TABLE OF CONTENTS

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

IMPORTANCE OF DATA QUALITY IN THE 2020s ENTERPRISE .................................................................................... 4

DATA QUALITY IN CRISIS .............................................................................................................................................. 8

TAKING RESPONSIBILITY FOR DATA QUALITY ........................................................................................................... 14

DATA QUALITY IN THE CLOUD .................................................................................................................................. 16

WEIGHING THE SOLUTIONS .......................................................................................................................................... 19

CONCLUSION AND RECOMMENDATIONS ....................................................................................................................... 23

SURVEY DEMOGRAPHICS .............................................................................................................................................. 24
EXECUTIVE SUMMARY

In an era in which businesses are depending on advanced initiatives such as artificial intelligence and edge computing, there’s a need to believe that the data behind these technologies can be trusted with the fate of the business. The timeliness, completeness, and cleanliness of the data may make or break the fortunes of a data-driven enterprise. However, while today’s enterprises see data quality as a top strategic priority, eight in 10 enterprise data managers say their organizations’ data quality efforts are lagging or problematic.

These are among the findings of a survey of 238 data managers that was conducted by Unisphere Research, a division of Information Today, Inc., in partnership with Melissa. A majority of these data executives, 71%, are in positions to either recommend or implement data quality solutions. Respondents represent a range of industries and company sizes. (For more details, see the “Demographics” section at the end of this report.)

“We rely very heavily on data, and hence always validate the accuracy of data with our tool before final publishing,” states a respondent, pointing to the vital role of data in enterprises, and the need to guarantee its trustworthiness. Data quality results from issues with manual data entry errors, OCR (optical character recognition) glitches, ambiguousness in data, incomplete fields, redundancy, and inconsistent formats.

The survey revealed the following trends shaping enterprise data quality:

- Most data managers are at least somewhat skeptical about the quality of data in their enterprises and recognize achieving data quality as a top corporate priority. Data quality initiatives are not widespread enough to enable consistent quality, and discovery is often through informal processes. The greatest risk factors to data quality come from project complexities and other unknown factors that may arise as new technologies and architectures are implemented. Gaining organizational support is another challenge.

- There are many parties involved in achieving data quality. User involvement in the process is considered crucial. Data managers need to step up and take an advocacy role along with their technical contributions.

- Most organizations have moved at least some of their data to the cloud, but this doesn’t mean data quality issues are being resolved. In many cases, cloud adoption has increased the number of issues faced.

- While custom, home-grown solutions have dominated responses to data quality issues in the past, most data managers recognize that outside help is required.

On the following pages are the results of this comprehensive survey on the state of data quality in an era of digital transformation.
IMPORTANCE OF DATA QUALITY IN THE 2020s ENTERPRISE

Data quality is considered essential to corporate performance going forward in the digital economy. However, close to half of data managers report they either don’t have an active data quality strategy or aren’t aware of one. Such strategies are critical, as customer and product data are on the line.

There’s no question that data quality ranks at, or near, the top of enterprise priority lists. Eighty-seven percent see it as “important” to “somewhat” of a priority, with half citing is as their top priority (see Figure 1).

Close to half of data managers, 47%, indicate their enterprises either do not have, or they are not aware of, an advanced, cohesive data quality strategy—meaning it includes documented requirements and rules for measuring success, and ongoing activities tied to business goals. This suggests that strong data strategies are not as widespread as they need to be in an era in which every company needs to compete on data analytics (see Figure 2).

In many cases, a comprehensive data strategy needs to be baked into the corporate culture. As one respondent put it: “We design and verify, so adjust going forward—this has minimized any overarching revisions required.”

In most cases, customer and product data is on the line, as cited by 70% of data managers in the survey, and these categories are also the fastest-growing data domains. Prominent as well is vendor data, which takes on more importance in an era in which companies are relying on cloud vendors and other third-party providers (see Figures 3 and 4).
Figure 1: How important is data quality to your organization’s data strategy?

- Top priority to ensuring the ongoing success of projects/initiatives: 50%
- Somewhat important among competing priorities: 37%
- Data quality is often an afterthought at our organization: 8%
- Data quality is never considered: 1%
- Don’t know/unsure: 4%

Figure 2: Does your organization have a data quality strategy with documented requirements and rules for measuring success, and ongoing activities tied to business goals?

- Yes: 53%
- No: 26%
- Don’t know/unsure: 21%
Figure 3: What domains of data are stored in your databases?

- Customer contact data (name, address, phone, and email) - 70%
- Product data - 69%
- Vendor data - 61%
- Industry specific data - 45%
- Unstructured data (not parsed and fielded) - 42%
- Mobile and Web data (IP address, geo location, IoT data, etc.) - 34%
- Offline data (proprietary data about the customer, user behavior, demographics) - 32%
- Don’t know/unsure - 7%
- Other - 1%
Figure 4: What data domains are growing the fastest in your databases?

- Customer contact data (name, address, phone and email)
  - 41%
- Product data
  - 41%
- Vendor data
  - 24%
- Unstructured data (not parsed and fielded)
  - 38%
- Mobile and web data (IP address, geolocation, IoT data, etc.)
  - 22%
- Offline data (proprietary data about the customer, user behavior, demographics)
  - 16%
- Industry specific data
  - 16%
- Don’t know/unsure
  - 9%
DATA QUALITY IN CRISIS

Most data managers are at least somewhat skeptical about the quality of data in their enterprises and recognize achieving quality as a top corporate priority. Data quality initiatives are not widespread enough to enable consistent quality, and discovery is often through informal processes. The greatest risk factors to data quality come from project complexities and other unknown factors that may arise as new technologies and architectures are implemented. Gaining organizational support is another challenge.

While generally is confidence in the data that moves through their organizations, this confidence is lukewarm. Less than one-third of respondents, 30%, indicate they are “completely confident” in the integrity, accuracy, and trustworthiness of the data that moves through their organizations (see Figure 5). Notably, those organizations with comprehensive data quality strategies (per Figure 2) are more than twice as likely to express complete confidence in the viability of their data (see Figure 6).

For a majority of respondents, 74%, the quality of data presents issues, at least some of the time. For 26%, the problem is severe, cited as a “constant, ongoing issue.” Only 21% could say data quality is a non-issue at their enterprises at this time (see Figure 7).

There are many ways that data quality issues are surfaced, including formal and informal processes. For 45% of respondents, such issues are uncovered sporadically, as they upgrade databases or move to new systems. Likewise, 43% are alerted to issues as new analytics initiatives are launched. For another 45%, such issues are uncovered on a more regular basis through data quality systems and tools. However, disturbingly, 43% report finding out about such issues through customer and partner complaints (see Figure 8).

Having a process for acknowledging and addressing data quality issues ultimately improves the ability to respond to customers. “Data quality reports are sent to an IT operations team for review, and correction. [We] can forward [issues to a] service center to help identify the correct data and in rare cases customer contact,” one survey respondent explains.

How often are these issues addressed? Respondents in this survey are almost evenly divided in terms of the frequency by which they engage in data quality efforts. For a large segment, of 40%, such projects are few and far between—having only taken place more than a year ago, if ever. Similarly, another 40% report having had a data quality project within the past year (see Figure 9).

Human error may be the leading cause of data quality issues. A majority of respondents, 58%, indicate their enterprise data is keyed in by company employees. Close to half, 49%, report customers entering data (see Figure 10).

The unknowns involved with technology development are whathamper data quality the most, close to six in 10 data managers report. Gaining organizational support is also a potential roadblock, with 42% citing issues with gaining internal support. In addition, the financial aspects of data quality initiatives can be challenging, with more than one-third of respondents struggling with calculating return on investment (ROI) for data quality investments and getting the right amount of funding. When it comes to budgeting, accounting for unknowns, as well as ROI, are the most difficult aspects of securing budgets for data quality projects (see Figures 11 and 12).
Figure 5: How confident are you in the integrity, accuracy, and trustworthiness of data at your organization?

- Completely confident: 30%
- Somewhat confident: 63%
- Not confident at all: 3%
- Don’t know/unsure: 4%

Figure 6: Confidence in Data—Based on Presence of Data Quality Strategy

- Completely confident
  - Have Data Quality Strategy: 42%
  - No Data Quality Strategy: 17%
- Somewhat confident
  - Have Data Quality Strategy: 55%
  - No Data Quality Strategy: 76%

BUILDING A CULTURE OF TRUST IN A COMPETITIVE ECONOMY: 2021 SURVEY ON DATA QUALITY was produced by Unisphere Research and sponsored by Melissa. Unisphere Research is the market research unit of Unisphere Media, a division of Information Today, Inc., publishers of Database Trends and Applications magazine and the 5 Minute Briefing newsletters. To review abstracts of our past reports, visit www.unisphereresearch.com. Unisphere Media, 121 Chanlon Road, New Providence, NJ 07974; 908-795-3702.
Figure 7: How often is data quality an issue at your organization?

- Data quality is a constant, ongoing issue: 26%
- Data quality is an issue that pops up once in a while: 48%
- Data quality is rarely an issue: 17%
- Data quality is never an issue: 4%
- Don’t know/unsure: 5%

Figure 8: How do you find out about data quality issues?

- From our data quality systems/tools: 46%
- During new database projects (mergers/migrations/consolidations): 45%
- Customer/partner complaints: 43%
- During new analytics projects (data science, machine learning, AI): 43%
- Employee complaints: 41%
- We never hear about data quality issues: 3%
- Don’t know/unsure: 7%
**Figure 9: When was the last time your organization had a data quality project?**

- Within the past 12 months: 40%
- 1–2 years ago: 22%
- 3–5 years ago: 8%
- 5+ years ago: 4%
- Never: 6%
- Don’t know/unsure: 18%
- Other: 1%

**Figure 10: How does data get added into your databases?**

- Keyed in by employees: 58%
- Importing partner data: 54%
- Automated data collection: 53%
- Direct customer data entry: 49%
- Third-party sources: 48%
- Don’t know/unsure: 9%
Figure 11: What are your biggest data quality challenges?

- Project complexities and unknowns: 59%
- Internal support for project: 42%
- Getting funds approved: 28%
- Accurate calculation for expected ROI: 35%
- Budget is not an issue: 7%
- Don’t know/unsure: 11%
- Other: 3%
Figure 12: What is your biggest challenge to securing budget for a data quality project?

- Project complexities and unknowns: 23%
- Accurate calculation for expected ROI: 21%
- Getting funds approved: 16%
- Internal support for project: 19%
- Budget is not an issue: 7%
- Don’t know/unsure: 12%
- Other: 1%
TAKING RESPONSIBILITY FOR DATA QUALITY

There are many parties involved in achieving data quality. User involvement in the process is considered crucial. Data managers need to step up and take an advocacy role along with their technical contributions.

There is a pressing need for automation to monitor and address data quality issues. Close to two-thirds of data managers, 63%, say that they work with users to ensure the quality of their data, followed by monitoring its performance within applications. A majority have also built their own customized systems to track and monitor for data quality issues (see Figure 13).

Data quality is an effort seen at the upper levels of IT organizations. Close to half, 49%, report their C-level IT executives (CIO/CTO/CDO) and VPs of IT are in charge of these initiatives. Close to two-thirds, 64%, also indicate that their DBAs are involved with these efforts (see Figure 14 and 15).

Every organization has different levels of responsibility. “Typically, it is the business analyst’s responsibility to gather requirements,” states one respondent. “During development and unit testing phases, the development team must work with other internal groups to have data cleansed.”

Figure 13: How do you address the quality of your data?

<table>
<thead>
<tr>
<th>Method</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>User validation</td>
<td>63%</td>
</tr>
<tr>
<td>Application validation</td>
<td>54%</td>
</tr>
<tr>
<td>Home-grown systems</td>
<td>51%</td>
</tr>
<tr>
<td>Third party tools</td>
<td>41%</td>
</tr>
<tr>
<td>Has not been addressed</td>
<td>1%</td>
</tr>
<tr>
<td>Don’t know/unsure</td>
<td>8%</td>
</tr>
<tr>
<td>Other</td>
<td>1%</td>
</tr>
</tbody>
</table>
**Figure 14: Who makes decisions at your organization about data quality projects?**

<table>
<thead>
<tr>
<th>Role</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-Level IT (CIO/CTO/CDO) and VPs of IT</td>
<td>49%</td>
</tr>
<tr>
<td>IT Directors/Managers</td>
<td>43%</td>
</tr>
<tr>
<td>C-Level Business Executives (President/CEO/COO) and VPs of Business</td>
<td>34%</td>
</tr>
<tr>
<td>Line-of-Business Managers</td>
<td>32%</td>
</tr>
<tr>
<td>Database Administrators</td>
<td>21%</td>
</tr>
<tr>
<td>Cross-Enterprise Committees</td>
<td>18%</td>
</tr>
<tr>
<td>Other</td>
<td>1%</td>
</tr>
</tbody>
</table>

**Figure 15: Are database administrators involved with data quality projects at your organization?**

- Yes: 64%
- No: 22%
- Don’t know/unsure: 14%
DATA QUALITY IN THE CLOUD

Most organizations have moved at least some of their data to the cloud, but this doesn’t mean data quality issues are being resolved. In many cases, cloud adoption has increased the number of issues faced.

A substantial segment of respondents, 42%, report their data quality challenges increased to some degree after moving to the cloud. Cloud does not appear to reduce the challenges already existing—only 8% report seeing issues alleviated through cloud, and 14% donot know if there was any effect (see Figure 16).

Just about every enterprise stores data in the cloud. For 22% of enterprises, this constitutes a majority of data (50% or greater) now maintained in the cloud (see Figure 17).

The large infrastructure-as-a-service and platform-as-a-service leaders also provide the most data storage support, the survey shows. Microsoft Azure and Amazon Web Services are the two cloud infrastructure platforms of choice when it comes to data storage, with just over half of respondents storing their data in each. Google Cloud Platform rounds out the top three (see Figure 18).

Figure 16: If you are using cloud-based solutions, how have your data quality challenges changed after moving to the cloud?
Figure 17: What percentage of your data is stored in the cloud?

- 100%: 4%
- 75% to 99%: 13%
- 50% to 74%: 15%
- 25% to 59%: 22%
- 10% to 24%: 19%
- Fewer than 10%: 9%
- None: 3%
- Don’t know/unsure: 14%
Figure 18: Which cloud-based solutions do you use to store your data?

- Microsoft Azure: 51%
- AWS: 52%
- Google: 21%
- Salesforce: 16%
- Private Cloud: 16%
- IBM: 12%
- Oracle: 13%
- None: 3%
- Don’t know/unsure: 9%
WEIGHING THE SOLUTIONS

While custom, home-grown solutions have dominated responses to data quality issues in the past, most data managers recognize that outside help is required.

For the most part, respondents have turned to home-grown solutions to address data quality issues. A number also assign the responsibility to business users. While most data quality efforts have been custom-built in the past, the largest segments of respondents are turning to outside assistance to move forward in the future (see Figures 19 and 20).

Email is a preferred method of communication with software vendors, as rated by 73% of respondents. Otherwise, live interactions are preferred—over the phone or live chat, as cited by 55% and 54%, respectively (see Figure 21).

The European Union’s General Data and Protection Regulation (GDPR) dominates as the data quality measure of most concern, as many companies have European customers or deal with European-based entities, and, as a result, 41% seek assistance with GDPR mandates (see Figure 22).
**Figure 19: How have you tried to solve your data quality issues in the past?**

<table>
<thead>
<tr>
<th>Method</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Built custom code for validation and consistency</td>
<td>58%</td>
</tr>
<tr>
<td>Business users have been responsible for validation</td>
<td>38%</td>
</tr>
<tr>
<td>Contacted clients for data validation</td>
<td>35%</td>
</tr>
<tr>
<td>Unable to find a solution</td>
<td>5%</td>
</tr>
<tr>
<td>None</td>
<td>4%</td>
</tr>
<tr>
<td>Don’t know/unsure</td>
<td>15%</td>
</tr>
</tbody>
</table>

**Figure 20: How do you envision solving your data quality challenges in the future?**

<table>
<thead>
<tr>
<th>Method</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third-party application integration</td>
<td>48%</td>
</tr>
<tr>
<td>Cloud-based API</td>
<td>46%</td>
</tr>
<tr>
<td>Open source tools</td>
<td>41%</td>
</tr>
<tr>
<td>Roll your own</td>
<td>34%</td>
</tr>
<tr>
<td>Don’t know/unsure</td>
<td>25%</td>
</tr>
<tr>
<td>Other</td>
<td>1%</td>
</tr>
</tbody>
</table>
Figure 21: How do developers at your organization prefer to interact with software vendors for product support?

<table>
<thead>
<tr>
<th>Method</th>
<th>Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email</td>
<td>73%</td>
</tr>
<tr>
<td>Phone</td>
<td>55%</td>
</tr>
<tr>
<td>Live Chat</td>
<td>54%</td>
</tr>
<tr>
<td>Live Video</td>
<td>39%</td>
</tr>
<tr>
<td>Forums</td>
<td>29%</td>
</tr>
<tr>
<td>Discord</td>
<td>4%</td>
</tr>
<tr>
<td>Social Media</td>
<td>12%</td>
</tr>
<tr>
<td>Wiki</td>
<td>12%</td>
</tr>
<tr>
<td>Other</td>
<td>3%</td>
</tr>
</tbody>
</table>
Figure 22: What privacy and security measures are you expecting when working with a data quality partner?

- GDPR: 41%
- HIPAA/HITRUST Compliance: 39%
- CCPA-Certified: 21%
- SOC 2.2 Compliance: 21%
- ISO Compliance: 27%
- None: 8%
- Don’t know/unsure: 23%
- Other: 4%
CONCLUSION AND RECOMMENDATIONS

The data in this survey demonstrates that data quality is considered critical to competing in today’s digital economy. The following are key steps organizations can follow for achieving this:

- Take a leadership role in advocating for data quality. Encourage and lead collaboration between enterprise teams.

- Measure and track key data metrics on an ongoing basis, especially for timeliness, consistency, and completeness.

- Monitor and ensure that incoming data meets standards.

- Don’t relegate data quality to cloud providers; this is always the responsibility of the individual enterprises.

- Automate as much of the data quality process as possible. Make data quality an integral part of a DataOps pipeline.

The survey results, drawn from today’s leading data practitioners, shows that data quality is considered essential to corporate performance going forward in the digital economy. At the same time, data managers don’t trust the state of their organizations’ data quality, and many do not have cohesive strategies to address issues that arise. The time has come for enterprise-scale approaches and the adoption of proper tools and methodologies to build data quality into all digital transformation efforts.
SURVEY DEMOGRAPHICS

Figure 23: What is your primary job title?

- Director/Manager of IS/IT or computer-related function: 13%
- Database Administrator (DBA): 11%
- Chief Information Officer/CTO/Vice President of IT: 10%
- Data Architect: 9%
- Director/Manager Analytics: 8%
- IT Consultant: 7%
- Data Scientist: 4%
- Executive management level for the business: 4%
- IT Operations Manager: 4%
- Systems Administrator: 4%
- Programmer/Developer: 4%
- Project Manager: 4%
- Director/Manager Application Development: 3%
- Chief Data Officer: 2%
- Analyst/Systems Analyst: 2%
- Manager of a business unit (other than computer): 2%
- Other: 11%
Figure 24: How many employees are in your entire organization? (Include all locations, branches, and subsidiaries)

- 1–500 employees: 34%
- 501–1,000 employees: 16%
- 1,001–5,000 employees: 15%
- 5,001–10,000 employees: 11%
- More than 10,000 employees: 25%
**Figure 25: What is your organization’s primary industry?**

<table>
<thead>
<tr>
<th>Industry</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>25%</td>
</tr>
<tr>
<td>Financial Services</td>
<td>11%</td>
</tr>
<tr>
<td>Healthcare/Pharmaceutical</td>
<td>11%</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>10%</td>
</tr>
<tr>
<td>Education (all levels)</td>
<td>8%</td>
</tr>
<tr>
<td>Business Services</td>
<td>8%</td>
</tr>
<tr>
<td>Government (all levels)</td>
<td>5%</td>
</tr>
<tr>
<td>Retail</td>
<td>5%</td>
</tr>
<tr>
<td>Insurance</td>
<td>4%</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>3%</td>
</tr>
<tr>
<td>Media/Entertainment</td>
<td>2%</td>
</tr>
<tr>
<td>Aviation/Aerospace</td>
<td>1%</td>
</tr>
<tr>
<td>Transportation</td>
<td>1%</td>
</tr>
<tr>
<td>Other</td>
<td>8%</td>
</tr>
</tbody>
</table>

0 20 40 60 80 100
Figure 26: What is your primary role in improving the quality of data at your organization?

- Implement solutions: 36%
- Recommend solutions: 35%
- Technical support: 12%
- Business user: 6%
- Financial approval: 3%
- Not involved: 5%
- Other: 2%