Alation

Customer Case Study: Allegro

How a European E-Commerce Marketplace Gained Faster Value from Data with Alation

allegro

- **Industry:** Retail
- **Headquarters:** Poznań, Poland
- **Interesting Fact:** More than 135,000 merchants participate in the Allegro online marketplace

Finding Data Across Systems and Business Units

In 1999, Allegro.pl debuted as an online auction site in Poland. Today, Allegro is an international marketplace platform that connects more than 135,000 merchants with over 22 million site visitors each month. It is the most popular shopping platform in Poland and the largest e-commerce platform of European origin.

Data plays a key role in the company's goal to create the best possible experience for both customers and vendors. Allegro uses analytics to understand customers and improve recommendations and offers. Data also drives the promotional programs they offer to vendors.

Allegro constantly analyzes data within their 20 business units, among them logistics, payments, and pricing. The ever-increasing volume and complexity of data required the company to change how they catalog data and produce data transformations, aggregates, and indicators. To better support data users within the Allegro ecosystem, the data and artificial intelligence team sought to improve data governance and usage, shorten the time needed to gain value from the data, and promote the creation of data-driven products.

Business Results

Accelerates Time to Value: Alation speeds the creation of reports, models, and data products

Boosts Analyst Efficiency: Users report a 24-minute weekly time savings

Improves Data Governance and Quality: Data consumers can see data ownership, usage permissions, and SLAs in Alation before using data

The company thus embarked on a digital transformation effort. They are moving from a centralized, monolithic, and domain-agnostic architecture to a distributed, domain-driven architecture, self-serve data platform design, and enterprise-to-enterprise (E2E) data product thinking. The transformation required a new approach to data organization and governance to ensure fast and easy data findability, ingestion, processing, and delivery.

> *"We wanted a single interface from which our data community could access various data sources. Additionally, we hoped to encourage communication among analysts through sharing queries and engaging in conversations and to facilitate faster analyst onboarding."*
>
> Marcin Cinciala
> Senior Data Engineer, Allegro

The team interviewed a wide range of data producers and users to determine their needs. It became clear that data was difficult to find and access, and users often needed to ask the same questions of data producers to determine who owned the data and whether it was up to date and trustworthy. "We wanted a single interface from which our data community could access various data sources," says Marcin Cinciala, Senior Data Engineer at Allegro. "Additionally, by enabling shared queries and conversations, we hoped to encourage analyst communication and facilitate faster analyst onboarding."

Allegro hoped to relieve the frustration of both data producers and users by implementing a data catalog to make data discoverable, addressable, and trustworthy. Although their teams of software engineers, data engineers, and data analysts produced documentation for their data products, they lacked a uniform approach to cataloging it. Information was spread across various systems and formats such as Confluence, GitHub README files, and Jira tickets. Lineage could only be traced by reading through the actual code of the data pipelines. When someone wanted to use a given data source, and do so conscientiously, they had to contact the developer directly to ensure its accuracy and validity.

Creating a One-Stop Shop for Data

Allegro chose Alation Data Catalog to combine intelligent metadata management with intuitive search tools to help analysts and other data users find, understand,

Alation

and safely use data. Other tools within the Allegro big data ecosystem feed information about the structure, use, and quality of data into Alation.

Alation has become the company's data integration hub. In the beginning, back in 2016, Alation was used by Data Analysts to write queries, share knowledge, comments, and for embedded data. Now, the single platform combines information about datasets and descriptions from other applications in one location. Table metadata, complete with the business context provided through articles, is easily accessible. Visual representation of data lineage completes the picture. To ensure consistency and completeness, data producers are thoroughly instructed on how to create good documentation. Allegro selects the datasets to curate in Alation based on the frequency of use or on specific attributes set by the data producers. For example, Alation is the primary catalog for datasets from Google BigQuery and Tableau as well as front-end and mobile events through virtual data sources.

"Alation enables our users to find the right data immediately and learn how it is used within the company," says Cinciala. "We want everyone who creates or works on data to ensure that it is also documented in our data catalog."

> *"Alation enables our data users to find the right data immediately and learn how it is used within the company. We want everyone who creates or works on data to ensure that it is also documented in our data catalog."*
>
> Marcin Cinciala
> Senior Data Engineer, Allegro

Allegro created an API to automatically load data documentation and information into the catalog directly from the data source, helping the company complete the data picture with relevant table details and data lineage. The API also helps guarantee the consistency and timeliness of the information in the catalog. Data owners and analysts no longer need to browse multiple applications to find descriptions of tables or datasets. Linked to specific tables and continuously updated with new user questions and answers, articles reduce data owners' workload of maintaining the datasets.

Alation

Boosting Data Governance, GDPR Compliance, Literacy, and Quality

To improve data governance and ensure the proper use of Allegro's datasets, Cinciala's team directs analysts to Alation for information about data ownership, necessary access permissions, business descriptions, and more. The team publishes documentation from other big data and AI–related areas of the business in Alation in addition to links to webinars and various types of reports.

To drive data literacy through wider adoption of the tool, Cinciala's team focuses on the quality of the metadata in the catalog. Whenever a software or data engineer adds a label to any table signifying it is ready to be added to Alation, Cinciala's team automatically checks how "complete" the metadata set is and automatically informs the person responsible for the table which sections in Alation still need to be filled in.

"We were definitely able to establish comprehensive data governance," says Cinciala. "By pushing metadata into Alation, we have a much better overview and full visibility into data with preconfigured content, like data categories, lineage, workflows, reports, and more. of the available data to base our decisions on. Moreover, Alation helps us to be GDPR-compliant by building a privacy-aware data culture." And Allegro.eu S.A., the holding company which engages in the management of e-commerce platforms Allegro, Ceneo, and others, manages even more sensitive data. Now that the data can be labeled, it is more searchable and can be controlled much better."

> *"It's much easier to encourage data consumers to use the system daily if we ensure that what we store in Alation is complete and up to date."*
>
> Marcin Cinciala
> Senior Data Engineer, Allegro

The team also measures the quality of documentation to ensure that tables include a business description, defined service level agreements (SLAs), and specified ownership, among other attributes. "It's much easier to encourage data consumers to use the system daily if we ensure that what we store in Alation is complete and up to date," notes Cinciala. "And it's easier to drive insight and knowledge as we are using a Data Mesh approach to building a decentralized data architecture by leveraging a domain-oriented, self-serve design."

Alation

Improving Efficiency with a Single Source of Reference

The catalog has become the central place for data users to quickly find information about the data they need. "Alation is our primary source of truth about datasets," says Cinciala. "It aggregates information from multiple repositories, presents that information to users, maintains the consistency of descriptions, and saves time in creating new solutions and analyses, all through a flexible user interface."

"We expect the data catalog to significantly accelerate the process of creating analyses, reports, models, and data products by reducing the complexity of the data search process."

Marcin Cinciala
Senior Data Engineer, Allegro

According to a user survey after implementing Alation, Allegro discovered that the tool saves 24 minutes weekly for the average user. "Producers who have documented a dataset well in Alation no longer have to answer the same question a dozen times," says Cinciala. "And data consumers no longer need to wander through Slack channels searching for the owner of the data."

Alation not only facilitates communication between data producers and consumers but also reduces onboarding time for new employees by making it easier for them to quickly understand the company's data. "We expect the data catalog to significantly accelerate the process of creating analyses, reports, models, and data products by reducing the complexity of the data search process," says Cinciala.

"Producers who have documented a dataset well in Alation no longer have to answer the same question a dozen times, and data consumers no longer need to wander through Slack channels searching for the owner of the data."

Marcin Cinciala
Senior Data Engineer, Allegro

Alation

Data documentation is no longer spread throughout multiple systems or on wikis. Requiring producers to catalog table metadata and descriptions in Alation has also improved the quality of the documentation. In addition, the API not only facilitates the automatic upload of documentation to Alation, but it also allows the datasets in the data catalog to be automatically enriched with custom fields from other tools and services, including active directory, service catalog, pager duty, and more.

"When we started looking for a data catalog, we wanted a single place for people to search for data," notes Cinciala. "I don't think we even considered the kind of system that we're developing now — a robust, API-fed application that integrates information about table metadata, data governance, data quality, and lineage."

Data Democratization to Drive User Adoption

Up to now, only the most popular 5,000 tables of data were visible to users. The plan is to extend the content of the data catalog and to push all data into Big Query so users can leverage 10 times as much data as they can use today. Users are also encouraged to hand in feature requests to drive Data Democratization - i.e. the process of making digital information accessible to the average non-technical user of information systems, without having to require the involvement of IT.

About Allegro

In 1999, Allegro.pl debuted as an online auction site in Poland. Today, Allegro is an international marketplace platform that connects more than 135,000 merchants with over 22 million site visitors each month. It is the most popular shopping platform in Poland and the largest e-commerce platform of European origin.

Data Environment

- Apache Hive
- PostgreSQL
- Google BigQuery