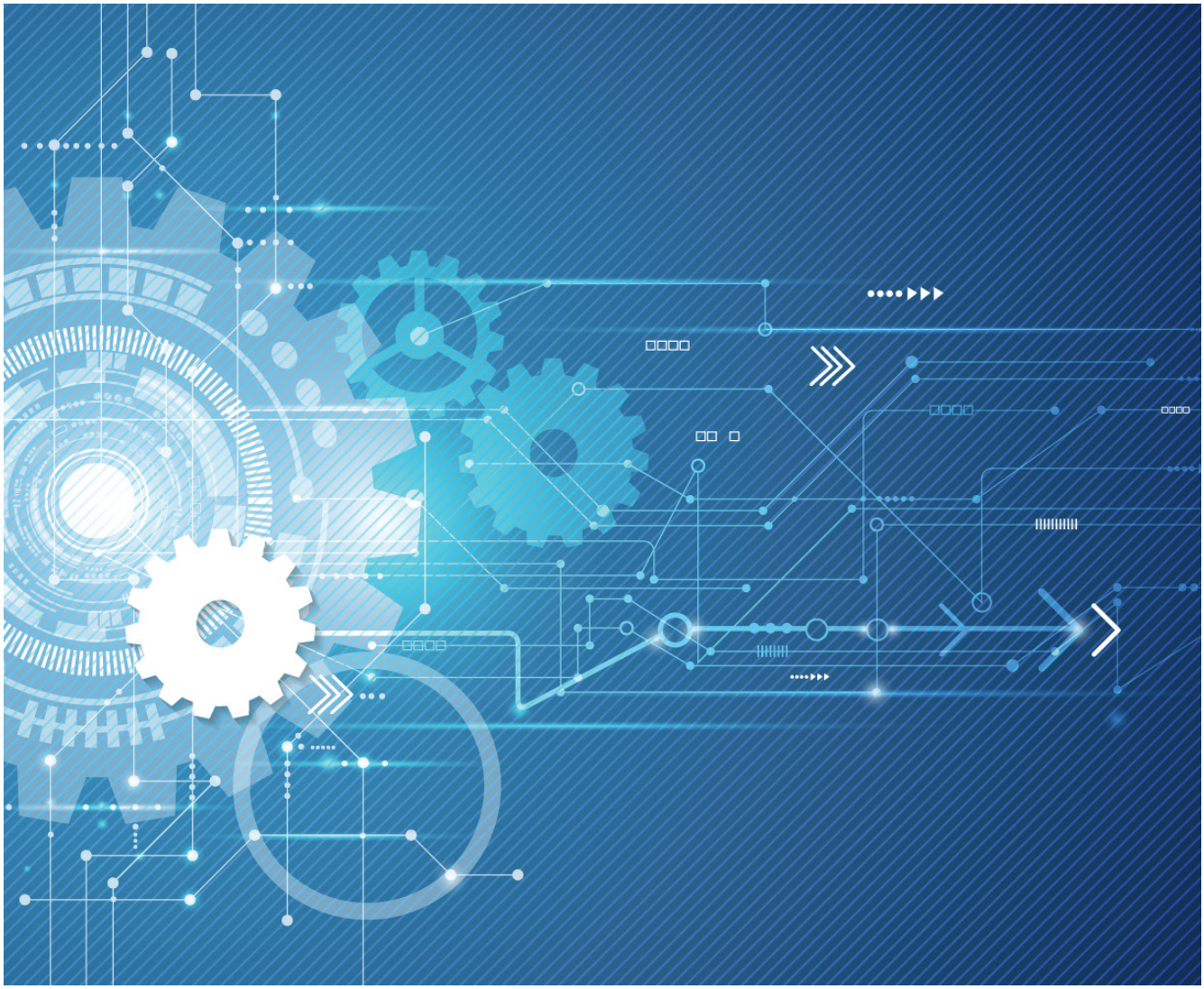
The 2020 State of Data Governance and Automation

Many organizations have started their data governance journeys to achieve data intelligence, but they have not automated their data operations to create sustainable and repeatable practices. Without an accurate, high-quality, real-time enterprise data pipeline, it can be difficult to uncover the necessary insights to make optimal business decisions.



erwin®
the data governance company

DATAVERSITY®

Executive Summary

Data governance is a complex but critical practice, and most enterprises have encountered difficulty in mastering all its requirements, including executive sponsorship, policy and process implementation, and technological support. Operationalizing data harvesting, cataloging, lineage, mapping, code generation and impact analysis presents tremendous challenges, especially when relying on traditional manual practices.

DATAVERSITY® partnered with erwin® in November 2019 to survey business technology professionals at organizations of various sizes about their data governance and intelligence operations, focusing on the role of automation in these efforts.

This research provides a holistic view into the needs of a broad cross section of stakeholders across a wide range of industries and organization sizes.

The survey had 15 questions, including four general demographics questions. Respondents were recruited primarily through a marketing email campaign. They received links to *The 2020 State of Data Governance and Automation Survey* and did not have a time limit answering the questions. No compensation was given for participation. The results and conclusions contained within this paper are representative of the survey respondents only, a group predisposed to interest in data governance.

Most of the 263 survey respondents are in data and/or information architecture roles (29 percent), followed closely by information/data governance professionals (25 percent). Business intelligence/analysts are the third-most represented job function (14 percent). Other roles include data scientists at 4 percent, executive management at 6 percent, and many others at 3 percent or less.

Among industries, government and technology are the two most-represented sectors at 12 percent each. Consulting was at 11 percent, with education, finance and insurance at around 7 to 8 percent each.

About one-fifth of respondents work in businesses with 101 to 1,000 employees (21 percent), with 1,001 to 5,000 employee businesses coming in second at 19 percent, and 10,001 to 50,000 employees coming in third. However, all ranges of business size were represented in the survey, with the lowest being 11 to 100 employees at 9 percent.

This erwin *2020 State of Data Governance and Automation* report is a follow-up to the *2018 State of Data Governance* report the company sponsored to explore attitudes on data governance ahead of the implementation of the General Data Protection Regulation (GDPR). It revealed considerable awareness of the importance of data governance but also several factors for concern – primarily an incomplete understanding of data governance and a lack of both budget and executive support.

Data Governance and Intelligence Maturity

This report shows a slowdown in fully implementing data governance as a core organizational capability – less than 15 percent have achieved that mark compared to 31 percent in 2018 [Figure 1]. But that's not to say that businesses haven't taken some important steps. In fact, more than 50 percent say they have deployed metadata management, data analytics and data quality solutions, and close to 50 percent have deployed data catalogs and business glossaries.

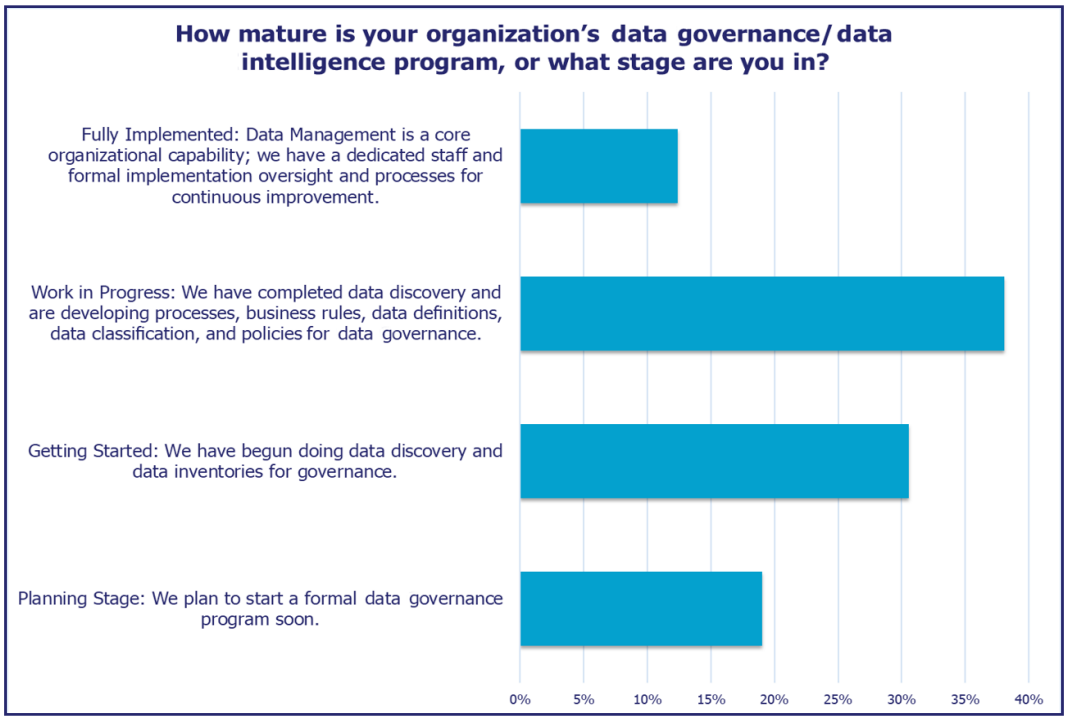


Figure 1: Data Governance and Data Intelligence Maturity

That said, deploying individual data governance elements does not constitute a strategy, much less a sustainable program, which is why this year's report explores the automation of data operations. And as we suspected, organizations seem to procrastinate in automating these processes and therefore aren't positioned to achieve data governance and intelligence at speed or scale.

Data Preparation, Governance and Intelligence Solutions Are at Work

In the last survey, 98 percent of the participating organizations considered data governance important or critically important. However, while they recognized the value of data – with 64 percent of respondents indicating they viewed data assets as somewhat or much more valuable to the organization than physical assets – 46 percent had no formal governance strategy in place.

Back then, more than 4 in 10 of the organizations in the survey were already managing more than 1 petabyte of data. Given the size of today's big data environment – with 2.7 zettabytes of data now in the digital universe – it's good to see that 67 percent of this year's report respondents have deployed data preparation, data governance or data intelligence solutions [Figure 2].

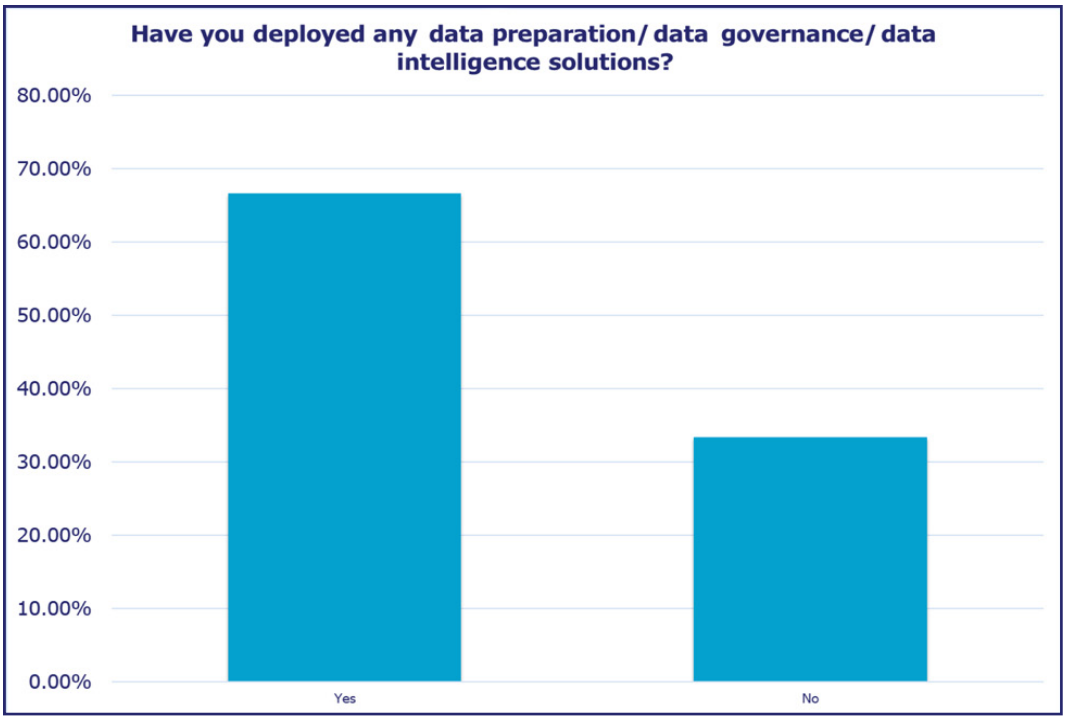


Figure 2: Data Prep, Governance, Intelligence Deployment

When asked to specify which types of solutions in the above categories had been deployed, data analytics topped the list with 65 percent, followed by metadata management (59 percent) and data quality (59 percent) [Figure 3]. Data catalogs and business glossaries round out the list of deployed solutions.

Of course, the effectiveness of analytics deployments is threatened when the enterprise hasn't first integrated and synchronized data.

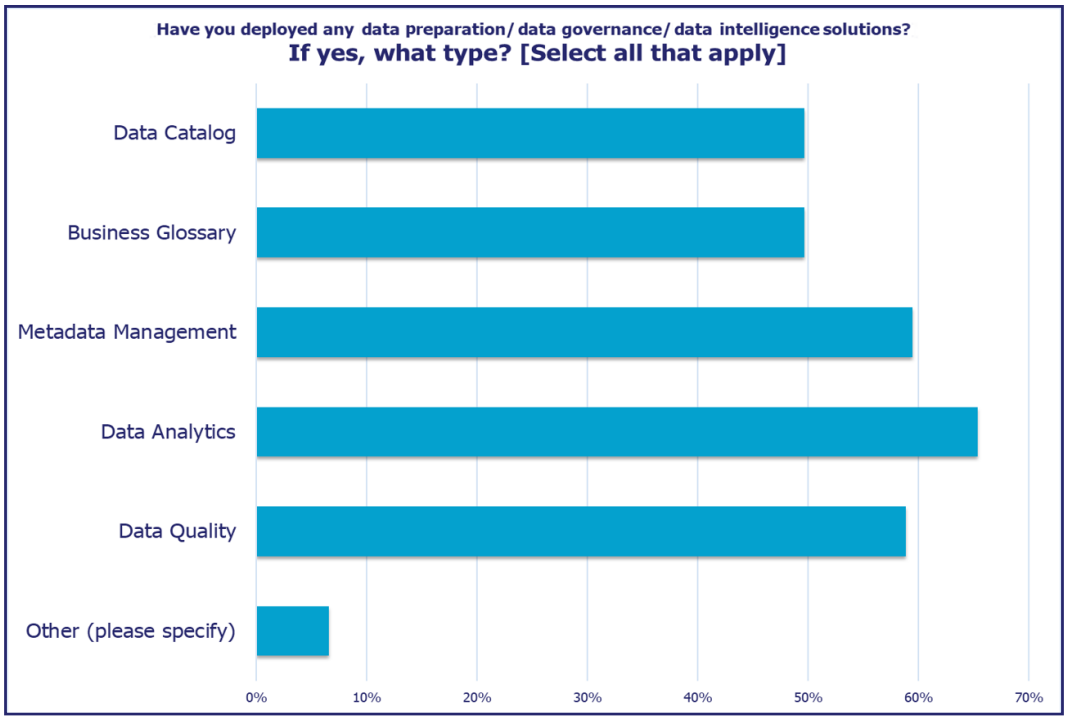


Figure 3: Platform Type

Of course, the effectiveness of analytics deployments is threatened when the enterprise hasn't first integrated and synchronized data. It must be:

- Complete— all required elements should be accounted for;
- Consistent—no conflicting information between data sets and systems;
- Linked—connected together by relationships across tables in different systems;
- And clean—duplicate entries must be eliminated.

There's some recognition of this because 58 percent of respondents report that understanding the quality of source data is the third most serious bottleneck in their organization's data value chain [Figure 4]. Automating and matching business terms with data assets and documenting lineage down to the column level are critical steps to optimizing data quality.

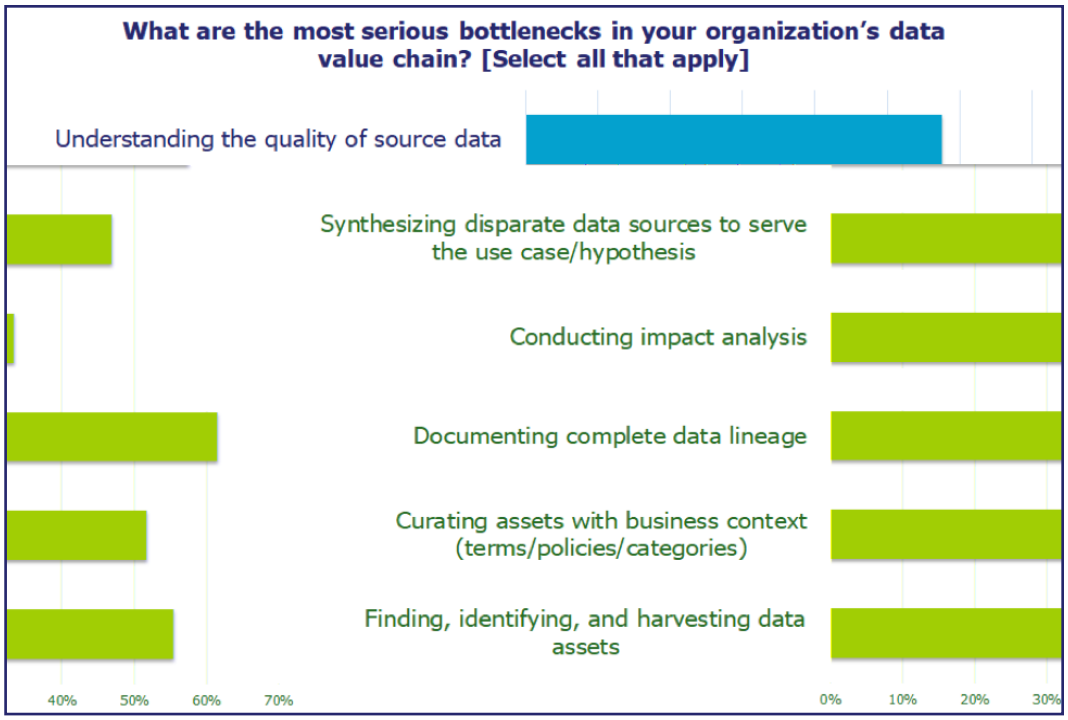


Figure 4: Most Serious Bottlenecks

Data Governance Drivers Shift After GDPR

Respondents in the 2018 report ranked regulatory compliance as the No. 1 reason to implement data governance (60 percent). That's not surprising given GDPR's effective date was looming and required all organizations handling the personal data of European Union residents to implement specific privacy and security controls or face significant penalties. Customer trust/satisfaction ranked second at 49 percent, followed by better decision-making at 45 percent.

In this latest survey, respondents say better decision-making is the primary driver of data governance (62 percent) [Figure 5]. Analytics is the secondary driver (51 percent), and regulatory compliance ranks third (48 percent).

Digital transformation and data standards/uniformity round out the top five data governance drivers with 37 percent and 36 percent, respectively. Digital transformation continues to be an important trend as businesses further embrace the opportunity to reinvent themselves to stay ahead of evolving customer needs and better compete with disruptive start-ups in their industries.

The maturity of data governance programs appears to have regressed with only 12 percent of respondents for this year's report declaring they are fully implemented as compared to 31 percent in 2018.

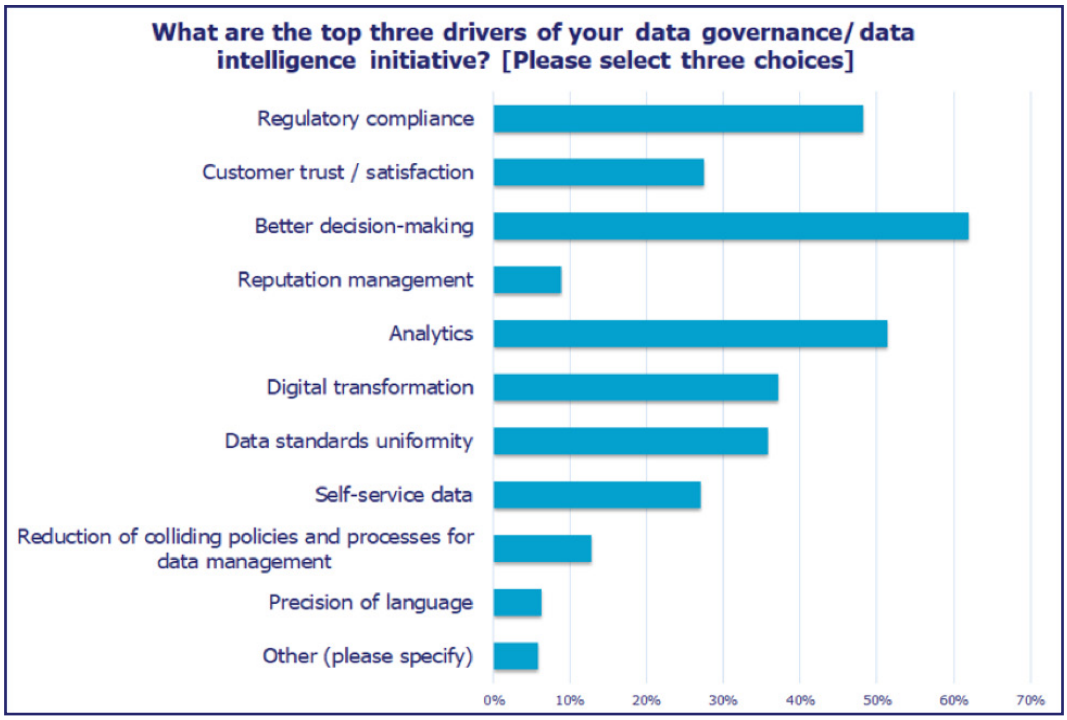


Figure 5: Top Governance and Intelligence Drivers

The shift to better decision-making as the primary driver may be indicative of organizations using data to improve their overall performance, rather than merely ticking off a compliance checkbox. Acknowledging data governance is about more than regulatory compliance and risk management may indicate that companies are becoming more confident in their data. That doesn't mean businesses aren't still aggressively engaged in ensuring compliance, but they may be comfortable enough to focus more attention on data-driven, revenue-generating initiatives, such as securely monetizing their information.

A word of caution for companies aiming to strike a better balance between regulation and monetization: More regulations are coming down the pike, including the GDPR 2.0 ePrivacy Regulation for protecting personal privacy across electronic communications.

Program Maturity Steps Backward

The maturity of data governance programs appears to have regressed with only 12 percent of respondents for this year's report declaring they are fully implemented [Figure 1] as compared to 31 percent in 2018. As concerning, 38 percent of current survey participants say their data governance programs are a work in progress and 31 percent are just getting started as compared to 42 percent and 21 percent in the prior survey.

Consider this possibility for the slowdown: In the last survey, 42 percent of participants said that "understanding the right approach" to data governance was one of the biggest challenges to their initiatives. Businesses still may be grappling with that and are now taking a closer look at whether they need to rethink the plans they've made – and perhaps even partially implemented.

Businesses are very much aware of the growing complexity of the data governance process. So, it's possible they want to strengthen what they've already started in order to account for the fact that every year brings:

- More data to handle (much of it unstructured)
- More sources (like IoT)
- More points of integration
- More regulatory compliance requirements

That's a lot to take in before continuing to build a data governance program or starting a new one.

Data Governance Blocked by Bottlenecks

The *2018 State of Data Governance* report revealed cost as the most significant obstacle to implementing a data governance initiative (58 percent).

Two years later, as we hone in on day-to-day operations and bottlenecks to effectiveness, 25 percent of respondents say length of project/delivery time is the most significant challenge. Data quality/accuracy is next at 24 percent, time to value at 16 percent, and reliance on developer and other technical resources at 13 percent [Figure 6].

Figure 6: Most Significant Challenge

…documenting complete data lineage is the number one bottleneck according to 62 percent of respondents.

Cost is the least of the challenges at only 10 percent. This may indicate that expectations of companies spending more on data governance – shown in the results of the previous survey – are being met. In the 2018 report, 15 percent of companies expected budgets to increase by more than 10 percent, and 17 percent said their budgets would grow by up to 10 percent in the following 12 months.

As was displayed in Figure 4, documenting complete data lineage is the No. 1 bottleneck according to 62 percent of respondents. Understanding the quality of source data is the next most serious bottleneck (58 percent); followed by finding, identifying, and harvesting data (55 percent); and curating assets with business context (52 percent).

The survey reveals that all but two of the possible bottlenecks were marked by more than 50 percent of respondents, in each case as obstacles to supporting the data value chain. Clearly, there's a massive need for a data governance framework to keep these obstacles from stymying enterprise innovation.

Current Appetite for Automation

Close to 70 percent of respondents say they spend an average of 10 or more hours per week on data-related activities, and most of that time is spent searching for and preparing data [Figures 7 and 8]. There may well be a connection between the long hours worked on data-related activities and the fact that the majority of respondents' data operations are not automated or only mildly automated.

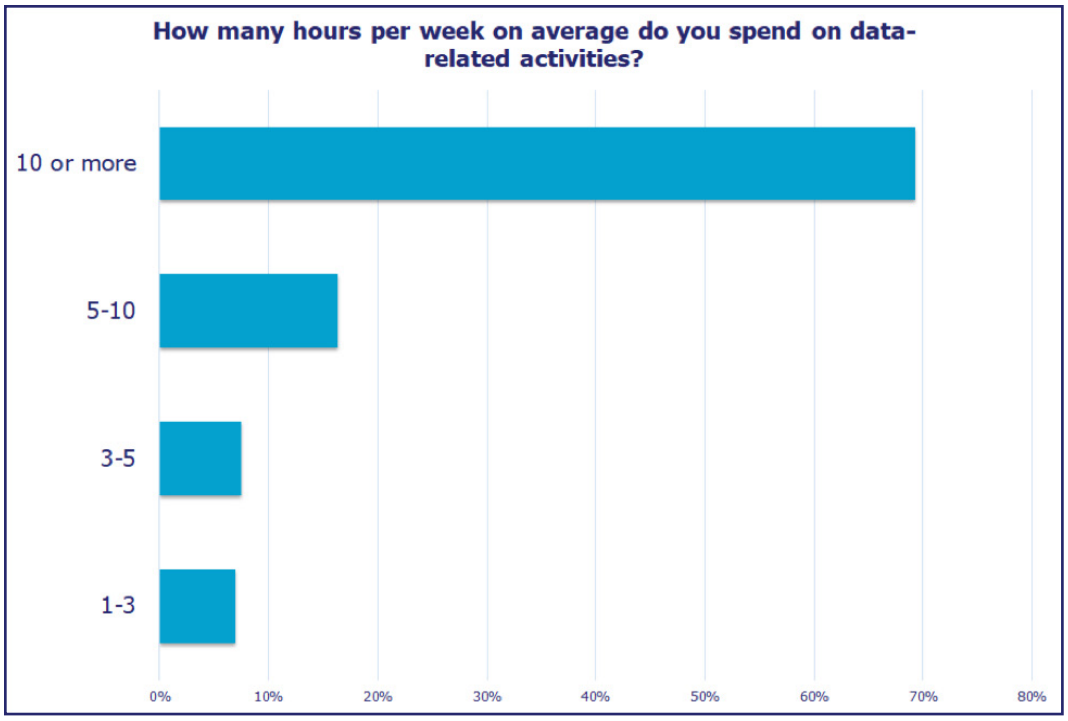


Figure 7: Hours of Work

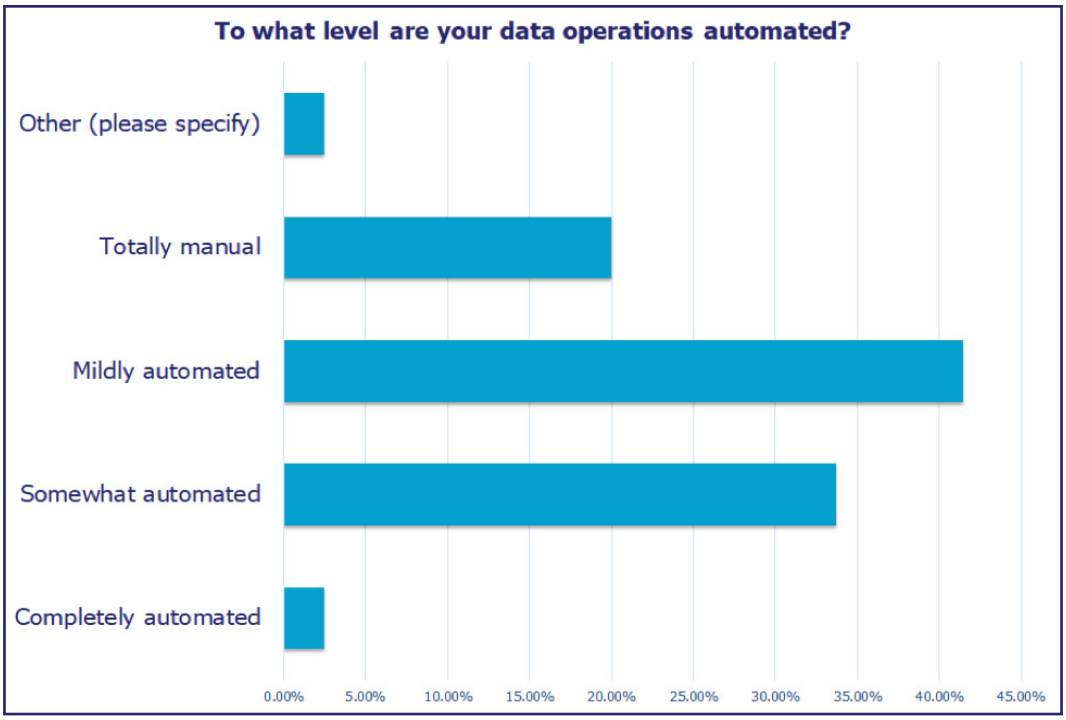


Figure 8: Level of Automation

It's not surprising, then, that 61 percent of respondents report that data cataloging – for finding, understanding and using data strategically – would be one of the most valuable data operations to automate. But so far, only 26 percent have automated it [Figures 9 and 10].

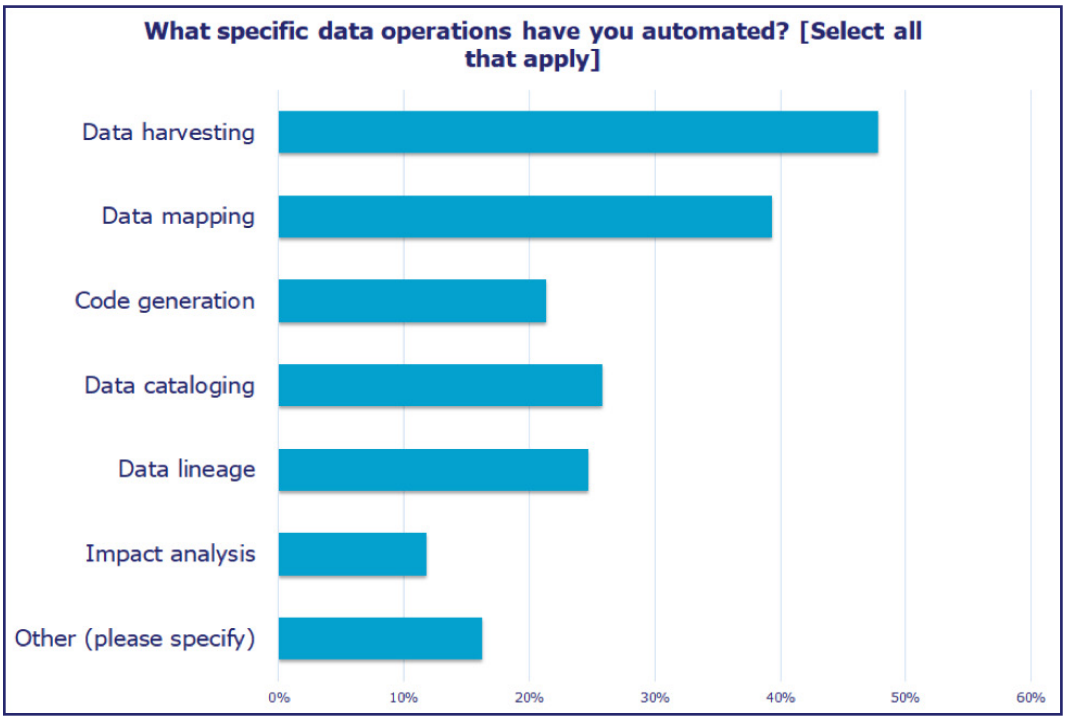


Figure 9: Data Operations Automation

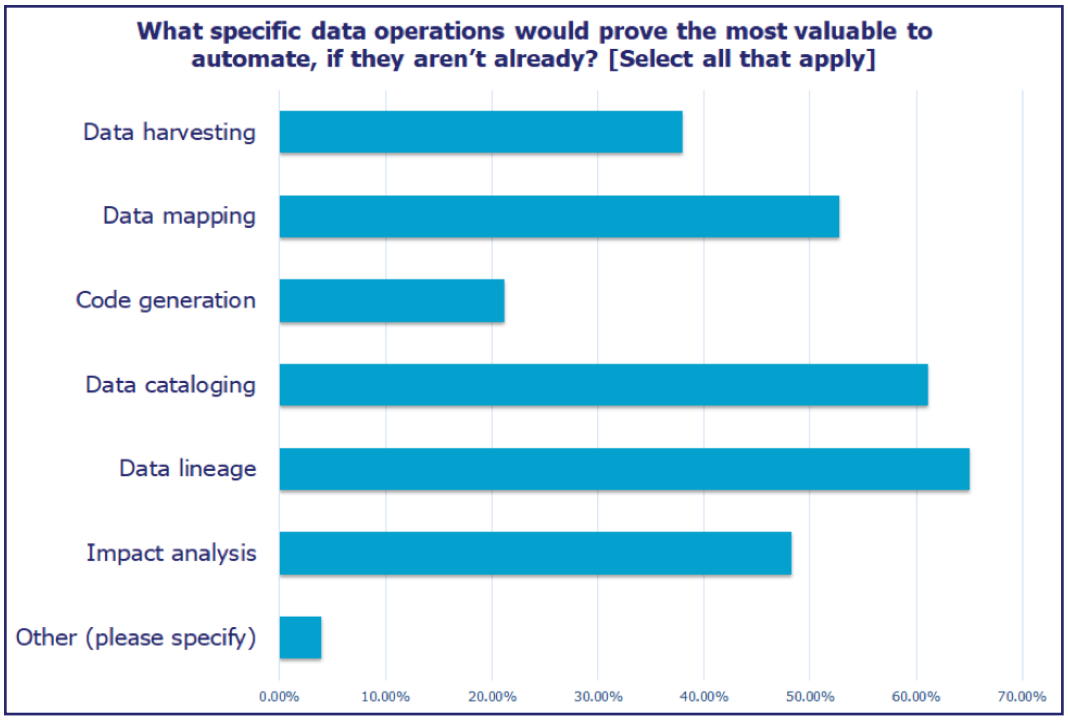


Figure 10: Operations Valuable to Automate

Data cataloging and impact analysis are two sides of the same data prep and governance coin. And like data cataloging, there is a significant spread between how many respondents already have automated the impact analysis process (12 percent) and how many haven't but say it would provide the most value for them to automate (48 percent).

Documenting complete data lineage is currently the data operation with the largest percentage spread between its current level of automation (25 percent) and being seen as the most valuable operation to automate (65 percent). Clearly, organizations want to move the needle there.

Data mapping follows data cataloging and data lineage in terms of the most valuable data operation to automate at 53 percent, with 39 percent having done so already.

Conclusion

The data from our survey suggests that companies are still grappling with the challenges of data governance, challenges that will only get worse as companies collect increasing amounts of data. Unless data is well-governed, downstream data analysts and data scientists will not be able to generate significant value from it.

Availability, quality, consistency, usability and reduced latency are all requirements at the heart of successful data governance, as well as the provisioning of a strategic data pipeline. These are also the benefits that can be realized through automation, whether it is rules-based or steeped in artificial intelligence and machine learning.

Businesses still depend too much on manual approaches to data management. Based on these results, we offer the following recommendations:

1. Don't ignore the complexity of data lineage, which is about more than tracking data flows at the lowest level in the lifecycle. It's a risky endeavor to support data lineage using a manual approach, and businesses that attempt it that way will find that it's not sustainable given data's constant movement from one place to another via multiple routes – and doing it correctly down to the column level. Adopting automated end-to-end lineage makes it possible to view data movement from the source to reporting structures, providing a comprehensive and detailed view of data in motion.
2. Automate code generation to alleviate the need for developers to hand code connections from data sources to target schema. Mapping data elements to their sources within a single repository to determine data lineage and harmonize data integration across platforms reduces the need for specialized, technical resources with knowledge of ETL and database procedural code. It also makes it easier for business analysts, data architects, ETL developers, testers and project managers to collaborate for faster decision-making.
3. Use an integrated impact analysis solution to automate data due diligence for IT and deliver operational intelligence to the business. Business users benefit from automating impact analysis to better examine value and prioritize individual data sets. Impact analysis has equal importance to IT for automatically tracking changes and understanding how data from one system feeds other systems and reports. This is an aspect of data lineage, created from technical metadata, ensuring that nothing "breaks" along the change train.
4. Put data quality first. Users must have confidence in the data they use for analytics. Automating and matching business terms with data assets and documenting lineage down to the column level are critical to good decision-making. If this approach hasn't been the case to date, enterprises should take a few steps back to review data quality measures before jumping into automating data analytics.
5. Catalog data using a solution with a broad set of metadata connectors so all data sources can be leveraged. This includes big data, ETL platforms, BI reports, modeling tools, mainframe, and relational data as well as data from many other types of systems. Don't settle for a data catalog from an emerging vendor that only supports a narrow swath of newer technologies, and don't rely on a catalog from a legacy provider that may supply only connectors for standard, more mature data sources.

6. Stress data literacy to ensure data assets are used strategically. Forty-seven percent of businesses have trouble synthesizing data sources for that purpose. Automation expedites the benefits of data cataloging. Curated internal and external datasets for a range of content authors doubles business benefits and ensures effective management and monetization of data assets in the long-term if linked to broader data governance, data quality and metadata management initiatives. There's a clear connection to data literacy here because of its foundation in business glossaries and socializing data so that all stakeholders can view and understand it within the context of their roles.

   This answers the need business and IT users have to discover what data is available and understand what it means in common, standardized terms given that these data elements may mean different things to different parts of the organization. The automation of common curation processes – such as name matching, categorization and association – will optimize governance as well as data pipeline and preparation efforts.

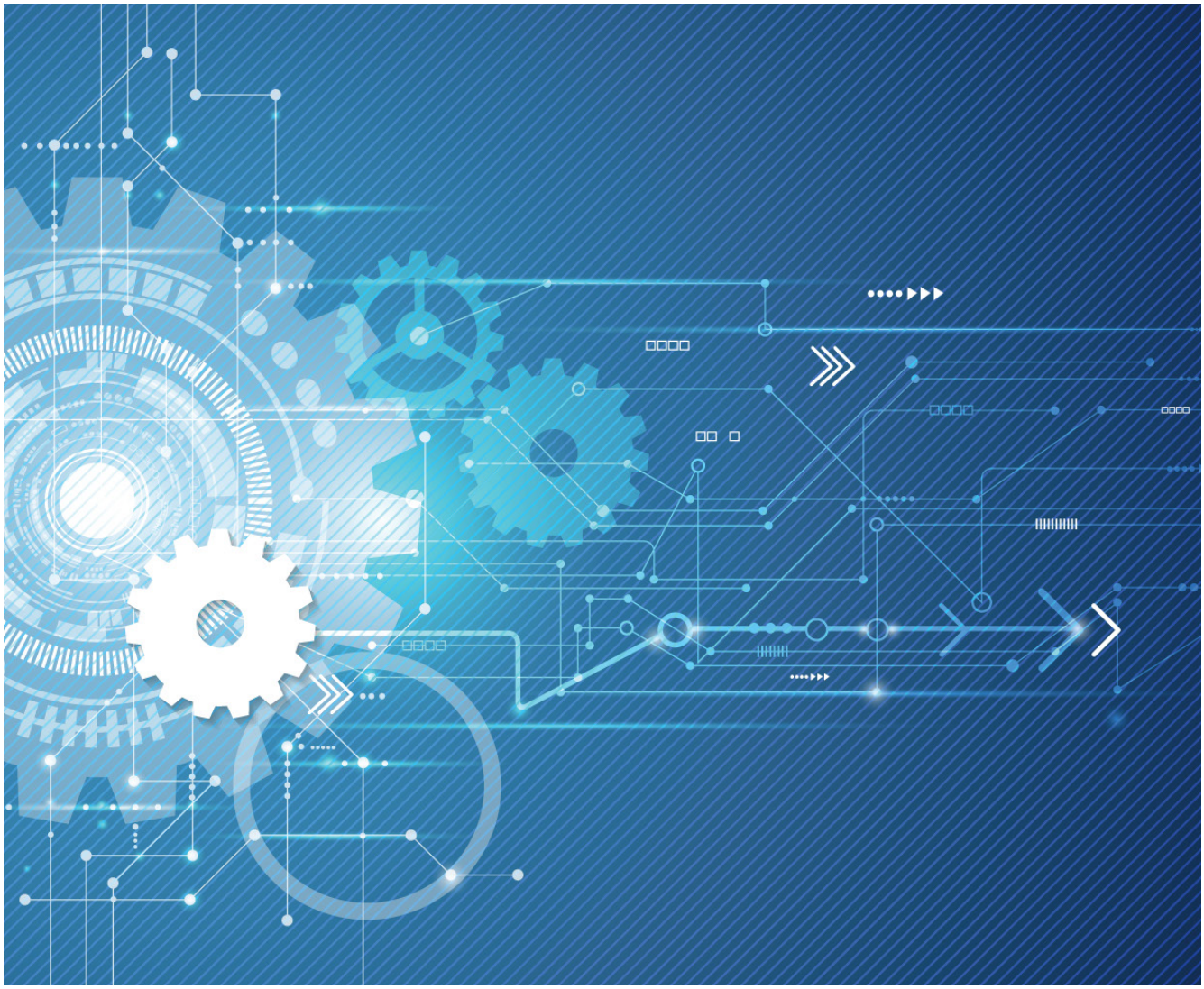
7. Make automation the norm across all data governance processes. Automation was not necessarily a prominent factor in promoting data prep and governance for either 2018 or 2020 report respondents. Too many companies are still living in a world where data governance is a high-level mandate and not a practically implemented one.

   To fully realize the advantages of data governance and the power of data intelligence, data operations must be automated across the board. Without automated data management, the governance housekeeping load on the business will be so great that data quality will inevitably suffer. Being able to account for all enterprise data and resolve disparity in data sources and silos using manual approaches is wishful thinking.

8. Craft your data governance strategy before making any investments. Gather multiple stakeholders—both business and IT—with multiple viewpoints to discover where their needs mesh and where they diverge and what represents the greatest pain points to the business. Solve for these first, but build buy-in by creating a layered, comprehensive strategy that ultimately will address most issues. From there, it's on to matching your needs to an automated data governance solution that squares with business and IT – both for immediate requirements and future plans.

erwin®
the data governance company

As the data governance company, erwin provides enterprise modeling and data intelligence software to help customers discover, understand, govern and socialize their data to mitigate risks and realize results. The erwin EDGE platform facilitates IT and business collaboration in driving actionable insights, agile innovation, risk management and business transformation. We help government agencies, financial institutions, healthcare companies and other enterprises around the world unlock their potential by maximizing the security, quality and value of their data assets. To learn more, please visit erwin.com.



erwin®
the data governance company

DATAVERSITY®