



## Market Insight Report Reprint

# Coverage Initiation: Anomalo automates data observability for data quality through unsupervised methods

March 16 2022

by **Paige Bartley**

Interconnectivity of data systems means that a data anomaly in one place can populate elsewhere almost immediately, resulting in a domino effect. Anomalo is focused on cloud-enabled data stacks, looking to build a dedicated data quality assurance 'layer' where unsupervised learning is used to automate monitoring and detection of issues.

451 Research

---

**S&P Global**

Market Intelligence

This report, licensed to Anomalo, developed and as provided by S&P Global Market Intelligence (S&P), was published as part of S&P's syndicated market insight subscription service. It shall be owned in its entirety by S&P. This report is solely intended for use by the recipient and may not be reproduced or re-posted, in whole or in part, by the recipient without express permission from S&P.

## Introduction

The desire for automation is a fact of modern data management, whether it be data integration, metadata management, data quality or other related initiatives. The current adoption of AI and machine learning suggests this reality. Based on 451 Research's Voice of the Enterprise: AI & Machine Learning, Use Cases 2022 survey, 76% of respondents report that 'purchasing applications with embedded ML capabilities' is a strategy currently used, or soon to be used, by their organization. Anomalo is looking to package that automation – specifically, unsupervised learning – via its offering to help build a data quality and observability layer for the modern data stack so that issues can be detected and remediated before they become problems in downstream enterprise data consumption.

The vendor's approach emphasizes automation, differentiated primarily through unsupervised data monitoring to automatically find significant changes inside sources of raw data. Via additional low-/no-code options for key metrics and validation rules, it seeks to help organizations achieve ease of use for data quality assurance, assisting data professionals with being proactive in kicking off workflows to remedy data outages and other data issues.

## THE TAKE

Borrowing from engineering principles, data observability technology has become a relatively defined subsegment of the data quality market, looking to push data quality monitoring and remediation further upstream in the enterprise data supply chain. Products in this realm seek to differentiate in various ways, but they all largely share the same objective: To monitor and detect data system anomalies before they spiral into problems for downstream enterprise data use and consumption. With today's volume, variety and velocity of data, it's only logical that data observability offerings try to apply automation techniques to their fullest extent.

This is what Anomalo is aiming to do with its unsupervised data monitoring, which can be augmented via low- and no-code options that can essentially be layered on top to guide the best fit for specific business use cases. While automation can be powerful, the company still has some education to do – today's data stack participants can span various personas and roles, and often there is no clear 'typical' end user for the vendor's product given the multi-stakeholder nature of data pipelines. Part of Anomalo's challenge will be to clearly define its purchase influencers.

---

## Context

The vendor was founded in 2018 by Elliot Shmukler (CEO) and Jeremy Stanley (CTO), who both met while working at Instacart. The impetus of Anomalo's technology was the idea that despite the deployment of cloud-enabled modern data stack tools (or perhaps indirectly due to them), huge volumes of data and diverse data sources continued to confound basic efforts at data quality assurance.

Extensive manual efforts were often required to maintain data quality for mission-critical purposes, and dependencies between systems could often trigger 'snowball' effects when data outages in one place would automatically populate elsewhere due to existing integrations. An example experienced by the founding duo while at Instacart was when a major grocery retail partner sent a data feed with an entire category of products missing, which was then automatically populated into the Instacart app, immediately affecting both the Instacart customer experience and the retailer's sales.

The founders saw a market opportunity for automation of detection of these potential data anomalies so that they could be addressed proactively before they affected the bottom line of the business. Unsupervised learning was chosen for its ability to automatically find significant changes inside the raw data and provide extensive validation with basically no setup – helping find ‘unknown unknowns.’

Anomalo is headquartered in Palo Alto, California, but maintains a remote-first model for its workers. The company has raised nearly \$39m in funding to date across a \$5.6m seed round and a \$33m series A in October 2021 led by Norwest Venture Partners. With just over 20 employees now, Anomalo has a small team with ambitious growth plans – it is looking to increase headcount to approximately 40 by the end of 2022.

Customers typically have complex, cloud-focused data stacks with complex and data-intensive use cases. Organization size is not necessarily indicative of data intensity, although there is often correlation. Reference customers include BuzzFeed, Discover, ID.me, OpenWeb and Substack.

## Products

Anomalo essentially has two offerings, which are tightly integrated and interdependent: the flagship Anomalo platform layer designed for unsupervised monitoring; and the Anomalo Pulse dashboard, which was introduced in March 2022 and is designed to give an executive-level view of data quality performance over time. Use of the latter is contingent on deployment of the Anomalo platform layer.

As its key differentiation, the Anomalo platform layer utilizes unsupervised ML methodology to connect to existing enterprise data warehouses, and is capable of monitoring relevant tables within minutes of deployment. Cloud data warehouses are most common with the vendor’s customers and in January, it announced a formal partnership with Snowflake.

The ML techniques of the Anomalo platform layer are capable of automatically learning the historical structure and patterns of data within the data warehouse, allowing the firm to alert regarding common data issues without the need to directly predefine rules or set thresholds. However, no- and low-code validation rules for data quality can easily be added and layered on top, helping customize for specific business needs. Rules may also be written in SQL, if desired. Additionally, low- and no-code options exist for defining and monitoring key metrics, available in the native Anomalo UI.

Fully automated functions of the Anomalo platform layer, powered by unsupervised learning, include consistent monitoring of data freshness, data volume, potential missing data, and data-warehouse table anomalies. Of course, monitoring and early detection is only as useful as the ultimate actions that are taken to remediate potential problems that have been detected. As such, Anomalo has rich notification capabilities, integrating with productivity and communications platforms such as Slack and Microsoft Teams so that appropriate stakeholders may be notified, and natively click through notifications in their preferred communications environment to conduct root-cause analysis of issues before continuing on to remediation actions.

Anomalo Pulse, the dashboard and analytics offering, provides a high-level view of data quality health and monitoring that is more easily digestible by multiple data stakeholders, including nontechnical roles. The dashboard offers data owners the capability to understand blind spots in data quality, data arrival times compared with internal or external SLAs, data quality trends, and ‘repeat offenders’ in terms of data issues. Time-series visualizations and deep-dive capabilities further help decision-makers drill into problems and decide what KPIs should be monitored moving forward. Anomalo Pulse dashboard is packaged with the Anomalo platform layer for all customers, ensuring that they get both monitoring capabilities and insight/analytics capabilities together.

## Competition

As it is focused on early monitoring, detection and remediation of data system abnormalities, Anomalo can be considered to be part of the emergent 'data observability' market subsegment, a grouping that 451 Research currently considers a part of the data quality sector in its annual Data Management Market Map. Some of its relevant peers include Acceldata, Bigeye (formerly Toro), Databand, Datafold, Lightup, Monte Carlo and Soda. Each provider has its own differentiators, but they all generally share a commonality of customer use cases focused on data quality outcomes.

Still, traditional data quality detection and remediation specialists cannot be ignored. Examples of vendors with robust data quality functionality include Atacamma, Atlan, Dactatics, Experian Data Quality, Global IDs, Hitachi Vantara, IBM, Informatica, Precisely (formerly known as Syncsort), SAS Institute, Semarchy, Syniti, Talend, TIBCO and Zaloni. Tamr, although primarily focused on data mastering, emphasizes data quality assurance.

With its focus on ML-enabled unsupervised learning, Anomalo might also potentially contend with homegrown ML models that are built in-house for data management use cases. Machine learning technology can be bought or built, and Anomalo is trying to make a strong case for purchase. However, data-intensive organizations with extensive internal engineering and data science resources may find that they can build and deploy models that help address some data observability and quality needs, albeit without the assured support of a packaged product. That being said, the majority of homegrown approaches deploy deterministic and rules-based methods rather than unsupervised models.

## SWOT Analysis

<p><b>STRENGTHS</b></p> <p>Anomalo is focused on leveraging unsupervised ML techniques for data system monitoring, allowing for rapid deployment and insight into potential data quality issues before they become problems. Additional options for metrics and validation rules allow organizations to customize to their individual needs.</p>	<p><b>WEAKNESSES</b></p> <p>The vendor is trying to automate away many of the functions that data engineering teams and other technical experts specialize in for a living. While there is a clear argument to be made that its product could help these roles refocus their resources toward more productive endeavors, the company may have trouble communicating that.</p>
<p><b>OPPORTUNITIES</b></p> <p>Enterprise data quality efforts continue to further 'shift left' with increasing application of engineering and observability principles. Organizations are seeking ways to catch potential issues with the data layer much earlier, before they manifest into downstream data quality problems. Anomalo's automated approach has appeal in this regard.</p>	<p><b>THREATS</b></p> <p>The entire data observability subsegment of technology providers could be considered to be attractive targets for some of the larger players in the traditional data quality remediation market, which could likely augment their portfolios by baking in early detection and monitoring. An acquisition could theoretically limit product utility for diversified data environments.</p>

## CONTACTS

### The Americas

+1 877 863 1306

[market.intelligence@spglobal.com](mailto:market.intelligence@spglobal.com)

### Europe, Middle East & Africa

+44 20 7176 1234

[market.intelligence@spglobal.com](mailto:market.intelligence@spglobal.com)

### Asia-Pacific

+852 2533 3565

[market.intelligence@spglobal.com](mailto:market.intelligence@spglobal.com)

[www.spglobal.com/marketintelligence](http://www.spglobal.com/marketintelligence)

Copyright © 2022 by S&P Global Market Intelligence, a division of S&P Global Inc. All rights reserved.

These materials have been prepared solely for information purposes based upon information generally available to the public and from sources believed to be reliable. No content (including index data, ratings, credit-related analyses and data, research, model, software or other application or output therefrom) or any part thereof (Content) may be modified, reverse engineered, reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of S&P Global Market Intelligence or its affiliates (collectively, S&P Global). The Content shall not be used for any unlawful or unauthorized purposes. S&P Global and any third-party providers, (collectively S&P Global Parties) do not guarantee the accuracy, completeness, timeliness or availability of the Content. S&P Global Parties are not responsible for any errors or omissions, regardless of the cause, for the results obtained from the use of the Content. THE CONTENT IS PROVIDED ON “AS IS” BASIS. S&P GLOBAL PARTIES DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE, FREEDOM FROM BUGS, SOFTWARE ERRORS OR DEFECTS, THAT THE CONTENT’S FUNCTIONING WILL BE UNINTERRUPTED OR THAT THE CONTENT WILL OPERATE WITH ANY SOFTWARE OR HARDWARE CONFIGURATION. In no event shall S&P Global Parties be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs or losses caused by negligence) in connection with any use of the Content even if advised of the possibility of such damages.

S&P Global Market Intelligence’s opinions, quotes and credit-related and other analyses are statements of opinion as of the date they are expressed and not statements of fact or recommendations to purchase, hold, or sell any securities or to make any investment decisions, and do not address the suitability of any security. S&P Global Market Intelligence may provide index data. Direct investment in an index is not possible. Exposure to an asset class represented by an index is available through investable instruments based on that index. S&P Global Market Intelligence assumes no obligation to update the Content following publication in any form or format. The Content should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. S&P Global Market Intelligence does not endorse companies, technologies, products, services, or solutions.

S&P Global keeps certain activities of its divisions separate from each other in order to preserve the independence and objectivity of their respective activities. As a result, certain divisions of S&P Global may have information that is not available to other S&P Global divisions. S&P Global has established policies and procedures to maintain the confidentiality of certain non-public information received in connection with each analytical process.

S&P Global may receive compensation for its ratings and certain analyses, normally from issuers or underwriters of securities or from obligors. S&P Global reserves the right to disseminate its opinions and analyses. S&P Global’s public ratings and analyses are made available on its websites, [www.standardandpoors.com](http://www.standardandpoors.com) (free of charge) and [www.ratingsdirect.com](http://www.ratingsdirect.com) (subscription), and may be distributed through other means, including via S&P Global publications and third-party redistributors. Additional information about our ratings fees is available at [www.standardandpoors.com/usratingsfees](http://www.standardandpoors.com/usratingsfees).