Alation

Automating Data Governance Workflows: A step-by-step guide

Introduction

Your organization has had "too much data" for decades. The rise of AI, generative AI, AI-driven automation, and other promising technologies is making those days seem quaint.

For organizations of all sizes, the need for good data is growing with volumes: exponentially. Generating the data is easy; it's the "good" part that separates the leaders from the laggards.

Simply having access to a large quantity of data is not enough. It's crucial that organizations focus on quality over quantity when it comes to their data.

This is where best practices come in. A good process can become a great one when it is automated properly. But as the saying goes, "a bad process automated is a really bad process." Don't be tempted to automate every aspect of governance without first evaluating its effectiveness and efficiency.

This guide is designed for data leaders and engineers exploring how they might implement automation to ease data management tasks, particularly as they relate to data governance and stewardship. We'll begin by contextualizing the need for data stewardship today, share examples and benefits of data governance automation, and conclude with a step-by-step guide for implementing a governance framework, as well as example automation use cases. Let's dive in!

The need for data stewardship is increasing

Data stewards are agents of data. They take accountability for curating data in an organization. It's less a role and more of a responsibility since data stewards can be subject matter experts, data analysts, data experts, or in other positions. They just need to understand the data they steward.

Data stewards are becoming increasingly critical to data cultures and data-driven organizations. Now, with the explosion of AI and its insatiable need for good, clean data, the need for data stewards has grown, too. Successful AI demands quality data. Garbage in will result in garbage out—and the failure of AI initiatives. Those who steward data well will succeed. Increasingly, organizations are leveraging automation to ease the burden of stewardship tasks.

Why is curated data so critical?

Gartner predicted a few years ago that by 2025, 95% of decisions using data would be at least partially automated. More recent surveys find that 79% of corporate strategists say AI and automation will be critical to success. Without good data, those goals are unattainable.

But good data requires data stewards who are, in fact, human. Scaling stewardship is a decidedly 1:1 relationship: More stewardship and curation require more stewards. That's because stewardship is tedious and time-consuming.

The pain of manual data stewardship

Data stewards manage data curation and maintain descriptions to ensure data is used correctly and complies with data governance policies. Stewards put data governance into practice and rely on solutions like the Alation Data Intelligence Platform as the central access point, where people can find policies that reference how data should be used.

But it's still a decidedly manual effort to check for missing data titles, update security classifications, manage upcoming policy expirations, and more. Those manual efforts not only lead to expensive manual errors, but they are also inefficient and sap productivity (and morale) from these agents of data who often have full-time positions and have been given data steward tasks as an extra responsibility

These manual, repetitive tasks can reduce the data steward efficiency, risk data going uncurated and misunderstood, and reduce the time available for higher-value, strategic work. Other risks include:

Risk	Pain
Time	Decreased productivity from manual monitoring of metadata completeness, compliance, content, and changes.
Cost	Higher costs from increased staffing needs and the likelihood of errors caused by overworked data stewards.
Accuracy	Reduced accuracy from inconsistent data handling that compromises downstream decisions and confidence.
Scale	Hampered scalability as data and workloads grow and manual processes become increasingly difficult to scale in tandem.

Stewards aren't scalable

Exponential data growth demands a scaling of stewardship powers. Throwing more people at the problem by hiring more data stewards is cost-prohibitive, especially as data volumes grow faster than people can effectively manage. Of course, the alternative — risking noncompliance by underfunding stewardship — will likely become expensive when it results in fines or governmental penalties.

The Cost of Non-Compliance

Non-compliance events cost an average of just under $6 million and GDPR fines begin at $11 million.

Average Cost of Compliance —————— $5.5 million

Average Cost of Non-Compliance —————— $14.8 million

GDPR Fines —————————————— >$11 million

Cost of a Single Non-Compliance Event —— $5.9 million

Source: Colligo, "The True Cost of Non-Compliance"

The challenge will only get worse. In a PwC survey, 78% of executives say their company has moved to the cloud in most or all parts of the organization. While the cloud adds speed and scale, it also brings complexity, near-infinite data volumes, and data flowing in from hundreds or thousands of sources. This is a much more challenging data estate to govern!

Add up the fast-moving advancements in data, systems, clouds, AI, and other data-related technologies and how fast they're advancing and generating new data and it's clear that human data stewards cannot keep up.

Automating data governance

Automation helps humans do more faster, and with greater accuracy. It also lowers costs, increases competitiveness, improves productivity, and increases employee and customer satisfaction. Combined with AI, [automation is expected to add trillions of dollars to the global economy](#).

Automation can also bring a world of efficiency to data management. Automating core data lifecycle processes removes slow manual work from humans, eliminating inefficiencies, giving people more time to work on higher-cognition activities, and reducing burnout.

For example, **compliance monitoring**, while critical to the safety and effectiveness of business processes, is often repetitive and time-consuming. Automation can significantly streamline these processes to improve compliance accuracy and coverage for fewer violations and more robust risk management. The results are cost savings across a resource reduction and potentially reduced fines or penalties.

Data quality monitoring is another example. Automation can instantly notify data stewards of potential issues within their purview to increase the scale and offer more effective coverage across a growing and diverse data landscape. Data stewards then spend less time fire-fighting problems late in the data cycle, giving users and developers more confidence in data quality.

Benefits of data governance automation

Simply by alleviating some of the manual effort required of data stewards, automation accelerates, decreases the cost of, and reduces risks of data governance.

Time	Cost	Compliance
Get appropriate, trusted, accurate data into decision-makers' hands faster to increase data's time-to-value.	Reduce the cost of data stewardship for direct financial benefits.	Improve compliance as the basis for better governance efforts.
- Identify critical gaps that could hamper decision-making efforts and prioritize them. - Guide data stewards to these more critical curation needs. - Improve processes to build policies and glossaries to build data users' confidence. - Manage policy and glossary review cycles, alert reviewers to past-due tasks, and remove reviews as a speed bump in the time-to-value path.	- Enable more curation and governance impact with fewer resources to manage growing data volumes and sources. - Drive increases in curation accuracy and quality by eliminating human errors and reducing tedious tasks that invite burnout.	- Help data stewards stay current with changing regulations and ensure policies are updated. - Monitor usage and alert data stewards to possible noncompliance issues. - Streamline data classification efforts to highlight missing or out-of-date classifications. - Block non-compliant data usage that could lead to penalties and fines.

A step-by-step guide for data governance

You're likely well aware of the need for data governance, and have established some level of active data governance in your organization. It's a continuous process to improve and scale those efforts, especially as new technologies like AI and generative AI increase the need and needed speed for data.

To help guide your journey, we've created a seven-phase cycle for active data governance that enables you to strive for continuous improvement. It's designed to ensure data stewards can focus on critical asset reviews and classifications, reduce tedious manual tasks, improve data quality, and support ongoing monitoring, measurement, and optimization of data governance efforts.

How Alation Powers Active Data Governance

Monitor & Measure

- Determine policy conformance
- Monitor curation analysis
- Measure usage & asset creation
- Continually assess data quality
- Run data policy audit reports

Drive Community Collaboration

- Promote trusted data use
- Leverage community knowledge
- Help address needs/questions with conversations
- Determine whether data is fit for purpose / intended use

Apply Policies & Controls

- Implement policies & controls
- Generate appropriate lineage charts
- Build data quality rules & reports
- Enable workflow for data objects, glossary terms, and policies

Populate Data Catalog

- Ingest metadata (technical, business, lineage)
- Analyze metadata (top users, popular data)
- Determine privacy details - data categories & characteristics

Establish Governance Framework

- Set vision & mission
- Organize people for data governance

Empower Data Stewards

- Recognize and assign stewards
- Automate stewardship processes
- Identify reviewers & workflow approvers
- Enable peer review of data assets

Curate Assets

- Document data quality rules
- Describe data, apply trust flags
- Surface descriptions, quality, etc. to users at point of data use
- Establish data policies, standards, and terms (glossaries)
- Curate data categories & characteristics
- Document record of processing activities

Continuous Improvement

A step-by-step guide for data governance

Step #1: Populate your data catalog and kick off curation

Get started by understanding the basics of your data.

- Ingest metadata (technical, business, lineage).

- Determine privacy details by asking which data categories and characteristics need to be masked and secured. Local regulations and industry standards should inform this.

- Analyze metadata to identify popular data and top users.

Step #2: Establish a governance framework

Before you get into the weeds of data documentation and curation, convene your council of data leaders and together align on your vision for data governance. What is the purpose of data for your enterprise?

- Document your data governance vision, mission, and team

- Disseminate your plan to the broader organization

Step #3: Assemble your team (with help from ML and AI)

Identify potential data stewards while leaning on technology to streamline and automate processes. Steps #2 and #3 can be run simultaneously.

- Recognize and assign data stewards based on data usage.

- Automate data stewardship processes to efficiently align expertise with activities, reduce resource burden, and increase the quality time available for curation (see next section for example, bot use cases).

- Identify reviewers and workflow approvers (supporting players).

- Enable peer review of data assets with help from bots.

A step-by-step guide for data governance

Step #4: Curate assets

Leverage your team's brain trust to add wisdom to the most used assets while adding key rules, business and otherwise, to the platform.

- Describe data and its top uses and apply trust flags for broader consumption

- Establish data policies, standards, and terms. Document data quality and business rules for company-wide reference and guidance

- Document a record of processing activities

Step #5: Determine, define, and apply your automation use cases

With your team of data stewards and bots in place, it's time to put them to work.

Completeness: Use automation to ensure new data in the catalog is complete:

- Define completeness: an asset needs a title, description, and domain

- The governance council determines who is responsible for each facet and who reviews/approves each facet

- IT leaders automate the bots to trigger notifications to the parties responsible for each step (adding new facets or reviewing & approving those additions)

A step-by-step guide for data governance

Compliance: Use automation to ensure new data in the catalog is compliant (for example, that private data is masked).

- Define privacy compliance using the privacy details identified in step #1.

- Create a rule that, for every asset added to the catalog, AI and ML will flag whether private data is included.

- Use bots to check assets with private data to ensure a business policy is attached. If not, this is a compliance risk.

- If a data policy is not attached, kick off an automation workflow for bots to either:

- Notify the data steward when there's an asset needing their attention or

- Automatically attach a business policy.

Step #6: Drive community collaboration and ongoing curation

You're now in the second phase of data governance automation, which begins to run independently. Here are some tips for maintaining the momentum:

- Promote trusted data use.

- Leverage the data catalog's conversations feature to communicate and promote data asset use cases and best practices.

- Document and update data quality rules as they shift, and leverage automation to create approval workflows to safeguard data quality.

- Create and share glossaries with details on standards, terms, and policies.

A step-by-step guide for data governance

Step #7: Monitor, measure, and optimize

Keep an eye on progress and execution to continuously improve your data governance efforts and impact. These best practices can help:

- Leverage consumption tracking features to track usage and rate of policy conformance.

- Monitor curation analysis.

- Measure data usage and asset creation.

- Continually assess data quality. Integrate data quality and observability tools in your data intelligence platform to make health metrics visible to your downstream consumers.

- Run data policy audit reports and share the results.

- Work with IT to build workflow automation that eases stewardship tasks.

Automation bots: Example use cases

Collecting data is easy. Curating it is another story. Automation bots in your data intelligence platform help ensure that new data is complete, with the requisite details. These example use cases illustrate how automation workflows leverage AI and ML bots to keep the right humans in the loop and the incoming flow of data properly curated.

Automation bots: Example use cases

Data creator

Rule: The custom field "data creator" must be populated within 5 days of a new table's publication.

Process:

- After five days pass with the creator field going unpopulated, a notification is sent to a mailing list of data creators requesting them to assign one

- After ten days pass, a follow-up notification sends to data creators and management for escalation

- Optional: A quarterly BI report automatically generates a demonstration of all tables with required fields empty after 10 days.

Data supervisor

Rule: Within two days of a data publisher being assigned, the "data supervisor" custom field must be populated.

Process:

- If two days pass and a supervisor still has not been assigned, a notification is sent to that table's data publisher to let them know a supervisor is required.

- This notification can be sent every two days until ten days have passed.

- If ten days have passed and a supervisor is still not assigned, it will be visible in a report to the platform team to manually escalate.

Get proactive on active data governance

Accurate, high-quality data has always been critical to decision-making confidence and impactful business outcomes. New technologies are also relying on your data to influence more strategic business decisions. Ensuring data is effectively and quickly curated gives your organization a unique competitive advantage, as it empowers a broader range of data users to make more informed decisions.

However, most organizations today have too much data for data stewards to manage. As data volumes continue to explode, human data stewards simply can't scale at the same pace. This risks cost, accuracy, and business momentum, not to mention potential fines and penalties from regulatory and privacy non-compliance.

Automation is here to help improve and scale data governance while keeping humans in the loop to ensure data quality is maintained and inadequate processes are eliminated. A data intelligence platform can serve as the foundation by automating many slow, error-prone manual tasks so data stewards can focus on what's most important to your organization.

Using the steps above, your organization can move forward, learn, and iterate to create an active, scalable data governance. Learn how Alation can help with our Workflow Automation solution.