Driving effective data governance for improved quality and analytics

Ever since the movie *Moneyball* popularized number crunching to win at baseball, it seems like everyone understands the benefits of data science. Whether it’s rendering new insights, improving decision making, or driving better business outcomes, enthusiasm for unlocking the power of big data has never been greater. With advanced analytics, AI, and machine learning, the ability of enterprises to optimize ever-greater amounts of data now ranks among the most important of all strategic endeavors.

Data is a strategic asset, and it’s essential to have a data governance approach to solving data quality issues. In Core Services Engineering and Operations (CSEO), we created a system that allows business groups at Microsoft to operate independently, while still driving data quality.

Simply put, data is the new currency of digital transformation—and with that power comes responsibility. “We actually have to think of data like code,” says Damon Buono, senior director for Enterprise Information Management at Microsoft. “The way forward is a data-driven culture—where everyone understands and agrees that data is a strategic asset.”

Figure 1 shows how data-driven decision-making cuts across organizational and functional boundaries.

**Driving a culture of data management**

Forging a modern data management strategy across disparate groups doesn’t come easy. Unlike top-down enterprises, Microsoft is famous for its decentralized and federated culture, rendering cross-company organizational change harder—though not impossible.

“There was no underlying understanding for why we need to focus on data management. Therefore, there was no overarching data governance,” Buono recalls. “If you don’t have a foundation of data quality and data ownership, you are just going to see bad analytics at a faster pace.”
It’s a governance gap that accentuates the risks inherent in the increasing volume and complexity of structured and unstructured data. We experimented with different approaches, like forming a centralized data governance board and initiating topic-based data governance. None yielded the results we were looking for.

“We realized over time that to drive cultural change, we have to drive data management visibility at the top for the quality and ownership questions to start being asked,” Buono says.

CSEO is chartered with driving Enterprise Information Management, which it defines as an “...integrative discipline for structuring, describing, and governing information assets across organizational and technological boundaries to improve efficiency, promote transparency, and enable business insight.”

What does that look like? Fundamentally, it involves cleaning up data today and implementing a framework for the future so that data is ready for use by business groups across Microsoft. Specifically, it includes assigning and implementing accountability and stewardship, ensuring accuracy and discoverability, and implementing role-based access to data.

Assigning and implementing accountability and stewardship

The first hurdle was establishing the new role of data steward for each of our organizations, with a primary focus on data quality. Our core mission begins with defining, communicating, evangelizing, and—ultimately—enforcing data quality standards across a spectrum of use cases. These range from lightweight and informal oversight of experimental data to stringent and active compliance monitoring of high-value information intended for customers, partners, or stockholders.

Like the traditional discipline of software testing with bugs filed against code, our data stewards perform a similar quality assurance function, managing a quality bar with standardized rules and data quality indicators within an overall measuring system. Data failing to meet the bar will prompt data stewards to investigate the nature of the quality gap and then work with teams to drive remedies appropriate to the intended use.

Rise of the data steward

It falls to data stewards to ensure that data management and governance policies are implemented and fully operationalized. Central to their “feet on the ground” mission is the role of trusted advisor, working directly with teams to solve immediate and long-term data management challenges. For example, this could mean helping implement controls that validate data during online transactions to comply with data quality standards and goals. When issues arise, data stewards own the root cause analysis to triage or escalate accordingly. Data stewards also manage data access, as explained later.

As shown in Figure 2, roles and responsibilities change across organizational levels in a manner that resembles governmental decision making. The local level of governance happens at the individual team level. The state level represents governance occurring at the organization level and the federal level denotes the overarching governance across all organizations. At the base of the pyramid, each business unit empowers stewards to manage the data that belongs solely to their functional unit. When data touches multiple functional or business units, ownership shifts to data domain stewards, driving overall data excellence. Any issues that remain unresolved get escalated to corporate decision makers responsible for enterprise strategy or—if necessary—to top executive levels.
Figure 2. Driving the data quality lifecycle from enterprise strategy to operations
Ensuring accuracy and discoverability

Ensuring that data across CSEO is accurate, discoverable, and accessible sits at the heart of our journey to raise awareness around data quality. It begins with a commitment to the highest quality data underlying critical business functions. Improving overall data quality begins with prioritizing data with the biggest business impact. Equally important is driving taxonomic consistency to ensure that different groups reference things in the same way, potentially boosting the overall customer experience. Figure 3 shows the full spectrum of data quality improvement activities, beginning with identifying problem areas through implementing preventative measures to avoid problem recurrence.

Figure 3. Driving the data quality lifecycle

Our efforts to drive data quality have evolved into the following best practices:

1. **Collect evidence of a problem.** Our data stewards continually monitor systems for data management anomalies and prioritize fixing issues with the greatest opportunity costs and business impacts.

2. **Explain problem.** They develop “problem statements” using data and metrics as the key to effectively communicating issue scope at the executive level.

3. **Spotlight problem with executives.** We hold regular data quality forums with senior leaders and executives to highlight and discuss the most critical issues.

4. **Drive resolution.** With executive approval, they secure the appropriate resources and expertise needed to drive resolution across teams, helping stakeholders understand the costs accruing from the data management gaps.

5. **Fix underlying gaps.** Discovering inconsistencies in our data management processes helps prioritize improvements to fill these gaps.

6. **Preventative action.** Regularly highlighting data quality issues with senior leaders and executives prompts teams to scrutinize their procedures and implement remedies to reduce the risk of recurring issues.
Implementing role-based access to data

Data that originates or is collected at the business unit level is owned by the business unit, not us. Implementing appropriate access happens at the business unit level in coordination with companywide IT policies. At CSEO, data access is defined according to specific roles. Our goal is to provide consistently managed processes around data security and integrity across the following roles:

- **Business data owners** have direct ownership over data for specific functions within a business unit. As part of the team implementing governance processes across internal groups, they coordinate with data stewards and data custodians to drive policies around the definition, access, and use of data. Business data owners are ultimately accountable for data quality and oversee regulatory compliance.

- **Data stewards** create, maintain, and implement policies and business rules to manage data and metadata in specific functional areas such as finance, sales, or operations. Stewards collaborate with domain data stewards and business owners to design data and usage processes, maintaining consistency with enterprise-wide data handling policies. Together with data custodians, data stewards manage data access along a “whitelisting” model, only granting permissions to users or groups with the delegated authority to access data for its intended use.

- **Domain data stewards** drive data quality issues for specific areas like customer master data spanning lines of business. They coordinate with local data stewards across one or more functional areas, but not the entire enterprise.

- **Data custodians** perform an essential data engineering role in the security and management of data as part of protecting enterprise data assets. They maintain data management infrastructure, coordinating with stewards and data owners to ensure compliance with data access and security policies. Whether they’re embedded directly within individual business units or given a broader remit within CSEO, they help drive consistent data handling procedures across the company.

- **Data consumers** are users or systems accessing data via applications, reports, and dashboards. They provide requirements to data owners and stewards on data quality needs and are also responsible for adhering to data-use policies.

- **Data producers** are users or systems who create data. They provide feedback to data owners and stewards on data input standards, such as implementing data entry controls to create quality data.

Next steps

Addressing data quality issues that span data stores across multiple organizations—with no defined owner—ranks among our highest priorities. To meet this challenge, the Enterprise Information Management team drives troubleshooting with vested executive authority to implement specific solutions. By using metrics and data to identify problems and articulate business impacts, our data management team is now getting traction dispatching the right individuals or teams to get the job done.

Already, the difference is being felt in major areas, including sales, marketing, finance, and services. No longer is the team stuck in reactive mode and logjammed with fixing data. Now they’re implementing end-to-end data management practices during the design/build phase of strategic programs—instead of retrofitting data later.

Microsoft is hardly alone in this endeavor; almost every large company is in the process of transitioning from old warehouse models to hyperscale, big data infrastructure in the cloud. There is now a growing body of work across industries around implementing and operationalizing data governance goals.

Working with Gartner and others, we’ve developed an internal data fundamentals playbook that incorporates industry best practices. Next steps include completing a maturity assessment toolkit leading to a measurable data health index.

“We want to evolve and learn by deploying these fundamentals first in a few strategic programs,” says Kandy Samy, a director in the Enterprise and Information Management team. “Then we’ll roll them out more broadly with a long-term vision of automating the data management practices and controls for easier adoption.”
Becoming a data-driven culture

Even as each group operates independently, we found that consistent data management across departmental and organizational boundaries is important. Technology alone cannot solve the data quality problem. People must collaborate and be involved. Data stewards are a key component of improved and operationalized data quality, while the responsibility to treat data as a core strategic asset extends to everyone across the company.

Underlying this integrated approach is a growing cultural awareness around managing data as a key strategic asset. As more resources are dedicated to driving data quality—from the executive levels to the field—we’re seeing a snowball effect that’s driving momentum to integrate data quality criteria into service development lifecycles across the company.

Data-driven decision making is a key part of the digital transformation. Ensuring that data is accurate, discoverable, accessible, linked, and well-managed sets us up for success. With quality data, we’ll have the necessary insights to drive our business forward using the full power of advanced analytics, machine learning, and AI.

For more information

Microsoft IT Showcase

microsoft.com/its