Mastering Data Governance with Collibra: A process-based perspective



By Marielle van den Boogaard
& Lars van der Goes

expertum
White Paper



Data Governance with Collibra

White Paper

February 2025
Marielle van den Boogaard & Lars van der Goes

Table of contents

What is the challenge? 3
Managing the challenge....4
Why we believe that Collibra is the solution....5
Not just for data: unifying governance for AI....7
The linking pin with analytics....8

What is the challenge?

With the ever-increasing amount and diversity of data in organizational landscapes, as well as the steady advance of Artificial Intelligence within the broader society, it becomes more and more important for companies to not only have a firm grasp on, but also a deep understanding of their data and data-products. For years, organizations have employed various solutions such as databases, data warehouses, data lakes and various types of reporting and dashboarding solutions to get the most of all the information they gather and store.

In more recent years, organizational data landscapes became even more complex because of the changing nature of that data (e.g. semi-and-unstructured data) as well as its dispersion across multiple, often differently branded, IT systems (e.g. SAP, Databricks, Microsoft, Snowflake, Amazon Web Services (AWS), etc.). These developments often result in data inefficiencies, low transparency, high costs, complex architectures and governance concepts, difficult-to-manage security policies and finally, the creation of shadow IT constructs to even further increase these aforementioned challenges: challenges of which every organization that deals with at least a decent amount of data will surely recognize at least one.

Where does the data come from? Is it validated data? Who is working with what data? What decisions do we make based on the data? The challenges in a data-driven organization, or in an organization on the journey to work more data-driven, can lead to a sense of chaos, of a lack of control.

In the past the advice would often be: fix the basis, bring the foundation in order, start at the beginning. In today's society there are many beginnings and fixing the foundation is a never-ending task that will make your Data- or IT department into a 'department of no', triggering even more 'shadow IT' or Microsoft Excel solutions.

To stop this inefficient circle, a fundamental change of mindset is needed. Organizations need to change from controlling the result, to controlling the process. Change from trying to stop the chaos, to start using these initiatives to improve business commitment for data-driven solutions. Move from fear to trust and allow teams to travel the paths from data to information in the way they think is necessary; not in the way management ideally wants them to: that is the key to an effective and efficient data landscape.

In this White Paper, we will explain from an Expertum point of view how this challenge can be addressed. We will use our years of experience with and knowledge of (SAP) data and analytics solutions to argue the best way forward in a data-driven world that is ever increasing in complexity.

Managing the challenge

Moving from fear to trust is easier said than done. How do you change years of (hidden) competition between business- and IT teams, both claiming the other makes reaching their goals more difficult or sometimes even impossible? How do you get them to embrace the challenges of your organization and not just the challenges of their department? How do you get them to work together?

The answer is transparency. Working in a shared environment with shared information, goals and standards, with shared ownership and responsibilities. Create a safe, collaborative playing field with clear boundaries, but freedom within the boundaries. Agree on security, privacy and quality standards. State them in data- and information principles[1], but stick to a few sentences only, create clear guidelines, allowing employees to make their own decisions and choices within those set boundaries.

Support these initiatives with smart tooling. Use a platform that allows the different teams in your organization to collaborate together. Make them want to use the solution because it will provide what they need. Use your different knowledge bases within your organization to improve business productivity while at the same time build your governance and control platform to stay on top of all those initiatives and ideas. Integrate data-, information- and AI products into one, interconnected overview allowing you to control the process of good data- and information management.

Most importantly: **do not start from scratch**. This is not only an important notice, but also a given fact in many of today's organizations. *Use* your SAP solution, your AWS or your Google analytics. Your organization already has many places where data and information live and co-exist. Do not reinvent the wheel, but onboard all data- and information sets, even though they seemingly contradict each other, and pick up the process from that point.

As a final tip, do not expect it to work instantaneously. Embrace the chaos, pick your battles and work from there. 'Eat the elephant in pieces', use agile work methods to identify and work on the issues with the highest priorities and realize results.

[1] For some good examples of data- and information principles, check Data Governance – How to Design, Deploy and Sustain an Effective Data Governance Program by John Ladley

Why we believe that Collibra is the solution

At Expertum, we wholeheartedly subscribe to the proposition that data governance should be transparent, cohesive and comprehensive, no matter what solutions a customer uses. This is a value that Collibra can provide in the context of various data landscapes, such as an SAP environment. What distinguishes the solution in our view is that with Collibra, you do not have to choose, as you can onboard all your data solutions as input for your data governance.

If you take your SAP data as an example, you can already use the context of your data to quickly set up the first set of (master)data elements. You can use your existing data model to fill the Collibra metadata. You can select the dataset you want to start with from the 'interface', and you can immediately use the format, function and business context to jumpstart your first data quality checks.

Collibra can work with all your datasets. It has many out-of-the-box connectors allowing you to quickly start integrating your data and improving data quality. It also allows you to start small and build up from there, focusing on the creation of **data confidence**; a situation in which every user understands the value of data and when and how it can be utilized in the most efficient way.

In the below graphic, data storage takes place in databases such as SAP HANA, Google Cloud or Azure, whereas the Build-Run-Operate steps can be seen in the context of (cloud) data warehouses, such as SAP Datasphere, Databricks or Snowflake. Collibra's role comes in at the Trust-Comply-Consume steps; where the broad understanding of and confidence in data (and AI), its compliance with (inter)national regulations and secure and governed consumption is paramount. Policies for these areas can be defined as well as rolled out across the entire data chain.



'Collibra's role in the evolving data landscape: Trust, Comply and Consume' – Adapted from Collibra (2024).

In technical terms, Collibra starts by creating a governed environment for organizational data, setting up a (flexible) governance platform and delivering high quality data by understanding where your data lives. What differentiates Collibra here is that it's not just flexible but also designed to be consistent across the organization. On the one hand, organization-wide policies can be easily rolled out and monitored; on the other, departments and lines-of-business can employ exceptions and nuances within this policy to tailor for their specific needs. Hence, we believe Collibra is flexible where it should be and strict where it needs to be: exactly what one would ask from a proper data governance solution.

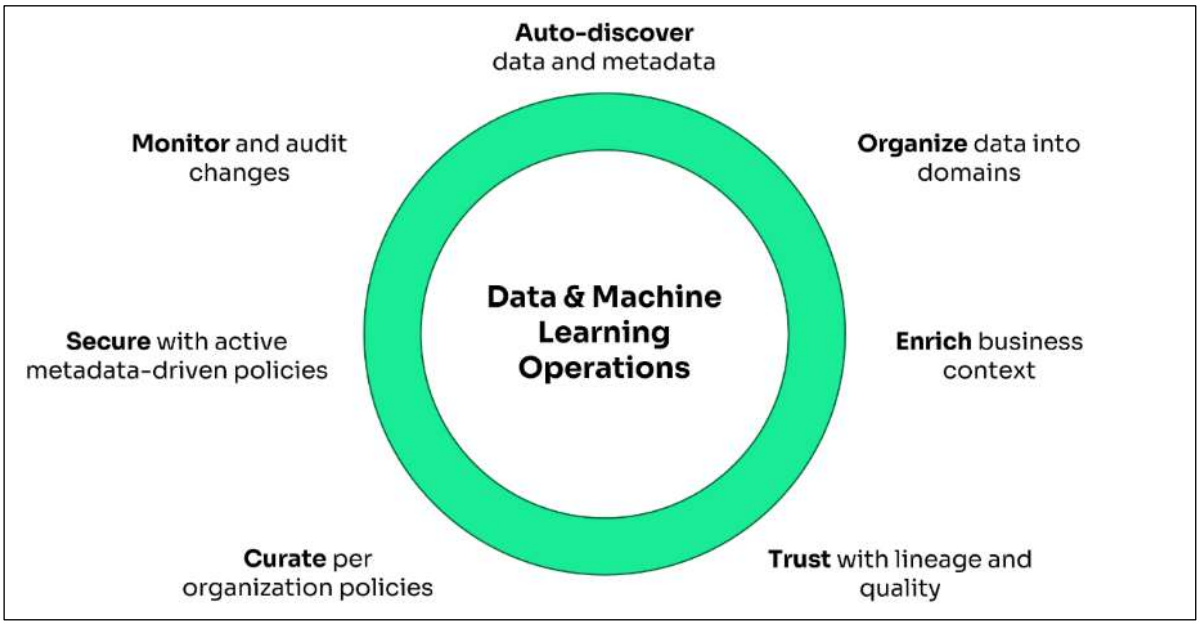
Once this is done, one can continue with improving organizational data by using the automated data lineage, identifying data touch points throughout the organization, showing where to connect business goals and grow as a unified team. After some time, one will have a continuous data quality process, supporting the organization in complying with the many data (privacy) laws a modern organization needs to adhere to, all while delivering high quality, compliant and trusted data to individuals and teams, enabling them to really work in a data-driven manner.

The takeaway here is that chaos will be harnessed into a myriad of data opportunities. Some will fail, some will succeed, but all will be part of a critical, governed data journey.

Not just for data: unifying governance for AI

One of the characteristics that makes Collibra different from alternative data governance tools is that it is not just about the data. Collibra creates collective libraries from data and from your data-products, this is a shared workplace where teams can collaborate using the same resources. It not only helps you to deliver trusted data but also helps you to turn this data into reports, dashboards and (predictive) analytics models. Your reports are stored in a shared library enabling visualization of your data lineage, making your organization faster in adapting changes and improvements. The use of collective libraries stimulate collaboration across teams.

Collibra also allows you to get a grip on **AI** initiatives, extending this newfound control to the most rapid-growing and disrupting, technology on the globe, that is ever-changing the way we as humans interact with each other and the world around us. The EU AI Act has emphasized the requirement for organizations to take their responsibility in this regard. Integrations with various partners such as SAP, AWS, Google Cloud, Snowflake and Databricks allow customers to leverage trustworthy, compliant AI solutions without the risks that might still be inherent to these native environments. Both business and technical stakeholders have a clear overview of and grip on the data and policies that define these AI applications within the Collibra Platform. And, although it might be 'just a marketing note'; Collibra has been marked as a Leader in the 2025 Gartner Magic Quadrant for Data and Analytics Governance Platforms, with only IBM and Informatica being in the same ballpark.



'Why data governance matters in the era of AI' – Adapted from Collibra (2024).

The linking pin with analytics

We have explained how Collibra offers a data governance solution that tailors to data quality, data lineage and AI systems through a unified platform that helps customers conform to (inter)national legislation, including the aforementioned AI Act. How does this interact with the domain of analytics and, more specifically, SAP analytics?

By leveraging unique integrations with SAP solutions such as SAP Datasphere and SAP Analytics Cloud, customers can govern and manage SAP assets within Collibra Platform. Customers can also utilize Collibra to track and trace their data through pipelines of views, tables, flows and source systems, all while keeping sight on the business semantics that accompany it (courtesy of the SAP Datasphere catalog).



The processes and users of the Collibra Platform - Adapted from Collibra (2024).

In practice, this means that users will be able to see what steps their data went through and what transformations were applied to it, all the way from the source to the target system. Moreover, Collibra is able to provide users with a single source of truth regarding their data from an observation, organization and governance point of view, and can be used to share specific data assets within a neatly authorized environment. This setup not only services traditional (technical) BI users (e.g. for data modelling), but also data scientists (AI model training), data engineers (data pipelines) and data analysts (reports and dashboards). Collibra is moving even closer to SAP, with a focus on AI governance and future developments that include support for additional SAP data sources such as S/4HANA, BW/4HANA and more in-depth integration with SAP Analytics Cloud. Of course, SAP's latest **Business Data Cloud** solution is also being integrated with Collibra. More about this collaboration can be read here.

Wrap up

Now that you have a concise picture of the challenges that the modern data landscape presents, and both the approach and solution that we believe you can tackle these with, it is time to take action. If you are ready to start addressing not only today's, but tomorrow's data governance challenges, please do not hesitate to contact us.