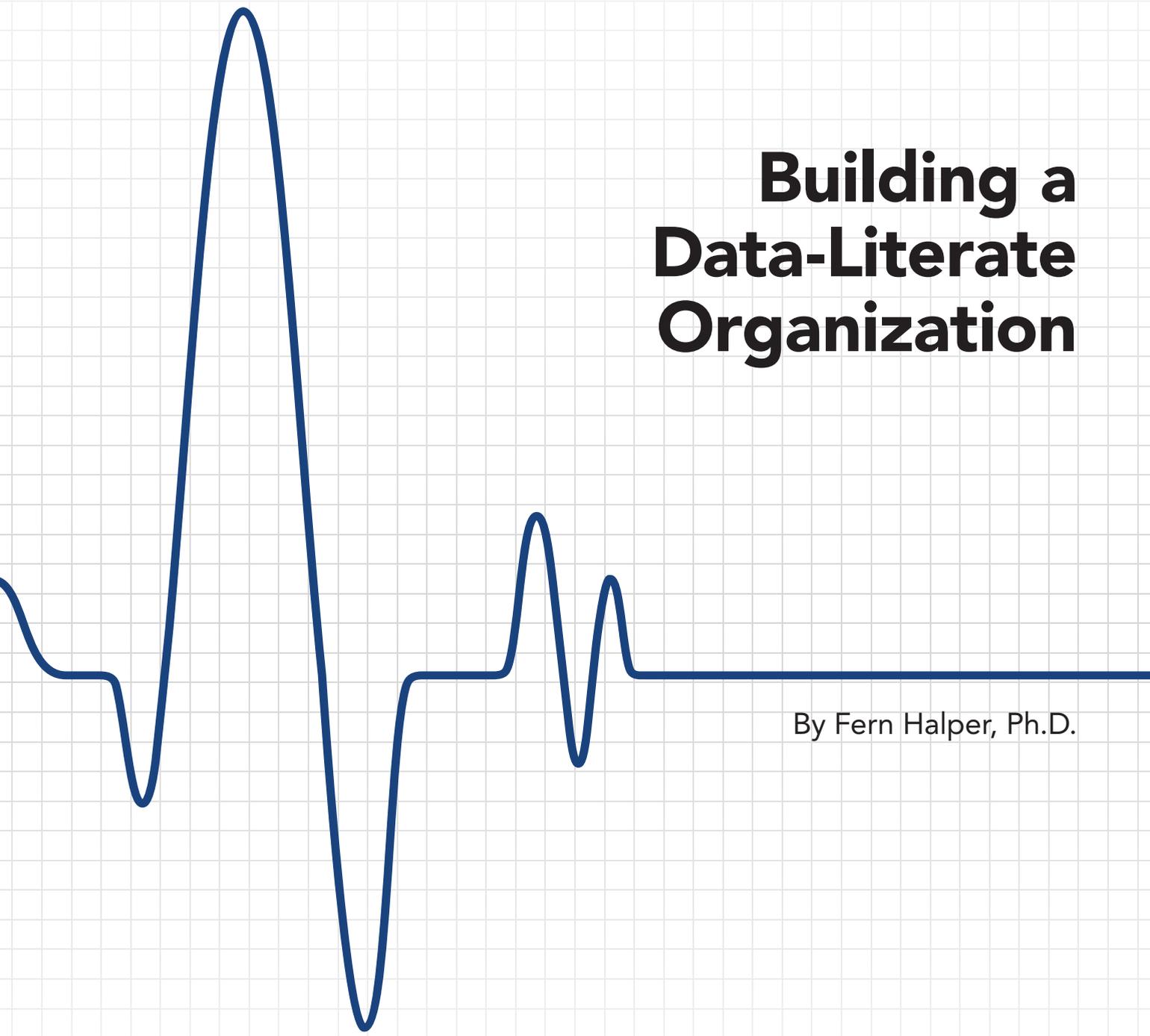




PULSE



Building a Data-Literate Organization

By Fern Halper, Ph.D.

Sponsored by:



Collibra



tableau





Building a Data-Literate Organization

By Fern Halper, Ph.D.

Table of Contents

- The Pulse: Organizations Move Forward with Data and Analytics Driving the Need for Data Literacy 3**
- Defining Data Literacy and Its Importance. 5**
 - Barriers to Data Literacy 6
 - Data Literacy Across the Organization 8
- Characteristics of the Data-Literate Organization 8**
- Governance and Data Literacy 10**
- Best Practices for Data Literacy 11**
 - Organizational Best Practices 11
 - Technology Best Practices 13
- Conclusion 14**
- About Our Sponsors 15**

© 2021 by TDWI, a division of 1105 Media, Inc. All rights reserved. Reproductions in whole or part are prohibited except by written permission. Email requests or feedback to info@tdwi.org.

Product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies. Inclusion of a vendor, product, or service in TDWI research does not constitute an endorsement by TDWI or its management. Sponsorship of a publication should not be construed as an endorsement of the sponsor organization or validation of its claims.

About the Author



FERN HALPER, Ph.D., is vice president and senior director of TDWI Research for advanced analytics. She is well known in the analytics community, having been published hundreds of times on data mining and information technology over the past 20 years. Halper is also coauthor of several Dummies books on cloud computing and big data. She focuses on advanced analytics, including predictive analytics, text and social media analysis, machine learning, AI, cognitive computing, and big data analytics approaches. She has been a partner at industry analyst firm Hurwitz & Associates and a lead data analyst for Bell Labs. Her Ph.D. is from Texas A&M University. You can reach her by email (fhalper@tdwi.org), on Twitter (twitter.com/fhalper), and on LinkedIn (linkedin.com/in/fbhalper).

About TDWI Research

TDWI Research provides industry-leading research and advice for data and analytics professionals worldwide. TDWI Research focuses on modern data management, analytics, and data science approaches and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of business and technical challenges surrounding the deployment and use of data and analytics. TDWI Research offers in-depth research reports, commentary, assessments, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

About TDWI Pulse Reports

This series offers focused research and analysis of trending analytics, business intelligence, and data management issues facing organizations. The reports are designed to educate technical and business professionals and aid them in developing strategies for improvement. Research for the reports is conducted through surveys of professionals. To suggest a topic, please contact TDWI senior research directors Fern Halper (fhalper@tdwi.org), James Kobielus (jkobielus@tdwi.org), and David Stodder (dstodder@tdwi.org).

Acknowledgments

TDWI would like to thank many people who contributed to this report, including our report sponsors, who reviewed outlines and report drafts. Finally, we would like to recognize TDWI's production team: James Powell, Lindsay Stares, Rich Seeley, and Michael Boyda.

Sponsors

Amazon Web Services, Collibra, and Tableau sponsored the research and writing of this report.

The Pulse: Organizations Move Forward with Data and Analytics Driving the Need for Data Literacy

As organizations look to digitally transform in order to become more data-driven and better compete, they are often ramping up their data and analytics efforts to better serve customers and enable efficient and effective operations. An important part of this effort is modernizing their data and analytics environments.

Modern analytics includes new tools that help organizations democratize analytics as well as become more proactive. Modern environments often use the cloud to support data and analytics. For instance, at TDWI, we see that implementing self-service—solutions that enable nontechnical users to be productive because they are easier to use, do not require coding, and do not require IT to set up all data access, queries, visualizations, and preparation—is *the* top priority for organizations in 2021.¹ Demand for advanced analytics, such as machine learning, also continues to increase.

In the survey for this report, the majority of respondents were at the self-service stage of analytics adoption or beyond (see Figure 1). Here 42% were using spreadsheets or a combination of spreadsheets, reports, and dashboards; 36% were using these tools as well as self-service analytics; and one in five respondents was also using advanced tools such as predictive analytics and machine learning (13%) or AI and natural language processing (8%).

Which statement best describes the analytics environment in your organization?

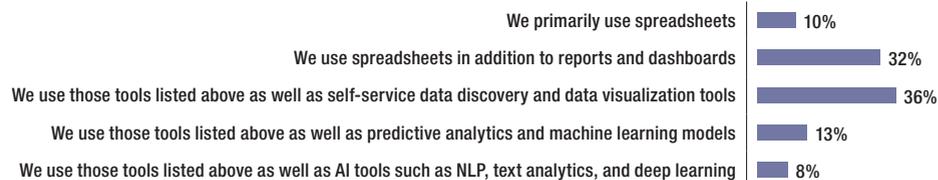


Figure 1. Based on 395 respondents.

On the data management front, most respondents use the cloud in some form (see Figure 2). For instance, 46% of respondents to this survey stated that they had a hybrid environment that is both on premises and in the cloud; another 16% were completely in the cloud. A small group claimed to use multiple cloud environments. At TDWI, we typically see that organizations are modernizing their data environments to support analytics activities. In fact, cloud data warehouses and cloud data lakes are already mainstream. Companies want to analyze more diverse and larger volumes of data and the cloud provides the flexibility and scalability to do so.

¹ Unpublished TDWI survey on data and analytics, 2021

Which statement best characterizes the data environment in your organization?

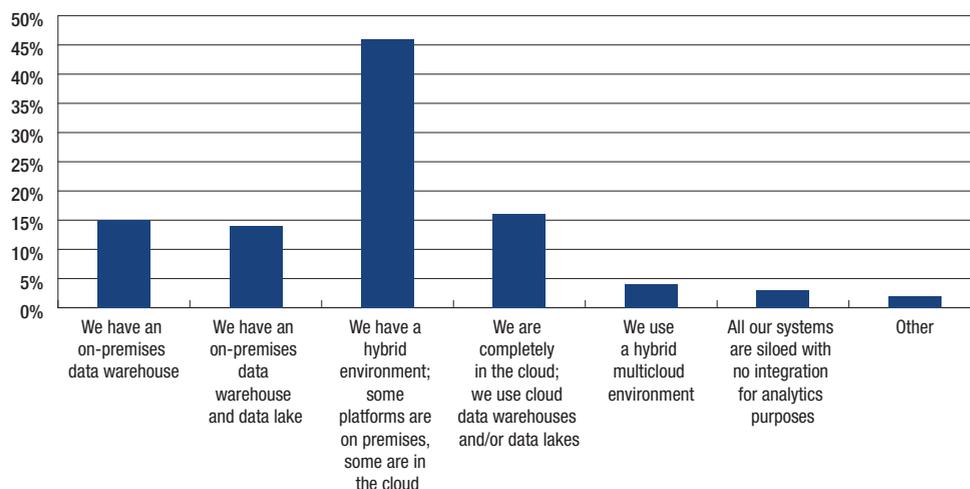


Figure 2. Based on 395 responses.

As part of the move to modernize and move forward with analytics (with the goal of becoming more data-driven), organizations realize they need to become more data literate. Data literacy involves how well users understand and can interact with data and analytics and communicate the results to achieve business goals. It includes understanding the business, framing analytics, critically interpreting the results, and communicating these results to others. Data literacy is not only for business analysts; this is a critical skill for individuals across a data-driven organization, although the level of literacy may vary by role.

In TDWI research, data literacy is often cited as a top priority. However, in this survey, the majority of respondents (68%) stated that their organization was only somewhat data literate, based on how well their users “understand data, visualizations, and analytics and can effectively interact with and communicate them to others to achieve results.” Twelve percent stated that they were not data literate. Only 20 percent felt their organization was data literate (all not shown).

In other TDWI research, when asked to rate the data literacy of individuals in their organization, the highest percentage said their organizations were “about average” (42%) in terms of users’ ability to consume, analyze, interact with, share, and discuss data in the course of carrying out their roles and responsibilities.² That means that although organizations want to modernize, there is still work to be done in order for them to make the best use of their data and analytics.

This TDWI Pulse Report uses recently collected data from a global survey to examine how data and analytics professionals view data literacy and why it is important, the characteristics of data-literate companies, and best practices for becoming a data-literate organization. The report also compares the three groups mentioned above (not data literate, somewhat data literate, and data literate) to illustrate similarities and differences with an eye towards best practices.

² See TDWI Best Practices Report: Faster Insights from Faster Data, online at tdwi.org/bpreports.

Survey Methodology and Demographics

In March 2021, TDWI sent an invitation via email to data and analytics professionals in our database, asking them to complete an online survey. The invitation was also posted in some of our virtual events as well as in publications from TDWI. The survey collected responses from over 600 respondents globally. A total of 395 respondents completed all of the questions and met TDWI’s validity criteria for this survey.

Sixty-nine percent of respondents came from the U.S. and Canada, 14% from EMEA, 13% from APAC, and the rest from other parts of the world. Respondents come from enterprises of all sizes and across all industries, with the largest percentages of survey respondents working in consulting/professional services and financial services (9% each) computer manufacturing (8%), education (7%), manufacturing (non-computers) (7%), and government (6%).

Respondents came from companies of all sizes. Roles included managers of IT, data, and analytics (21%), business analysts (13%), directors/VP of IT and analytics (12%), data engineers (11%), and a host of other data and analytics titles.

Defining Data Literacy and Its Importance

Different groups in an organization often define data literacy differently. For instance, on the IT side, data literacy might include understanding the data, providing metadata, and defining data elements. For an analytics team, it may include the ability to understand the data *and* analyze it and communicate the results so different groups can understand them. We asked respondents for their own definitions of data literacy as well as reasons why it is important. Representative responses are listed in Table 1.

IT-focused roles	Analytics-focused roles
“Understanding what data an organization owns, where to find it, what the data standards are, and knowledge/skills to consume and work with that data”	“Being able to effectively communicate the results of data so informed decisions can be made”
“Data definition knowledge”	“Data literacy is the ability to derive meaningful information from data”
“The sweet point between reading data, working with data, and communicating the data to others”	“Understanding the impact of data on organizational processes, having data and information assets defined and managed, and using information to measure progress and making decisions”
“Understanding of where data is located and how it can be used effectively for analytical work (whether it is a visualization, cube, simple report, or advanced analytics) in the user’s tool of choice”	“Ability to read, write, and communicate data in context” [Note: this is from Gartner’s definition of data literacy]
“Ability to understand what the key data elements for a particular working area are, knowing what they mean, and knowing how they impact the company”	“How well users understand data, visualizations, and analytics and communicate them to others to achieve results”

Table 1. Representative definitions of data literacy based on 395 respondents.

When respondents describe data literacy, there are important commonalities. For instance, some respondents (often IT) talk about the need to understand what data means, find it, and access it. They talk about the definition of data. All of that involves understanding and accessing data so you can use it. Others (often those with analytics titles) talk about the ability to derive meaningful insights from data to help make decisions. They discuss the ability to perform and interpret analytics and the ability to communicate findings. In other words, once you can access and understand data, data literacy includes the ability to interpret analytics in a meaningful way and communicate your insights.

Barriers to Data Literacy

Although most organizations would agree that data literacy is important, there are definite barriers to becoming data literate. These include organizational barriers such as culture and talent as well as technical barriers such as putting a data infrastructure in place to support analysis.

Data infrastructure. One of the top barriers to data literacy is siloed data, i.e., when data managed by one group may not be available to others. Nearly 35% of respondents said data silos are a problem. Data silos undermine data quality and integrity because the same data is often stored in more than one place, and it gets out-of-date or transformed in a way that compromises its integrity. Data silos limit data sharing. Silos also can delay delivering insights. All of this can get in the way of building a culture that supports the widespread use of data for decision making. At TDWI, we often see organizations trying to unify their environment for analytics to overcome the silo issue and build a foundation for data literacy, often deploying a cloud data warehouse or a data lake (or unifying the two). In fact, as we will see, organizations that are data literate are more likely to use the cloud than those that are not.



One of the top barriers to data literacy is siloed data.
Nearly 35% of respondents said data silos are a problem.

Skills and talent. Almost tied for first place in terms of obstacles was people not making use of available software for insights because they lack the skills (34%). In other words, the organization may have BI and analytics tools available, but many people cannot use the software because they lack the skills—either because they lack data literacy skills or because they lack skills in the tools themselves. For instance, in a separate question (not shown) we asked how widely used certain tool types are. Unsurprisingly, spreadsheets were the most widely used tool across all organizations, but especially across organizations that are not data literate. If that is the case, many organizations are not moving forward towards the widespread use of other kinds of analytics tools.

Becoming data literate is about more than software literacy, of course. Individuals in organizations need foundational skills to explore and understand data. This includes skills such as recognizing well-structured data, thinking critically about data and understanding variables and variable types and distributions, and understanding how to aggregate data. With these

skills, individuals can interpret data and analyses presented on dashboards and visualizations more effectively, regardless of whether they are creating the analysis themselves. This, in turn, provides individuals with the confidence to make data-driven decisions.

Tools. Other respondents felt they didn't have the right tools to find, access, and prepare data for analytics (30%). This is a function of data access and understanding data rather than using analytics tools. This was an important part of the definitions respondents gave for data literacy. Additionally, in a separate question, only about a third of respondents agreed with statements that individuals in their organization understood what data was available to them or had access to the data that they needed (not shown). The rest either were on the fence or disagreed with these statements. Clearly, there is room for improvement in the area of data understanding. The good news is that there are tools that can help, which we discuss later in this report.

Culture. Of course, some organizations simply do not have a culture of relying on data to make decisions or inform strategy. Close to 30% of respondents cited this as an obstacle to data literacy. Often, this stems from the top of the organization, where executives make decisions based on gut instinct rather than relying on data. If the culture is not geared towards data literacy, there likely won't be much funding for programs to build literacy or, for that matter, analytics. Unsurprisingly, we will see that organizations that are not data literate are more likely to be less analytically mature than those that are data literate.

Data governance. Interestingly, data governance ranked at the bottom of the list of obstacles for analytics and data literacy, especially when it comes to compliance. This may be because the other obstacles loom larger when it comes to data literacy.

What top obstacles are your organization facing in its analytics journey? Please select up to three challenges.

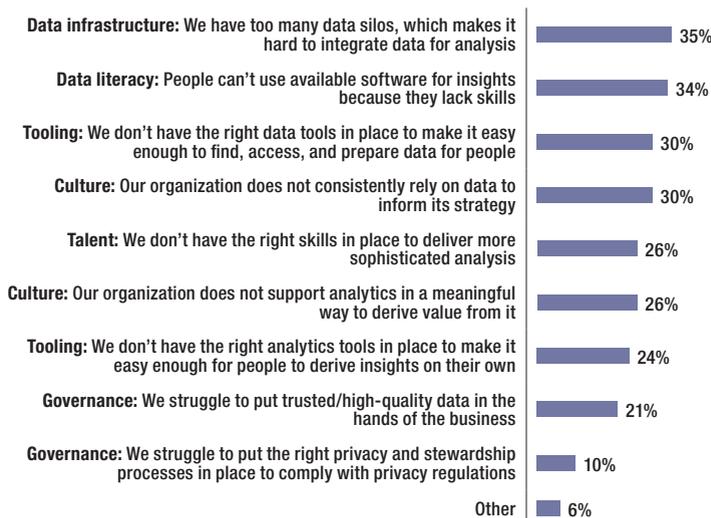


Figure 3. Based on 395 respondents. Up to three responses allowed per respondent.

Of course, organizations should understand the full breadth of data literacy, too. As we saw earlier, IT thinking it is one thing and analysts thinking it is another can cause issues. This should be part of the recommended training described later in this report.

Data Literacy Across the Organization

Data literacy barriers are reflected in respondents’ perceptions of who is literate in the organization (see Figure 4). Unsurprisingly, the analytics team and the IT team ranked at the top of the list for data literacy. Finance and marketing were also considered to be somewhat data literate. At TDWI, we often see marketing organizations making use of advanced analytics to understand customers.

However, certain groups, such as the executive team, ranked lower on the list, with only 53% of respondents stating they were data literate enough to do their jobs. This is a conundrum. On the one hand, we will see that executive support is needed for the organization to be data literate. On the other hand, about half of the respondents either don’t think their executive team is data literate (24%) or don’t know (21%). What’s more, often those in executive roles are more likely to call out the fact that executives weren’t data literate enough (not shown). Executives are often needed for funding, evangelizing, and making important decisions based on data. Executives who aren’t data literate can impede the organization’s ability to become data literate.

Do you believe that the following groups in your company are as data literate as they need to be to do their jobs effectively?

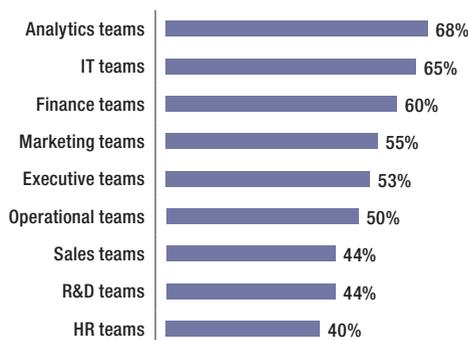


Figure 4. Based on 395 respondents. Only “yes” responses are shown.

Characteristics of the Data-Literate Organization

What does it take to be data literate? To answer this question we compared the three groups of survey respondents mentioned earlier in this report (who categorized their organizations as either data literate, somewhat data literate, or not data literate) across organizational and technical factors to understand characteristics of the data-literate organization. Figure 5 illustrates the characteristics of the data-literate organization based on significant differences between the groups.

Respondents were presented with statements about aspects of data literacy in their organization and asked if they agreed, disagreed, or were neutral.

- **Leadership.** Data-literate organizations were more likely than somewhat-literate or not-literate organizations to have leaders who encourage and reward data use (84% versus 43%).

These leaders are more likely to have confidence in the data used for decision making (81% versus 51%). They buy into the principles of data literacy.

- Culture.** Data-literate organizations were more likely than somewhat-literate or not-literate organizations to have a culture that encourages experimentation and innovation (77% versus 53%). In data-literate organizations, data is viewed as an asset. The culture is collaborative and results-driven (92% versus 46%). Analytics goals are tied to management goals (73% versus 40%).
- Data governance.** Data-literate organizations were more likely than somewhat-literate organizations to have a data governance plan in place (63% versus 46%). They were more likely than somewhat-literate or not-literate organizations to have easy access to trusted data (68% versus 21%). They were also more likely to use technology to help with data governance such as a data catalog. That isn't to say that there isn't room for improvement in data governance for data-literate organizations. We will see that even data-literate organizations still face issues with data governance.
- Infrastructure.** Data-literate organizations were more likely than somewhat-literate or not-literate organizations to have a data strategy in place (63% versus 25%). They were also more likely than not-literate organizations to use the cloud (as part of a hybrid environment in this case) for data management (48% versus 26%). They were more likely than organizations that aren't data literate to know what data was available to them for analysis (68% versus 25%).
- Analytics maturity.** Data-literate organizations were more likely than somewhat-literate or not-literate organizations to perform more advanced analytics. They have the skills in place for analytics (a +26 point spread). They were more likely to have the tools in place for analytics, as well.

The Data-Literate Organization

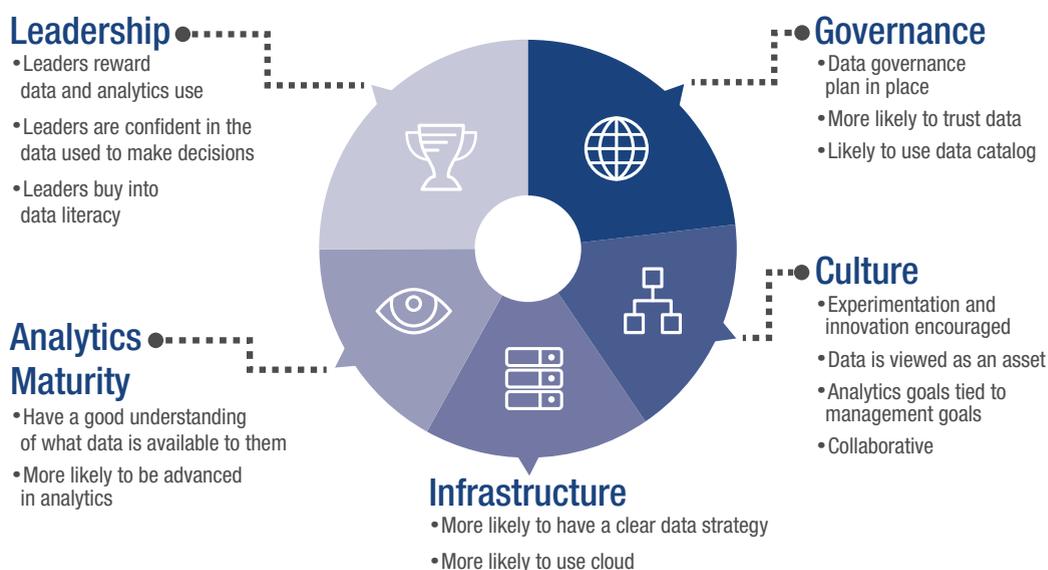


Figure 5. The characteristics of a data-literate organization, based on a statistical analysis across three groups: data literate, somewhat data literate, and not data literate.

Specific characteristics are not that surprising and are part of a virtuous circle. As organizations begin to reap the benefits of data and analytics deployments and build data literacy, they tend to put more data and analytics in place. The success builds on itself.

Governance and Data Literacy

As discussed above, organizations that are data literate were more likely than those that aren't to have a data governance plan in place. Building trust in data (e.g., knowing that it adheres to standards and is complete, accurate, and timely) is important to data literacy. If people lose trust in the data—for instance, by depending on a chart in a presentation that turns out to be based on inaccurate or poor-quality data—they will be reluctant to try again. That would be detrimental to growing data literacy.

At TDWI, we often see that governance is a top priority for organizations, especially as they advance into new platforms. In this survey, 51% of respondents had a data governance program in place and another 39% were implementing one this year. Only 10% had no plans for a data governance program (all not shown). Yet, the majority of respondents in this survey either disagreed or were on the fence about whether they thought their organizations were dealing well with important aspects of data governance. These include data quality, completeness, compliance, confidence, lineage, and whether the data governance program instills confidence in the data (see Figure 6). This was the case regardless of the level of data literacy. Data governance still has room for improvement, which is probably why it is always a high priority in TDWI research.

Please rate the following statements related to trust in data as either disagree, neutral, or agree.

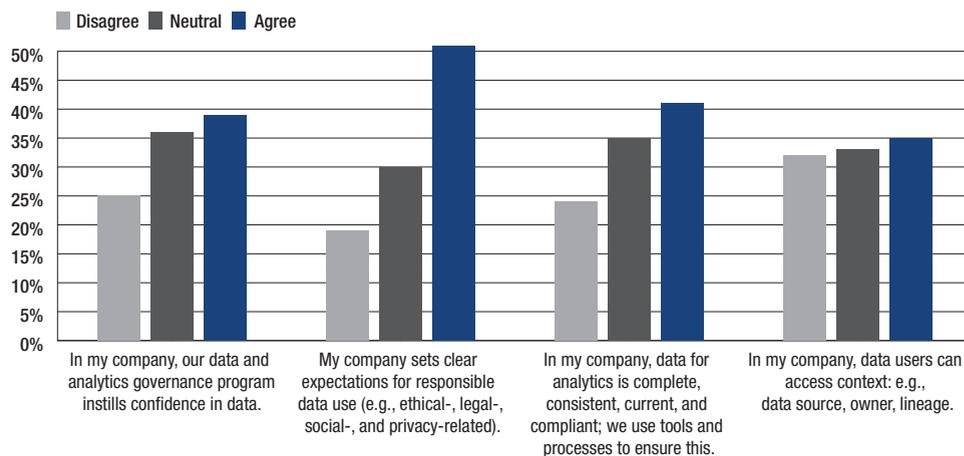


Figure 6. Based on 395 responses.

Best Practices for Data Literacy

We previously discussed the characteristics of organizations that felt they were data literate. They tended to be mature with regard to technology. They also had strong leadership and a data-driven culture. Given these traits, what steps are organizations taking to become (and remain) data literate?

Organizational Best Practices

Organizations are employing several best practices to become more data literate. Training, of course, is a top best practice for enabling a data-literate organization (Figure 7). Although 22 percent of respondents stated that their organizations did not offer any training, others were making use of different kinds of training.

- **Scheduled training.** Many respondents to this survey were using scheduled training, either in person or virtual, offered by external experts (39%), in-house experts (36%), technology vendors (36%), or some combination of the three. This training is important because it can help build data literacy skills (e.g., critical thinking, number sense, interpreting outputs, communicating with data) as well as help more individuals in the organization learn to use new tools. Nineteen percent of respondents stated that their organizations even offered support for university-level training. Of course, different roles in the organization may require different kinds of training.

Twenty-five percent of respondents mentioned that they have a literacy enablement team that is responsible for training (Figure 8). This group was more likely to have scheduled training in their organization, have a data governance plan in place (and trust their data), and set up processes. In other words, they take data literacy seriously.

- **Casual training.** Some organizations also make use of casual training such as lunch-and-learn sessions or office hours (40%). This training is typically offered by experts in the organization who may be part of a center of excellence or an analytics enablement team. This can be helpful because not everyone needs the same level of training or learns the same way.
- **Portals.** Some organizations provide online portals to help individuals in their organization learn about data literacy (30%). These portals may contain online courses offered by internal or external providers. They may contain documentation about how to use certain pieces of software.
- **Hiring for the skills.** Of course, organizations may also look externally for people with particular skills, such as data science or data engineering. Here, 30 percent of respondents stated that they often hire externally.

Training can help build data literacy skills and help individuals
in the organization learn to use new tools.



How is your company helping to build competency in data literacy? Please select all that apply.

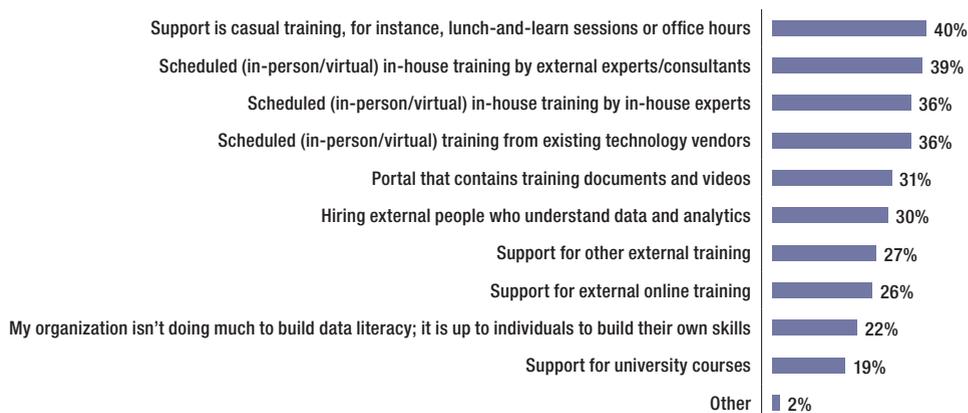


Figure 7. Based on 395 respondents.

Advancing the culture is another key component of building a data-literate organization (Figure 8). Often this comes from the top with executive buy-in. Thirty-seven percent of respondents said their executives were using data and analytics to make decisions. Data-literate behavior, then, is being modeled from the top. About 32% have hired a chief data officer or chief analytics officer who can help evangelize the use of data in decision making (29%) and help create a culture that supports learning.

In addition to enablement teams, some organizations have centers of excellence (18%) or innovation boards (8%). Such teams can accelerate the growth of data literacy in a company by putting data literacy plans in place that include defining goals, creating assessments, and launching programs for various groups across the organization.

A collaborative environment is also important for data literacy and advancing an analytics culture. Changing the culture does not happen overnight, but there are steps organizations can take to become more collaborative. For instance, 36 percent of respondents have set up processes to share information such as monthly meetings, internal analytics fairs, or newsletters to share success stories. In this way, individuals can come together, share information, and learn from each other. This starts to build a collaborative environment. For example, if someone reads a success story in a newsletter, they may reach out to the successful team and see if what they did can work for their own organization. This success will build on itself. At TDWI, we often see programs start small and then grow with more people becoming involved.

What is your organization doing to advance its data and analytics culture? Please select all that apply.

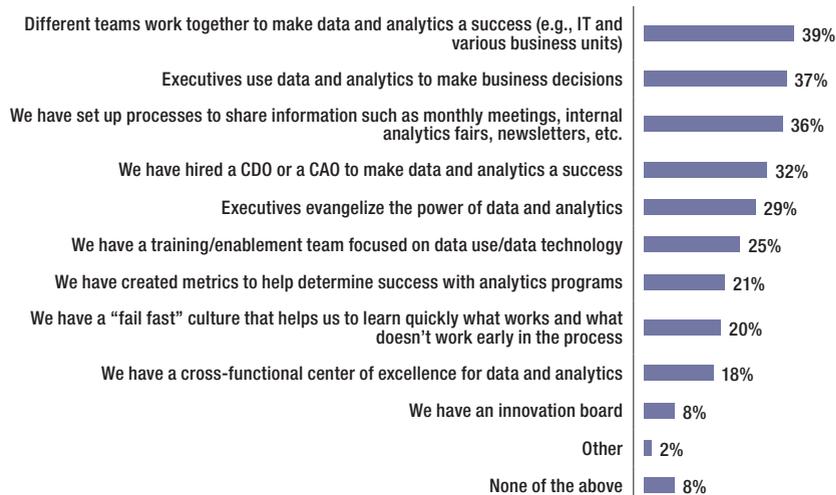


Figure 8. Based on 395 responses.

Technology Best Practices

As organizations work to become more data literate, there are technologies that can help make analytics easier to perform, provide trusted data, and help govern data. Some technologies that can help support data-literate organizations include (Figure 9):

- The cloud data warehouse.** Although the cloud data warehouse itself doesn't improve data literacy, it is foundational to the operational model of the company and to opening up data to more users. Many companies want to have a common data platform from which to perform analytics, and they are often moving to the cloud for this. In this survey, 34% of respondents stated that their cloud data warehouse provided governed access to data and opened up data to more users.
- Discovery and visualization tools.** Likewise, easy-to-use data visualization and discovery tools can help organizations explore data more efficiently and effectively. Thirty-two percent of respondents cited these tools as helping to open up data to more users. These tools can help organizations glean insights faster and democratize analytics (i.e., open analytics to more users). This can help drive data literacy (with the caveats described earlier).
- Modern data quality tools.** Modern data quality tools are often augmented to automatically find data issues. These tools can help ensure trusted and high-integrity data. Twenty-eight percent of respondents were using modern data quality tools and felt they helped provide governed access to data.
- Data catalogs.** Modern catalogs have features that can help improve users' ability to find data and build trust in that data. For instance, some data catalogs can deduce metadata. Others have rating features so users can see how colleagues regard the data. Some allow the data steward to certify the data as trustworthy. Others allow users to post comments about the data and the use cases for which they used it. In this survey, 28% of respondents were using

a data catalog and felt it provided governed access to data. At TDWI, we see implementing data catalogs as a top priority.

What tooling does your organization currently have to provide governed access to data and open analytics up to more users? Please select all that apply.

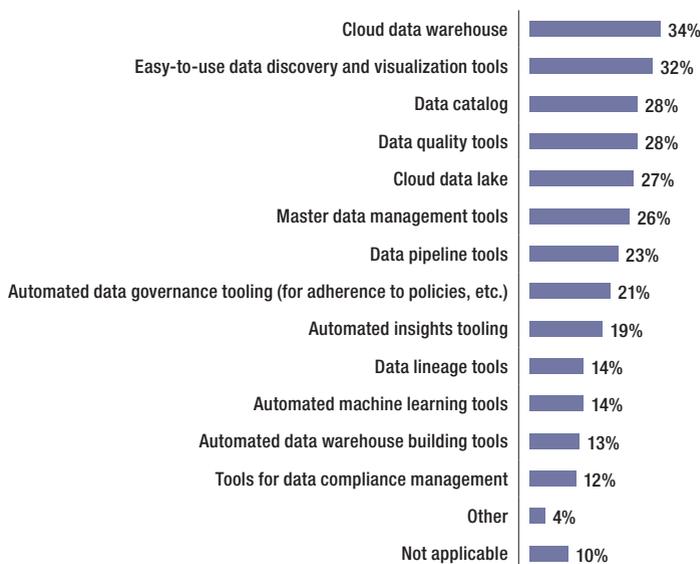


Figure 9. Based on 395 respondents.

Conclusion

Organizations must become data literate to be competitive. Data literacy enables companies to both understand their data and get the most value from it. Even though data literacy is important and is often a top priority for organizations in TDWI research, today many organizations have not embraced data literacy.

Getting to a higher level of data literacy will require work from both an organizational and technological perspective, including implementing a foundational data infrastructure, analytics tooling, and the proper governance. It will require training and a cultural change. The effort is worth it. Companies that are data literate are often more mature analytically, and more mature companies are typically more likely to measure value—as either a top- or bottom-line impact.

About Our Sponsors



For over 15 years, Amazon Web Services (AWS) has been the world’s most comprehensive and broadly adopted cloud platform. AWS has been continually expanding its services to support virtually any cloud workload, and it now has more than 200 fully featured services for compute, storage, databases, networking, analytics, machine learning and artificial intelligence (AI), Internet of Things (IoT), mobile, security, hybrid, virtual and augmented reality (VR and AR), media, and application development, deployment, and management from 80 Availability Zones within 25 geographic regions, with announced plans for 18 more Availability Zones and six more AWS Regions in Australia, India, Indonesia, Spain, Switzerland, and the United Arab Emirates. Millions of customers—including the fastest-growing start-ups, largest enterprises, and leading government agencies—trust AWS to power their infrastructure, become more agile, and lower costs. To learn more about AWS, visit aws.amazon.com.



Collibra is “the Data Intelligence company.” We accelerate trusted business outcomes by connecting the right data, insights, and algorithms to all data citizens. Our cloud-based platform connects IT and the business to build a data-driven culture for the digital enterprise. Global organizations choose Collibra to unlock the value of their data and turn it into a strategic, competitive asset. We have a diverse global footprint, with offices in the U.S., Australia, Belgium, France, U.K., and Poland. For more information, visit collibra.com.



Tableau helps people see and understand data. As a leading analytics platform, Tableau offers visual analytics with powerful AI, data management, and collaboration. From individuals to organizations of all sizes, customers around the world love using Tableau’s advanced analytics to fuel impactful, data-driven decisions. For more information, please visit www.tableau.com.



research

TDWI Research provides industry-leading research and advice for data and analytics professionals worldwide. TDWI Research focuses on modern data management, analytics, and data science approaches and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of business and technical challenges surrounding the deployment and use of data and analytics. TDWI Research offers in-depth research reports, commentary, assessments, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.



**Transforming Data
With Intelligence™**

A Division of 1105 Media
6300 Canoga Avenue, Suite 1150
Woodland Hills, CA 91367

E info@tdwi.org

tdwi.org