Collibra

Collibra Best Practices on Community Structure Design

Collibra EAS

22/09/2020

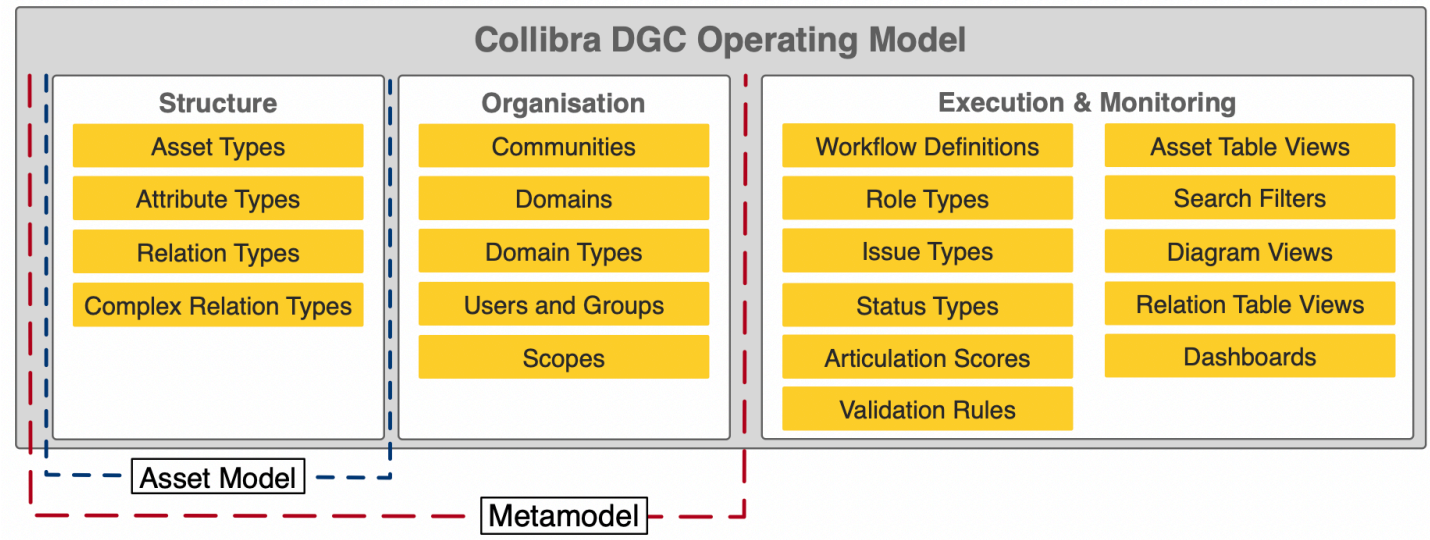
Collibra

Table of Contents

Introduction

Robust organization structure, understanding of the owned organizational assets and the process of governance is the foundation of Data Governance Centre (DGC) configuration. In Collibra terminology we call this Operating Model (OM).
While Collibra DGC offers a high degree of flexibility and configuration, the resilient and viable OM is the necessary foundation to address possible future challenges and avoid pitfalls. Thus, we believe that Operating Model Design Procedure (OMDP) is a necessary key step in building Data Governance Foundation (Building Block of Data Driven Organization per Collibra Prescriptive Path (CPP)).

The Collibra Operating Model consists of 3 high-level components/concepts:



In this document we will introduce you to Collibra Best Practices on Organizational Concepts, e.g. how to approach Your Organization Structure, translate that to Community structure, what are the different options representing the Ownership and Stewardship concepts.

The Vision (Architecture Vision)

In this chapter we will discuss the general pre-requisites, activities and questions to ask while shaping the vision of Your Data Driven Organisation.

1 Formalizing the Data governance

Before proceeding to software implementation, or implementation of Your Data Governance Organization in any tool, do an exercise of shaping your (architecture) vision. Assess you As-Is situation and set a high-level target of To-Be organisation, put your general concerns on paper.

- *Are you a large organization with a lot of stakeholders?*
- *Is the Chief Data Officer (CDO) assigned? Do you even need one?*
- *Are you centralized and reporting to CDO?*
- *What is your current organisation structure? Is this a target structure?*
- *Do you have different business units that are not related to each other in any way?*
- *What are the typical data problems/challenges your company wants to solve?*
- *Are the roles and responsibilities who entitled to work with data defined?*

Questions like this can help to create a "decision path" and start formalising your *ownership and stewardship* models of your organization

Ownership Model shows how information grouped in organisation and who own this information. Defining of data owners usually helps to understand which Communities will be created in DGC (see. *2.2.1 Ownership Concept*). For example, if known that information is grouped in your company by business units (Finance, Marketing, Sales etc.) and owners also known (CFO, CMO, CSO etc.) then very probably that final metamodel will contain (Sub)Communities like “Finance”, “Marketing” and “Sales”. CFO will be assigned on Finance Community with role Owner, CMO will owns all data within Marketing Community, CSO will owns Sales Community etc.

Name	Domain Type	Owner
Enterprise		Bob Brown
Compliance		Laura Fontan
Finance		Dmytro Lugovyi
Information Technology		John Fisher
Marketing		Olivier Van Hoof
Sales		Tudor Borlea

Stewardship Model shows how people treat information (asset types) and how they govern this information (people and processes).

In DGC it will helps with defining of domains with asset types and processes around them (see. *2.2.2 Stewardship Concept*). For example, if users identify two types of finance reports (e.g. “Strategic” and “Operating”) and it’s known that there two persons who govern this information on everyday basis (e.g. Jack and Mary) and all questions regarding quality of the data in these reports should be forwarded to them then very probably we will have two domains within Finance Community with names “Strategic Reports” and “Operating Reports”. Jack will be assigned as Steward on Strategic Reports domain and Mary will govern Operating Reports. Jack and Mary will be that people who will help us to find correct data in the company to import into DGC and will help to identify processes which will be applied to Reports.

Name	Domain Type	Owner	Business Steward
Enterprise		Bob Brown	
Compliance		Laura Fontan	John Fisher
Finance		Dmytro Lugovyi	
Finance Business Glossary	Glossary		
Finance Operating Reports	Report Catalog		Mary Smith
Finance Strategic Reports	Report Catalog		Jack Daniels
Financial Data Elements	Data Usage Registry		
Information Technology		John Fisher	
Marketing		Olivier Van Hoof	
Sales		Tudor Borlea	

Note*: it’s absolutely fine if existing roles in your organization will be mapped to Collibra DGC out of the box (OOTB) roles with another names, e.g. Data Manager will be mapped to DGC OOTB role Business Steward. Permissions and responsibilities are more important than names of these roles.

1.1 Organization Patterns

We've seen a lot of different implementations of organisation handling their data. In order to support you in your decision-making process how to better organize Your enterprise we've summarized few of the concepts with it's pross and cons.

Note: Historically, data has been collected, managed and owned at the level of individual departments. Each department has developed their own procedures, data formats, source system and terminology for their own needs. And now we need to get all of this information to work as a whole as a one peace for the benefit of the company. And it means that the best approach during the design phase will be not to describe existing information landscape but think about how this information should be structured and how do we want to govern it in the future.
For example, instead of the many glossaries for each functional area we may want to create one new Enterprise Business Glossary domain with all business terms that are used in these functional areas but well known across the organisation.
Or another example, quite often we have to find and assign owners for the information or change current ones, and it changes our Ownership Model and the way how information will be structured in DGC and how it will be governed in the future.

1.1.1 Ownership Concept

In our nature to group objects by their characteristics. Talking about information, grouping helps us to search, understand and of course manage it.

Variety of the information gives us an option to group it in many different ways. And to be able to govern this information effectively we need to find the best way of grouping, taking into account information itself, business use cases, ownership model etc.

There are examples of the common approaches to how information can be grouped in DGC:

- By Function
- By Subject Area
- By Business Process
- By Geography
- By Business Unit
- By Product

Each of these approaches has its own pros and cons in gathering, importing and later governance. For example:

By function (*E.g., marketing, sales, legal, finance*)

Pros	Cons
• Often easy to find and set-up	• Sensitive to changes. If Sales division will be divided into "Corporate Sales" and "Private Sales" then it will be quite difficult to split business terms between these two domains. • Assets owned across functions (i.e. Customer Data Domain) and identify one person who will take responsibility for the asset might be challenging

By Subject Area (E.g., customer, address, fleet)

Pros	Cons
• Pragmatic, easy to implement • Group people working on same subject area across silos	• Incomplete, no contextualization • Hard to assign responsibilities

By Business Process

Pros	Cons
• Detailed, complete, key concepts contextualized by process; cross-functional process traversal removes silo thinking	• High time and effort cost • Sensitive to changes

There is no one correct way to group information. Different companies, different information, different use cases and different result in the end. But in most cases users (stewards) already know the answer. Because information already has been grouped many times, in different documents, other tools etc. So, we just only need to talk to people and answer will be found shortly.

1.1.2 Stewardship Concept

Data Grouping is the first step. But before we will import this information to DGC and govern it there, we have to also consider company organization structure, ownership/stewardship model, existing process and of course the specifics of DGC tool. Because all of this will have an impact of the effectiveness of the data governance in the end. So, considering all this we can define three common approaches for Data Governance Structure – *Centralised, Decentralised and Federated*.

1.1.2.1 Centralised

Works well for

- single organisation at corporate level
- small to medium businesses with limited amount of business lines

Pros	Cons
• Rapid problem resolution • High resource efficiency • "Guarantees" global visibility • High levels of standards enforcement • High level of rule automation leading to data quality	• Larger prioritisation queue • Local dependencies on central group (e.g., time zones) • Can be bureaucratic and taking too much time to get things done • Perceived as being cumbersome versus agile • Has a tendency to be perceived as communicating versus collaborative

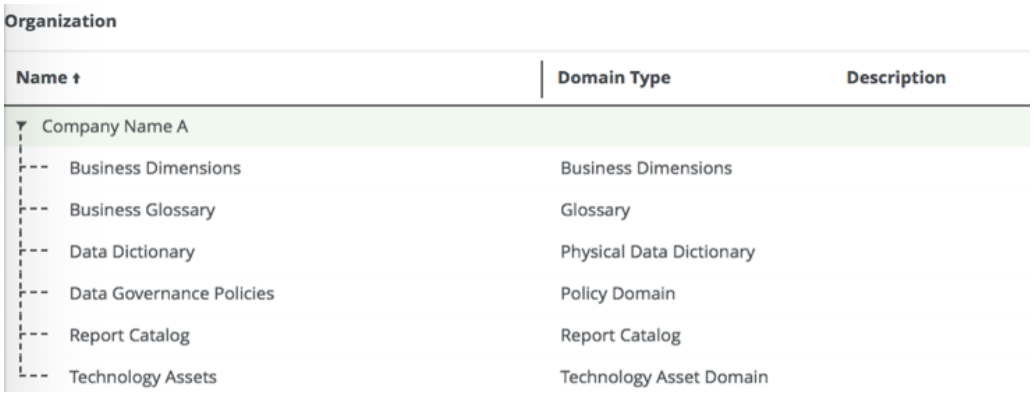
1.1.2.1.1 Centralised Generic Approach

Centralized Generic Approach - the simplest flat structure with one community and domains which are centrally managed. Every domain is configured for assets of the particular type(s) and stores all assets of this type across the organisation.

Specifics:

- Roles assignment will have to be micromanaged - all permissions will be defined on the asset level.
- Restrictions on workflow assignments – it won't be possible to apply different processes (i.e. "Legal Business Term Approval", "Finance Business Term Approval") to the same asset types ("Business Term") if they will be stored in the same Domain.

Example



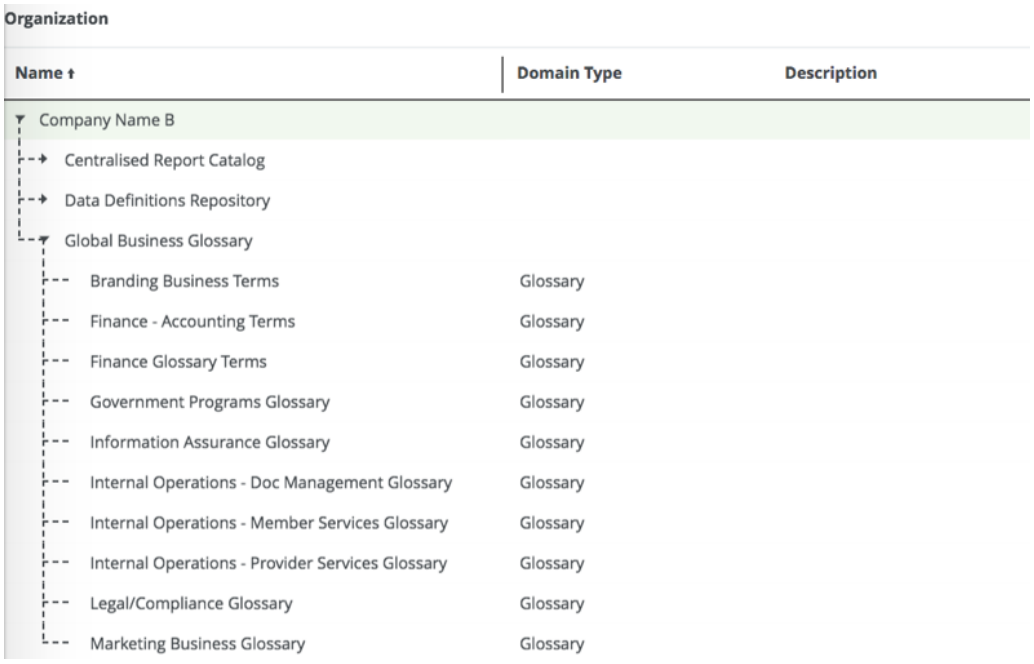
1.1.2.1.2 Centralised Segregated Approach

Centralized Segregated Approach - this structure is similar to the previous example but domains can be divided into several (logic can be very different) and then grouped under the one sub-community (see screenshot below). All sub-communities and domains are centrally managed.

Specifics:

- Problem with the location for the “global” assets that doesn’t belong to any particular business area – “Revenue” Report, “Region” Business Term, “Month” Dimension etc.

Example



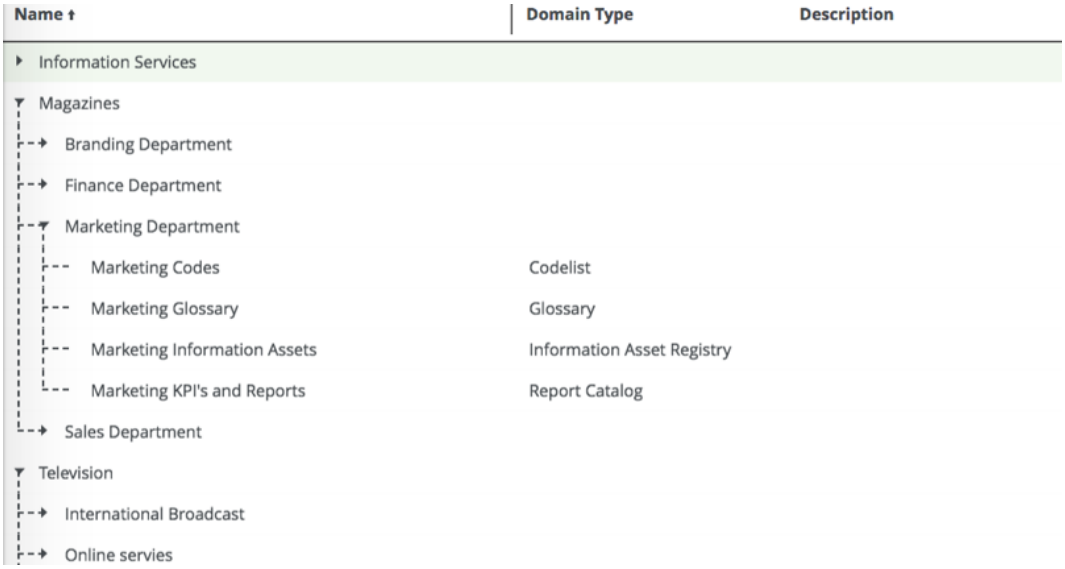
1.1.2.2 Decentralised

Works well for

- large companies with business units that have own information governance programs with no corporate alignment
- companies where each division is generally set up as an independent organization with complete control over own branding, marketing, sales, and finance

Pros	Cons
• Rapid response to local needs • Perceived as being more agile • Ownership aligned with individual business operational functions • Perceived as being in operational control • Communication is the norm not collaboration	• Local standards make it so the data and information is not shareable across the enterprise • Reporting and terminology in specific business operation and local vernacular • Quick to react and band-aid symptoms rather than solve root-cause problems • Little to no collaboration outside the local • Communication with other locals and across the enterprise becomes difficult because we lose the common language

Example



1.1.2.3 Federated

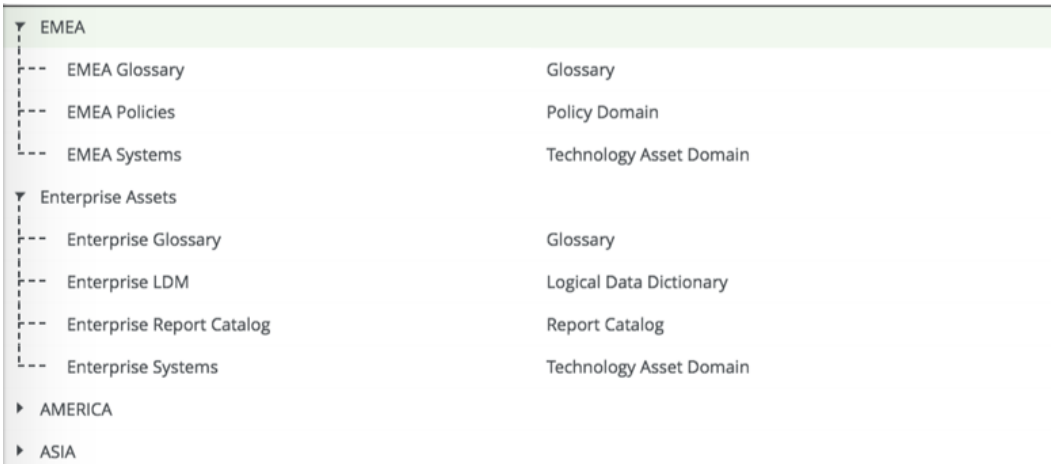
Federation provides uniform governance, preserving an enterprise view critical to top executives and at the same time it gives business units enough autonomy to govern data meeting their needs without delay or enterprise intervention. At the same time, it minimizes overhead and redundancy, saving costs through economies of scale.

Works well for:

- large companies with business units that have their own information governance programs with corporate alignment. Companies which are mature enough to decide which data they want to govern centrally and which locally (and which in a hybrid way)

Pros	Cons
• Accommodates local needs in timely response • Tighter alignment with business governance • Maintains standards enforcement on the enterprise level • Specific SLA's can be created to respond to enterprise & local needs • Automate rules to reduce the risk of creating duplicates (data quality improvement)	• Recognises what is shareable versus not shareable data & info across the enterprise • Need to be careful to not add too much bureaucracy, making it harder to get things done • Can be perceived as less agile and less responsive than decentralised • When the asset promoted to the enterprise level, ownership identification can be tricky •

Example



Commonly used Federated Processes

- **Adoption**: Local communities can adopt an asset from enterprise whose standard attributes are managed at the enterprise level and any changes made in the local communities will be overridden by enterprise.
- **Promotion**: Assets in Local communities can be prompted to enterprise communities once majority of business areas agreed on the common standardized properties(attributes/relations)

Rarely used Federated Processes

- **Specialization**: Assets in local area communities can take the properties from enterprise and specialize those properties to their needs.
- **Generalization**: Assets with defined for different purpose with different Properties which semantically mean same thing or generalized in the enterprise communities.

Important: Like with data grouping, data governance structure can't be absolutely correct or incorrect. Every single company is unique so, every time when we are preparing structure we have to consider specific of the company, its information, processes and so on. And don't forget about scalability. Data Governance Structure is not something that can be easily changed in the future. Think about the future.