Alation

Open Data Quality Initiative

Alation has launched an Open Data Quality Initiative. This encompasses both a product upgrade and the launch of a new partnership program, which welcomes all data quality (DQ) and data observability vendors to the Alation platform. Now, with this framework:

- Alation users will be able to see vital DQ metrics within their workflow, across numerous data sources, so they can see what data is trustworthy and make more informed decisions about the data they use.
- Customers will get the flexibility and freedom to select the best DQ solution for their business and industry.
- Data quality partners are all welcome on our platform, and each get a suite of services to make integration easy.

Are you a customer who codes? Join our developer community!
Are you a data quality vendor? Partner with us!

By integrating DQ visibility into the point of self-service for data consumption — and giving customers the flexibility to partner with their DQ vendors of choice — Alation customers can design the best DQ program for their business and ensure that those DQ metrics are guiding accurate self-service analytics, focused data governance programs, and compliant usage at all levels of the business.

The Data Quality Challenge

How do you know what data to trust? Data quality information signals what to trust and what to avoid. Yet for too long, knowledge about data quality has been siloed within IT. This means analysts must exit their workflows and communicate with IT in order to assess an asset's quality. This not only wastes time, it leads to inaccurate and potentially dangerous analytics results.

Further, as data collection has exploded, many data users may be familiar with "analysis paralysis," where an avalanche of choice makes determining the best asset increasingly difficult. Data consumers need to be able to immediately determine data trustworthiness as they are evaluating potential data sources.

To make the best decisions and data models, data consumers need to discover data and its trustworthiness within the same workflow. How can we deliver DQ information to those who need it faster, as they are working with that data?

The Alation Data Quality Solution: An Open Framework

With this initiative, Alation releases a product called the Open Data Quality Framework (ODQF). The ODQF integrates data quality information into an asset's profile. Now, users can assess DQ at the point of discovery and more quickly determine whether that data is of sufficient quality for decision making.

Historically, the challenge with data quality tools is that they often live in silos within their organization. This limits the effectiveness of the DQ tool, as it depends on various data consumers to step outside of their normal workflows to check whether or not the data is fit for consumption and analysis. However, when DQ tools are combined with a data catalog, DQ alerts can effectively reach their intended audience and prevent costly mistakes.

With this new integration, data consumers no longer need to exit their workflow to locate DQ data or wait on IT to deliver DQ information. Analysts and other non-technical users will now see valuable DQ information alongside the data in the catalog, at the points of search and consumption.

Now, the data catalog will present captured data health, including DQ metrics and warnings for data consumers. Users will see rules, metrics, and conformance in more detail, at the column level. With this information, data users can collaborate directly with data experts and stewards on data quality issues.

By creating a one-stop shop for data consumers to view data quality recommendations and lineage, the platform will break through the previously siloed knowledge around data quality and bring it to all the relevant stakeholders. By surfacing data quality metrics directly within the data catalog, business users, business analysts, and data scientists glean useful data quality information at the same time as they search for information.

Why Integrate Data Quality into the Data Catalog?

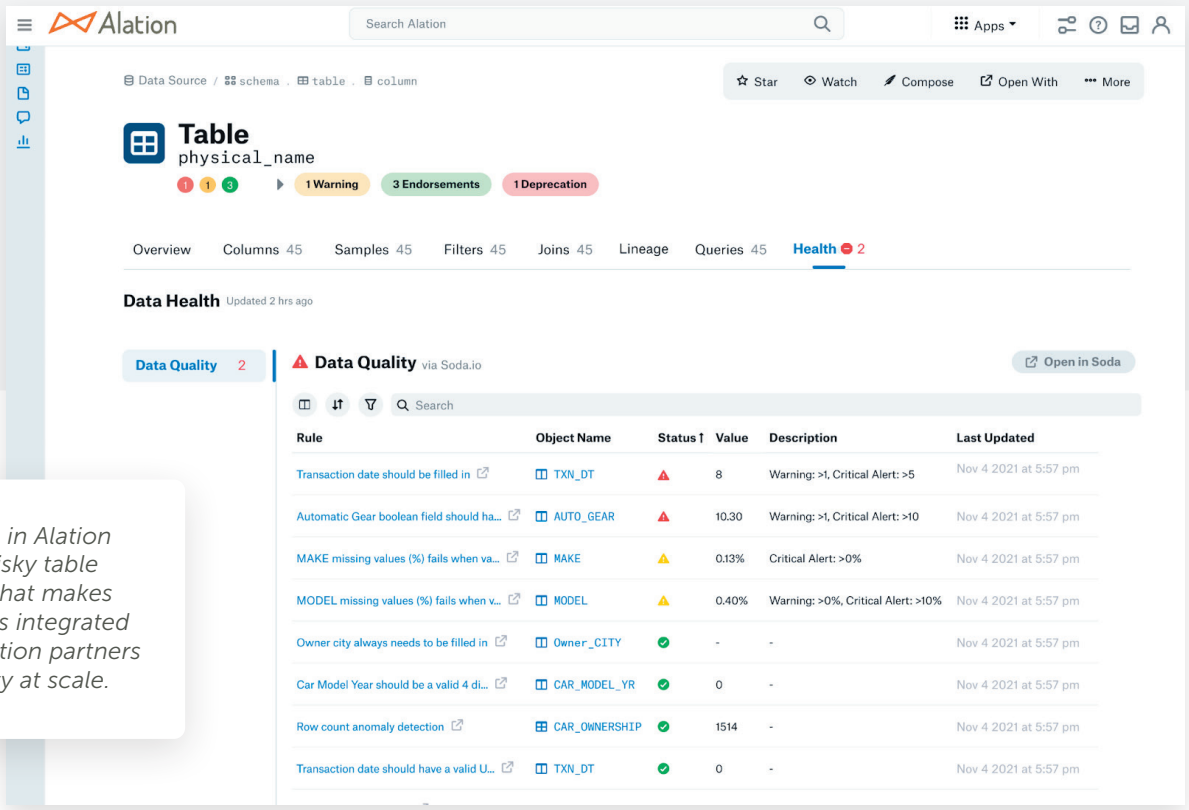
- **Determine data trustworthiness with ease.** The data catalog will present captured data health — including DQ rules, metrics, and warnings — to data consumers. With this information, data users can directly collaborate with data experts and stewards on data quality issues.
- **Integrate lineage with DQ data.** Since the data catalog captures lineage data, this integration will empower users to combine DQ information with lineage and profiling data — enabling them to troubleshoot DQ issues with a lineage map. (Note this will be available in subsequent upgrades.)
- **Increase analyst and data scientist productivity.** Data scientists and analysts spend much of their time on determining that data that is trustworthy for an analytic or data model. By surfacing DQ data into their workflow, Alation will cut down their time spend significantly.

How Does It Work?

As a first step, the Data Quality API grants customers the choice and flexibility to select the right data quality solution for them. The Data Quality API can be used to connect to any data quality solutions, including in-house built data quality solutions.

The Open DQ API Makes Adding DQ Data to the Catalog Easy

- The open API allows DQ data to be consumed across multiple data sources — from cloud data warehouses to business intelligence tools — with minimum configuration
- Connect to partner tools and other ecosystem players
- Pull DQ results, metrics, and warnings directly into Alation, so they are surfaced to users at point of discovery and consumption



Data quality tabs in Alation flag potentially risky table columns... but what makes them risky? Rules integrated from DQ integration partners judge data quality at scale.

Aggregate DQ Data Results Are Presented to Users

- Alation persists data results, so users have clear visibility into what data has issues and what is trustworthy
- Users can note DQ status changes as they occur in real time

Combine with Alation Native Data

- All data consumers have immediate access to DQ rules, metrics, and conformance — at the column level
- API integrations complement Alation's out-of-the-box DQ capabilities, including data profiling for at-a-glance view into data distribution (min, max, nulls), as well as features like:
    - **Trust Flags:** Allows the data community to endorse, warn, and deprecate data to signal if data can or can't be used

- **Data Profiling:** Statistics such as Min, Max, Mean, Null can be applied to certain columns to understand the shape of it
- **SmartSuggestions:** As you query in Compose, Alation's SQL editor, pop-up suggestions actively show you relevant data to use, powered by AI

Alation also provides a starter kit for DQ vendors who are using the API. The starter kit includes the API, documentation, an onboarding walkthrough, and integration best practices.

Conclusion

To trust data and use it correctly, data consumers need easy, obvious access to DQ information. By integrating DQ metrics into the discovery and consumption process, Alation not only accelerates the productivity of data consumers, like analysts — it frees up time for engineers and IT, who no longer need to play DQ middlemen.

For too long, data-driven companies have not had the freedom to design their modern data stack. Vendor lock-in is common. Yet, industries have rapidly diversified needs around data quality. They need the flexibility to custom-design a solution atop a data catalog platform; only then can they address their DQ pain points, which are specific to their industry and use cases.

With this framework, we welcome all DQ and data observability vendors to Alation with a starter kit. This kit includes open API, how-to docs via our dev portal, an Alation partner onboarding walkthrough, integration best-practices, and co-marketing support. This is because we want to make your DQ solution successful.

And as for the future? Data quality will increasingly add value to data governance initiatives. Further automation of governance, with help from DQ, is coming. In the future, the data catalog will integrate data quality rules into data governance policies. And with the open API, the Alation Data Governance App will be able to relate data quality rules to overarching data quality policies captured in Alation. This is part of the trend of providing integrations to data lineage tools and connectors to most popular data sources.

Learn more about the Alation Data Quality Framework:

- Read the blogs: Alation 2022.2: Open Data Quality Initiative and Enhanced Data Governance and a Q&A on Alation's Open Data Quality Framework
- Check out the press release
- Watch the data dialog: Why an Effective Data Quality Program Includes a Data Catalog
- Apply to become an Alation data quality partner