

# UNFINISHED BUSINESS: TAKING ON THE DATA QUALITY CHALLENGE IN THE AGE OF AI

By Joseph McKendrick, Research Analyst  
Produced by Unisphere Research,  
a Division of Information Today, Inc.  
November 2023

Sponsored by

**melissa**

Produced by

 **UNISPHERE**  
RESEARCH

---

**TABLE OF CONTENTS**

*EXECUTIVE SUMMARY* ..... **3**

*THE GROWING DATA CONFIDENCE GAP* ..... **4**

*AI SHINES ATTENTION ON DATA QUALITY* ..... **7**

*EYE OFF THE BALL?* ..... **11**

*MORE DATA IN THE CLOUD, BUT NOT A PANACEA* ..... **17**

*LESS MANUAL, MORE AUTOMATED DATA ENTRY* ..... **21**

*RECOMMENDATIONS* ..... **26**

*DEMOGRAPHICS* ..... **27**

## EXECUTIVE SUMMARY

Data quality has never been as important as it is now. Businesses are aggressively moving into artificial intelligence (AI)—both generative and operational—which requires massive amounts of accurate and timely data. The rise of large language models (LLMs)—both publicly available as well as contained within enterprises—to support business decision making and customer communications means data is being pressed into service in new and highly demanding ways as training data and real-time streaming feeds. The models need to support growing demand for intelligent customer-facing applications such as chatbots and conversational interfaces, as well as intelligent assistants for internal enterprise operations.

One industry survey shows that more than one-third of organizations (34%) are looking at integrating LLMs to add greater intelligence to their applications and systems. Data is also being pressed into service behind the scenes to support operational AI behind automated systems, robotics, and devices. There are two sides to the rise of AI. At the same time it is exacerbating the need for data quality, AI-based tools are emerging to help address overall data quality.

Add the growing push into data monetization—leveraging the value of data assets—as well as reliance on networked edge devices, and it's evident that organizations depend on data for growth and the ability to compete. Data lakes and data lakehouses pose challenges, as they end up as repositories for unvetted data from many sources.

However, ask data administrators and managers about the state of their data, and it becomes evident that enterprises are ill-prepared for the times ahead. Confidence in the integrity, accuracy, and trustworthiness of data has been falling in recent years, a new survey of 202 data decision makers by Unisphere Research, a division of Information Today, Inc. finds. The survey,

conducted in partnership with Melissa, reports no letup in the growing confidence gap in the data needed to support next-generation initiatives. This is the third survey in this annual series, and we will show trends that have evolved in the data quality space since the first survey in 2021.

The survey uncovered the following trends shaping data quality:

- Confidence in data quality is slipping. Only 23% express full confidence in their organization's data—down 7 percentage points from a similar survey conducted 2 years ago.
- Next-generation initiatives—particularly advanced analytics and AI—are shining greater attention on data quality.
- Confidence in data quality is slipping, yet organizations have taken their eye off the ball when it comes to addressing data quality issues. Lack of organizational support for data quality efforts—along with determining ROI—may be at the root for lagging progress in such efforts.
- The cloud is not a panacea for data quality issues. While close to half of executives say a majority of their data is now cloud-borne, this has not resolved data quality issues.
- Data entry is becoming more automatic, and less dependent on human input. Going forward, there will be more reliance on third-party or automated sources, such as mobile and web data.

On the following pages are the results of this survey.

---

## THE GROWING DATA CONFIDENCE GAP

**Confidence in data quality is slipping. Only 23% express full confidence in their organization's data—down 7 percentage points from a similar survey conducted 2 years ago.**

Trust is the glue that holds data-driven enterprises together, and decision makers will not move forward with machine-produced results if they cannot trust the integrity, accuracy, and trustworthiness of the underlying data. Essential to this trust is data quality, which refers to the accuracy, consistency, completeness, and timeliness of data. This quality suffers from inaccurate information, outdated contacts, incomplete records, and duplicate records. With low or no data quality, there is low or no trust. Without trust, AI and intelligent operations cannot function.

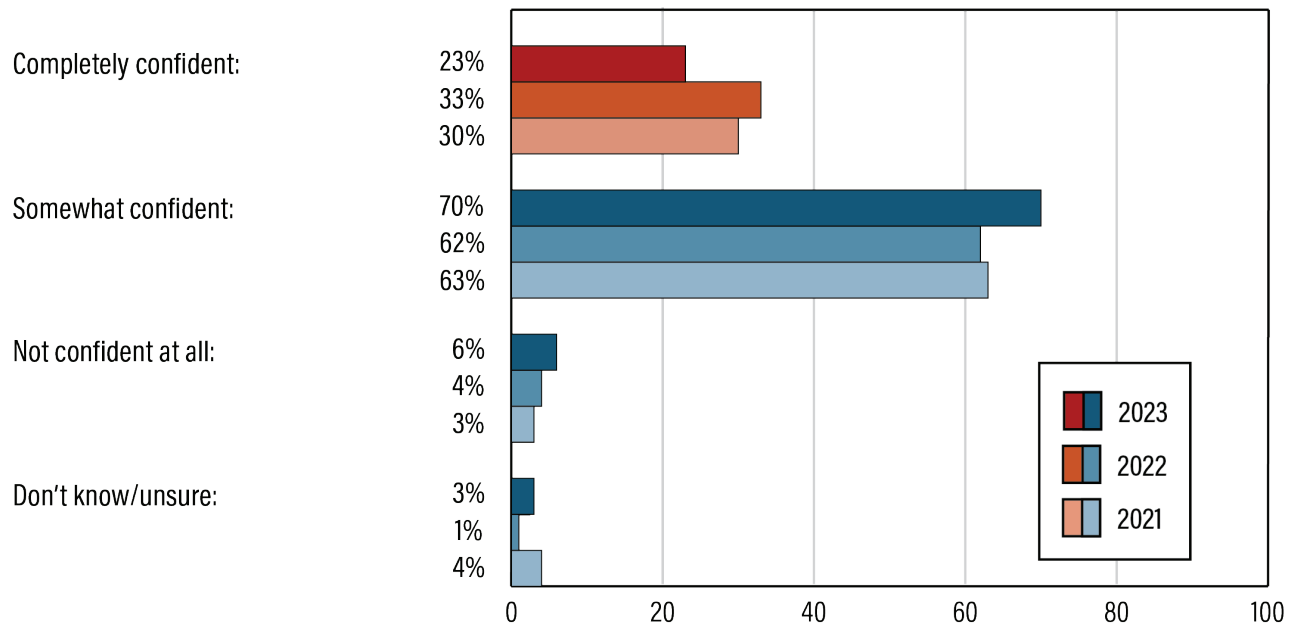
Still, even in the age of AI, survey after survey continues to show that trust in the insights delivered can be a hit or a miss with the C-suite. Executives continue to resist data analytics for higher-level decision making, preferring to rely on gut-level decision-making.

Unfortunately, among the individuals responsible for directly overseeing organizations' data assets, the data is not capable of ensuring such trust. Full confidence in data quality is at a low ebb at a time when it is more crucial than ever. Only 23% express full confidence in their data, down from 30% 2 years ago and 33% last year. (See Figure 2 on page 5.)

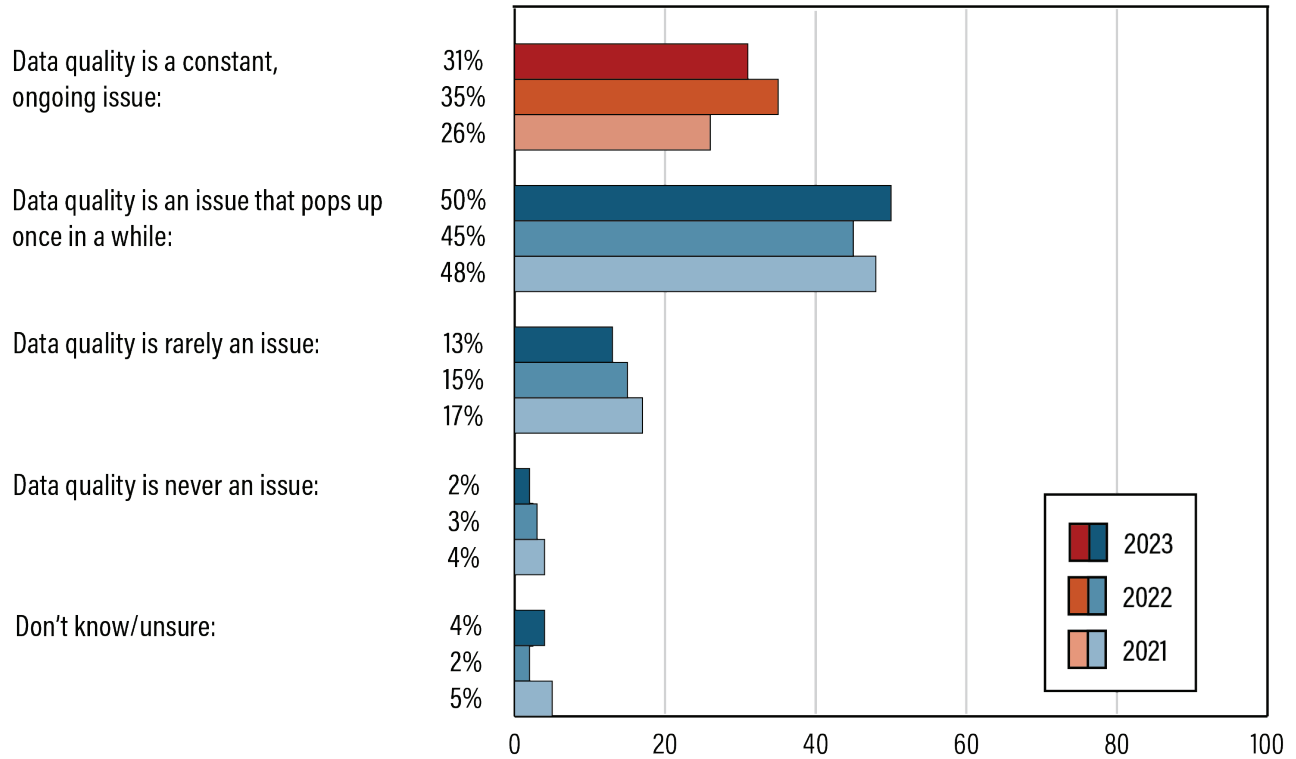
In addition, data quality is likely impeding efforts to move forward with data-driven initiatives essential to the 2020s' enterprise. Close to one-third say data quality is a constant, ongoing issue, up from 26% 2 years ago. (See Figure 2 on page 6.)

Why does data quality appear to be waning, and issues associated with data quality on the rise? Evidence from this survey suggests it is likely tied to the growing demand for AI or analytics workloads, which require the best data at a moment's notice. As we'll explore in the next section, the push to build AI capabilities is exposing weaknesses in companies' data supply chains that urgently need to be addressed.

## Figure 1: How confident are you in the integrity, accuracy, and trustworthiness of data at your organization?



## Figure 2: How often is data quality an issue at your organization?



---

## AI SHINES ATTENTION ON DATA QUALITY

**Next-generation initiatives—particularly advanced analytics and AI—are shining greater attention on data quality.**

Data analytics and AI projects now expose data quality issues over any other situations, surpassing more standard database upgrades or changes. A majority, 57%, report they find out about issues through implementation of next-generation data projects—up from 43% 2 years ago. This points to the criticality of data quality to the success of AI and analytics initiatives. (See Figure 3 on page 8.)

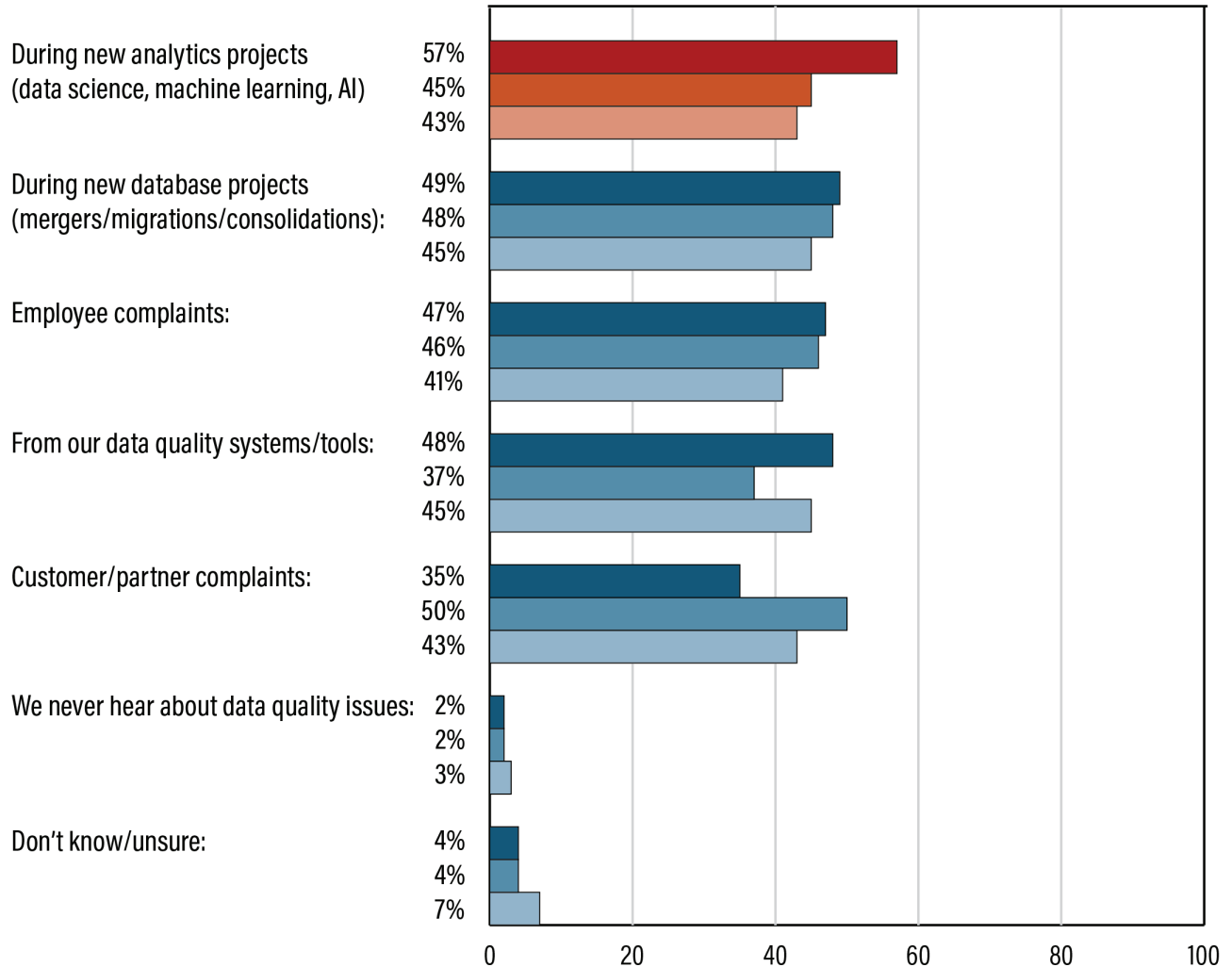
How do data managers and administrators monitor and discover quality issues within their datasets? The most prevalent means of discovery is still through human engagement over automated or systems discovery. The leading means addressing data quality is through user validation, as has been the case in previous surveys. Typically, dealing with data quality issues has

involved time-consuming manual processes. (See Figure 4 on page 9.)

The survey shows that there are efforts to step up automation to address data quality, while still relying on human input. There has been an acceleration in efforts to build custom solutions to address data quality issues, from 58% 2 years ago to 64% today. Still, 40% expect business users to take greater responsibility for the quality of data they oversee, a trend that has also accelerated somewhat over the past 2 years. (See Figure 5 on page 10.)

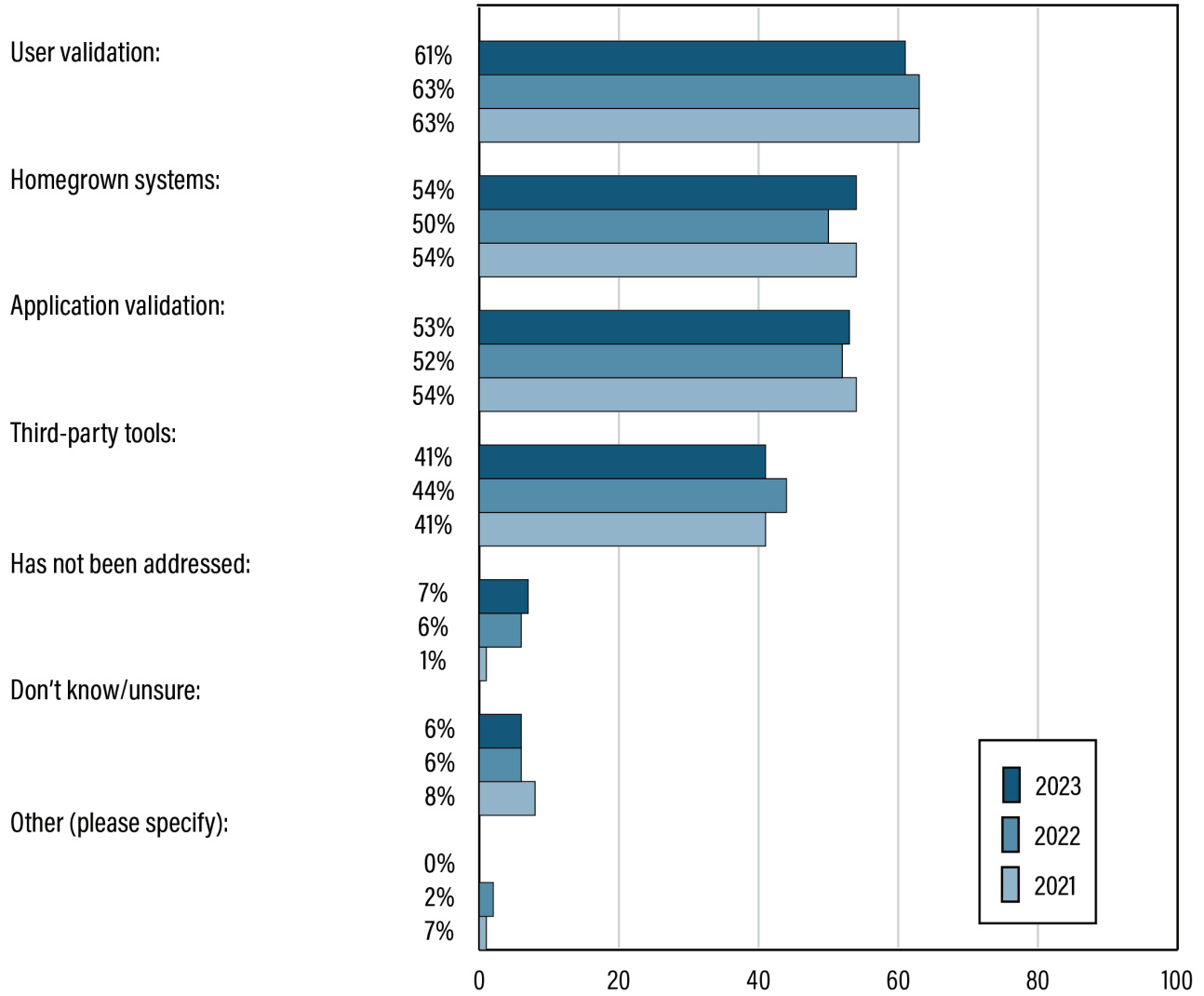
“We have quality control processes at multiple steps in the data integration process and throughout systems,” one respondent reported. “Regardless of this, we do not have the resources to fully achieve quality data to the most specific granularity. We rely on the business to report issues that are missed by our standardized quality control processes.”

**Figure 3: How do you find out about data quality issues? (Please select all that apply.)**

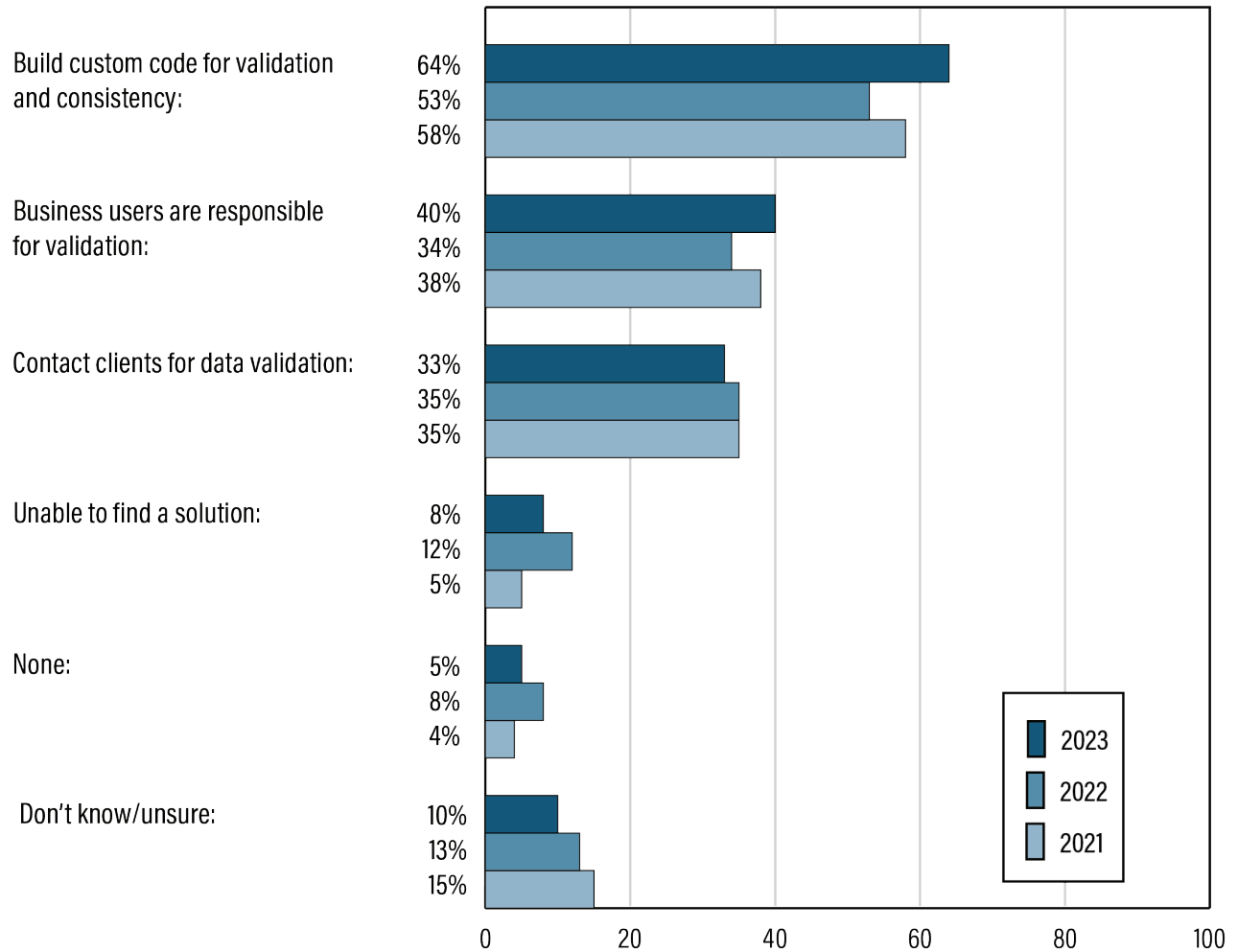




**Figure 4: How do you address the quality of your data? (Please select all that apply.)**



**Figure 5: How have you tried to solve your data quality issues in the past?**



## EYE OFF THE BALL?

**Confidence in data quality is slipping, yet organizations have taken their eye off the ball when it comes to addressing data quality issues. Lack of organizational support for data quality efforts—along with determining ROI—may be at the root for lagging progress in such efforts.**

The presence of a formal data quality strategy puts the organizational stamp of approval on quality assurance through data governance, policies, integration, and security. These are all essential ingredients to assure that the data that supports existing and new applications is accurate, complete, consistent, timely, and fair.

Confidence in data quality is slipping, yet organizations have taken their eye off the ball when it comes to data quality issues. Just over half indicate they have a formal data quality strategy, unchanged from previous surveys. (See Figure 6 on page 12.)

Efforts to address data quality issues are few and far between as well. It's notable that for close to one in four enterprises (22%), it's been 3 or more years since their last data quality initiative. Only 38% can say they've undertaken such a project in the past year. (See Figure 7 on page 13.)

At the same time, data quality has slipped as a priority—only the top of list for more than a third of companies, down from half in previous surveys. Instead, it has slipped to a priority that is competing with other enterprise concerns. Over the past year, the

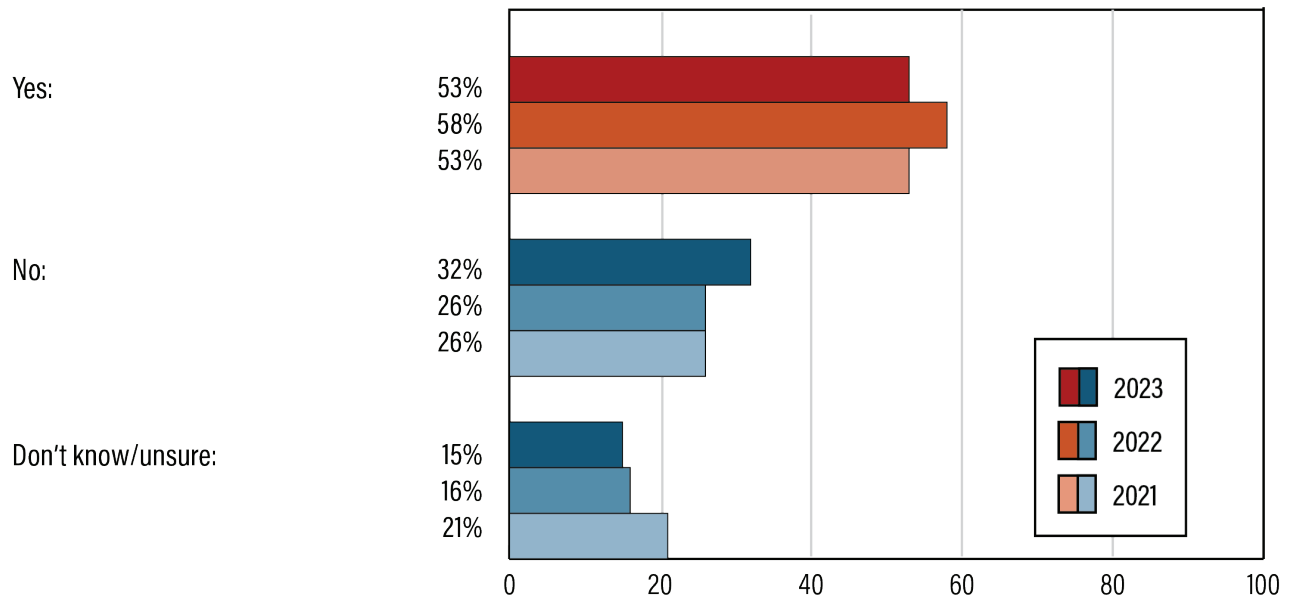
economy has been rocky, with heightened inflation rates, supply chain issues, and worker skills shortages becoming top of mind for many business leaders. (See Figure 8 on page 14.)

With such a panoply of competing challenges, it has grown more difficult for data quality proponents to secure organizational support for their efforts. In a leaner economy, demonstrating ROI is vital, the key to getting executives' attention. Unfortunately, it's often difficult to predict and plan data quality initiatives, and this has continued to be the leading challenge, cited by 58%. Fifty percent cite lack of internal support, up from 42% 2 years ago. Close to that, 45%, cite issues with determining ROI, up from 35% 2 years ago. (See Figure 9 on page 15.)

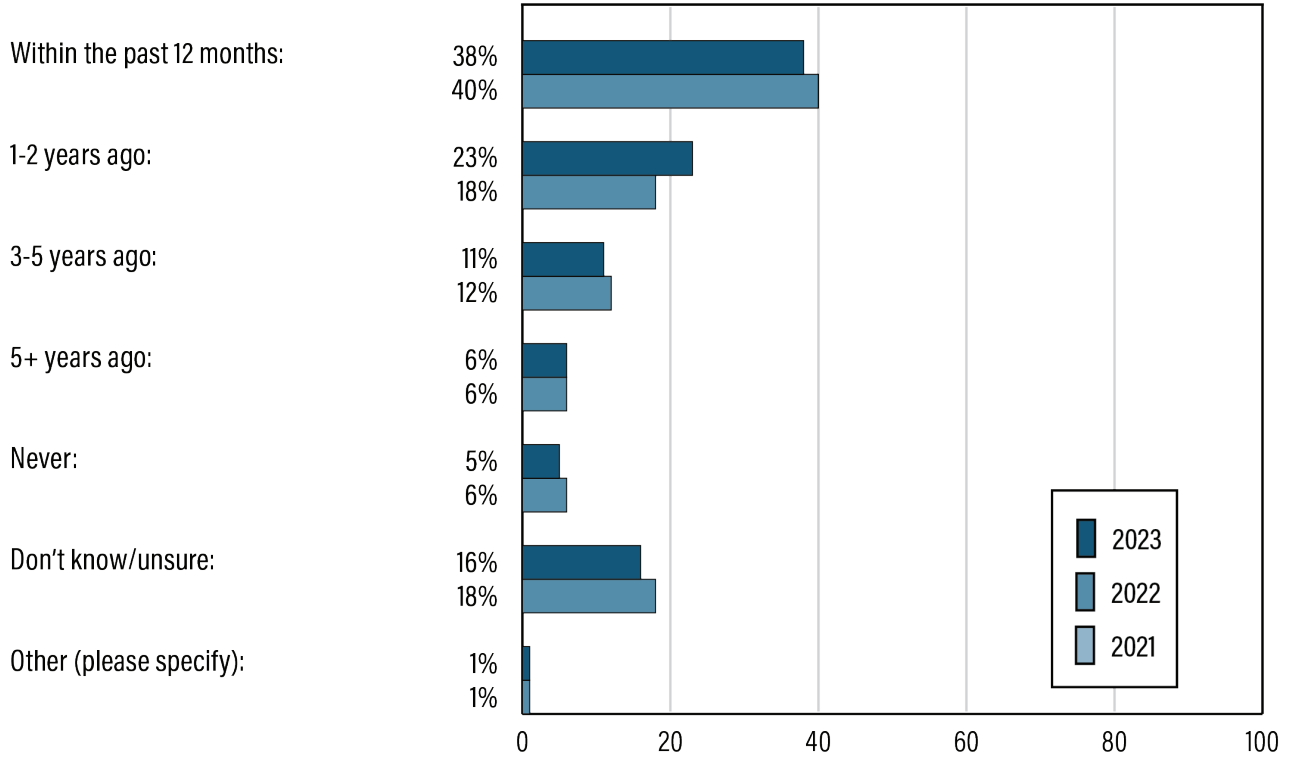
Gaining funds to move forward then, continues to be hampered by issues with project complexities, and unknowns continue to be the greatest challenge to data quality efforts. Calculating ROI is the second roadblock to gaining adequate budget. (See Figure 10 on page 16.)

"Data quality rules have thus far been prescribed by data governance for data certification purposes," a respondent said. "These tend to be rules around basic data hygiene. We are actively seeking to move to a risk and control-based approach integrated into our data feature lifecycle, with both proactive and detective controls."

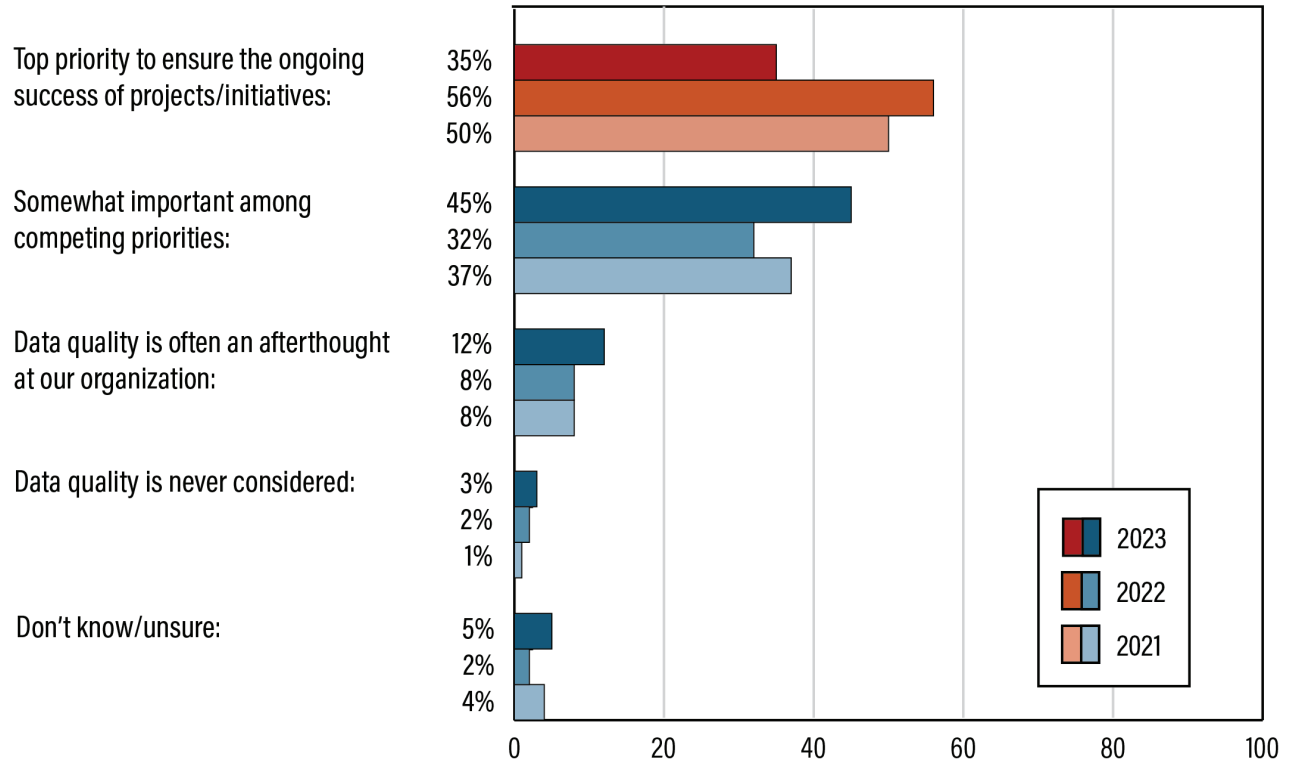
**Figure 6: Does your organization have a data quality strategy with documented requirements and rules for measuring success and ongoing activities tied to business goals?**



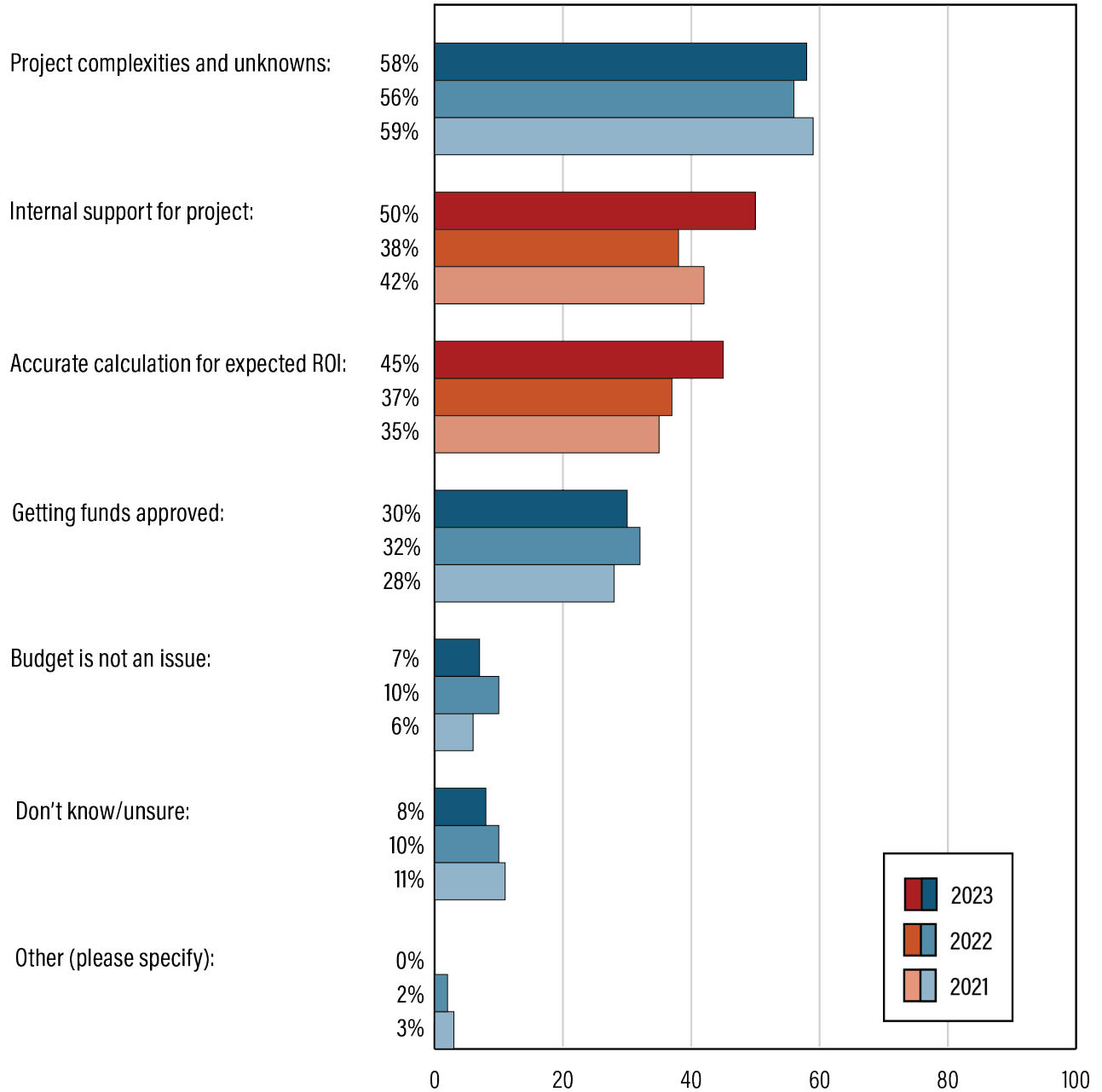
**Figure 7: When is the last time your organization had a data quality project?**



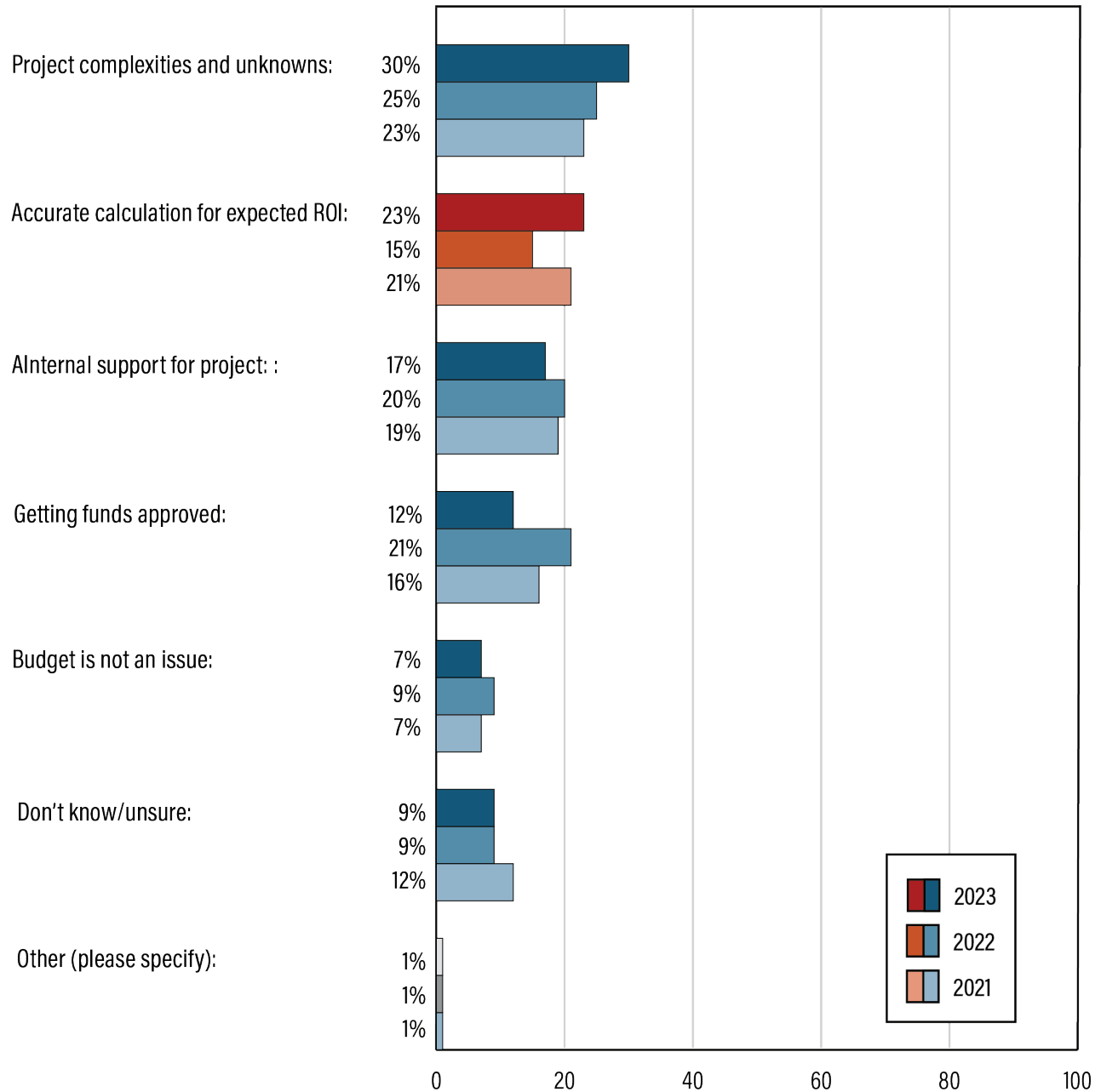
**Figure 8: How important is data quality to your organization's data strategy**



**Figure 9: What are your biggest data quality challenges? (Please select all that apply.)**



**Figure 10: What is your biggest challenge to securing budget for a data quality project?  
(Please select one.)**





---

## MORE DATA IN THE CLOUD, BUT NOT A PANACEA

**The cloud is not a panacea for data quality issues. While close to half of executives say a majority of their data is now cloud-borne, this has not resolved data quality issues.**

As with many things in the information technology space, there is an assumption that moving to cloud services can provide data managers and administrators some relief from their data quality challenges. AWS Glue Data Quality, for example, is a service designed to reduce manual quality efforts by automatically computing statistics, recommending quality rules, monitoring, and providing data quality alerts to help identify missing, stale, or bad data. Microsoft's Azure data quality framework provides tools designed to define data sources and destinations, configure ETL pipelines, apply data quality techniques, monitor the data quality metrics, and establish data quality governance.

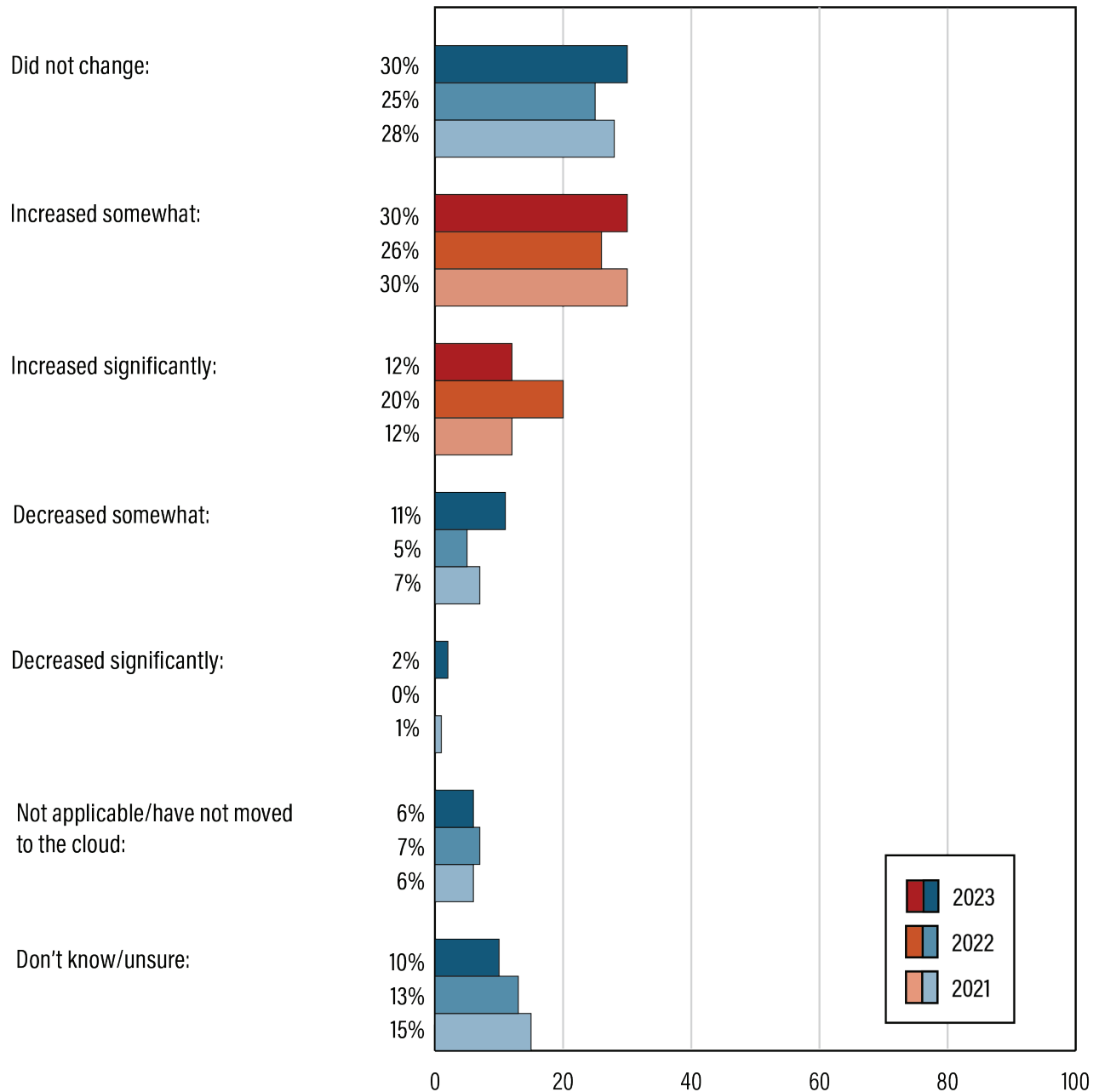
Still, data managers and administrators acknowledge that data quality needs to start at home, and that their issues persist, even

after data has been migrated or generated in the cloud. While the use of cloud for data storage and management has expanded, moving to the cloud has not resolved data quality issues, showing little or no improvement from previous surveys. Forty-two percent report an increase in issues, in line with previous studies. (See Figure 11 on page 18.)

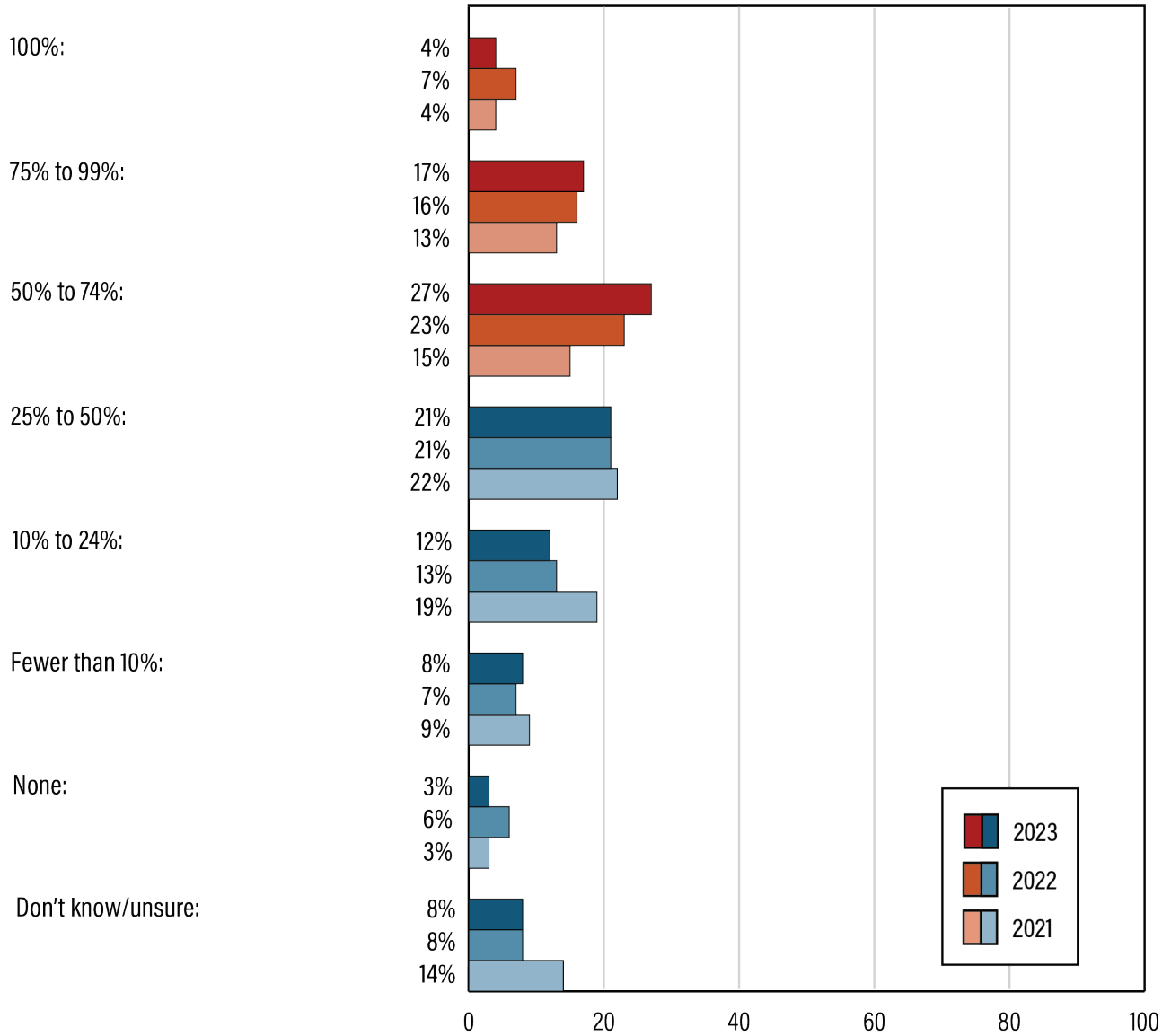
Data migration to the cloud continues unabated. Overall, there is more data that has been moved to the cloud: 48% say a majority of their data (>50%) is in the cloud, up from 32% 2 years ago. (See Figure 12 on page 19.)

Not only is more data going to the cloud, but it's going to a greater variety of clouds. AWS and Microsoft Azure remain top cloud choices—but Google Cloud Platform, Oracle, and Salesforce are moving up fast. Google adoption is up 62% over the past 2 years, Oracle is up 69%, and Salesforce has jumped 44%. (See Figure 13 on page 20.)

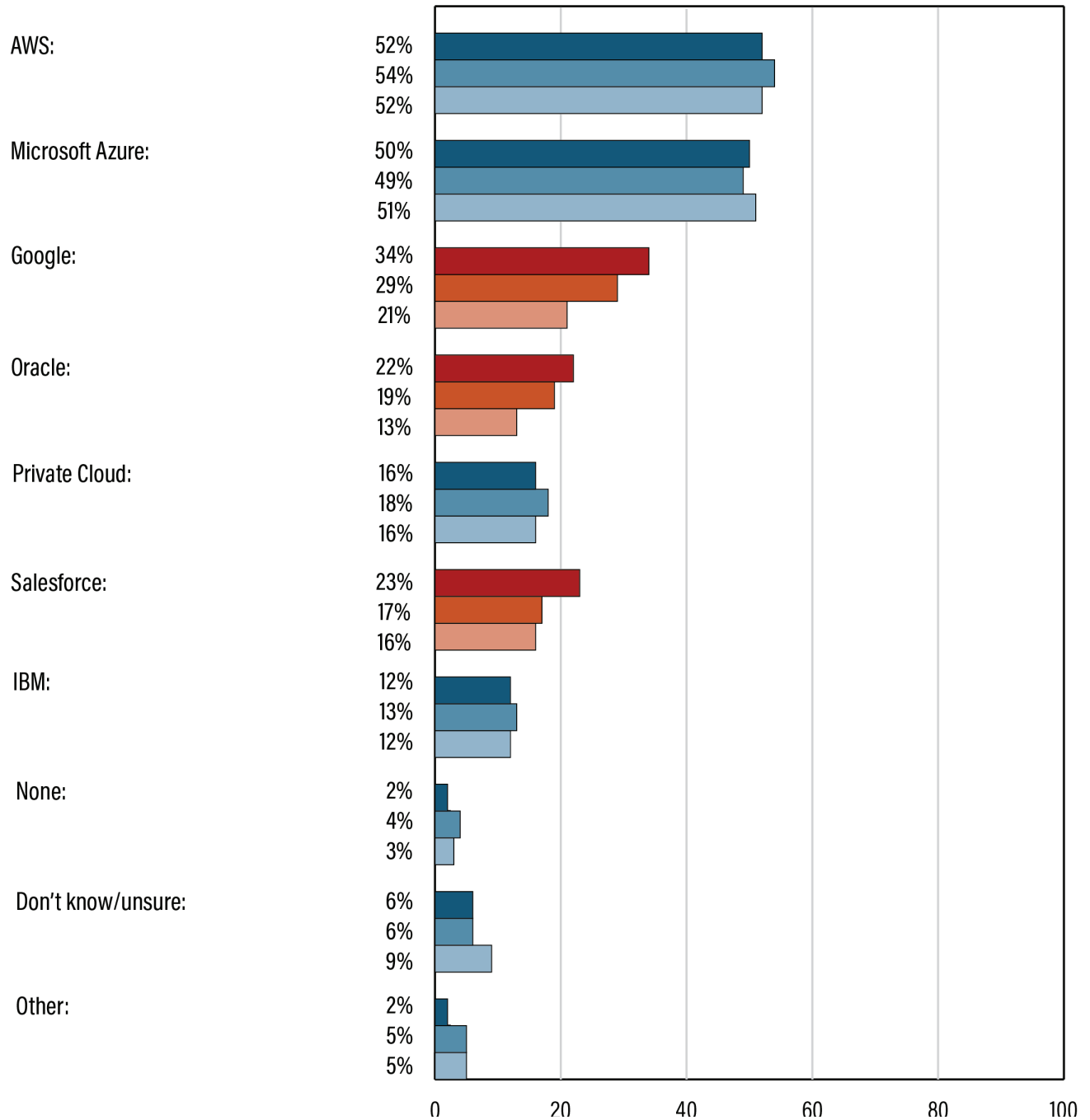
**Figure 11: If you are using cloud-based solutions, how have your data quality challenges changed after moving to the cloud?**



**Figure 12: What percentage of your data is stored in the cloud?**



**Figure 13: Which cloud-based solutions do you use to store your data?**



---

## LESS MANUAL, MORE AUTOMATED DATA ENTRY

**Data entry is becoming more automatic, and less dependent on human input. Going forward, there will be more reliance on third-party or automated sources, such as mobile and web data.**

There is an element of progress in the battle to maintain data quality—there is less reliance on error-prone human data entry, the survey shows. Data entry is becoming more automatic. There is more reliance on third-party sources and automated data collection and less on manual keying by employees—down from 58% 2 years ago to 53% in the current survey. Reliance on direct data entry by customers is down slightly—from 49% to 46%. (See Figure 14 on page 22.)

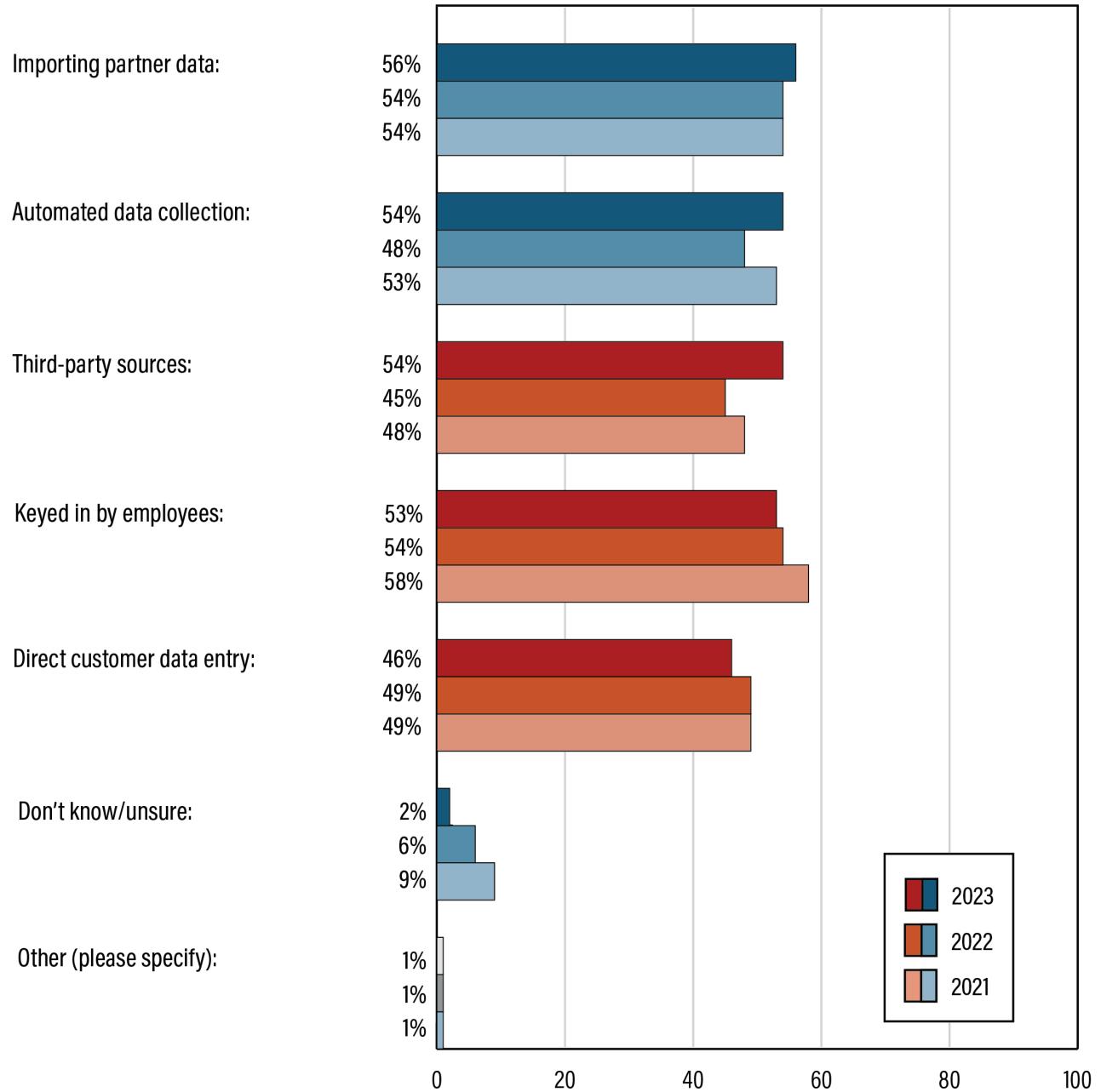
“Data quality has always been neglected in the organization,” a respondent said. “We have always had many departmental systems, hosted in different ways and places. Lately, we are acting more professionally regarding our data, starting by seeking solutions to our most common problems. One of the main

problems is the incorrect entry of the collected information, made in many different places. This manually entered data is riddled with wrong, inaccurate, and inconsistent information.”

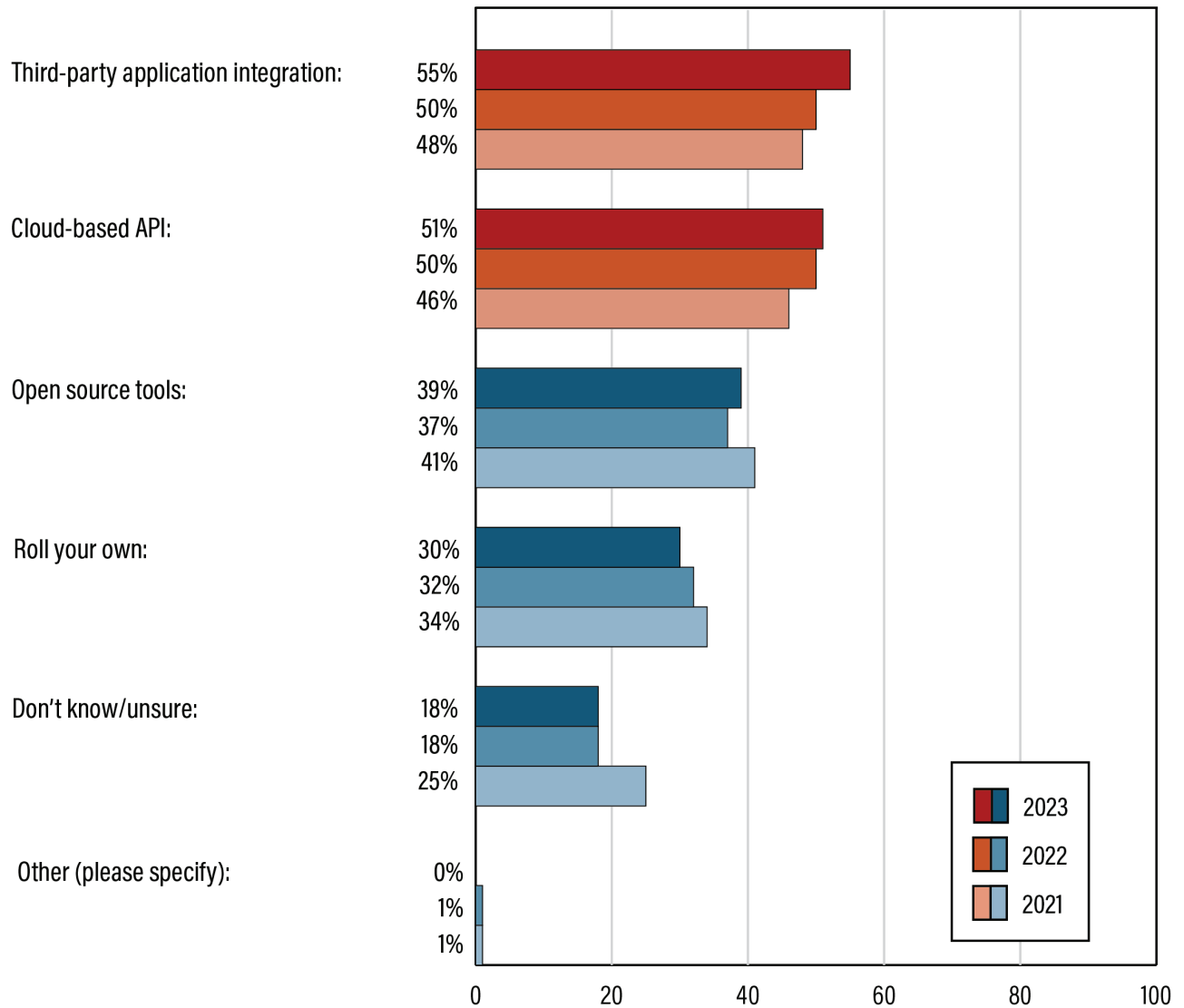
Going forward, there will be more reliance on third-party or automated sources. A majority of data managers and administrators, 55%, say data issues can be mitigated by direct integration with third-party applications—rising from 48% in the 2021 survey. Cloud-based APIs also serve as a trusted data channel, rising from 46% to 51% in recent years. (See Figure 15 on page 23.)

Customer contact data and product data are the types of data seen the most within enterprises. But there are some fast-growing data domains also emerging, particularly mobile and web data, encompassing IP addresses, geolocations, and IoT data. (See Figures 16 and 17 on pages 24 and 25.)

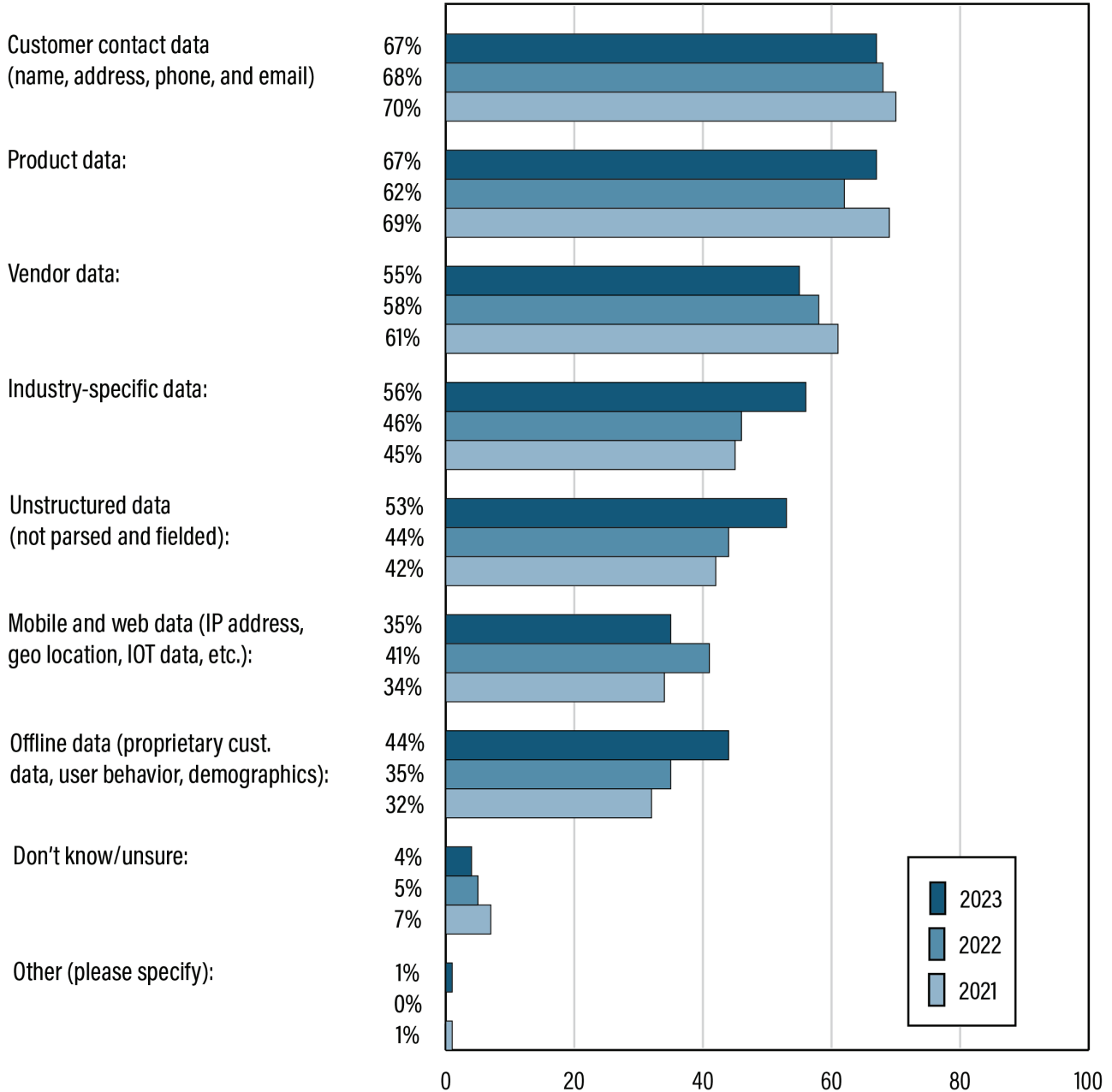
### Figure 14: How does data get added into your databases?



**Figure 15: How do you envision solving your data quality challenges in the future?**

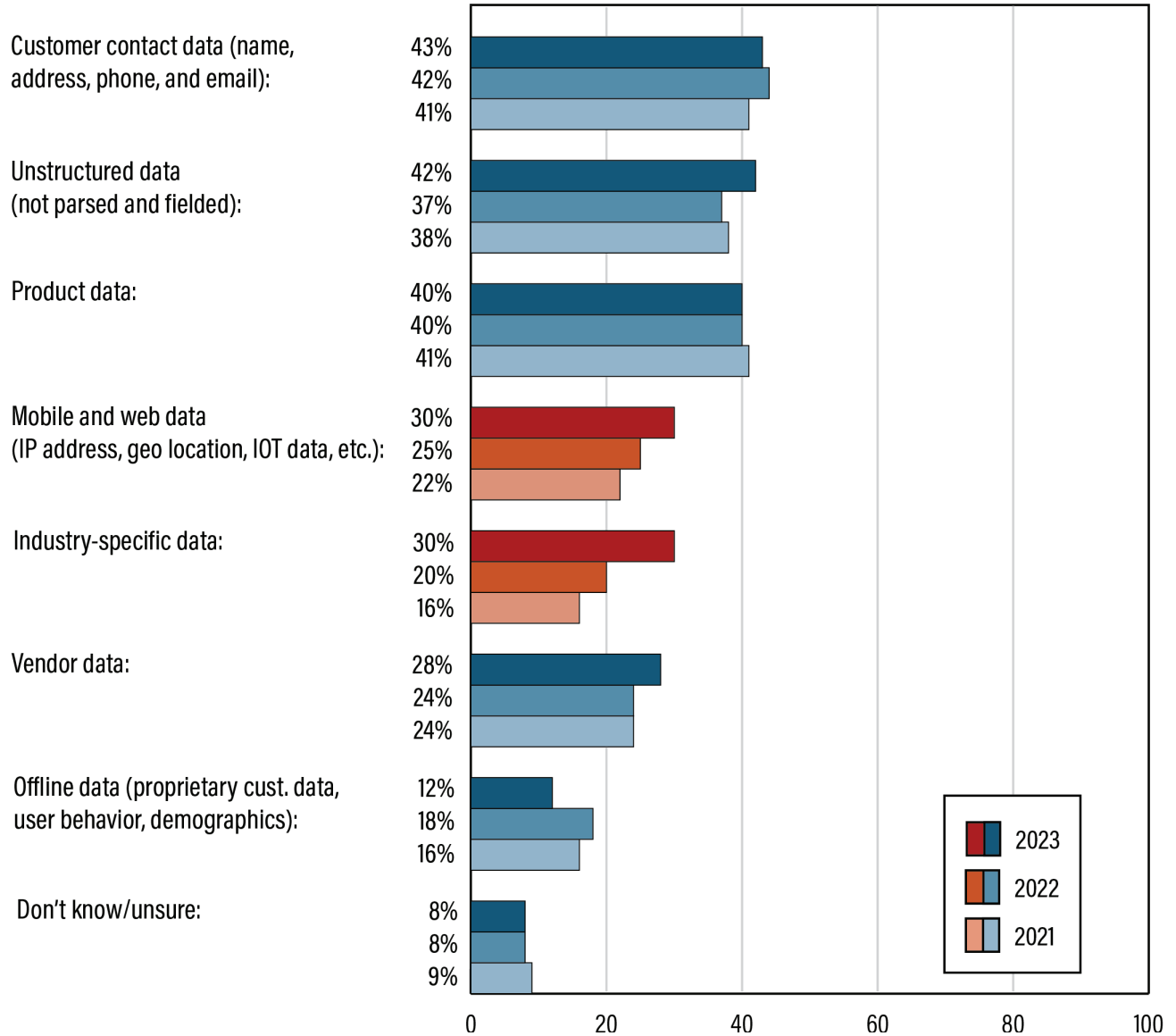


**Figure 16: What domains of data are stored in your databases?**





**Figure 17: What data domains are growing the fastest in your databases?**



## RECOMMENDATIONS

This survey, now in its third year, finds that perceptions of data quality are slipping, even though data quality has increased in importance due to the rise of analytics, AI, IoT, data monetization, and other next-generation initiatives.

Data administrators and managers are concerned that the quality of their enterprise data is deteriorating, with only 23% currently expressing full confidence in their data—down 7 percentage points from a similar survey conducted 2 years ago. The need is urgent, as initiatives as advanced analytics and AI, are shining greater attention on data quality.

In addition, lack of organizational support for data quality efforts—along with determining ROI—may be at the root for lagging progress in such efforts. Fifty percent cite lack of internal support, up 8 percentage points from 2 years ago.

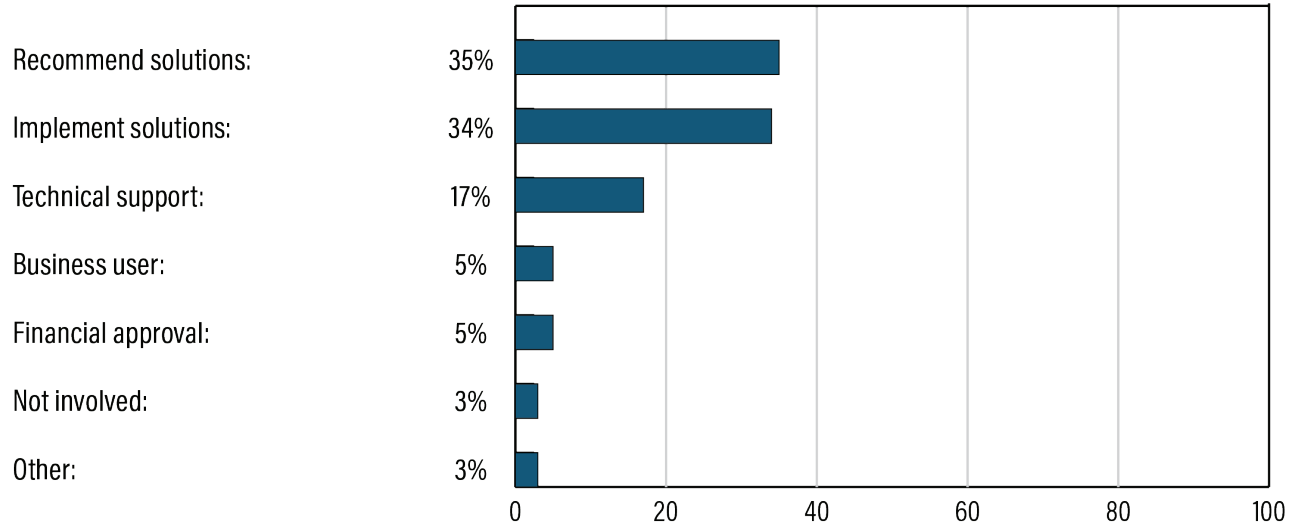
Consider the following courses of action when seeking to increase data quality levels and prepare enterprises for the coming era in AI and intelligence:

- Data quality needs to be a shared responsibility. Promote ownership and responsibility for data assets beyond the IT department, from anyone who touches the processes that involve data, such as AI. Often, corporate cultures discourage such shared responsibility, and the silos that have resulted need to be torn down.
- Don't outsource data quality to the cloud. The cloud is not a panacea for data quality issues. While close to half of executives say a majority of their data is now cloud-borne, this has not resolved data quality issues.

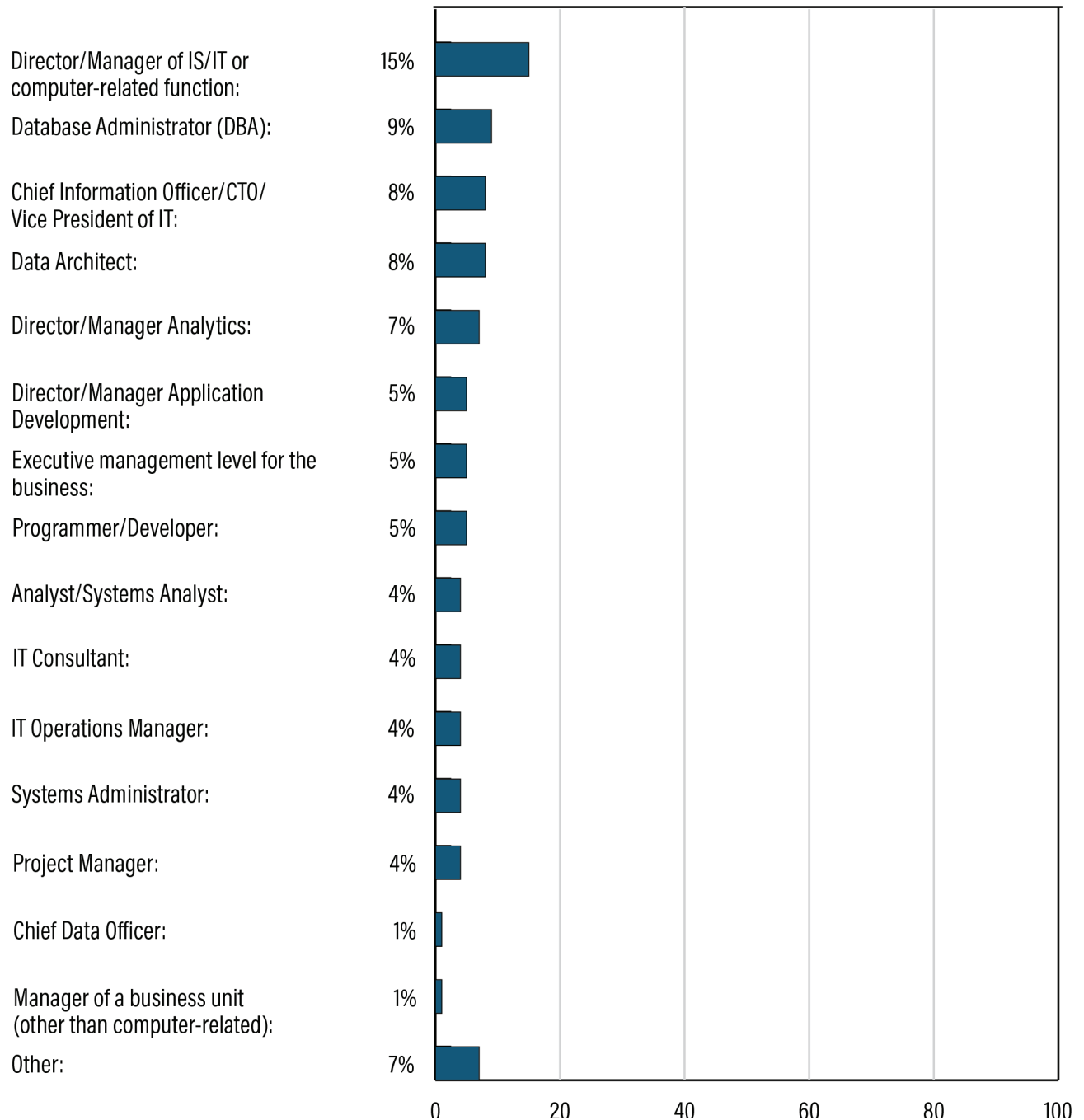
- Step up efforts to eliminate data bias as well as overall quality issues. Ensure fairness in data-driven decisions through greater transparency and maximum observability of data pipelines and decision-delivery chains.
- Investigate and advocate for the latest technology solutions. The technology to assure greater accuracy, consistency, and timeliness in data keeps improving each year as providers recognize the real-time needs of enterprises. Recognize that AI, which we found to be exacerbating data quality needs, also can offer solutions to addressing data quality in a more automated fashion as well. “We are using AI to find gaps, duplication, and data fabrication,” a respondent said. “And using rules and AI to fill gaps in data or reject data.”
- Foster greater awareness and training for data quality at all levels. Data quality needs to be everyone's job. Promote or advocate for incentives to assure the best possible data.
- Review or audit results from data-dependent applications on a regular, systematic basis. “We have worked on data governance policies,” a respondent said. “We profiled our data, cleaned, standardized, and validated. We have also implemented data audits as part of our internal compliance.”

Many operational courses of action are already being driven by AI-based algorithms, and, increasingly, many of the strategic decisions made in the C-suite will be AI-assisted. If the output of data-assisted decisions is based on reliable, unbiased, and transparent data, this will provide business leaders with the confidence they need to make decisions based on AI and advanced analytics.

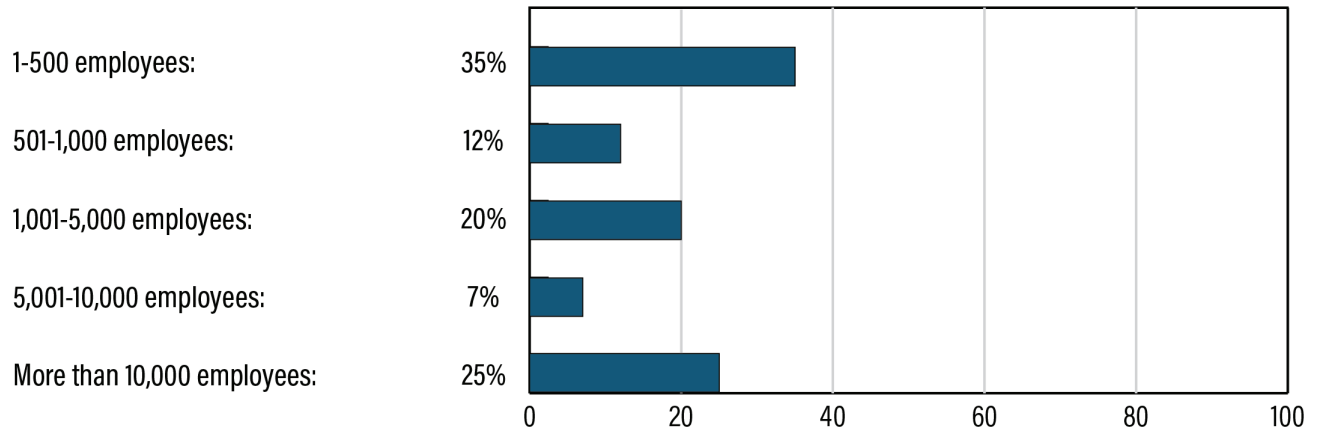
## DEMOGRAPHICS

**Figure 18: What is your primary role in improving the quality of data at your organization?**

### Figure 19: What is your primary job title?



**Figure 20: How many employees are in your entire organization?  
(Include all locations, branches, and subsidiaries.)**



### Figure 21: What is your organization's primary industry?

