Bloor

InComparison

Comparing
Data Governance Solutions –
Alation vs Ataccama vs Collibra

Data governance is the process of managing and overseeing an organisation's data assets. It includes defining policies and procedures for data collection, storage, use, and disposal. Data governance also involves ensuring that data is accurate, complete, and reliable.

Introduction

Ever since organisations started entering their data into computers, there has been an awareness of the wisdom of paying attention to the governance of that data. Data dictionaries were created in the 1960s as part of the first database management systems. Indeed, the author of this document recalls working on the IMS Data Dictionary in his first job in 1983. However, for a long time, these data dictionaries collected mostly technical metadata and were largely passive, disconnected from the live operational database applications. Although some companies tried to produce glossaries of business definitions, it was difficult to keep these up-to-date without a connection between the data dictionary and the operational systems.

In more recent times there has been a resurgence in interest in data governance, as companies struggle to keep pace with the explosion of data sources that they have to deal with, both inside and outside the enterprise. In some industries and countries, regulation has demanded that companies get a better grip on their data assets and the quality of their data, with GDPR being just one example. Most companies struggle with competing data definitions and data duplication, as well as variable data quality, to the extent that only about one in three executives trust their corporate data, according to several independent surveys. Efforts to exert control of the data landscape through ERP projects and master data management projects have proved difficult to fully implement and complete, despite considerable investment. The average large company still has hundreds of different applications, according to various surveys, and this number is increasing as companies move gradually from on-premise data centres to private and public cloud computing.

It has become clear that business ownership is essentially a prerequisite for successful master data management initiatives, and this realisation has led to the introduction of a new generation of data catalog/data dictionary products. The modern data catalog is not merely a repository of technical metadata but a business glossary, documenting data assets and the relationships between these assets. Such products allow the classification and discovery of data and for policies to be developed and enforced for that data. A modern data governance product supports business data ownership and stewardship and the monitoring of data quality and usage.

In this document, we examine three of the leading vendors in that market: Alation, Ataccama and Collibra.

Alation

Alation is a modern data catalog product that has an extensive range of connectors to data sources, including SAP, Snowflake and Databricks. The catalog is a central repository to store relationships, data definitions and usage patterns. Users are able to search the catalog using a range of criteria and Alation provides rich metadata about the data assets, such as the ownership that has been assigned to each data element, the lineage of the data and its usage. Data can be classified and policies can be established for that data, with supporting workflow. Alation encourages business users to collaborate on data projects, for example sharing assessments of which data sources are most reliable. A business glossary is available that, once it is populated, contains definitions of data elements. There is a natural language interface to the catalog, enabling users to discover their data without the need to learn a reporting tool or a data access language like SQL. One strong feature of Alation is its ability to read the logs of databases so that it can determine the levels of usage of data sources by business users; for example, which systems are most heavily used and which may be under-utilised.

Alation has an artificial intelligence software layer ("Allie") that can, in its latest version, generate data content descriptions using generative AI, based on a foundation model using Meta's LLaMA large language model. It deploys pre-processing and post-processing to ensure the quality of the descriptions produced and assesses the reliability and quality of these in a traffic light form based on its post-processing. Users can also easily export data to third party reporting tolls such as Tableau.

Alation does not have its own data quality technology but instead relies on interfaces to third party specialist products such as Experian and Montecarlo. Similarly, Alation can interface to Manta for data lineage, though it is also building its own capabilities in this area.

Top Alation strengths and weaknesses

As well as getting an impression of the product via briefings and a demonstration, I examined a large number (well over a hundred) customer testimonials.

Common themes – likes

- Many customers commented on the ease of use and power of the search capability.
- Customers seem impressed with the ease of set-up and installation of the product.
- Customers like the wide variety of connectors to other related applications.
- Several customers commented on the user interface, which has recently been revamped, and now seems generally well-received by business users.
- The core data catalog gets praise for being highly functional.
- Customers like machine learning capability to analyse the ways that data is used.

Common themes – dislikes

- Multiple customers commented on the levels of bugs and sometimes unpredictable performance.
- Several customers had issues with the data lineage capabilities, which is provided by a third party product; Alation is actually working to build its own functionality in this area.
- Some customers commented on a number of somewhat immature areas of product functionality that now are in the product roadmap.
- While Alation is very good at accessing relational data sources, some customers report that it is trickier to deal with non-relational sources such as data lakes. There is a well-proven connector to Databricks, however.
- Some customers would like to see a greater degree of customisation within the product.
- Some customers find the product pricing to be rather opaque.
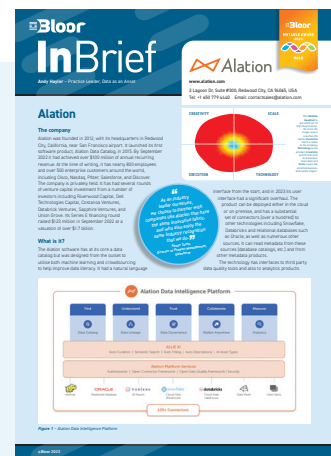
Business glossary	😊
Compliance monitoring	😊
Data catalog	😊
Data classification	😊
Data lineage and impact analysis	😊
Data ownership and data stewardship	😊
Full data quality functionality	😞
Master data management hub	😞
Policy management	😊
Workflow management	😊

Key

😊 Full functionality

😐 Some functionality

😞 No functionality

Ataccama

ataccama

Ataccama ONE is a data management platform that extends well beyond data governance functionality. Indeed, it began as a specialist data quality product, later adding master data management. Consequently, this product is broader in scope than the other two being considered, which were designed from the ground up to be data catalog and data governance products.

Although it is a broader solution, Ataccama nonetheless has the full set of functionality that you would expect from a data governance product, with a fully searchable data catalog and business glossary at its heart. You can classify data, for example personally identifiable data or other sensitive data, and can define policies about that data. Data ownership can be assigned and workflow can be set up, for example, to allow changes to a business glossary definition if and only if the business owner of that data approves it. The product has a suite of dashboards and monitoring capabilities, including a visual knowledge graph that can show data relationships.

The product's features can be mapped to an overall data quality lifecycle, with the steps: "document and define", "understand", "improve", "prevent" and "govern". For example the data quality firewall fits within the "prevent" stage, while governance policies, stewardship and workflow fit within the "govern" stage.

Ataccama can be deployed in either the cloud or on-premises and although the majority of its roughly 350 customers are on-premises, the vast majority of new deployments are in the cloud.

Top Ataccama strengths and weaknesses

As well as getting an impression of the product via briefings and a demonstration, I examined a large number (over a hundred) customer testimonials.

Common themes – likes

- Many customers commented on the intuitive user interface, which certainly comes through during product demonstrations.

- The data profiling capabilities of the product were remarked upon positively by customers, which makes sense given the long data quality heritage of the product.

- Customers consistently praised good performance on high data volumes.

- Customers found it easy to structure workflows using the visual setup configurations.

- A common theme through customer comments was that of strong technical support and excellent professional services staff.

- Another common theme was that customers appreciated having a wide range of data management tools available in one place with a consistent user interface.

Common themes – dislikes

On the less positive side, there were reservations expressed by customers concerning:

- A learning curve in getting to grips with rule expressions.

- A desire for clearer error messages.

- Some desire for more flexible/powerful reporting, though to be fair this can easily be handled by exporting data to a specialist reporting tool like Tableau.

- Several customers commented on the complexity of the integrated development environment.

- Some customers wanted dashboards that could be more flexible.

- Some customers (location-specific) found issues in finding contractors with experience in the product.
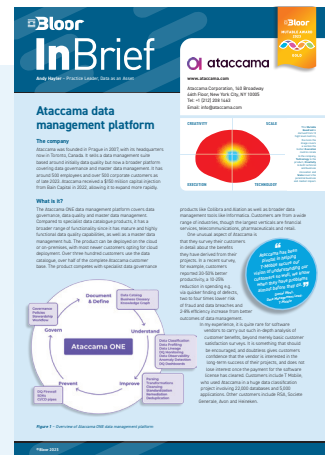
Business glossary	😊
Compliance monitoring	😊
Data catalog	😊
Data classification	😊
Data lineage and impact analysis	😊
Data ownership and data stewardship	😊
Full data quality functionality	😊
Master data management hub	😊
Policy management	😊
Workflow management	😊

Key

😊 Full functionality

😐 Some functionality

☹ No functionality

Collibra

Collibra is a platform that allows organisations to manage their data assets. From the start it was aimed at business users rather than the IT department, and it is based around a centralised data catalog that documents data assets, relationships and definitions. Policies around data can be defined and data can be classified. For example, some data can be identified as being sensitive, and access rules can be defined for that particular type of data, for example, to require data masking for personally identifiable information. Business users can be assigned ownership of data, and

support is provided for data stewards in the form of workflow and task management.

Collibra has built-in support for compliance and regulation, enabling reporting relevant to regulations like GDPR and CCPA. The product also has extensive dashboards and reports. A natural language interface is provided. Collibra is a mature product, having been first developed in 2008. It has been deployed to over 700 customers by 2023, with nearly half of these being large-scale companies.

Collibra has its own data lineage features and in February 2021

introduced its own data quality capability in the form of Collibra Data Quality & Observability. This includes the use of machine learning to identify anomalies and to auto-generate data quality rules.

Top Collibra strengths and weaknesses

As well as getting an impression of the product via briefings and a demonstration, I examined a large number (over a hundred) of customer testimonials.

Common themes – likes

- Many customers comment on the sheer range and depth of functionality that the product offers for data governance.

- Generally speaking, customers seem to find that the user interface for business users is quite well put together and that business users find the product reasonably easy to use.

- Some customers seemed impressed with the range of connectors and ease of integration to connect to data sources and other products.

- Customers like the ease of exporting data to other platforms e.g. for regulatory reporting purposes.

- Some customers praised the quality of the training materials for Collibra.

- Customers seem to like the ease of setting up and customising workflow.

Common themes – dislikes

- The data lineage capabilities of the product felt able to be improved upon by certain customers, though to be fair some others were perfectly happy with it.

- The data quality and profiling capabilities were identified as needing further refinement by some customers.

- The degree of integration and consistency between the core product and the acquired features of data lineage and data quality were an issue for some customers.

- The product is highly functional and some customers found that this could be rather daunting for casual users; presumably, greater use and control of user permissions could be used to reduce this issue.

- Some customers felt that the product was quite expensive, with somewhat opaque pricing.
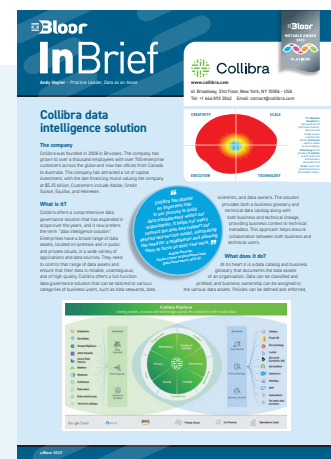
Business glossary	🙂
Compliance monitoring	🙂
Data catalog	🙂
Data classification	🙂
Data lineage and impact analysis	🙂
Data ownership and data stewardship	🙂
Full data quality functionality	🙂
Master data management hub	☹️
Policy management	🙂
Workflow management	🙂

Key

🙂 Full functionality

😐 Some functionality

☹️ No functionality

Analysis and conclusions

The three data governance solutions covered in this report are all strong products, each with hundreds of corporate customers. Ataccama differs in scope from the other two in that it grew up as a data quality vendor, later adding master data management and data governance capabilities. Hence it is a more complete data management solution than the specialist approach of Collibra and Alation, which were designed from the ground up as data catalog and governance solutions. All three products have the broad functional coverage that you would expect from a data governance product: a core data catalogue, business glossary, policy management, data lineage and impact analysis, workflow management and support for data ownership, and the ability to report and monitor on such things as compliance and overall data quality.

Customers considering a solution first need to decide whether they are looking for an overall data management solution from a single vendor or whether they prefer a more best-of-breed approach, choosing a specialist data catalog product and separately considering a best-of-breed data quality solution, and indeed a master data management product. Depending on the use case, some particular feature may be of overriding importance, and this may indicate a best-of-breed approach where each element (data catalog, data quality, master data management) will be evaluated on its merits and selected accordingly. Such an approach contrasts with one where a broader solution is chosen, in which case you trade off specific features against higher integration efforts. Clearly, if you choose separate technologies for these three related areas then there will be some level of integration effort, and the user experience will be somewhat more disjointed than if you go with a solution from a single vendor.

There is no right or wrong answer here. You need to carefully document your own specific requirements and criteria, weight these by importance, and then evaluate candidate products against your criteria using your own data. This last point is critical: vendors will carefully select demonstration data sets that show off their best features, but these may or may not be relevant to your specific needs. Always test using your own data and with full production volumes of data, to ensure that there are no surprises when you deploy. You may find it helpful to use the services of an independent third party who has experience in the market to help you structure your evaluation process.

	Alation	ataccama	Collibra
Business glossary	🙂	🙂	🙂
Compliance monitoring	🙂	🙂	🙂
Data catalog	🙂	🙂	🙂
Data classification	🙂	🙂	🙂
Data lineage and impact analysis	🙂	🙂	🙂
Data ownership and data stewardship	🙂	🙂	🙂
Full data quality functionality	☹️	🙂	🙂
Master data management hub	☹️	🙂	☹️
Policy management	🙂	🙂	🙂
Workflow management	🙂	🙂	🙂

Key

🙂 Full functionality

😐 Some functionality

☹️ No functionality

About the author

ANDY HAYLER
Practice Leader: Data as an Asset

Andy is an established software industry authority, an independent strategy consultant advising corporations, venture capital firms and software companies. He is the founder of Kalido, which under his leadership was the fastest growing business intelligence vendor in the world in 2001. Kalido was recognised as an innovator in data warehousing, and then launched arguably the first true master data management product, a market which at the time did not exist but is now a well recognised and fast growing industry. Andy was the only European named in Red Herring's "Top 10 Innovators of 2002". He was a pioneer in blogging with his award winning "Andy On Enterprise Software" blog.

Andy started his career with Esso, working in a number of technology roles before moving to Shell. He was Technology Planning Manager of Shell UK, then Principal Technology Consultant for Shell International. He later established a global information management consultancy, which under his leadership grew to 300 staff.

Bloor overview

Technology is enabling rapid business evolution. The opportunities are immense but if you do not adapt then you will not survive. So in the age of Mutable business Evolution is Essential to your success.

We'll show you the future and help you deliver it.

Bloor brings fresh technological thinking to help you navigate complex business situations, converting challenges into new opportunities for real growth, profitability and impact.

We provide actionable strategic insight through our innovative independent technology research, advisory and consulting services. We assist companies throughout their transformation journeys to stay relevant, bringing fresh thinking to complex business situations and turning challenges into new opportunities for real growth and profitability.

For over 25 years, Bloor has assisted companies to intelligently evolve: by embracing technology to adjust their strategies and achieve the best possible outcomes. At Bloor, we will help you challenge assumptions to consistently improve and succeed.

Copyright and disclaimer

Bloor

Bloor Research International Ltd

⊙ 20–22 Wenlock Road, London N1 7GU, United Kingdom

☏ +44 (0)1494 291 992

✉ info@Bloorresearch.com

🌐 www.Bloorresearch.com