Guide to

Agile Data Governance

How to Utilize Agile and Iterative Approaches to Develop Effective and Transparent Data Governance Initiatives.

Passerelle
DATA TO ACTION

talend
A Qlik Company

SqlDBM

ALTR

Why Agile Data Governance?

Nearly 4 out of 5 businesses globally are looking to implement big data management, and for good reason. Data creation is expected to grow to 181 zettabytes in 2025 – a 150% increase over data creation in 2023*.

Growth in data volume is driven by a combination of factors, including:

- Cloud Computing and Storage: As businesses and individuals migrate services to the cloud, the volume of data stored and processed grows.
- Increased Internet Usage: The International Telecommunication Union (ITU) reported that 5.3 billion people were using the Internet in 2022
- about 66% of the world's population.** This figure represents an increase of 24% since 2019. As more people get online, there will be a corresponding increase in the amount of data being created and shared.
- IoT (Internet of Things) Devices: Smart home and IoT devices are driving immense data growth – with more than 14 billion active endpoints in 2022.
- Regulatory Compliance: Compliance initiatives require the creation and retention of extensive records, such as transaction logs, customer information, and audit trails, to show adherence to legal standards and industry regulations. Documentation requirements lead to significant increases in data volume, creating both a repository of historical information and data management challenges.

*Duarte, Fabio. "Amount of Data Created Daily (2023)." *Exploding Topics,* April 3, 2023, https://exploding-topics.com/blog/data-generated-per-day. Accessed Sept. 5, 2023.
**ITU Development Statistics. https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx. Accessed Sept. 5, 2023.

With exponential data growth comes an imperative to keep data secure and ensure it can be used to fuel better marketing, forecasting and operational efficiency. For these reasons, Data Governance has never been more essential to data-intensive industries.

In this Guide, we will outline the pillars of Data Governance. In a metaphorical "Carrot or Stick" scenario, Data Governance is too often viewed through a punitive lens – how to prevent unlawful use and satisfy regulatory requirements. But the rewards of a robust Data Governance posture are equally important – creating a framework to access data at scale and use data as a product throughout your organization.

With this Guide, you will understand the fundamentals of Agile Data Governance. By looking at Data Governance as an iterative process, you can work within a common project delivery mindsight, starting small and building Data Governance into the foundation of your data management practices. We've also provided five questions that can help you identify the first use case for your Agile Data Governance journey.

We hope you enjoy your Guide. If you have any questions, we are happy to help with a complimentary 30- or 60-minute consultation or needs assessment. Request yours at www.passerelledata.com or with the QR code.

Guide to Agile Data Governance

DATA GOVERNANCE PRIMER

Data Governance should be at the heart of every decision involving your data estate. For that to happen, Data Governance shouldn't be viewed as a heavy lift – but as an integrated, strategic component of data management. With a focus on Data Governance, you can unlock the ROI potential of organizational data by enabling employees to best know, trust, and use data in scale to drive revenue and efficiencies.

In this Guide, we explore how to introduce the concept of Agile Data Governance into your organization and provide challenger questions to help identify a first use case. But first, let's talk about what Data Governance is (and what it is not).

Data Governance is typically applied across five areas of focus – Data Availability, Data Quality, Data Security, Data Usability and Data Auditability. While many useful tools help build a Data Governance program, most solutions aim to address one or two of these areas. Data Governance is not a silver bullet – Data Governance can't be solved with one tool or through one initiative. It should be viewed as a holistic concept, and programs should be organically developed based on the needs and structure of an organization.

Data Governance Guiding Principles

At a surface level, these Guiding Principles for Data Governance might look simple – it would be hard to find anyone who would argue that data shouldn't be trusted or accessible. Applying these principles takes intention, discipline and tools that add scalability, observability, and ultimately, sustainability to your Data Governance program.

- Data should be ingested, stored and processed to promote accessibility.

- Data can be trusted, known and measured.

- Data is relevant to business use cases.

- Data is secure, observable, auditable and compliant throughout its lifecycle.

Data should be managed to promote accessibility.

Making data available to the right user at the right time is a critical driver of digital transformation and a central tenet of Data Governance, and to do that, more and more organizations are turning to the cloud. According to a report by Gartner***, most enterprise IT spending will shift to the cloud by 2025. Cloud adoption will only increase as organizations try to keep pace with data availability, cost savings, and scalability.

While cloud adoption will continue to increase, adopters of a cloud–based data management strategy have quickly come to understand that cloud utilization is not enough, especially when it comes to establishing Data Governance programs. To support data accessibility, data–minded organizations should focus on the structure of their data ecosystem, with an eye toward business use cases.

Create a Governed Data Lake

A data lake is a central repository for all data – unlocking data from siloed systems and creating a landing zone for other data applications. Creating a data lake is often seen as the primary goal of a digital transformation initiative; in reality, creating a data lake is just the starting point. Data Governance should address how a data lake is administered and maintained.

For data to be usable within a data lake, care should be taken during the ingestion, integration, and transformation process. A data lake can become a swamp if improperly managed and organized. To prevent muddying the waters of your data lake, a metadata management system should be used
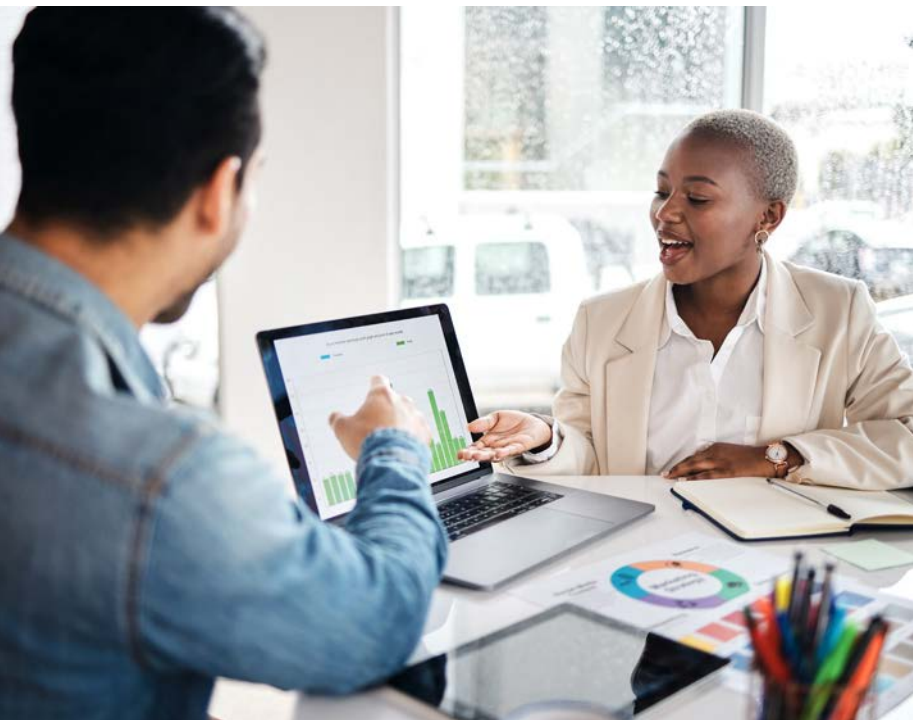
to keep track of data. Regular data quality checks and monitoring should ensure data is accurate and up-to-date. Finally, data should be regularly cataloged, curated and endorsed, so it can be found easily, understood and trusted by anyone in the organization.

Ensure Data is Timely

Functional Data Governance ensures data is as timely as it needs to be. Most organizations don't need real-time data ingestion, but that doesn't mean they don't need to ensure data is timely and current.

Change Data Capture (CDC) helps ensure data in the data lake warehouse is updated as changes are made to the source system. CDC tools track and capture changes made to a database, helping organizations identify and capture only the changes made to data, instead of processing an entire data set. With CDC, organizations don't have to wait for a full data refresh to identify and capture data changes.

Provide Data & Analytics for Business Use

Building robust business analytics for getting key data insights with properly defined KPIs (Key Performance Indicators is a straightforward focal point for emerging Data Governance programs. Analytics highlight the importance of access in Data Governance programs – with robust analytics, data end-users don't have to wait on reports from IT and sift through erroneous or duplicate data to get exactly what they need. Relevant analytics removes human error from data decision-making.

Data is Known

Building a data literacy program can help drive Data Governance and will help identify data owners and stewards, if they are currently unknown. During the establishment of a data literacy program, you will want to identify and engage with relevant stakeholders, including IT personnel, data management teams, business analysts and data users.

Identify what (Metadata) you want to capture with your stakeholders in your business terms/glossary, data dictionary, including data element

names, business terms and definitions, data types, accepted ranges and values and data sources. Your data dictionary will help define how data is maintained and who is responsible for maintaining it. Once your data dictionary is in place, every relevant stakeholder should be educated and have access to the data dictionary. Data literacy builds scalability into your data management by eliminating tribal knowledge and building a common understanding of basic data stewardship.

Data is Trusted

In the last decade, data-centric industries have invested massively in point solutions for online application platforms, customer relationship management (CRM) and fraud detection. While each serves a functional purpose, they also create multiple versions of the truth, which can have far-reaching implications across business lines and operational teams.

These sprawling data ecosystems leave business users with unpalatable options, resulting in incomplete customer pictures, and shadow data marts/data warehouses. There is a better way. To fully leverage the benefits of their data, without sacrificing the agility of self-service reporting, organizations can create a Single Version of the Truth with a well-governed enterprise lake warehouse. With an enterprise lake warehouse, business users across an organization can confidently access trusted data, knowing they are pulling the most relevant and complete data, and getting a complete, 360-degree view of customers and their enterprise operations.

Data is Relevant to Business Use Cases.

Moving data to an enterprise data lake warehouse is just the start of a data estate modernization initiative. To make data usable to different use cases and business users, your data architecture must support integration into various applications.

Data transformation is THE essential bridge from data lake to business use cases, ensuring your data can be trusted, observed, and acted upon as it moves through your organization. Data availability and accessibility, enabled by data transformation, helps support the most critical indicator of a successful Data Governance program – a Data Governance Culture. To promote a Data Governance Culture, champions must exist throughout your organization – from IT to customer service, sales, and senior leadership.

The easiest way to show the value of Data Governance is to make data available for any use case – which often means making it available to third-party data science applications and software platforms. But that's not as easy as it sounds – data applications are diverse, including ML/AI models, analytics dashboards, and marketing platforms. The data required to drive

customer support use cases will be completely different from the data required for laser-focused marketing, or KPI-driven C-suites.

Data integration and transformation are essential to ensuring data consumers have access to relevant and ready-to-use data.

Data is Secure, Observable, Auditable and Compliant throughout its Lifecycle.

It has always been imperative for organizations to keep data secure. But putting data behind a lock and key doesn't make it more valuable. As data ecosystems grow and applications become more diverse, data security, observability and compliance are foundational in data management.

An important first step is to define Data Governance policies and procedures to be applied across your organization. When putting those processes in place, the right technology tools can help automate and scale Data Governance – reducing the burden on your technical team.

How to Build Your Data Governance Posture

Data Governance might be seen as a heavy lift, and even viewed as competition for other valuable resources in IT. It's time to abandon that mindset. Regardless of where you are in your data estate modernization journey, Data Governance should be top of mind.

If you are just getting started, the first step should be establishing a clear understanding of data assets and how they are used throughout the organization. This includes identifying all data sources, classifying data based on sensitivity and importance, and defining ownership. Once you've established a glossary of data in the organization, you can determine how you want to control data – establishing who can access data, how they will access it, and the security, auditability, and visibility measures in place to enforce your Data Governance strategy.

If you have a floundering Data Governance program, take a step back and envision what you could improve. Tools that can automate manual processes help you troubleshoot faster, and ultimately free up resources for higher-value work that pays for itself and helps create a scalable Data Governance framework. Integrated Data Governance keeps data secure while making it accessible and observable throughout the data lifecycle. By integrating your Data Governance program throughout your data catalog, ETL/ELT and data stack, you can build practical workflows and formalized Data Governance roles, setting a strong, scalable foundation for your organization.

When Data Governance initiatives fail or stall out, it is not because they aren't viewed as important. Even with executive and stakeholder buy-in, Data Governance can be seen as too big a lift and might be deprioritized in favor of initiatives with more immediate ROI.

In the first chapter of this Guide, we learned about the guiding principles of Data Governance, and why organizations should look at Data Governance to unlock the potential of organizational data.

But how can organizations without the appetite for a large Data Governance initiative get started? What if the framework to successfully develop a Data Governance posture is right under your nose?

With Agile Data Governance, organizations can utilize the iterative, collaborative approaches that might already be in place for product development and project management for more effective and transparent Data Governance initiatives.

What is Agile Methodology?

First, a primer in Agile methodology.

When the Agile Manifesto**** was created in 2001, it transformed software development and project management from rigid, linear planning to a flexible and iterative approach. Agile has grown beyond software development in the years since, with Agile

****Beck, Kent, et al. "Magile for Agile Software Development." 2001. *https://agilemanifesto.org/*. Accessed Sept. 5, 2023.

principles adopted by teams and organizations of all sizes across various industries. Today, large technology companies like Google, Microsoft, and Spotify, alongside smaller startups and even non-tech companies use Agile methodologies to deliver products and services more efficiently and effectively. Agile is now considered a mainstream approach, and its principles have also influenced other business practices, including project management and organizational leadership.

The Agile Manifesto is comprised of four key values that serve as the philosophical foundation of Agile methodology:

- Individuals and interactions over processes and tools: Agile methodology prioritizes the human element and communication over rigid adherence to tools or processes.

- Working software over comprehensive documentation: Agile methodology aims to develop software that works and meets users' needs, rather than producing extensive documentation.

- Customer collaboration over contract negotiation: Collaboration is seen as more beneficial than negotiating fixed contracts, facilitating a better outcome.

- Responding to change over following a plan: Agile values adaptability and being able to respond to changes, over sticking strictly to a predefined plan.

The Blueprint for Agile Data Governance

It's time to start thinking about Data Governance as an agile and iterative process, with an organically growing and maturing Data Governance program. Rather than a rigid set of rules and procedures, applying Agile methodology to data governance initiatives can help make the process more adaptable and effective.

Start with a Vision, and Iterate

You can create an initial vision for your data governance program but understand this vision will evolve. An Agile approach focuses on iteratively and incrementally refining this vision. Start with high-priority areas and evolve your strategy as you gain more understanding and encounter new challenges.

Create Cross-Functional Data Governance Teams

Data governance is a cross-departmental effort and requires input from various roles: data stewards, data owners, data users, IT, and business stakeholders. Like Agile development teams, these individuals should work

together regularly to ensure a holistic approach to data governance.

Work in Iterations

Break down the data governance program into smaller, manageable objectives. Each iteration should have a clear goal and result in tangible progress toward the overall data governance objectives.

Adopt User Stories

While Agile methodology in software development prioritizes customer success and engagement, Agile Data Governance considers the needs of data users. User stories can be a helpful tool for understanding the needs of different data stakeholders. For example, a user story might be "As a data analyst, I want to have access to clean, consistent sales data so that I can generate accurate reports," or "As the Director of a business line, I need to be able to cross-reference data in my ERP and CRM."

Frequent Delivery and Review

At the end of each iteration, deliver some tangible improvement to the organization's data governance. This could be a new data quality rule, a cleaned dataset, or an improved data governance policy. Then, review progress and adapt future plans based on feedback and results. Regularly reassess and adjust your data governance practices based on the results. Reflect on what's working, what isn't working, and how you can improve.

Leverage Automation

Similar to Agile's emphasis on technical excellence and useful design, use automation in your data governance program to improve efficiency and consistency.

Prioritize Transparency and Collaboration

Encourage active participation and open communication among all stakeholders. This includes clear documentation, shared definitions, crowd-sourcing of business definitions and usage associated with a data asset, open discussions about data quality and security, and collaborative decision-making. While Agile Data Governance will always require buy-in from organizational leadership, it also ensures a bottom-up implementation approach that will ensure stickiness beyond the initial implementation.

Ready to get started? In the next chapter, we will explore how to develop the first use case for Agile Data Governance, including some challenger questions to identify Data Governance priorities and tools that can help accelerate implementation.

The biggest obstacle to establishing a Data Governance posture can be scale. Metaphors about boiling saltwater or eating large mammals aside, most organizations on a Data Governance journey will find the most success by starting with smaller use cases and scaling out. We call this Agile Data Governance.

Can Agile Data Governance work for you? Ask yourself these challenger questions to identify Agile Data Governance opportunities in your organization.

Are there issues or bottlenecks in your current data management process that could be alleviated through improved Data Governance?

As discussed in Chapter 1, one of the primary functions of Data Governance is accessibility – or making data available to the right user at the right time.

Identifying common issues and bottlenecks can pinpoint immediate opportunities for accessibility improvements and produce quick wins. These issues could include inconsistent data definitions, data quality problems, lack of data ownership or accountability, or problems with data access and security problems.

To get started, you should audit your existing infrastructure to determine where data is stored. Taking data out of siloes allows you to create domain-specific data marts - promoting usability and supporting domain-specific data products with a governed

data federation. Data accessibility initiatives can provide immediate ROI – the cost-savings associated with replacing brittle ETL systems often pay for modernization tools. Eliminating multiple data warehouses also improves security and control functions.

Is there a common terminology or data dictionary across the organization?

A lack of common understanding of data can lead to inconsistency and misinterpretation. Establishing a business glossary that provides a common language and interpretation of data across the organization is an easy first use case and will introduce Data Governance to stakeholders throughout your organization. While tools exist to automate and manage data dictionaries, the first iterations of your data dictionary can take place in Excel. Most importantly, you will want to document the following:

- Identify your Data: Understand your data and identify all the data elements within your organization, spanning across different databases and systems.

- Classify your Data: Group related data elements into categories. This might include grouping by data type (e.g., numeric, text), by business function (e.g., sales, marketing), by sensitivity (e.g., public, confidential), or by any other characteristic that is meaningful for your organization.

- Define your Data: Provide a clear, concise definition for each data element. This should be in plain language that is easy for anyone in the organization to understand. Include any business rules or constraints that apply to the data.

- Identify Owners and Stewards: Every data element should have a designated owner and steward responsible for the data's quality and use. Who are the organization's data stewards, and how effectively are they managing data? Establishing clear data stewardship roles can be a use case in itself.

- Develop a Standard Format: Your data dictionary should have a consistent format that is used for all data elements. This might include fields for data name, definition, data type, owner, steward, related data elements, etc.

Is there any department that has started on its Data Governance journey?

Maybe you don't have to reinvent the wheel! The concept of Data Governance has been a hot topic since the early 2010s, thanks to innovations in big data management, increased regulatory pressures, and exponentially growing data volumes. Especially in larger organizations,

there might be business lines that have started Data Governance initiatives you can build from. If you have a successful Data Governance initiative underway, even if it is one department, document a use case to expand on their work, learn from their experiences, and apply it to other areas. For example, if your finance department has created its own data catalog, the rest of the organization could immediately benefit.

Are compliance requirements challenging to meet due to data quality or management issues?

Regulatory compliance can be one of the biggest drivers of Data Governance initiatives. If you are in a regulatory-intensive industry, such as banking or healthcare, you might want to identify a use case around improving Data Governance to meet compliance requirements. For example, if your organization needs banking regulatory compliance, improving Data Governance could help ensure that all personal data is appropriately identified, secured, and managed. First steps in a Data Governance use case for compliance include:

- Data Classification: Classify data based on sensitivity to ensure the right security controls are applied to the right data.

- Data Quality Management: Poor data quality can lead to non-compliance. Good Data Governance helps ensure data accuracy and integrity.

- Establish Data Policies: Create clear, organization-wide policies for how data is handled, stored, and deleted to support compliance with data privacy regulations.

- Access Controls and Auditing: Implement strong access controls and keep audit trails to ensure only authorized individuals can access sensitive data, which many regulations require.

- Eliminate Manual Processes: Manual data processes in regulatory reporting can be inefficient and error-prone. Automated reporting will decrease the burden of regulatory reporting and let your data team focus on activities that add value to your organization.

Are you maximizing the use of your metadata?

A Data Governance use case could involve implementing a metadata management solution if current management is poor or non-existent. This will help users understand where data comes from, how it changes over time, and how it is connected to other data.

With metadata management you can:

- Understand Your Data: Metadata, or "data about data," provides essential context for understanding what data represents, its purpose, source, format, and relationship to other data.

- Improve Data Quality: Metadata can track when data was last updated, by whom, and if any validation rules were applied, which can help assess its accuracy and reliability.

- Track Data Lineage: Metadata helps trace the journey of data from source to current form. Observable data lineage is crucial for troubleshooting, impact analysis, and ensuring the integrity of data over time.

- Maintain Security and Compliance: Metadata can provide information about data sensitivity, helping to enforce appropriate security measures and ensuring regulatory compliance. For instance, metadata supports privacy regulations by identifying personally identifiable information (PII).

- Foster Accessibility: Metadata makes data more easily searchable and accessible, ensuring the right data can be found and used when needed.

- Promote Integration: By providing context and standardizing descriptions, metadata helps integrate data from different sources, making it easier to share, collaborate, and gain insights across various data sets.

While no technology should be viewed as cure-all when establishing your Data Governance posture, most organizations can benefit from tools that make it easier to implement and adopt Data Governance principles. This Agile Data Governance toolkit includes tools that facilitate collaboration and automation of critical tasks.

Data Rocket

Data Rocket is an automation framework driven by metadata, built in Snowflake using Talend's Data Fabric. With Data Rocket, you can easily assemble any data type from any source system, such as Relational database system (RDBMS, Flat files, Semi-structured data files (such as XML, JSON, etc. and API data service. Data Rocket's Governed Dynamic Ingestion ELT Process, with data pipeline observability dashboards, provides a resilient, replicable and automated way to move data to the Snowflake Data Cloud.

Built on Talend and Snowflake

Talend

Talend supports Agile Data Governance with a comprehensive end-to-end data ETL/ELT platform that can move any data from any source and offers robust data cleansing, transformation, and

integration capabilities. Talend's data management guarantees the reliability and accuracy of data, equipping businesses with the tools they need to drive data-driven decisions confidently.

Snowflake

Snowflake seamlessly integrates the flexibility of a data lake with the structured utility of a data warehouse, offering organizations a best in breed cloud data platform. Snowflake's stringent security protocols ensure that sensitive data is protected while still readily accessible to authorized users. This accessibility, coupled with Snowflake's dynamic metadata management and real-time data-sharing capabilities, allows for a nimble response to changing business needs without compromising data governance standards. The result is a versatile data infrastructure where quality, security, and agility coexist, empowering businesses to make swift, informed decisions.

Data Rocket Accelerators

Automation and acceleration tools build Agile Data Governance into your data stack. These components include:

- Data Quality Watch – profile and measure quality of data based on integrity, timeliness, consistency, completeness and popularity.

- Audit and Control Framework – enable targeted, expedited troubleshooting and decreased business outage time with historic and real-time data ingestion information within Snowflake.

- Mastered Data Framework – eliminate duplicate data to create a Golden Record in your organization.

- Snowflake Watch – quickly gauge key volumetric stats with customizable visual dashboards, and drill down to the query- and user-level for deeper insights.

ALTR

ALTR is an automated solution for controlling data access and ensuring security, specifically tailored for safeguarding sensitive data from source to cloud to data users. It enables organizations of all sizes to quickly

distribute more data to the appropriate users, with enhanced security measures. ALTR's SaaS interface offers a Shift Left® approach to data access governance and data security – allowing initiation of robust data governance and data security capabilities to occur at the source and ensuring policies remain attached to the workload throughout the data journey to the cloud.

ALTR offers a scalable data governance approach, unparalleled observability, and patented data protection mechanisms. With interactive usage heatmaps and comprehensive drill-down dashboards, organizations gain insight into data usage patterns, including who is accessing the data, when it's being accessed, and how frequently. These capabilities allow organizations to send alerts, slow, or stop abnormal data access attacks in real-time.

All of ALTR's features are available on Snowflake Partner Connect- start for free and upgrade for more databases and users when you're ready.

SqlDBM

Effective metadata management is the cornerstone of organized and reliable data handling within any organization. A well-constructed data model acts as a guiding blueprint, outlining how data should be structured, documented, and utilized. This structured approach not only fortifies various facets of data governance but also upholds the integrity and trustworthiness of the data itself.

SqlDBM is a cloud-based and model-driven metadata management platform. Tailored for the Snowflake Data Cloud environment, it empowers your entire team, including database developers, architects, engineers, project managers, and data stewards, to seamlessly collaborate on data modeling, governance, and monitoring. With just a web browser, you can collectively analyze and visualize a project from inception to completion, ensuring clarity and ease of collaboration.

Data Catalog Tools

Modern data catalog tools can support Agile Data Governance with automated metadata management and discovery. Data catalog tools streamline operations and build trust in your data by making data definitions, lineage, and usage searchable.

The right data catalog tool for you will depend on specific requirements, existing architecture, and future scalability needs. With a data catalog tool, you can expect some or all of the following features:

- Collaboration: Many modern data catalog tools support real-time

collaboration, where users from across your organization can annotate, discuss, and share insights about data. Collaboration features can streamline data stewardship roles and improve Data as a Team Sport initiatives.

- Data Lineage: Data catalog tools should support robust data lineage capabilities and provide clear insights into how data moves and transforms within an organization.

- Automation: Look for automation tools that can simplify tasks like data discovery, metadata extraction, and lineage visualization.

- AI Integration: As generative AI becomes more commonplace, data catalog tools will continue to build on machine learning to speed up data discovery, making recommendations based on user behavior and usage patterns.

Assembling the right toolkit doesn't have to be intimidating. If you're ready to get started, Passerelle can help you assemble the Agile Data Governance toolkit that's right for you, identify your first use case, or optimize the tools you currently use. Contact us for a complimentary consultation.

www.passerelledata.com