Data Sharing Agreements

Leveraging **data sharing** between producers and consumers under a governed environment is one of the most relevant outcomes of a data governance program. And, of course, designing and implementing the right framework and ecosystem to achieve this goal is not an easy task.

IT-led silos-based data governance

While in the world of silos, data consumption is mostly under control at some level by system owners (if some governance has been put in place), when evolving to a scenario where data is being treated as a strategic asset, this approach is not viable anymore for various reasons.

Data is not an IT asset

So system owners of departmental systems should not be the only ones controlling data access for consumption even they might have business knowledge. They should manage availability, performance and quality within their systems so they will be involved in data security and data access management, but they have a limited view of the business needs (quality, context, purpose, …) and the whole data map within the organization (golden sources, data reuse, …) so they are not ready to make businessoriented decisions.

New complex added-value architectures

They break silos down by gathering a high amount of data from different sources and domains so people implementing data security within those architectures are usually IT experts with limited knowledge about business needs.

Regulations and normative

(Both internal and external) around data need to be applied not only from an IT perspective but also from a business approach and it has its implications when managing data access for sharing and consumption.

The key element to bridge the gap between business and IT worlds for data sharing and consumption management

Data as an strategic asset

Nowadays, data is being democratized more and more across any organization (both internally and externally) because of the need of **using data to make business decisions** but there are lots of aspects to address when implementing a data governance framework where data ownership and stewardship take special relevance.

From consumer's perspective

- Data context needs to be well understood to get insights when analyzing data, so all metadata and data lineage around datasets should be clearly defined.
- Data quality is very important when using data to make decisions and not every case of use requires the same level of quality.
- SLAs for data must be fulfilled to ensure the right execution of business processes.
- Most of the data consumers do not want to be involved in any non-compliance issue so they will need to know which policies, regulations and normative apply to the data they are consuming.

From producer's perspective

- Data knowledge from producers is important to be shared together with data because it will ease the work for data analysts and scientists when using data.
- Some data need to be prepared, certified, or grouped before exploitation for a proper consumption.
- Data owners and stewards should know who is using the data under their responsibility and for what purposes.
- As part of their functions, data owners and stewards must be aware of data quality issues or performance degradation in SLAs so they may take action to solve any problem related to the data under their responsibility.
- Data producers must ensure that regulations and normative around the data under their responsibility are met so they need to be aware of which policies should be applied to which data.

Consumer's perspective

data analysts
data engineers
data scientists

Producer's perspective

data owners
data stewards
data engineers

Oversight perspective

C-level
Data Office
Audit
Legal
Compliance

From an oversight perspective

- ROI should be maximized in data initiatives which means that data reuse and synergies among areas should be promoted and leveraged.
- Data governance implementation must power efficiency and productivity for people working with data.
- Automation should be a focus to reduce costs and operational risks.
- Data access management homogenization and standardization eases the control and oversight of data sharing and consumption.
- Regulations and normative around data must be met so compliance and audit analysis and assessments should be enhanced.

The irruption of Data Sharing Agreements

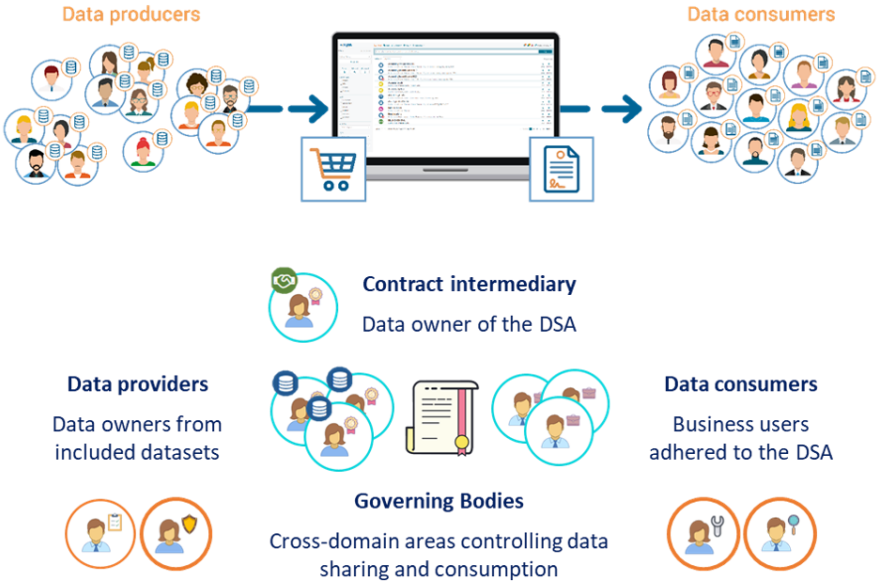
In this context, Data Sharing Agreements are not just related to regulation or compliance. DSAs become the key element to bridge the gap between business and IT worlds for data sharing and consumption management. Moreover, they enable data access management standardization while offering data stakeholders a new experience for information sharing.

But what exactly is a **Data Sharing Agreement**? You may find several definitions but, in this document we mention one which has been implemented in Anjana Data from its very beginning.

> ***In Anjana Data, a DSA is a flexible, adaptable and fully customizable logical group of datasets (understanding datasets as physical data assets – persisted or not, structured or not) designed to establish a global data access model while easing data democratization and sharing.***

A DSA has several characteristics:

- As a logical object, a DSA is technology-agnostic, it may include one or more datasets of any kind (tables, views, topics, files, …) and it may have its own specific metadata, so data producers or data brokers may prepare better-quality data for consumption by grouping it under the same business concept (i.e. customers data for marketing campaigns).
- The DSA will gather all metadata and lineage from the datasets included so data stakeholders will have deep knowledge of which data is being shared from both business and technical perspectives.
- As a data asset, DSAs must have their own procedures and policies for creation, modification, deprecation, and expiration which should be included within the global data governance model and framework, defining the roles and responsibilities for every DSA stakeholder. In Anjana Data, this is natively implemented through fully customizable roles-based validation workflows thanks to our fully integrated BPM.
- Each DSA may have different levels of sensitivity and policies for data access and consumption based on included datasets and an expiration date may be set to automatically revoke data access permissions at some time.
- Every DSA has a Data Contract attached to it specifying the licensing terms and the conditions of use for the data within the DSA. A Data Contract is a multiparty legal agreement between data producers, brokers and consumers including all requirements from every stakeholder of the DSA.
- When asking for permission to access some data, the corresponding Data Contracts must be read and signed by any data consumer so he/she explicitly confirm that is aware of the licensing terms, the conditions of use and the regulations and normative that must be met.
- As all DSAs are being handled as objects within the metamodel where all metadata and relations are stored, versioning of DSAs and active management of impacts are natively supported (i.e. a modification of a dataset may carry the deprecation of all DSAs including that dataset).



Surrounding DSAs for data sharing enhancement

Now that we have explained the nature and importance of DSAs, we are in a good position to take full advantage of them by including some additional concepts and features of Anjana Data which will allow us to unlock all their potential.

Data Portal and Marketplace with Shopping Cart

The Data Portal is the place where all governed Datasets and DSAs will be published along with all their metadata and lineage information to foster **data democratization**.

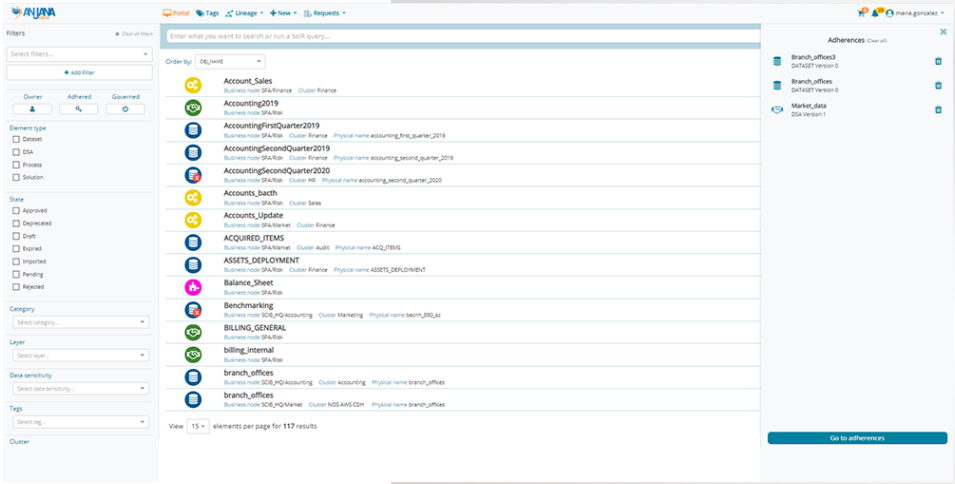
Thus, in this Portal:

- Users will easily find the data under their responsibility and those to which they have access to by using **quick filters**.
- An **advanced Google-like search engine** with dynamic filters is at data consumers' disposal allowing them to look for the data assets they need using both **functional and technical language.**

Based on their permissions, users will have the possibility to review all the details and context about any data assets including data lineage, also allowing navigation between related objects. In addition, users will be able to query a sample of the data assets if allowed by the data owner (dataset fields marked as PII will always be masked when plotted).

Different types of data assets may be included in the Shopping Cart (Datasets, DSAs, Business Terms, …) and the application will recommend to the user different options (including the "most recommended one") for data access requests.

With this new data sharing experience, **the access request process will be standardized** and always be performed using the corresponding available DSAs which is implemented as an adherence request from a data consumer to one or more DSAs. Every data consumer requesting adherence to an existing DSA will be required to read and sign the corresponding **Data Contract** related to the DSA and any data consumer may request the creation of new DSAs if needed so the data producers may assess the requirement to then publish new DSAs for the consumption of specific data under some conditions.



On top of the Data Portal, the Marketplace with Shopping Cart brings to data consumers a whole new experience for information access request as if they were in an Amazon-like e-commerce of data.

Relations between objects within the metamodel

One of the most important characteristics of data assets is that relations between the different data assets will give us really valuable information.

Then, in Anjana Data, relations may be established between any object within the **Business Glossary** and the objects within the **Data Catalog** at any level. Business concepts, reports, metrics, KPIs, …, may be linked to DSAs so data consumers may discover which data assets are ready for consumption when looking for business entities.

Taking advantage of those relations, the corresponding business entities may also be included within the Shopping Cart and the application will identify which DSAs are the best for the **data consumers** to have access to consume the required data assets.

Extended data lineage and impacts management

Data lineage is one of the most heterogeneous aspects of data governance because there are plenty of types of data lineage (business, logical, technical, …). But, any type of data lineage is based on **existing relations within data assets**.

In this context, relations between DSAs, Datasets, business entities, owners, consumers, … are also represented as part of the data lineage map so the users may have a **comprehensive view of the whole data map**. Data consumption by users is being natively included in the data lineage map using a 3D advanced graph-based library so any user may quickly know which users have access to what data through what DSAs.

Furthermore, as we know every relation between objects, **impacts will be actively managed** when declaring a modification in any data asset so the user will know who is going to be impacted from a change because of the related DSAs and all the stakeholders will be notified. Taking advantage of the versioning of the objects, **DSAs may be automatically versioned** when the customizable versioning rules are met after some changes to the related objects.

And, the **deprecation or expiration** of any DSA may be manually enforced, automatically set by date or even configured when some conditions are met (i.e. as an impact derived from a change in a related Dataset).

Customizable roles-based validation workflows with messages and alerts:

Open-source BPPM Activiti has been natively integrated within Anjana Data for **validation workflows management**. Thanks to it, validation workflows are fully flexible and customizable based on the different roles defined within the governance model. Thus, the approval of any validation step or even the entire workflow may be configured to be automated based on the role who has raised the request along with any metadata attributes within the DSA or the related Datasets.

In addition, a **notifications module** is also included within the solution and linked to the BPM so collaboration between the different stakeholders will be enhanced. Both BPM and notifications module may be integrated with external tools through the API and other actions may be configured (i.e. notifications through email).

Active governance for data access management

Moving from a passive governance into a proactive and preventive one is one of the key differentiators of Anjana Data. By the **native integrations with data platforms**, Anjana Data is able to automate the management of data access permissions and data structures for a bunch of technologies.

This is a **change of mindset** about how data governance is usually understood because the organization may take advantage of putting data governance at the front of the data value chain to automate a lot of common technical processes.

With this new approach, data governance turns into the key enabler for real DataOps implementation since everything may be configured to be automated based on the metadata stored within Anjana Data's metamodel.

As a result of this, **low-level fine-grained policies for data access and consumption may be directly enforced** from Anjana Data to the corresponding identity and data access management systems as defined within the DSAs.

Preventive data governance by monitoring users' activity in data platforms

The native integrations from Anjana Data with data platforms extends data governance capabilities not only to actively manage data access and data structures but also to monitor users' activity generating **dynamic lineage**.

This dynamic lineage is being automatically generated by **extracting and interpreting execution audit logs** from different technologies used for data processing within data platforms and then linking it to the specific governed objects as defined in the metamodel.

By analyzing this information, data stakeholders may:

- **Indetify non-governed data assets or processes**: those elements which have not been declared within global Data Catalog.
- **Detect data security breaches**: processes or queries executed by users which have not signed the corresponding data contracts.
- **Anticipate data governance issues**: processes not meeting SLAs, degradation of service, data governance KPIs monitorization, ...

Data sharing operationalization

Understanding and implementing DSAs as the key element to bridge the gap between business and IT worlds for data sharing and consumption management brings a new opportunity to **operationalize data sharing between producers and consumers** while ensuring effective and efficient data governance.

Example: A data scientist working on a new algorithm

- A data scientist is working on a new customer retention algorithm and she needs some historic data about inquiries.
- She goes to the Data Portal and look for the different possible data assets containing this information until she spots some datasets with valuable and good-quality information (based on the metadata information and sample data queried).
- The data scientist put those datasets in the Shopping Cart ready to request access but none of the available DSAs meet her requirements, so she requests the creation of a new DSA specifying her needs.
- The corresponding data stakeholders are notified as defined in the data governance model so they may assess and coordinate the creation of the DSA with the corresponding Data Contract including the licensing terms and the conditions of use.
- The creation of the DSA will follow the corresponding approval workflow as defined in the governance model and, as soon as the new DSA is approved, it will be published in the Marketplace being available for the data scientist.
- The data scientist will be notified and she may request access to the data she needs by requesting adherence to the new DSA. In order to do that, she will be required to read and sign the corresponding Data Contract.

- Once the request is approved, data access permissions will be automatically generated and the corresponding data access policies will be enforced so the data scientist will be able to query the data she needs by using the available technologies for data exploitation.
- During the whole process, every action will be tracked for audit purposes and data governance monitoring.

In the **traditional method**, even having some tools covering parts of the process or automating some steps, all **this process may last weeks or months** and for sure there would be some gaps.

But, taking advantage of the vision and approach of Anjana Data for DSAs implementation and operativization, your organization will be able to reach the next level in its data strategy, allowing the implementation of **real DataOps**, which may reduce the time spent for this process in days, hours or even minutes.

If you are interested in learning more, please visit our website and request a demo at **anjanadata.com/request-a-demo**