (avoidance of multiple data entries within systems, reduction of reconciliation tasks)

Global and significative improvement of data quality

- Minimization of risk, through compliance with applicable regulations.
- Reliable decision-making based on data

3.5 Roles and responsibilities

3.5.1 Chief Data Officer (CDO)

Is(are) responsible for:

- Ensuring a Domain Data Owner (Master Data Owner) has been identified
- Developing the road map of the development and deployment of Master Data in collaboration with Data Domain Owners.
- Promoting the usage of master data in Business processes and IT systems.

3.5.2 Data Domain Owner (DDO)

Is(are) responsible for:

- Promoting the usage of master data in Business information and processes.
- Identifying which master data elements are key to the success of this business strategy.
- Ensuring the master data is used as much as possible in Business processes and IT systems.

3.5.3 Data Owner (DO)

Is(are) responsible for:

- Defining the Master Data access rights
- Defining the expected Master Data quality levels in collaboration with relevant stakeholders.
- Ensuring a unique definition of the Master Data is documented in the Business Glossary.
- Ensuring the usage of the right data source for the creation of master Data, (except if the MDM is the system source)
- Contributing to the development of the conceptual data model in the MDM

3.5.4 Data Producer (DP)

Is (are) responsible for:

- Providing accurate and up to date data when they are the source.
- Updating data in the source system when data quality issues are detected on Master Data.

3.5.5 Business Data Steward (BDS)

Is (are) responsible for:

- Monitoring the quality of data provided by the data producers
- Liaising with the data producers and data owners to ensure corrections are made

- Ensuring updates are done only in the source system and replicated toward the reference system (e.g. SAP, MDM).

3.5.6 IT Enterprise Architect

Is (are) responsible for:

- Ensuring usage of Master Data to support operational data as much as possible before the starting of the project.
- Ensuring Master Data is integrated during the implementation phase of the system/ solution.
- Ensuring the conceptual MDM data model is compliant with data architecture rules and standards.
- Promoting and ensuring the development of interfaces between MDM and other IT systems.

3.5.7 Data Governance Lead

Is (are) responsible for:

- Ensuring the development of Data Checks within MDM and interfaces to ensure data quality
- Developing performance indicators related to Quality of Master Data and Quality of Master Data Management

Example of Master Data Quality indicators

- *Number of Duplicated master Data records*
- *Number of missing data records in relevant source systems*
- *Number of Inconsistent master data records in relevant systems*

Example of Master Data Management indicators

- *Number of Days to correct a master data record in the source system from the time of detection*
- *Number of Master Data used and not used*
- *Number of obsolete master data record (inactive records)*
- *Number of active users with the right to update master data in relevant systems.*
- *Number of systems sources not consuming master data.*

3.6 How does data become master data?

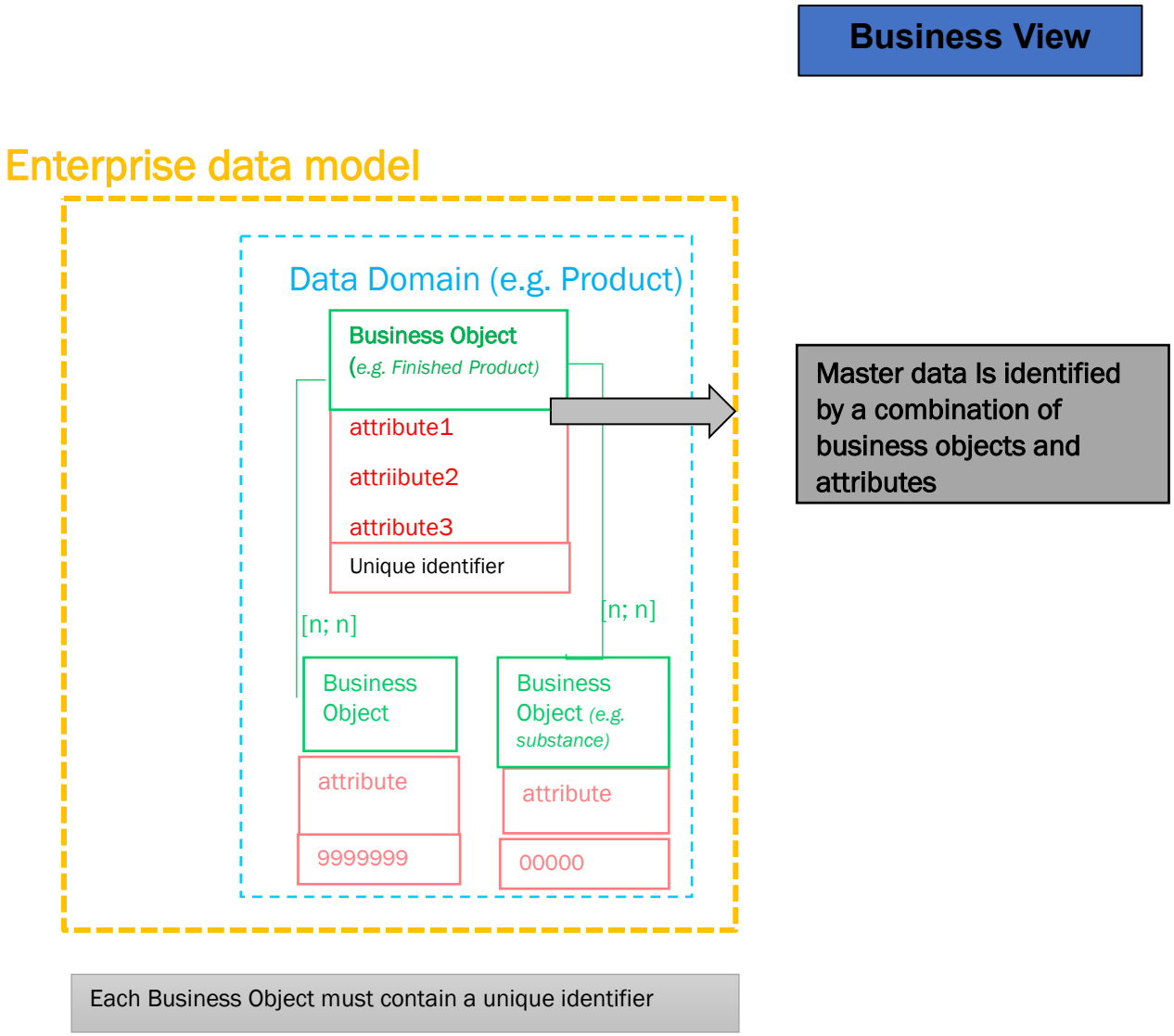
Data is commonly aggregated from back-office systems (e.g. SAP, RIM) into an MDM central repository (hub) and then validated against industry standard (e.g. IDMP) and business rules and then normalized using standard terminologies before becoming "Master Data".

The MDM is also sometimes used as a system source, when data is not yet digitalized (i.e. Non structured data contained only within documents)

These data can then be shared internally and externally ensuring data consistency checks.

3.7 Levels to categorize master data

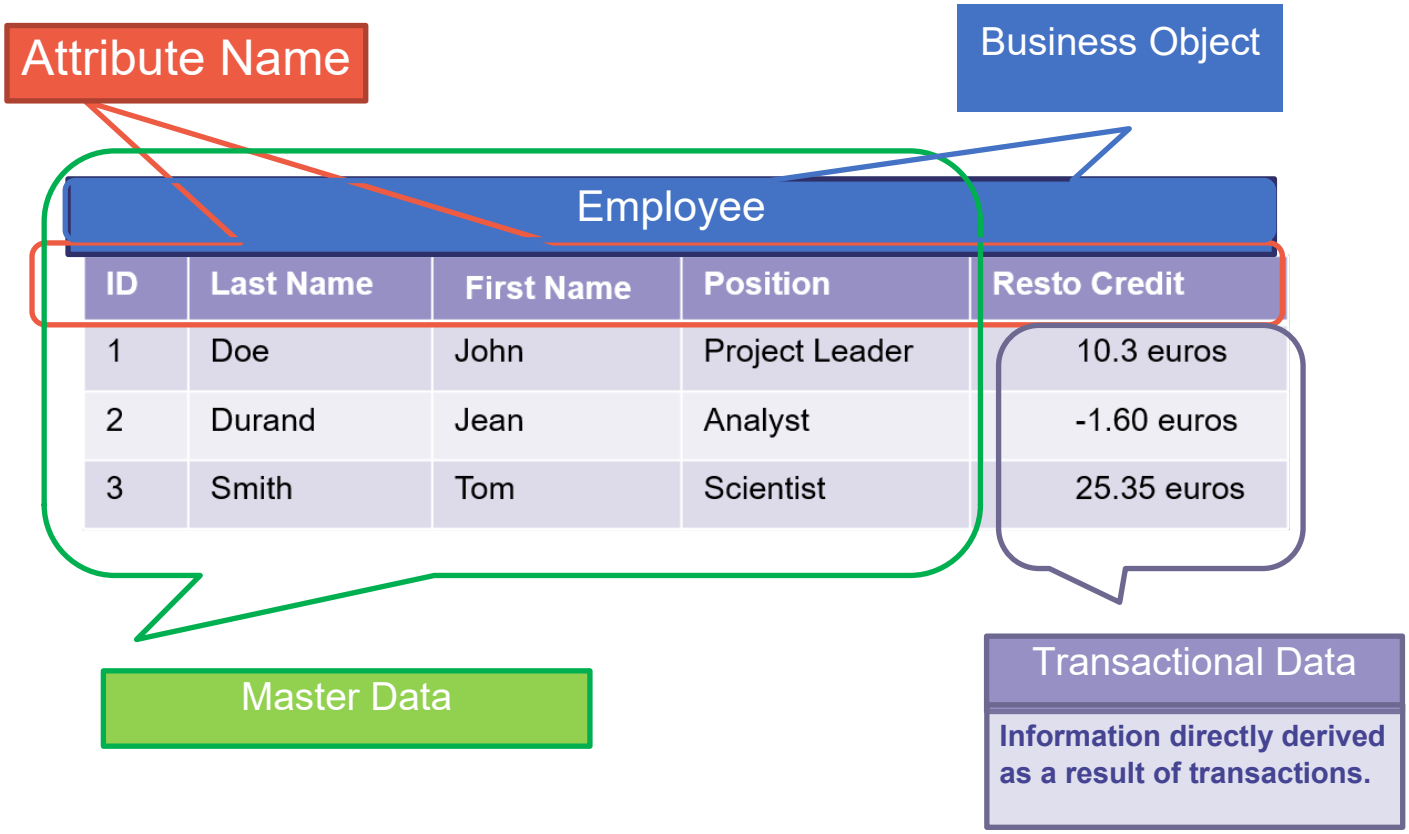
Standardizing the terminology to categorize our master data avoids misunderstanding between Business and IT and with our external partners

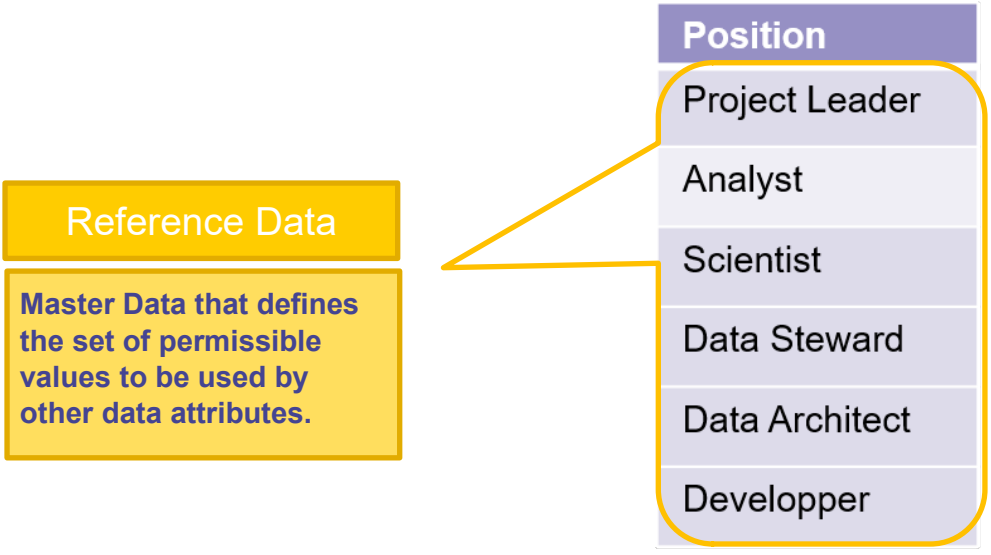


3.8 Master Data Terminologies and definition

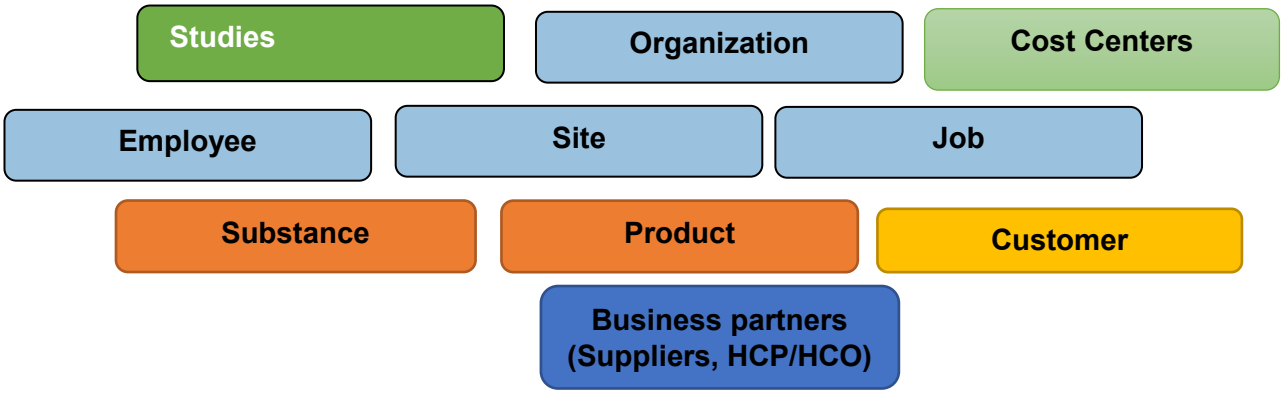
Name	Alias	Definition
Attribute	BO characteristics Data elements	A Business Object attribute is an element that constitutes and defines a business object.
Business Object	Business entity	A Business Object Is a meaningful concept for stakeholders within an organization. A business Object is uniquely identified by one or a combination of several attributes. e.g.: finished product.
Data Domain	Class	A Data domain regroups business objects logically linked together (e.g. Product, study, ...)
Enterprise data model		An Enterprise data model refers to the logical inter-relationships between different data domains involved at the enterprise level. e.g.: Ipsen, Sanofi, Pierre-Fabre, Servier

3.9 Example of Master Data Categories





3.10 Common examples of Master Data in Pharma (not exhaustive list)



Reference Data

- **Countries**
- **Currencies**
- **Exchange rates**
- **Languages**
- **Unit of measures**

Don’t miss our video

YouTube

https://www.youtube.com/watch?v=Jd2yitGIZ5M

4 Data Governance tools

4.1 Introduction

The terms Business Glossary, data catalog and data dictionary are often used interchangeably within the industry, even though they have distinct meanings.

Therefore, this can lead to confusion on what is meant by these terms.

During several DataGov Lab workshops, we have started to define these tools, their contents and the associated Use cases and decided to focus on Business Glossar(ies) and Data Catalog(es).

The Business Glossary and Data Catalog are part of the Data Governance program of all companies who are part of the DataGov Lab.

4.2 Data Governance tools – Takeaway/Summary

- **A Business Glossary** contains the list of important Business terms (the speech community), a business definition and the Data Domain Owner of the definition. This tool is owned by Business.

- **A Data Catalog** is a collaborative tool containing the list of data available within the company and the data owner associated to the data.

- **A Data Dictionary** contains the description of the data (e.g. data type, format, length, defaulted values, ...) within an application (data warehouse, data Lake, …). This document is created and owned by IT.

- **Data desk** stores any data, whatever its format. It provides the physical access or link to the physical data in accordance with Confidentiality, Integrity and Availability requirements, as well as whether access requires completion of training beforehand

4.3 Business Glossary and Data Catalog: Why these Systems are required?

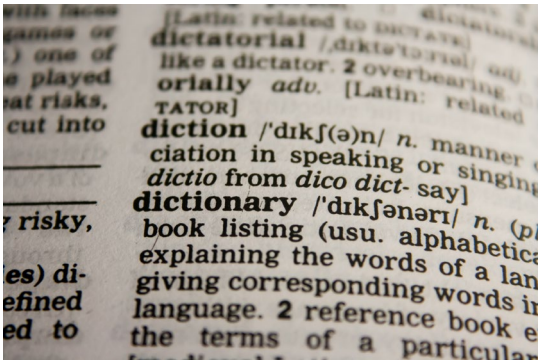
Dedicated systems are really expected to offer functionalities that cannot be fulfilled with excel.

Furthermore, the tools available on the market allow links between the Business glossary and Data Catalog tools, which is not possible to do with excel.

- **Access rights management** (ability to create user groups and roles)
- **Audit trail** (Ability to track old value, new values, date of creation, last change data and the user responsible (creation, update, delete)
- **Review and approval workflow**
- **Customizable dashboard (number of data / definitions, etc.)**
- **Ability to do query and export**
- **Search functionality (**Quick search bar allowing users to perform “Google-like” semantic search

5 Business Glossary

5.1 Introduction



The Business Glossary is a tool that ensures consistent terminology (data definitions, vocabulary) across the enterprise/organization, allowing the use of the right data for the right purpose by data consumers.

The common language held in the business glossary democratizes data knowledge.

A technical expert can understand the business meaning and business is able to determine exactly the name of the data required in a business process. Another objective of this tool is to eliminate any confusion about the meaning of a Business Term

The most difficult part for the Data Governance team, is to demonstrate the added value of a Business Glossary within the enterprise.

It becomes a success when Enterprise functions adopt it in their current Business activities.

With a Business Glossary, the Business community should be able to get answers to the following questions:

- Is the definition available, approved in the Business Glossary?
- Is the definition coming from an international standard?
- What is the Business term to use in my report, system?
- Is the definition used in my process is consistent with the enterprise one?
- Who is the owner of the term and definition?
- Are there synonyms, acronyms describing the Business term?

The Business glossary is a living tool as the content matures and evolves. The language also changes over time. It takes a continuous effort to ensure it is up to date and synchronized with the most appropriate and recent guidelines and international standards.

5.2 What is a Business Glossary?

Examples of definition

Gartner

Summary

The business glossary is the semantic foundation for logical data warehouses and business analytics. To generate interest and sustain effort, adopt both a top-down approach (such as dedicated funding and defined projects) and a bottom-up, grassroots approach (using gamification and storytelling).

Business Glossary definition

"The Business Glossary is an approved, governed compendium of business term names and definitions.

The process developed will define how the organisation creates, approves, updates, and promulgates consistent business terms and definitions, fostering shared data usage across the organisation."

DAMA Southern Africa

DAMA Version 2

Business glossaries have the following objectives:

- Enable common understanding of the core business concepts and terminology
- Reduce the risk that data will be misused due to inconsistent understanding of the business concepts
- Improve the alignment between technology assets (with their technical naming conventions) and the business organization
- Maximize search capability and enable access to documented institutional knowledge

ZEENEA

What is a business glossary?

While data dictionaries are useful to technical users, a business glossary is meant to bring meaning and context to data in all departments of the enterprise.

A business glossary is therefore **a place where business and/or data terms are defined**. It may sound simple, however, it is rare that all employees in an organization share a common understanding of even basic terms such as "contact" and "customer".

Colibra

What is Collibra business glossary?

The **Business Glossary business** case envisions a systematic process to reach agreement and alignment between all stakeholders on their **Business** Assets and how they relate to Data Assets and Technology Assets. It envisions to answer questions such as: what is the meaning of a **business** term, e.g., "Customer"?

IBM® InfoSphere™ **Business Glossary** provides users with a Web-based tool for creating and managing standard **definitions** of **business** and organization concepts by using a controlled **vocabulary**. ... You can use terms to classify other objects in the metadata repository based on the needs of your organization.

5.3 Business Glossary -DataGov Lab definition

A Business Glossary houses agreed-upon definitions of business terms and relates these to data.

Each term is associated to metadata such as synonyms and responsibilities on data (e.g. Data owner).

A Business glossary is often called a Data glossary, within this document only the term "Business Glossary" is used.

5.4 Benefits of a Business Glossary

The expected benefits of a Business Glossary are described below:

- Enables a common understanding of the core business concepts
- Enables access to the standardized terminology of the company
- Minimizes the misuse of data due to inaccurate understanding of the business terms.
- Improves alignment between cross functional teams (e.g. Business and Technical teams)
- Supports induction of newcomers
- Acts as a reference for the reporting and metrics (KPI) within the enterprise

5.5 Scope of the Business Glossary

As it's impossible to define all single data elements used within the enterprise/ organization, the Business Glossary should contain the important business terms (most meaningful) for the enterprise/organization.

It is recommended to start with a small scope of business definitions to identify the business needs of the enterprise/organization, onboard the stakeholders and set up the maintenance processes.

The small and first scope can be:

- Starting with the master data of the enterprise Data domain per Data domain (e.g. product, supplier, study...)
- Gathering business definitions already described in different documents/ referential of the enterprise
- Starting with a big project and having the project infrastructure to support the business glossary development,
- Focusing on Data used by different business functions and available in the Data Warehouse, Data Lake, dashboards, Key Performance Indicators (KPI), …

When the first draft of the Business glossary is completed with success, it can be extended according to the following steps:

- Identify the main Data Domains of the enterprise (e.g. study, product, supplier, …)
- Determine with each Business Domain owner, the main Business terms from each Data Domain (20-30 per Data Domains).

When several definitions of a business term appear to be the root cause of data pain points, this business term will be prioritized for its inclusion in the Business Glossary. The outcome may well indicate that the same term is used to describe different data elements within different parts of the organization and further refinement is required to create different Data elements. Indicating the name where the data is used in the data label can solved the issue.

Example: Delivery address. Is it the customer's address, the warehouse location, the suppliers' address

5.6 Business Glossary Ownership

The owners of the Business Glossary tool are really the Business functions responsible for identifying the important business terms and their definitions.

Technical functions are not part of the process of creation and maintenance of the Business Glossary.

5.7 How define a Business Term?

The pragmatic approach is to start by using international standards. Examples are provided below:

- **Industry terms:** Schema.org.
- **Finance**: services the Financial Industry Business Ontology (FIBO)
- **R&D - Studies:**
  - ❖ CDISC (SDTM, SEND, BRIDG, CTR)
  - ❖ clinicaltrial.gov
  - ❖ National Cancer Institute (NCI)
- **Genetic terms:** Glossary of genetics terms from National Human Genome Research Institute https://www.genome.gov/genetics-glossary
- **Product**: IDMP based on 5 ISO standards
- **ISO standard** e.g. ISO 3166 for the list of countries

As there can be differences across the different standards these will need to be identified as synonyms for appropriate mapping.

5.8 Rules to create a Definition

- Definition must be understood by everyone (Business and IT/technical people)
- Definition (and Business term) must be unique with a unique definition
- Definition must be precise, concise and unambiguous
- Definition must be stated in the present tense
- Definition must be stated in the singular
- Definition must be stated in a descriptive sentence
- Definition should avoid acronyms and abbreviations
- Definition must not contain the words used in the term (tautology)
- Definition must state what the concept is, not what it is not
- Definition must be stated without embedding definitions of other data or concepts
- Definition must not define the data with the words that make up the term. Avoid cheeseburger definitions (i.e. a cheese burger is a burger with cheese)

5.9 Business Glossary contents

5.9.1 Mandatory attributes

- **Terms Definition**

The best practice to define a term is to align it to an international standard (e.g. IDMP, CDISC). However, sometimes this definition must be adapted to the sponsor or context within the organization. In practice, both definitions, (standard and the adapted definition within the enterprise), are collected in the business glossary.

- **Classification of the Business terms**

The best practice is to organize a categorization in a manner that makes finding definitions easier. This categorization will allow the association of a business term to multiple classes.
Generally, Data Domains and subdomains are used to classify terms but other classifications such as processes name, data life cycle indicator (e.g. study set up , study conduct, study close out) can be added to contextualize the definition.
In practice, as some business teams can belong to several categories, it is not recommended to add too many classification levels.

The Business terms can be related either to Business Objects, attributes, List of Values (LOV) however this classification is not recorded in the Business glossary.
The terms and the associated definitions must be independent of the IT system model in which the terms are recorded.

When the Terminology is standard and well known in the enterprise (e.g. ISO 3166 country code), only the reference of the terminology is recorded without the detail of each value. For some less known standard terminologies it may be appropriate to provide a relevant definition description.

- **Versioning of the definition.**

The versioning of a definition is mandatory to track the changes done on the business definition. It is generally done automatically within a Business Glossary tool.

Example: 0.1 for the first draft definition and 1.0 for the first approved definition

- **Status of the definition.**

The status of a definition is mandatory to track its life cycle.

Some rules will determine at which status the terms /definitions should be published.

Example: Proposed, draft, approved, deprecated

- **Name of the Actors**

The name of the Data Domain Owner must be recorded in the Business Glossary to be contacted in case of questions or issues.

Other roles can be recorded in the business Glossary, (e.g. Data Owner, Data Steward) depending of the data maturity within the company.

The Owner's function name can also be added in the Business Glossary however company re-organizations will increase the requirement to update this more frequently.

The Owner's role within the function is also optional as the role is not an indicator of data ownership but it can facilitate the research of a new data owner when the previous one leaves the company or change of role.

- **Synonym**

Even if the objective of the Business Glossary is to define the "business term" to use in the enterprise, some other terms are frequently used. Therefore, the accepted term should be agreed on and described in the business glossary.

- **Acronym**

Sometimes a business term is only known by an acronym. Adding acronyms in the Business glossary is of immediate value and generates positive word of mouth associated with the glossary.

5.9.2 "Nice to have" attributes

These attributes can be helpful but most of the DataGov Lab members are not using it today.

- **Language**

This attribute is company dependent. Multiple languages can be used (e.g. French and English). It can be regulatory requirement to have local language for some departments (e.g. Tech Ops).

- **Metadata description**

Metadata, such as the name of the technical attribute and the format), can be added in the Business Glossary. This is especially valuable for master data to ensure this information can be exchanged within the company. These attributes, largely shared within the enterprise, should be the reference to use for the setup of a new IT system.

- **Related terms**

Related terms describe other terms associated with the definition

Example (non-exhaustive list)

- Is calculated from
- Is a Synonym to
- Is a type of
- Has a
- Has Type
- Is of

e.g. Bill of Material (BOM) definition is associated to a formula and pack.

5.9.3 Out of scope attributes

Data lineage is not a functionality expected in the Business Glossary but is in the Data Catalog.

If the Business Glossary and Data Catalog are connected, (interface or linked within the same tool), each data within the data catalog is to be linked to the definition in the Business Glossary when it's applicable.

5.9.4 Link between the Business Glossary and Data Catalog

Linking a Business definition recorded in the Business Glossary with data recorded in Data Catalog generates immense value.

It allows a better understanding of the data available within the enterprise and provides the best way to identify data which is aligned to/not aligned to the Enterprise definition.

5.10 Roles and responsibilities

Today, due to a lack of data maturity within the enterprise, Data Governance leads are generally responsible for creating and maintaining the Business Glossary.

Data Governance leads may help by proposing some definitions from international standards which are then updated if needed by Business. The classification of the terms is also generally done by Data Governance leads today.

Later on, as the level of maturity increases, the Data Domain owner, Data Owner and Data Steward should lead or oversee the business glossary process with obviously a Data Governance oversight.

5.10.1 Requester

Any collaborator can request a new term.

5.10.2 Business Data Steward (BDS)

Is (are) responsible for:

- Gathering information, agreeing on the definition between all stakeholders and propose the adding of the term within the Business Glossary.
- Classifying the Business Terms and definitions within the Business Glossary.
- Collecting feedback and questions from the business community.

The Business community should be able to provide feedback or address questions about a Business term/definition.

5.10.3 Data Owner (DO)

A data owner is a person or organization that has the authority on the Business Term and its definition.

Is (are) responsible for:

- Providing a definition and manage any changes of the definition.
- Approving the definition only after consultation and consensus with other Data Owners, and where necessary, this may require escalation to Domain Data Owners.

Note: The data owner has the right to define but also to change the definition An impact analysis must be performed to ensure a change of definition has no downstream negative impact on information flows, business processes or reporting and analytics. Changing the definition may impact on data values selected. Chances are, if the definition needs to be changed, this means a new data element is required.

5.10.4 Data Governance Lead

Is (are) responsible for:

- Defining the Business Glossary maintenance process and ensure the process is followed.
  The procedure should clearly describe:
  - *the process for the business terms approval*
  - *the publishing of the business Glossary*
  - *the maintenance of the Business Glossary to keep terms up to date as the business language changes.*
- Sharing Business Glossary within the enterprise/organization.
- Making the Business Glossary easily accessible to all employees.
  The optimal location would be the internal company web portal.
- Promoting the usage of the Business Glossary during the Business projects set up.
- Monitoring the usage of the Business Glossary.
- Identifying the Enterprise data using the business Glossary definition as well as the exceptions.
- Developing and publish KPIs.
  A series of Key Performance Indicators (KPIs) should be established, not only to demonstrate progress in the business glossary development, but also engagement of the user community.
  The KPIs should measure and track information on the number of approved business terms & definitions and the number of visitors during a period and feedback received.

5.10.5 Data Domain Owner (DDO)

The ***Data Domain owner*** is the owner of the definition within the enterprise.

The DDO will be the ultimate accountable for the definition. This will help in cases where the same term with the same meaning transcends different data domains.

Obviously, this ownership requires consultation and consensus with the other stakeholders of the enterprise.

Example: *the term "Study identifier" can be supported by several stakeholders such as R&D for interventional studies, Medical affairs for observational studies, Commercial operations for HEOR studies.*

5.10.6 Business Glossary Administrator

Is (are) responsible for:

- Managing access rights and roles
- checking the proposed term and definition does not already exist, meets the criteria for a good definition,

Note: The Administrator role is not the Approver role

5.11 Which tool?

Generally, the exercise starts in the enterprise within an excel file to understand what is needed.

After a period, when stakeholders are trained and onboarded, pilots have been carried out within a business area and processes are established, a dedicated tool becomes mandatory to facilitate the following:

- the sharing of information (easy access to all users)
- the set-up of review and approval workflows by stakeholders.
- The tracking of changes

5.11.1 Business Glossary and MDM tool

An application/IT tool connected with the business glossary enriches the information. this link should be done at the level of business object / application object.

A Business Glossary interfaced with the MDM can be helpful, (not mandatory,) to retrieve definitions from Business Glossary and push them in the MDM.

It can be also used to retrieve metadata information (e.g. data name, format) from MDM to enrich the Business Glossary.

5.12 Business Glossary Template

Business Glossary- Template DataGov Lab		
A	Business Term	Name of the data
B	Enterprise's Business Definition	Definition of the Business Term within the enterprise
C	Synonym	A word or a phrase that means exactly or nearly the same as the defined Business Term
D	Acronym	Abbreviation formed from the initial letters of other words and pronounced as a word (e.g. ASCII, NASA) used to describe the Business Term
E	Standard Business definition	Definition of the business term within international standard (e.g. CDISC, IDMP)
F	Functional Domain	Functional domain of the data
G	Sub Domain	Category of the functional Domain
H	Status of the definition	Status of the definition (e.g. Draft, Approved)
I	Versioning of the definition	Versioning of the definition (e.g. V1.0)
J	Data Domain Owner	Name of the person responsible of the term
Optional		
	Related Terms	Other terms associated with this term

5.13 Business Glossary Example

A	B	C	D	E	F	G	H	I	J	K
Business Term	**Enterprise's Business Definition**	**Synonym**	**Acronym**	**Standard Business definition**	**Functional Domains**	**Sub Domain**	**Status of the definition**	**Versioning of the definition**	**Data Owner**	**Related terms** *Is a type of* *Has Types*
Dose	*The amount of drug administered to a patient or test subject at one time or the total quantity administered.*			*The amount of drug administered to a patient or test subject at one time or the total quantity administered. (CDISC-SDTM)*	*Study*	*Clinical study*	*Approved*	*v1.0*	*Dupont Jean*	*Dose* ***Has a*** *UNIT*

5.14 The Future/dream tool – what would this look like?

A tool connected to international standards with the capability to automatically propose one or several definitions for a dedicated business term.

This dream tool should also automatically propose classification of Business terms and reclassifications to ensure the enrichment of the Business glossary.

5.15 Business Glossary reference

- **ISO/IEC 11179**
  The ISO/IEC 11179 Metadata Registry standard is an international ISO standard for representing metadata for an organization in a metadata registry. It documents the standardization and registration of metadata to make data understandable and shareable. Wikipedia

6 Data Catalog

6.1 Data Catalog introduction

The data catalog serves as a single point to locate information across the enterprise. It contains some information about the data, (business definition, confidentiality, GxP, privacy criteria, quality criteria), as well as the owner of the data within each data source (application, documents, reports).

A Data Catalog can be compared with an online shopping platform. The platform should present all the objects available for buying. Obviously as there is a large choice of different products, a buyer will generally start by selecting a category (books, music, home, health and beauty, …) and then sub-categories to select one or several products. The categorization is therefore essential to easily find the object you are searching for. When you have selected the correct category, you can see a description of several models and several vendors as well as recommendations from the other users.

When the correct product has been selected, the buyer will add this to their shopping basket and proceed to the payment section.
Once the product has been delivered, the user can return to the platform to provide their recommendation/critique about the product (example: data Quality assessment).

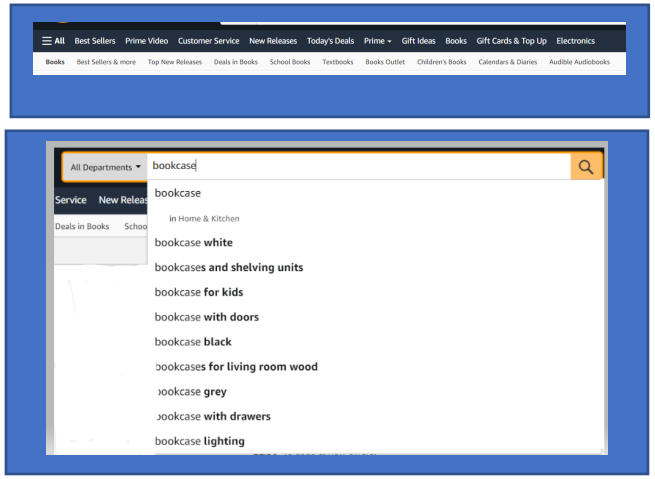
DataGov Lab

Data Catalog- An Online shopping Platform

Collaborators

Each collaborator contributes to the Enterprise Data Catalog

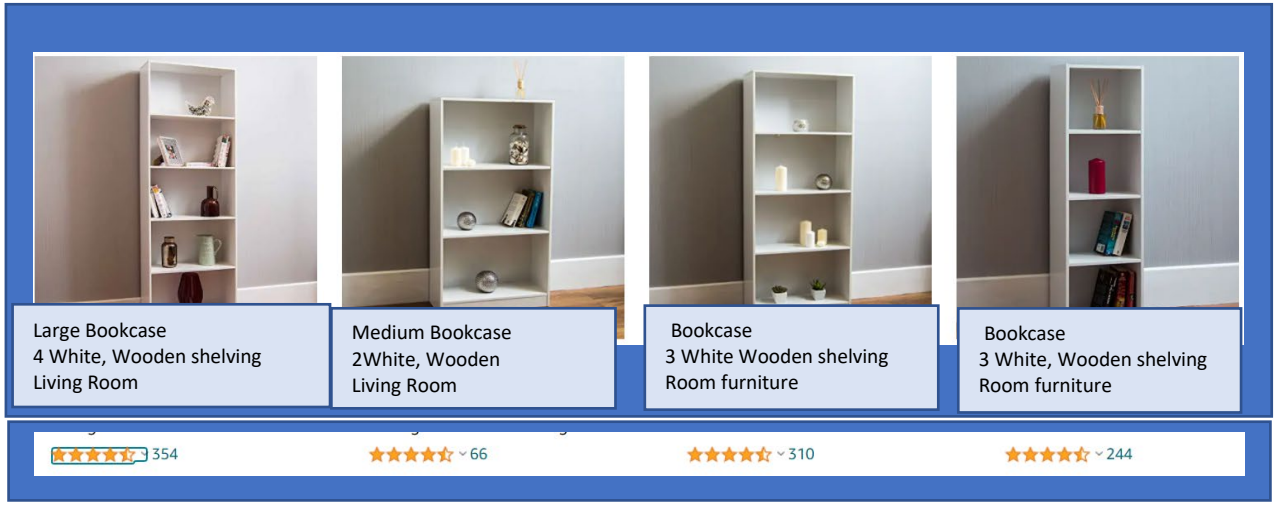
Recording critical data available within systems, documents or reports



Consumer

I search for a bookcase.

Providing data name, Definition, Business rules and Tagging data



I search for a specific white bookcase.

I consult the instructions and advice from other users.

I can ask Data Catalog Questions and feedback to data owner

6.2 What is a Data Catalog?

Colibra

A data catalog inventories and organizes all of an organization's data assets. A data catalog uses metadata to help data professionals discover, understand, trust and manage their data for governance or business purposes.

6.3 Data Catalog -Datagov Lab definition

A data catalog maintains an inventory of data available in the company's information landscape (adapted from Gartner)

6.4 Benefits of a Data Catalog

- Find and understand a relevant dataset for the purpose of extracting business value.
- Facilitate identification of master data.
- Facilitate data privacy guidelines (e.g. GDPR) compliance.
- Facilitate confidentiality classification awareness according corporate policy.
- Facilitate the usage condition (authorization to send data outside company even if public data).
- Facilitate impact analysis when data or system changes.
- Support IT for architecture, System design, interfaces and reporting development and development decisions.
- Facilitate identification of External databases (RWD, SPOR) and content supporting critical business activities.
- Facilitate quality assessment awareness.
- Facilitate the characterization of critical data to apply security measures.

6.5 Scope of the Data Catalog

Ideally the Data Catalog should contain structured and non-structured data.

It's easier to start in the same time as the setup of a new IT system as resources from different functions. i.e. (IT and business working in parallel) are available).

The most difficult part is to collect the relevant data from IT systems which are already in place. Some systems, such as ERP (SAP,) require multiple resources as each user may have expertise in only one area and only knows a single piece of the information.

Completing the Data Catalog with non-structured data created in key dashboard/reports or processes will be very helpful for the future users.

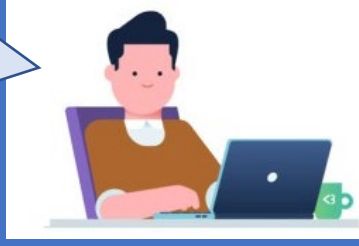
6.6 Data Catalog Tool Ownership

Depending on the company organization, the owner of the tool should be a corporate function such as the Data Architecture.

6.7 Users of the Data Catalog

The Data Catalog is a collaborative tool used by the different business functions and Technologies such as IT, security, digital, architecture. Each of these functions should categorize the data according their needs.

I use Data Catalog to validate/ support the architecture to be designed in a business project.



IT Architect

I want to know the quality of the data.

I want to identify the external data which cannot be used by all company users.

I want to know the level of confidentiality of the data

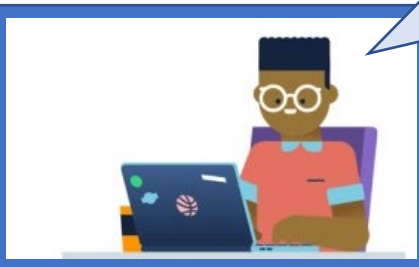
I know where I can find a required data

Data Consumer

I want to identify metadata in the information systems to be able to identify future referential

Master Data Manager

I understand what data are available in the data lake and I know the data owner to contact



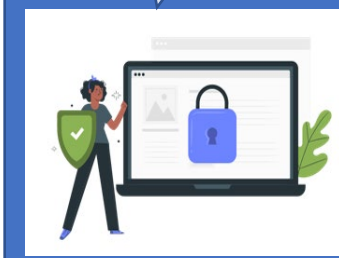
Dara Scientist

We want to identify systems and data impacted by a change.

I want to identify the critical data to apply security measures

Data Governance Lead

IT Business Analyst



Security Team

I want to identify the systems involved in the management of the personal data of a given patient in order to be able to ensure the exercise of patients' rights over their information.

Data Privacy Officer

6.8 Data Catalog contents

6.8.1 Mandatory attributes

- **Type of Data Source**
  The objective is to collect the structured and non-structured data to have a holistic view of the data within the enterprise.
  The type of data source will vary, for example it may be an IT system, a document or a report

- **Data Source name**
  This should be the application name, the document name or the report name.

- **Business Object and Data label**
  The data label is mandatory as well as the Business Object label.
  Technical names within the source can also be referenced.

- **Data usage**
  As a minimum, the data catalog must describe where the source is located and where the data is used (system, reports) and ideally the associated process name (e.g. SOP reference) .

- **Data Classification**
  The Data Catalog should allow the classification of important information related to data. (e.g. confidentiality, personal data, sensitive data, ..) in order to assess and evaluates :
  Examples:
  - ❖ Data Privacy compliance (e.g. Personal data Yes/no)
  - ❖ Security Compliance requirements (e.g. Confidentiality levels)
  - ❖ Regulatory Compliance (e.g. GXP data Yes/no)

6.8.2 Nice to have attributes

- **Data Quality indicators**

Depending of the maturity of the data governance within the enterprise, data quality indicators should be recorded in the Data Catalog. These indicators are very useful for the data consumers.

Some Processes should be put in place to define how to calculate, collect and record these indicators.

6.9 Data Catalog Functionalities

6.9.1 Nice to have functionalities

Depending of the maturity of the enterprise, the following functionalities should be used in the Data Catalog tool.

- Ability to Interface the Data Catalog with the systems/ applications of the enterprise to automatically extract the System's metadata. The metadata will be up to date, preventing a manual revision / update at each upgrade or change of the system.

- Ability to automatically identify the same data across landscape. The duplicates are then analyzed manually.
- Ability to execute an impact analysis if a change must be done in the Systems landscape.

6.10 Roles and responsibilities

6.10.1 Business Data Steward (BDS)

Is (are) responsible for:

- Contributing to the definition of the Data Quality indicators
- Assessing conformance to business rules, service level agreements etc

6.10.2 Technical Data Steward (TDS)

Is (are) responsible for:

- Completing the technical part of the Data Catalog. such as:
  - the metadata description (e.g. Technical Name, format, Business Object.)
  - The source of the data

6.10.3 Data Owner (DO)

Is (are) responsible for:

- Approving the definition of the data used in the system
- Defining the Categorization of the data in terms of confidentiality, Privacy, etc ..

6.10.4 Data Governance Lead (DGL)

Is (are) responsible for:

- Monitoring the completion of the Data Catalog and compliance with instructions.
- Maintaining Data Catalog processes and guidelines
- Training data actors on the Data Catalog

6.10.5 IT Architecture

Is (are) responsible for:

- Defining the process of maintenance of the Data Catalog and ensuring the process is followed.

6.11 What should be our Future/dream tool

- A Machine Learning data catalog to automatically discover, create inventories, profile and tag data and create semantic relationships between distributed and siloed data assets.

6.12 References

Data Catalogs Are the New Black in Data Management and Analytics -Gartner 2017