ARTICLE

DATA MARKETPLACE

HOW TO BUILD A DATA MARKETPLACE FOR YOUR ORGANIZATION

By Mario de Francisco Ruiz, CEO at Anjana Data

"HOW TO SUCCESSFULLY BUILD AN INTERNAL DATA MARKETPLACE TO YOUR ORGANIZATION"

SEPTEMBER 2021

WHAT IS A DATA MARKETPLACE?

The concept of Data Marketplace is something relatively new in the field of Data Management and comes from e-commerce and typical portals born on the Internet for ***buying and selling products and services****.*

Just as we buy almost anything on Amazon, rent vacation homes on Airbnb or search for cars on *Wallapop*, the purpose of a Data Marketplace is to facilitate the exchange of data between producers and consumers in a controlled environment.

A Data Marketplace can be **external**, i.e., it can be used to exchange data between different organizations or even at the individual level (e.g., buying and selling data, sharing data for research, open data portals, ...), or it can be **internal**, where the information is made available only in the controlled environment of an organization or group of organizations.

On a practical level, the latter is based on the construction of a large controlled bazaar where **all information assets are made available to potential consumers** by different producers within the governed environment of an organization.

What are the benefits for an organization of building an internal Data Marketplace?

According to **Gartner**:

- Data teams are twice as likely to generate **measurable benefits** from external data sharing and 3 times more from internal sharing of data assets with their partners and collaborators.
- By 2022, 90% of corporate strategies will explicitly consider **information as a critical** business asset and data analysis as an essential competency.
- By 2023, organizations that promote data sharing will outperform their competitors on the majority of **business value metrics**.
- Data Sharing is a **Business Necessity** to Accelerate Digital Business (December 2020)
- Organizations that list data sharing as a Business necessity rather than a Data Management function will perform better in **Digital Business** and be more successful than their competitors (February 2021).

Thus, an internal Data Marketplace aims to simplify the entire **data value chain** and the required processes for the **sharing of governed and quality data** among the different stakeholders of an organization, providing the following **benefits**:

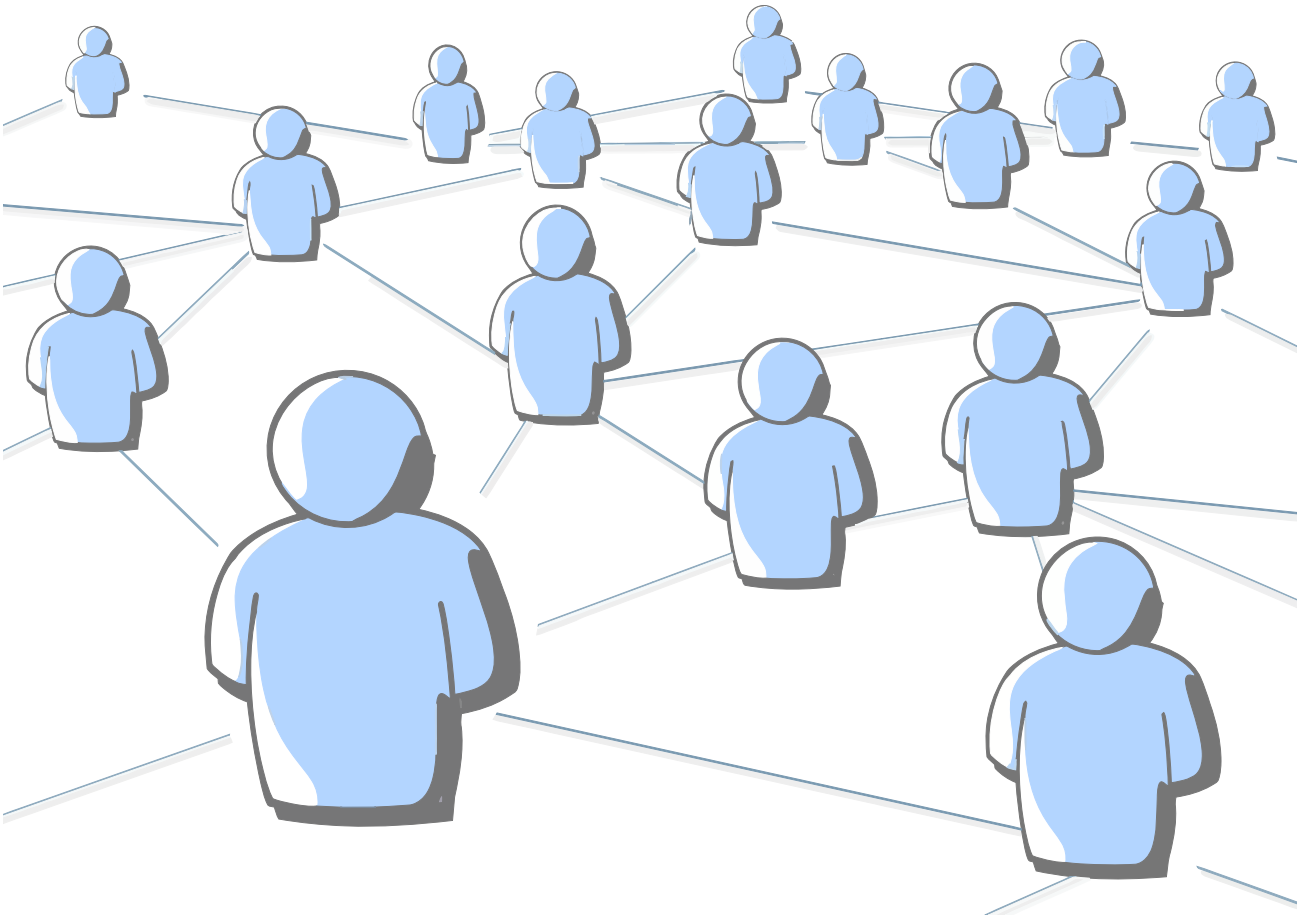
1. Greater understanding of the data assets available in the organization and greater contextual information facilitating self-service of data for better and faster decision making.
2. Facilitates the achievement of the objectives set by the organization thanks to the sharing of quality governed data.
3. Increased efficiency and maximization of synergies between projects, areas and departments thanks to the reuse of data and processes derived from the democratization of data.
4. Improved employee productivity and satisfaction: less time and effort spent searching for, understanding and requesting access to data; less time and effort spent preparing and cleaning data for exploitation; less time and effort spent fixing errors in data generation and processing.
5. Increased quality of data shared by producers and made available for use by third parties.
6. Increased transparency and trust in data assets and related processes by stakeholders.
7. Possibility of discovering new uses for the data available in the organization and generating new products and services.
8. Reduction of time-to-market and time-to-value in the launch of new products and services based on data analysis.
9. Cost reduction (maintenance, infrastructure, operations, development, impacts, ...) thanks to the simplification and automation of common technical processes and manual data management processes.
10. Prevention of potential impacts and reduction of operational risks derived from the processes inherent to the data lifecycle (manual errors, security breaches, ...).
11. Facilitates data monitoring and governance through automation and decentralization of processes, standardization of data access management and integration with demand management.
12. Easier to value and monetize data based on data usage.
13. It favors regulatory and normative compliance in the use of data, facilitating auditing and evidence gathering.

As we can see, building an internal Data Marketplace can bring a great number of benefits to an organization, but it is very important to always outline a **clear strategy**, starting by identifying some primary goals and defining the corresponding **metrics and KPIs** which will allow us to measure the impact of these benefits and whether we are really achieving them.

Who are the participants and stakeholders of a Data Marketplace within an organization?

__Data__ is an asset that has a characteristic that makes it __unique__ and differentiates it from any other commodity: the more and better it is __shared__, the more its __value__ and the potential impact it can have for an organization increases.

Ideally, a Data Marketplace should be used by the vast majority of data stakeholders in a **Data-Driven** organization.



In this context, a Data Marketplace is an ecosystem that has to take into account a very wide number of participants and stakeholders since its purpose is to serve as a **one-stop shop** for data governance and a central hub for **data sharing** within an organization. In this sense, a Data Marketplace cannot be defined thinking only from one point of view, but one must try to put oneself in the shoes of each and every one of the stakeholders that will be part of this new ecosystem.

Mainly, the participants and stakeholders of a Data Marketplace can be **classified into 3 groups**, each with their roles, functions, responsibilities, interests and corresponding benefits:

Producer's perspective  ***data owners*** ***data stewards*** ***data engineers***	• Control over the data production processes (data quality & availability). • Data preparation and certification for data sharing and consumption by others. • Knowledge about which owned data is being used by who and for what purpose. • Full audit to support normative and regulatory compliance.
Oversight perspective  ***C-level*** ***Data Office*** ***Audit*** ***Legal*** ***Compliance***	• ROI maximization for data initiatives. • Productivity and efficiency boosting. • Processes and tasks automation for costs reduction and operational risk mitigation. • Unified and homogeneus vision of the company's data usage and consumption. • Normative and regulatory compliance. • Possibility to value and monetize data.
Consumer's perspective  ***data analysts*** ***data engineers*** ***data scientists***	• Gain deeper knowledge about available data along with its context and meaning. • Better decisions driven by the access to better-quality and well-managed data. • Control over the data production processes (data quality and availability). • Agility in data consumption and exploitation leading to self-service.

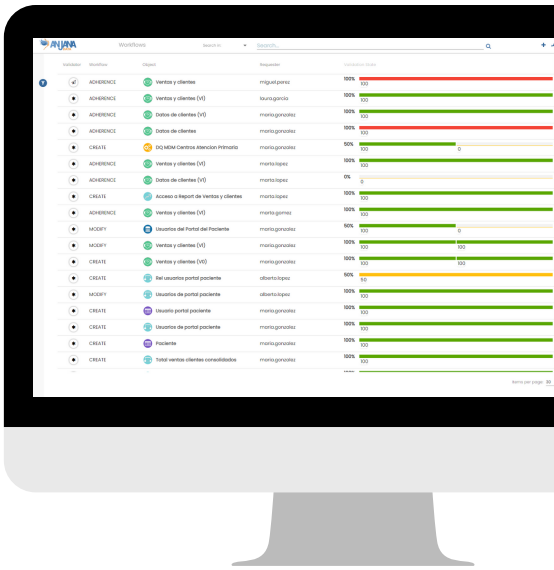
Historically, the roles of **producers/suppliers and consumers** of information have been performed by IT profiles with knowledge and skills in data development and exploitation (software/data/IT/BI developers/engineers/architects, operations/data technicians, ...) but as data start to gain weight within organizations and begin to be identified as a strategic asset, this paradigm begins to change, producing the following effects:

- **New roles** with functions and responsibilities over data assets within specific **data domains** related to business areas (data stewards, data owners, ...), transferring the responsibility for them from IT profiles to **business profiles**.

- Business profiles are beginning to have more training and greater capabilities for the exploitation and analysis of available data, moving from being mere consumers of prefabricated reports to become **data analysts**. In addition, these profiles have more and more and better technologies at their disposal that ease access, analysis and consumption of information in different ways, so their **"appetite for data" is growing exponentially** in a very accelerated manner.

- **New profiles** practically nonexistent until recently with **hybrid knowledge and capabilities** between technology, data and business such as data scientists, which are usually more attached to business areas than IT. Likewise, this type of **data specialist** profiles usually have a wide range of technologies within their reach that bring data much closer to the business and provide them with a set of previously unthinkable capabilities.

*All this means that we are no longer talking about **technological solutions** for technical IT profiles, but rather that it is becoming an obligation for technological solutions related to data to be designed to facilitate their **adoption by business profiles**, offering them a set of capabilities to which they did not previously have access, basically because they did not need them.*



Similarly, in terms of **cross-related roles**, we also find different profiles, both technical (IT architects, DBAs, systems technicians, CISO, CIO, CTO, ...) and business (data architects, compliance, legal, audit, ...), which are not left out of the new paradigm and have undergone a **significant evolution** while **new roles** have appeared that did not exist before (CDO, DGO, Data Office, ...). All of them also need to have their needs met from a data governance point of view and must be considered as stakeholders in the Data Marketplace ecosystem.

What are the elements of a Data Marketplace?

*The first and most important aspect of a Data Marketplace is to **abstract data governance from the underlying technologies** used for data capture, integration, storage, processing and exploitation.*

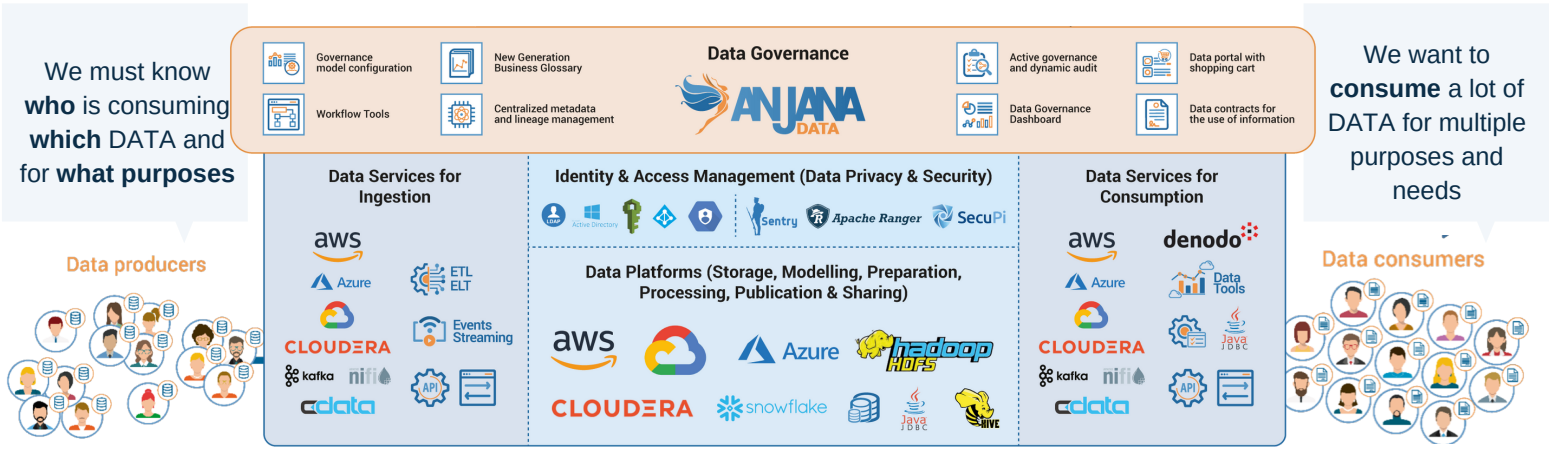
These technologies are usually designed from a **technical point of view**, where the most important variables considered are usually performance, integration capabilities, processing capabilities, data volumes, fault tolerance, high availability, etc. However, many of them do not take into account data governance, metadata management or interoperability capabilities from a more **functional point of view**.

This, together with the needs of an organization to have a **transversal vision of data**, much closer to the **language of the business**, facilitating its access and exploitation by potential consumers, means that these technologies have to be complemented by a higher layer that allows the implementation and operationalization of an **effective and efficient data governance with a technology-agnostic vision**.

"...DATA MARKETPLACE IS TO ABSTRACT DATA GOVERNANCE FROM THE UNDERLYING TECHNOLOGIES"

*From a Data Governance perspective data sharing should be leveraged thanks to the establishment of the corresponding **policies and procedures** while ensuring data compliance and stewardship*

- ✓ **Data sharing** needs must be satisfied in an **agile** way.
- ✓ Data access **policies and procedures** must be defined and followed.
- ✓ Data **stewardship, security and privacy** must be properly managed.



DATASETS

- Stored in various **repositories**
- Supported by different **technologies**
- Available in multiple **formats**

We **can divide the elements of a Data Marketplace into 3 well-differentiated blocks** according to their purpose:

Data Platforms Layer: Technologies where data live and move

- Technologies for data movement and ingestion.
- Technologies for data storage.
- Technologies for data processing.
- Technologies for data consumption and exploitation.
- Technologies for data integration.

Security Layer: Technologies that control access to data at the physical level

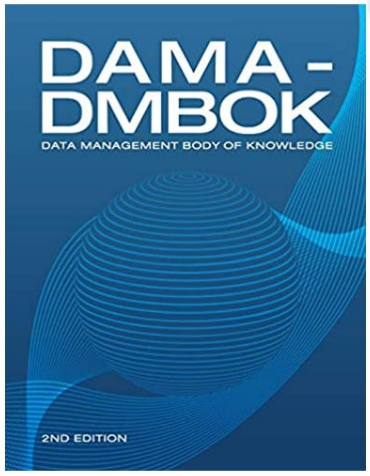
- Identity Management Systems.
- Data security and permissions management systems.

Data Governance and Metadata Management Layer:

- Centralized metadata repository.
- Reference Metadata Management.
- Complete data asset lifecycle management with versioning support.
- Governance model based on data domains, roles and permissions.
- Operational model based on automatable workflows for policy and procedures implementation.
- Data portal with Google-like search engine and advanced filters.
- One-stop-shop integrated with demand management.
- Data access management through Data Sharing Agreements and Data Contracts.
- Automation of common technical processes over Marketplace technologies and data platforms.
- Complete data quality lifecycle management.
- Collaborative, interactive and intuitive environment for non-technical users with messaging, notifications and alerts.
- Global, hybrid and extended view of data lineage and traceability along with knowledge graphs of the data ecosystem.
- Support for both internal and external auditing.
- Continuous monitoring and improvement of both Marketplace processes and data governance implementation.

The first two layers are usually quite well covered by a multitude of technologies in those organizations that want to be Data-Driven but the last layer is the one that is usually left behind and is the one that the Data Marketplace paradigm focuses on, especially with a multi-platform vision and a metadata-centric approach.

***Metadata-centric** approach allows **technical abstraction** in data management and the decoupling between data governance and the technologies for data capture, storage, processing and exploitation thanks to a centralized metadata management*



"An organization can't be Data-Driven without being Metadata-Driven"

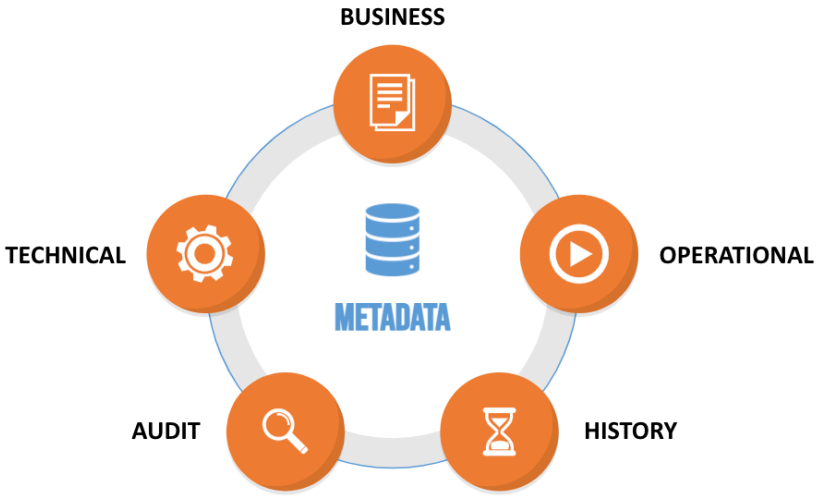
DAMA-DMBOK

DECLARATIVE METADATA

Metadata declared by end-users, passively extracted from technical platforms as built-in configuration or inferred from analytical algorithms

EXECUTION METADATA

Metadata actively obtained after the processes execution over the technical platforms or generated from the users activity and interaction



How to build an internal Data Marketplace with Anjana Data and make the most of it in an organization?

Building a **Data Marketplace** is not an easy task and therefore it is necessary to incorporate **innovative**, **disruptive and differential technological solutions** into the technological stack of an organization, which may offer not only the appropriate features and functionalities but also a series of **value-added capabilities** that allow the organization to make the most of this new paradigm.

In this regard, Anjana Data is the Data Governance solution for the Big Data, Multi Cloud and Data-Driven era, designed to help organizations with the effective and **efficient operationalization of their data strategy t**hanks to the implementation of a **proactive and preventive data governance** with a focus on extended metadata management and automation of common technical processes.

PROACTIVE AND PREVENTIVE DATA GOVERNANCE

- Putting Data Governance at the forefront of **the data value chain**
- **Collaborative** approach empowering different stakeholders
- Integration with **demand management**
- Incremental and iterative approach by **use cases**
- **Democratisation** and governed self-service
- **Monitoring** for continuous improvement
- Abstracting **data management** from technologies and platforms
- **Automation** of common technical processes
- **Integration** with other technologies and pieces
- **Metadata repository** at the heart of the data ecosystem

As we have seen through this article, one of the most important aspects of **proactive and preventive Data Governance** consists of **anticipating needs and making processes more efficient,** based on **business requirements**, which then have to be implemented on the technical layer. In order to **implement this new approach effectively** and obtain the maximum possible **value**, one of the keys lies in the automation of common technical processes through the activation of metadata.

This requires a **bidirectional integration** between the different layers of the Marketplace (Governance Layer, Security Layer, Data Layer) and this is where we come across the concept of **DataOps**, which is becoming more and more relevant nowadays.

PROCESSES AUTOMATION

METADATA AND LINEAGE HARVESTING

Automatic discovery and assisted importation of metadata and technical lineage for a multitude of information systems.

DATA SAMPLING

On-the-fly querying for the presentation of information samples and data statistics.

ACTIVE MANAGEMENT OF PERMISSIONS

Automation in the concession of privileges and the implementation of data access policies.

DATA STRUCTURES MANAGEMENT

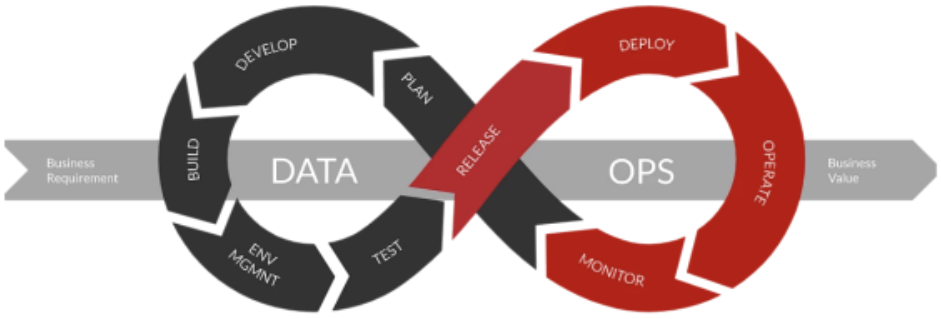
Automation in the creation, modification and deletion of data structures and storage points.

EXTERNAL AUDIT OF PROCESSES

Monitorization of data platform activity logs to obtain dynamic lineage and control data access.

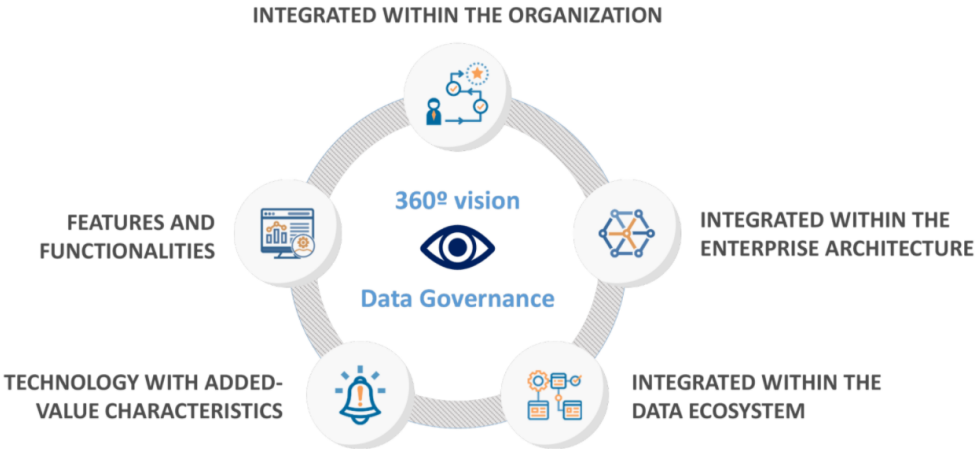
OTHER INTEGRATIONS

Different integrations with platforms, systems and technologies for supporting data management.



DATAOPS NEEDS PROACTIVE AND PREVENTIVE DATA GOVERNANCE WITH PROCESSES AUTOMATION

Starting from the most basic requirements of data governance, Anjana Data allows an organization to gradually evolve to cover in a cross-cutting manner the different requirements of the different stakeholders of a Data Marketplace, and thus become the **backbone of an organization's data ecosystem**.



Thanks to its **differential features and functionalities**, its innovative **vision**, its state-of-the-art **architecture** and a totally disruptive **pricing model**, Anjana Data is positioned as one of the best options currently available in the market to build a Data Marketplace and get the most out of it in an organization.

But, as we know, technology alone is not the solution, so let me give you a bunch of tips to ease the implementation of your Data Marketplace and to achieve success in this initiative:

Open your mind and think about **maximizing business results**, not about the underlying technical processes or only about regulatory compliance (e.g. profit increase, cost reduction, efficiency increase, ...).

Involve the participants and stakeholders of the governance model as soon as possible and involve them in the change, looking for **strategic allies**.

Define a **collaborative governance model** based on roles and data domains, with clear responsibilities and functions for all stakeholders (e.g. data stewards for business areas, data architects as cross-cutting roles, data engineers for technical asset management, ...).

Define **agile and flexible policies and procedures** that can be easily adapted and evolve over time as needed (e.g. data contract management policy, data access request procedure, dataset technical metadata change procedure, ...).

Define a **metamodel** that allows you to bring together all the knowledge and context of your organization's data assets from both a **business and technical** point of view according to **current needs** but also thinking about **future needs** (e.g. reports, metrics, concepts and quality rules for the Business Glossary and repositories, datasets and processes for the Data Catalog).

Define **processes for data lifecycle management** along with metadata enrichment and activation from a centralized metadata repository (e.g. sandbox generation processes for advanced analytics are configured based on metadata from the centralized repository obtained from assets stored in the Data Lake and available data contracts).

Promote **metadata enrichment** by data producers and data providers as well as **feedback** regarding their understanding, comprehension, potential uses and new needs from consumers.

Develop and implement **bi-directional integrations** between the Data Marketplace Governance layer and data platforms to improve the management of data assets from a governance perspective agnostic to data capture, storage, transformation, processing and exploitation technologies (e.g. automation in metadata extraction, active governance for data access permissions management, audit log monitoring, ...).

*Don't be afraid, test, iterate, evolve and look for support along with the **best allies** to achieve the success of the initiative.*

Design a specific **communication plan** and facilitate **change management**.

Start with the implementation for one **use case** and then **scale** to other data domains, systems or projects.

Separate **new production** from **stock** from the data point of view and apply the **governance-first** or **governance-by-design** concepts for all new requirements, uses and projects.

Provide and enhance **employee training** in Data Management practice, Data Literacy and analytical capabilities.

Build a **one-stop-shop** for data integrated with **demand management** that facilitates the day-to-day operations of all stakeholders.

Promote **proactive and preventive data governance** with a 360° view and **best practices** in the field of data management.

Define **taxonomies** for data assets to help with data management (e.g. semantics and ontology, information sensitivity, applicable regulation, ...) and promote their use.

Relate business assets to each other and to technical assets, creating **semantic maps and knowledge graphs** to facilitate their understanding.

Identify high-use and critical data and work on improving its **quality** and on generating logical groupings of data to facilitate its **sharing** in a regulated way.

Promote among data producers and providers the creation of **reusable logical groupings of data** regulated by **standardized contracts** to make them available to potential consumers.

Involve legal areas to ensure **compliance** with **data contracts** and the inclusion of specific clauses for both producers and providers as well as consumers.

Create dashboards, metrics and alerts for **monitoring** the Data Marketplace and its processes with a focus on **continuous improvement**.

Work on the creation of **data valuation** algorithms based on their criticality; their acquisition, maintenance and availability cost; their quality; their different uses; and the risk that their storage and use represents for the organization.

Incorporate to your technological ecosystem **differential, scalable and interoperable solutions** with **multi-platform** and **multi-cloud** vision and that offer the necessary **flexibility** and **adaptability** capabilities to face current and future challenges with guarantees.

CONCLUSION

As we have seen throughout this article, successfully building an internal Data Marketplace for an organization is a major challenge in terms of **people, processes and change management**, but it also may bring a large number of benefits that can make the difference in terms of achieving the goals set by the data strategy.

Finally, **the incorporation of new technology is not the main objective** in this case, but it has to serve as a support and as a means in this whole process of change and evolution of the data paradigm of an organization. However, **having the right technology will allow us to achieve the established goals**, so it must have the necessary capabilities and characteristics to help us in the implementation and operativization of our internal Data Marketplace

info@anjanadata.com | www.anajanadata.com

SEPTEMBER 2021