Enabling Governance for Insight

Trust in Data with Tableau
and Alation's Data Catalog

Today organizations are struggling with managing their data under competing demands. The competitive advantage that comes from giving broad data access to an organization drives the widespread adoption of self-service tools like Tableau. While this empowers business users, it has made the role of IT departments and data stewards that much more critical in ensuring the accuracy of insights in a self-service environment.

Thankfully, it is possible to strike this balance of governing data without losing the agility and speed of self-service analysis. Tableau is often used to both create data visualizations and trusted data sources, two critical types of data assets that can be shared, governed and managed through a central Tableau Server. By combining this power of Tableau with a data catalog like Alation to proactively recommend data best practices within the workflow of users, organizations get the best of both worlds. This modern approach to data governance is differentiated by delivering guidance where it is most impactful.

When analysts and data scientists have access to data governance policies and best practices directly within the flow of their analysis, the result is both more consistent compliance and more broadly adopted best practices. Data governance becomes consistent and reliable, driving compliant and accurate insights. With the combination of Tableau and Alation, organizations can not only balance the demands of agility and governance, they can actually optimize for both at the same time.

The Data Decision Effect

The most successful data-driven organizations fully embrace self-service workflows. With self-service data access, consumers of varying levels of technical sophistication are able to easily find or create their own data assets and draw their own conclusions. The advantage

of this approach is that business managers are more apt to base decisions on actual data rather than their own intuition. Hands-on experience with the data leads to data-driven decisions.

Tableau provides data analysts and line-of-business practitioners with a self-service tool for data analysis and visualization, allowing users to draw their own conclusions from the data. The good news is that data is accessible, approachable and immediately useful.

But with data accessibility comes greater risk and ever-increasing complexity. Organizations embracing self-service analytics need to successfully manage the sheer number of reports that are being generated. The single source of truth is no longer stored in a single, physical data warehouse, but spread throughout a set of distributed systems.

From an IT Management perspective, this challenge is often labeled as a free-for all, the wild west of big data. And IT teams have a point to be made, when business discovery takes over without governance it often results in data or insights that can't completely be trusted. So how to tame this wild west without destroying the spirit of self-service data?

Traditional Data Governance Approaches

Traditional data governance, which has primarily focused on compliance with government regulations, is typically a manual process. A Data Governance Council is created with the responsibility of manually logging all data assets in a centralized inventory of data. IT employees are required to document every data asset in that inventory and formulate data policies that protect the organization from risk. These policies often restrict data access with a heavy hand in order to minimize the risk of non-compliance. While useful for regulatory purposes, what gets lost in this approach is the documentation of best practices that would increase appropriate access to data.

Users of data in this framework of *Governance-for-Compliance* are often asked to memorize the rules, take a test and self-enforce compliance with top-down policies. The result? It is not always, but often inconsistent and incomplete adherence to policies. A situation emerges that is like taking a driver's license exam and then being expected to navigate city roads without stop signs or stop lights. Analysts are told effectively to just remember the "rules of the road" and self-police.

Perhaps even more concerning to organizations, the impact on business users of a pure *Governance for Compliance* approach can be a self-sustaining perception that using data is complex, fraught with risk and perhaps not worth the effort.

Data Governance for Self-Service

When data governance becomes a barrier to using data, all parts of the organization lose. IT teams see a reduction in the use of the data assets that they have spent time and resources building. Business teams see a drop in data-driven decision making. And the organization loses the competitive advantage of key insights.

In a Governance for Insight framework, you need not choose between compliance or usage. Instead, you promote both best practices and adhering to the rules of using data consistently in the organization. These policies for data usage are inserted directly into the self-service visualization environment as interventions in a user's workflow. By bringing the communication of policies and best practices directly into the user's workflow, you make it easy to ensure greater accuracy and compliance while still allowing for a free-flowing approach to data visualization.

Government regulation prohibits correlating gender and income in the same workbook? A recommendation notification during workbook creation is pushed to the user communicating to them that data policy and explaining why adherence is critical.

This column in the database includes a deprecated metric for revenue? A deprecation notification for that column is shared with the workbook owner to indicate to the user that they may want to remove that column from their visualization.

This approach to data governance delivers guidance where it is most impactful -- directly within analyst workflows. The result is both more consistent compliance and more broadly shared best practices.

Governance for Insight requires the engagement of the users of data. Data analysts, data scientists and data engineers must be engaged as stewards of their own data.

Alation's Governance for Insight Framework

Alation has been built from the ground-up for Governance for Insight. Alation integrates with and compliments Tableau's existing data governance capabilities so that organizations can truly balance the demands of agility and compliance, achieving both at the same time.

Alation connects to Tableau Server, slotting easily into an enterprise Tableau environment. Alation 4.0 was built with the support of Tableau's product and development teams to support Governance for Insight within Tableau in the following ways:

- Alation natively supports data governance processes in Tableau through Tableau Projects. Data stewards and data governance teams can use Alation to publish and automatically update the association of Tableau Data Sources with data governance approved Tableau Projects. This ensures a managed and consistent workflow for approving and updating data assets for use in Tableau.

- Data Stewards using Alation with Tableau can use Tags to communicate the context of data governance policies and best practices directly to data consumers in Tableau. Fields indicating the data policies applied and level of trust and validation of any data source can be maintained in the Alation Data Catalog. Tags are automatically synchronized between Alation and the Tableau Server so that they are immediately available to Tableau users as filtering criteria for finding and understanding Tableau Workbooks and Data Sources in the Tableau application workspace.

- Alation automatically delivers deep data impact analysis. Alation makes Tableau deeply data-pipeline aware by notifying Tableau data consumers by email when underlying changes in the source database or file structure has potentially affected the accuracy of their analysis. Consider a "profit" field in a Tableau Workbook. If a defining field in the database table or Hive on Hadoop structure is changed and the "profit" field is no longer valid, it will be flagged in Alation as a deprecated column. This notification is automatically pushed to all Tableau users who have used that column in a Data Source, Project or Workbook.

- Alation also enables Tableau users to search for their data across highly distributed big data environments through the Alation Data Catalog and generate workbooks directly from Alation. Along with samples of every data set available across systems, this single point of reference provides detailed descriptions of the data and its uses stewarded by both expert curators and the users of the data. This increases the ease of access for self-service analysis to a wide variety of database and file storage systems.

Alation & Tableau in Action: An Internet Analysis Story

For one large internet company, it was important to separate the wheat from the chaff. Not knowing which source of data to trust was a problem for GoDaddy. The domain name provider initially started using Tableau as their primary BI analytics tool in 2013.

With thousands of workbooks that had evolved organically, employees weren't always sure where to put their trust.

In Tableau Data Sources, there was a single source of truth containing all the data structures for analysis, but GoDaddy wanted to implement a reliable way that the IT department could communicate their seal of approval on critical data sources. This kind of governance was needed in order to deliver a completely scalable enterprise solution.

Today GoDaddy manages the work of over 800 Tableau users in Alation, curating Tableau assets for communication, sharing and re-use throughout their environment.

"

Tableau has been instrumental in allowing more of our teams to get hands-on with data. However, as tens of Tableau users turned into almost one thousand, a lack of formalized data catalog or data glossary made it difficult for any user to understand the appropriate usage and intent of the data. **Using Alation with Tableau enabled our users to understand the context and lineage of our data and ensure the accuracy of their analysis.** With the Alation data catalog governing our data in Tableau, we can confidently scale our data initiatives, and use the information to provide better experiences for our customers.

Sharon Graves,
BI Applications Admin,
GoDaddy

Establishing Trust in Data

It is possible to create an environment where risk is greatly minimized and analysts are given freedom to act on the data they have. Business users and data analysts can discover trusted assets and evaluate the level of trust for any asset they find.

Here are the five simple steps for setting up a Governance-for-Insight Framework with Alation in Tableau:

1. Create a Tableau Project. Create and name a Tableau Project - "Alation Certified". This will be a Project to which only the Alation application's user-id can add items. Similarly create Projects called "Alation Warned" Alation Deprecated" and "Alation Un-evaluated"

2. Set-up custom groups and permissions so that any table's flags in Alation, which indicate Endorsements, Deprecations, and Warnings, can be edited only by a given table's Data Stewards. Ensure that the Data Steward field in Alation can be edited only by the central Data Governance group.

Custom Field	Viewable By	✏ Editable By
📅 Approval Date	*Everyone* ➕ Add Rule	👤 **Stewards** people set ➕ Add Rule
{ } Policies	*Everyone* ➕ Add Rule	*Everyone* ➕ Add Rule
👥 Stewards	*Everyone* ➕ Add Rule	👤 **Governance Team** group ➕ Add Rule
☰ Approval Status	*Everyone* ➕ Add Rule	👤 **Stewards** people set ➕ Add Rule
☰ Domain	*Everyone* ➕ Add Rule	*Everyone* ➕ Add Rule
☰ Quality Score	*Everyone* ➕ Add Rule	👤 **Stewards** people set ➕ Add Rule
A Notes	*Everyone* ➕ Add Rule	*Everyone* ➕ Add Rule

Builtin Field	✏ Editable By
A Table Title	*Everyone* ➕ Add Rule
A Table Description	*Everyone* ➕ Add Rule
{ } Table Flags	👤 **Stewards** people set ➕ Add Rule
{ } Table Tags	*Everyone* ➕ Add Rule

Custom Permissions set such that a central governance group can specify a set of Stewards per table who become the only individuals able to set that table's Approval Status, Approval Date, Quality Score, or Flags (Endorsements, Warnings, and Deprecations).

alation.com

3. Create trustworthy sources using Alation Compose or a CREATE VIEW statement in any other SQL tool.

    -   If using Alation Compose, write a SQL query joining the appropriate source tables and selecting the appropriate columns (combined with the appropriate functions) or

    -   Use the Alation Data Catalog to find the relevant source tables and then construct and run a CREATE VIEW statement using another query tool. The Alation Data Catalog will automatically detect this new table during its next scheduled extraction

4. Publish the view as a Tableau Data Source. You can publish the view as an "Alation Certified" Project through the Alation UI.

5. Extract data. (Alation's automatic extraction will produce a page in Catalog for the Tableau Data Source and parse its lineage all the way back through the underlying tables). Alation can be configured to automatically re-categorize Tableau Data Sources into different projects if issues are flagged with any underlying data.

Now, Tableau authors can create workbooks using certified sources, a policy that can be enforced by IT.

When a new workbook is created, Alation will extract it and its full lineage (including the Tableau Data Sources to which it's connected). Stewards can verify the lineage and also make additional semantic checks (for instance, ensuring that the title and axis labels in each view are appropriate given its measures and dimensions). If everything checks out, they can use the Alation UI to transfer the workbook into the "Alation Certified" Project, where it will remain unless something is flagged upstream (as described above).

Maintaining Trust

Tableau Server viewers can visit the "Alation Certified" Project for a set of Workbooks that are totally trustworthy, all the time. With the strictest policy, Alation can remove Data Sources and Workbooks from the "Alation Certified" Project whenever a single link in its lineage lacks a Steward endorsement.

In addition to leveraging their own knowledge and that of subject matter experts, Stewards can leverage the rich usage information Alation Catalog surfaces to make decisions on what to flag.

Consider two tables with similar names and identical columns. If most analysts have historically queried one (which many workbooks reference), but some workbooks connect to the other—which has not been queried at all in months—that could be a warning sign.

A deeper investigation might turn up a staleness issue and lead a steward to deprecate the old table in favor of its fresh twin. All downstream workbooks would then be automatically removed from the "Alation Certified" project until their authors can "rewire" them to reference the fresh source. Of course, when workbooks point to Tableau Data Sources, not raw data, the process can be streamlined—a single "rewiring" can put all the affected Workbooks back into Certified status.

.

Alation Catalog automatically propagates deprecation information through lineage to downstream data assets.
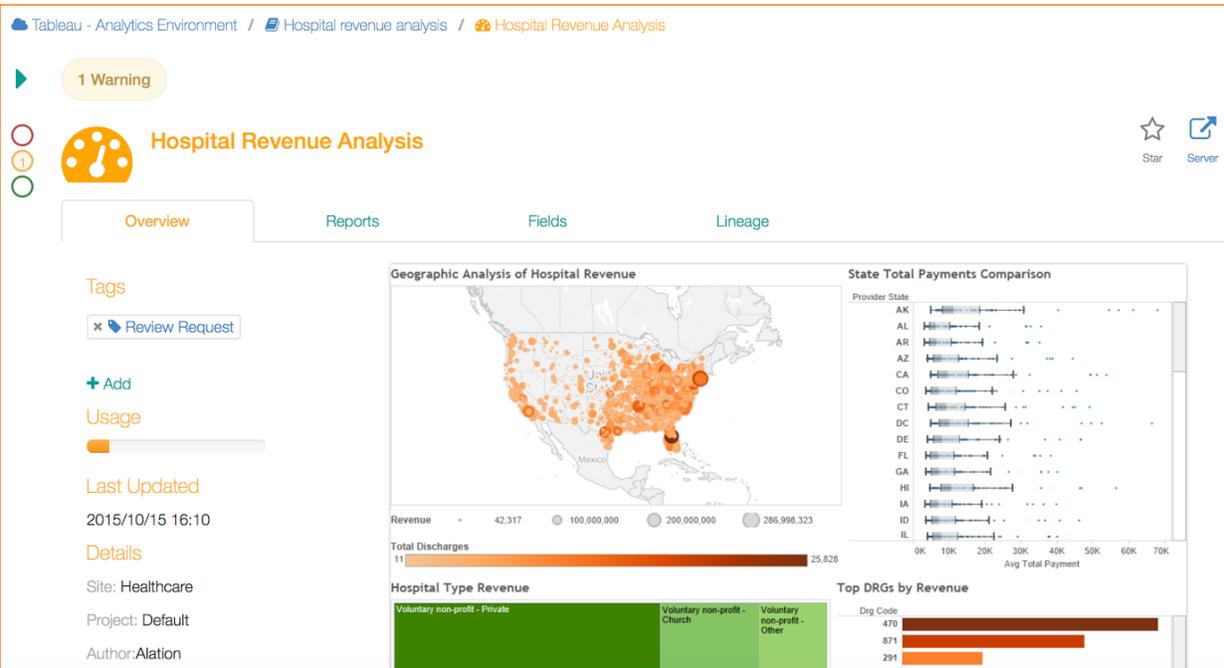
One of the four views in this Dashboard is built on a poor quality data asset. Looking at the Lineage can help a consumer figure out which.

The red arrows show a deprecation flagged against the upstream table `hospital_gen_info` being propagated down to one of four views in a Tableau dashboard.
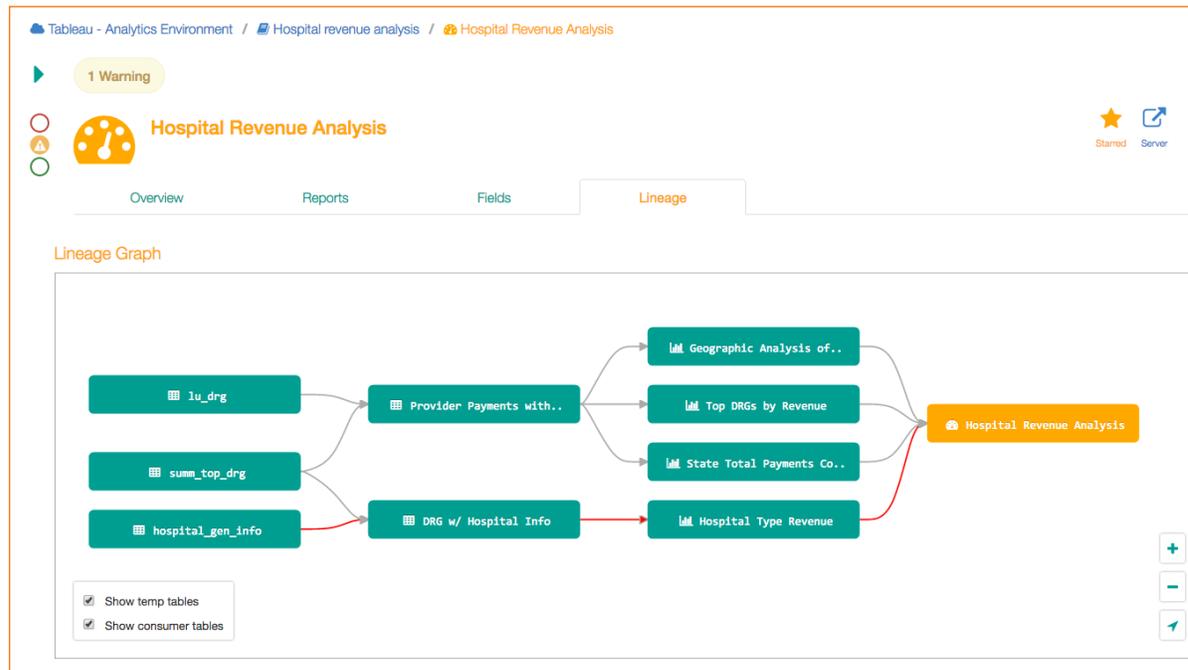
Additional Best Practices: Tableau Tags and Projects

Tags are a lightweight way of finding, filtering, and annotating objects on Tableau Server. A given Data Source or Workbook may have many tags. They can be a rich source of additional information surfaced from the Alation Data Catalog to end-users in Tableau. For instance, they can provide clarification about the source of a Certification-Classification, or they can convey other semantic information like a Data Domain. Since tags are open, mutable, and often ephemeral—and since they are not very prominently displayed—they should be used (both manually and in Alation integrations) for information which is nice-to-have but not necessary-to-see. Viewers should consider tags to be helpful but shouldn't necessarily assume content in tags is 100% reliable as it could have been added by anyone.

Projects, by contrast, have more constraints and are more prominently displayed in every screen in Tableau Server. Every Data Source and Workbook in Alation is in exactly one project at a time and admins can restrict who's allowed to add assets to a given project. This makes them ideal for defining and displaying the core certification status of each Data Source or Workbook.

The Best of Both Worlds

By following these best practices, your organization gets the benefits of both governance and self-service in Tableau Server:

- Tableau workbook authors have as much autonomy as ever but can work faster, more accurately, and with more confidence.

- Your analysts and business users have greater trust in the numbers they see. And by using Alation Catalog, IT gains even more confidence and control while analysts gets more accurate insights.

- Stewardship teams and IT can use the Alation Catalog to see a birds-eye view of which Tableau workbooks and data sources are being used

- Data consumers can discover popular/interesting assets and easily re-use—or build upon —existing materials, be they raw tables, or Tableau workbooks and data sources.

By leveraging the techniques and technologies described above, an organization can attain both the risk-mitigation derived from governance and stewardship and the upside-potential that self-service insights can promise.

About Tableau

Tableau Software helps people see and understand data. Offering a revolutionary new approach to business intelligence, Tableau allows you to quickly connect, visualize, and share data with a seamless experience from the PC to the iPad. Create and publish dashboards and share them with colleagues, partners, or customers—no programming skills required. See how Tableau can help your organization by starting your free trial at tableau.com/trial.

About Alation

Alation is the first data catalog built for collaboration. With collaboration, analysts are empowered to search, query and collaborate on their data to achieve faster, more accurate insights. Alation automatically captures the rich context of enterprise data, including what the data describes, who has used it, and the fit between the data and different types of analysis. Alation's catalog is generated and updated using machine learning and improved through human collaboration between analysts, stewards, experts and business users.

Alation is funded by Andreessen Horowitz, Bloomberg Beta, Costanoa Venture Capital, Data Collective and General Catalyst Partners. Customers include eBay, Albertsons, Square and some of the world's largest finance and retail firms. For more information, visit alation.com.

Contact Us

info@alation.com
(650) 799-4440