IDC

SPOTLIGHT

Sponsored by: Precisely

Building trust in data is foundational to creating enterprise intelligence to compete in a digital-first world. Organizations are working to close the data trust gap by improving data integrity using data enrichment and intelligence.

Improving Data Integrity and Trust through Transparency and Enrichment

June 2022

Written by: Stewart Bond, Research Director

Introduction

We are at a point in history when there is universal acceptance that data — its creation, management, analysis, and use — is a foundational asset for every organization. IDC research regularly shows that executives openly articulate the need for their organizations to be more data driven, to be "data companies," and to increase their enterprise intelligence. A global IDC survey fielded in the summer of 2021 indicated that 77% of organizations have a senior-level executive responsible for leading enterprise intelligence, which refers to an organization's capacity to learn and ability to synthesize information and deliver insights at scale.

AT A GLANCE

KEY STATS

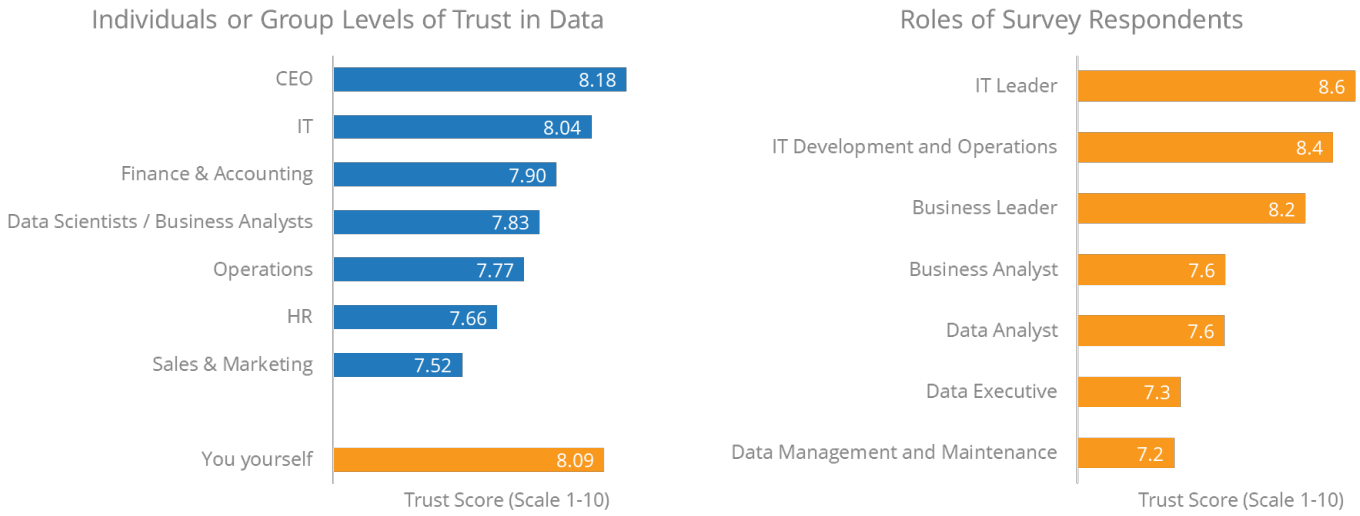
- 63% of data practitioners believe they are expected to make data-driven decisions.
- But only 30% strongly believe their actions are driven by data analysis.
- Even fewer, only 27% of practitioners completely trust the data they work with.

This need to be more data-driven at the executive level translates into expectations at the practitioner level. Nearly two thirds (63%) of data practitioners that responded to the most recent IDC *Data Culture survey* believe they are expected to make data-driven decisions, but less than one third (30%) strongly believe their actions are driven by data analysis. The most recent IDC survey on *Data Trust* indicated that only a quarter of the people (27%) noted that they completely trust data. Clearly, gaps exist between expectations placed on the use of data and the ability to trust and use data in delivering data-driven business outcomes.

Figure 1 illustrates the level of trust in an organization by the role. The chart on the left illustrates how much respondents believe individuals or groups in the organization trust data, including the respondent themselves. The chart on the right illustrates the roles of the people who responded to the survey, and how much each role trusts the data. People that are closest to the data trust it the least, while those that need to make decisions with data, trust it the most. This is because people that are closest to the data see the quality and integrity issues in the data, and spend time and effort to improve the integrity, so that those who need to make decisions with the data, can trust it.

FIGURE 1: ***Data Trust in the Organization***

Q To what extent do you believe these individuals or groups trust the data in the organization ?



Source: IDC Data Trust Survey, December 2021, N=500

Another interesting phenomenon the survey identified is that as data gets further away from where it is created, the level of trust degrades. Data is trusted the most inside of business applications, where much of the data about the business is created. As data moves into operational and analytical data stores, the level of trust starts to degrade because as data moves, its context, meaning, and shape can become distorted. In the most recent IDC survey on *Data Culture*, more than three-quarters of people who are making data-driven decisions expect to know where data came from (lineage), how clean or dirty it is, what it means to the business, and where the data is located.

The most recent IDC *Data Trust Survey* also told us that organizations are not performing well in enabling self-serve data access, identifying the lineage of data, locating the best data, and profiling data to calculate quality scores. Metadata management, the source of intelligence about data, is also the one capability that is implemented the least by survey respondents. It's no wonder only 27% of these respondents fully trust data. Intelligence about data can provide a path back to its source to improve the level of trust in data, because it provides transparency into the integrity of data in the organization.

Improving the integrity of data requires data intelligence to be actionable, to enable data governance and quality disciplines within organizations. Data governance is not only about access and authorization control, but it is also about quality control. Data quality attributes include duplication, accuracy, consistency, correctness, timeliness, precision, accessibility, and believability. Each of these attributes represents metrics that can be measured and monitored by data quality processes, to direct data stewards to the data that needs the most attention. These attributes also drive acceptable data use and movement polices into the organization to help maintain control.

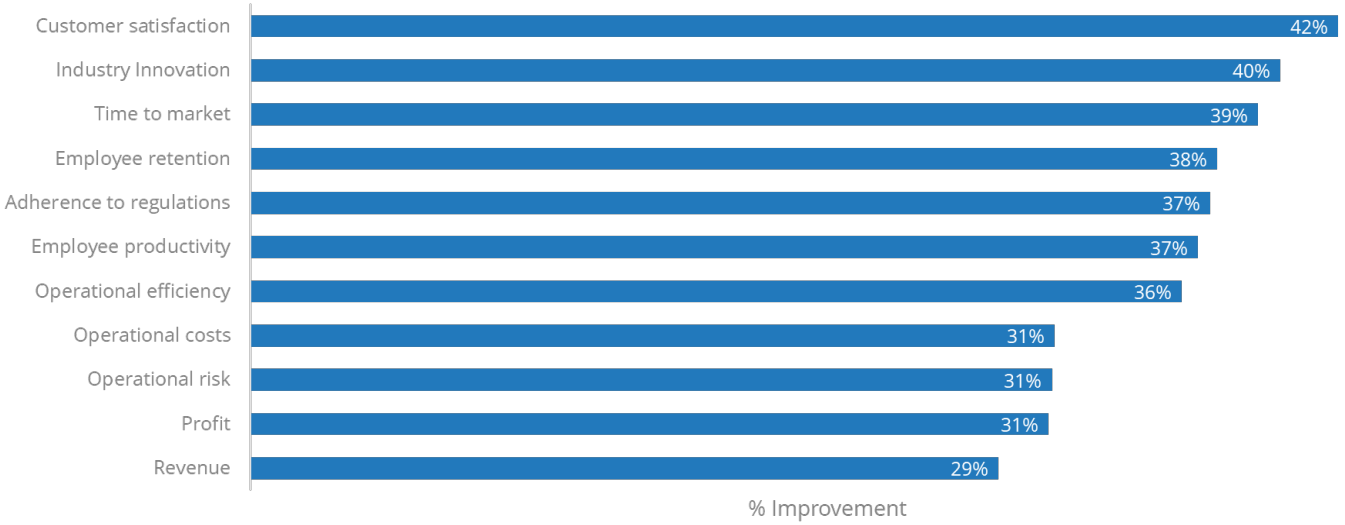
Benefits of Improving Data Integrity

Investing in data technology to improve integrity is a difficult business case for many organizations, as reported by 60% of respondents to the IDC *Data Trust Survey*, but it is worth it. Nearly two-thirds (63%) of the same respondents report that higher levels of data integrity are having positive impacts on business metrics. Figure 2 illustrates the magnitude of

improvements to business metrics. The highest level of improvement was reported in customer satisfaction, at 42%, and the lowest level of improvement was in revenue, but at 29%, which is still a significant improvement for any company.

FIGURE 2: ***Business Metric Improvements***

Q *How much positive impact has the level of data quality and trust had on each of these metrics?*



Source: IDC Data Trust Survey, December 2021, N=500

Conversely, these metrics can also be negatively impacted by lower levels of data integrity as noted by over 11% of respondents, with operational efficiency and costs being cited the most.

Improving intelligence about data provides transparency into the integrity of data being used by the organization, and it builds trust to help build the data culture within an organization. IDC has defined a new generation of data-native worker: Generation Data, or Gen-D for short. People in Gen-D are not just those that have "Data" in their title, but it includes many other roles distributed throughout the organization, all part of the data culture. Capturing and curating centralized but accessible intelligence about data that is easily searchable by data users, regardless of where they are in the organization, or level of technical skills, can improve the data discovery experience. Knowing where the best data is to solve a business problem, where it came from, what level of quality it has, who uses it the most and what it means to the business can go a long way to helping improve data literacy and finding the signal in the noise.

Improving the integrity of data requires intelligence and enrichment to deliver context as it relates to the business process, the regulations it is subject to, and location awareness. Many enterprises enrich data with demographic, descriptive, and behavioral attributes. In addition, every asset, every transaction, every customer, and every worker will have some relationship to the physical world, as people's actions and movements happen in specific places. However, most business analytics have focused on what happened and who was involved -- but have not considered where these events occurred. Understanding the place (and point in time) helps inform the “why's” that business analysts and strategists are trying to answer. In the past, location data has been the domain of a group of specialists tasked to analyze the physical footprint of facilities and the current location of enterprise assets, among other things. Now, people in a broader variety of roles have seen the power of quality location data.

Trends in Improving Data Integrity through Transparency

As business and IT teams work to leverage data as a strategic business asset and create a competitive advantage, they should factor in the following trends that will drive increased need for data transparency and integrity:

Organizations that have implemented DataOps have seen a 49% improvement in their ability to deliver data projects on time.

- **DataOps:** IDC defines DataOps as a combination of technologies and methods with a focus on quality for consistent and continuous delivery of data value. DataOps borrows from the lean and agile principles of DevOps, including continuous development, testing, and deployment of data applications, but adds statistical process control to continuously test data separate from testing of applications. Organizations that have implemented DataOps have seen a 40% reduction in the number data and application exceptions, and a 49% improvement in the ability to deliver data projects on time. Solutions that improve the integrity of data through intelligence and enrichment should also support continuous DataOps methods and technologies.
- **Data observability.** Related to DataOps, data observability is continuous monitoring and testing of data pipelines for operational, behavioral, and measurable anomalies, alerting operational and analytical staff of potential issues, to prevent bad data from being consumed. Data observability can alert on such anomalies as when data volumes are lower or higher than normal, or when there has been a significant spike in data access within a given period, or that values within a data field are falling outside of an acceptable tolerance, or values are an order of magnitude different from normal observation. Data observability needs to rely on data quality, data cataloging, data lineage, and context to identify issues, and guide data practitioners in root cause analysis and issue resolution.
- **Cloud migration.** Organizations have been on a steady pace of migrating data to the cloud, desiring to take advantage of the scalability and elasticity of cloud computing. Data and location intelligence can be leveraged to help organizations make decisions about what data to move to the cloud, and when, including understanding what data may be subject to regional regulations. The use of cloud also increases data distribution, allows for more data diversity, and permits data to be more dynamic because in some cases, fewer controls exist. Centralized intelligence about what data is where, across hybrid and multi-cloud environments, will offer organizations the ability to exercise greater control over data assets.
- **Artificial intelligence automation:** This approach is a necessity in modern data environments. The scale of data distribution, diversity, and dynamics can only be controlled through automation. Artificial intelligence, trained by machine learning, automates the discovery of data quality issues in real-time using anomaly detection, inferring relationships among entities held within data, and automating the collection of intelligence about data to derive context, location, and business meaning. Leveraging third party data for enrichment and reference can improve the quality of data, subsequently improving the outcome of artificial intelligence, trained by machines and data.
- **Environmental, Social, and Governance (ESG):** An organization's response to ESG issues is an increasing concern that can only be addressed through transparency. Transparency into how an organization uses data, relationships between entities within the data, and into any bias that has been built into analytics is critical in meeting ESG reporting requirements. Data reliability and accreditation is the most cited capability that organizations are using today to monitor, manage, and report on ESG risks, according to November 2021 IDC survey on the topic of

governance, risk, and compliance. It's not surprising that data integration, internally and externally, is impacting changes to GRC practices within companies. Within three years, the second-most cited capability that will be required is data ingestion and consolidation, next to the ability to monitor environmental and social activity. Data, and transparency into data, is at the core of importance for ESG initiatives now, and into the future.

Considering Precisely for Improving Data Integrity through Transparency

Precisely's legacy is in high volume data integration at scale, and through the addition of Trillium Software and Pitney Bowes Software and Data business, proven data quality, location intelligence, and data enrichment capabilities are part of the solution offerings. Precisely has recently added Infogix, Winshuttle, and PlaceIQ into its portfolio, bringing needed transparency through data intelligence by adding shared data cataloging, governance, self-service, and mastering capabilities.

The blend of data integrity software and data, differentiates Precisely's ability to integrate, verify, and govern enterprise data – and enrich it with context. The Infogix acquisition also added a rich set of strategic services to the Precisely portfolio, providing develop, design, delivery, and guided maturity of enterprise data strategies.

The modern data environment is not just new data technologies, but it continues to include legacy environments where significant time, effort, and money has been invested. Organizations wanting to improve data integrity are not interested in rip-and-replace, nor can they afford to throw the past out just to support the present and the future. The past is a part of the present and will continue to be a part of the future. Precisely now has a strategic blend of technology and services within its portfolio to meet the needs of organizations with legacy and modern data environments that are wanting to compete in a digital first world.

Challenges

The Alewife fish swim in unison within large schools, making many small fish appear as one large fish, protecting themselves from predators and providing an advantage when competing for food. Turbulence in the water or weaker fish in the school can break the optical illusion and the whole school can suffer. Precisely has assembled a school of products that can logically operate together, appearing to be one solution while functioning independently. Each new acquisition can gain strength from the larger company, but for the school to stay together in the turbulent waters of the market, Precisely needs to focus on integration towards a unified platform and provide an exponential model of value for customers.

The new Precisely Data Integrity Suite (announced at Trust '22) is bringing tighter integration across the portfolio with seven software-as-a-service modules with a modern user experience and shared metadata management. The Suite promises to simplify the adoption of data integrity capabilities and speed the customer's ability to trust their data. Precisely is delivering on the vision it has painted for the market of integrating their capabilities into a modular, interoperable suite that delivers exponential value by working together seamlessly.

Conclusion

Building trust is foundational in how organizations are creating enterprise intelligence to compete in a digital-first world, but there is a significant gap between expectations and reality in the level of data trust. DataOps connects data culture with data and technology, introduces continuous application and data testing to improve trust in both the data, and the outcome, helping to close the gap. Data intelligence provides the necessary transparency, backed by enrichment,

providing the necessary context to deliver integrity. Precisely has been closing gaps in its own portfolio, to help its customers close the data trust gap, and to the extent it can overcome the challenges noted above, it has significant opportunity for success.

About the Analyst

Stewart Bond, *Research Director, Data Integration and Data Intelligence Software*

Stewart Bond is Research Director of IDC's Data Integration and Intelligence Software service. Mr. Bond's core research coverage includes watching emerging trends that are shaping and changing data movement, ingestion, transformation, mastering, cleansing and consumption in the era of digital transformation.

MESSAGE FROM THE SPONSOR

More About Precisely

Precisely is a global leader in data integrity, providing accuracy, consistency, and context in data for 12,000 customers in more than 100 countries, including 99 of the Fortune 100. Precisely's data integration, data quality, data governance, location intelligence, and data enrichment products power better business decisions to create better outcomes. Learn more at www.precisely.com.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
idc-insights-community.com
www.idc.com