

THE POST-RELATIONAL REALITY SETS IN: 2011 SURVEY ON UNSTRUCTURED DATA

By Joseph McKendrick, Research Analyst
Produced by Unisphere Research, A Division of Information Today, Inc.
June 2011

Sponsored by  **MarkLogic™**

Produced by  **UNISPHERE®**
RESEARCH
Delivering Certainty
Thomas J. Wilson, President

TABLE OF CONTENTS

<i>Executive Summary</i>	3
<i>Where Unstructured Data Lives</i>	5
<i>Organizational Issues With Unstructured Data</i>	11
<i>Managing Unstructured Data</i>	17
<i>Technology for Unstructured Environments</i>	24
<i>Conclusions and Recommendations</i>	31
<i>Demographics</i>	32

EXECUTIVE SUMMARY

Many refer to the huge influx of data into today's enterprises as "Big Data"—terabytes and gigabytes of bits and bytes that are overwhelming many companies' information technology infrastructures. Across the information technology industry, there is agreement by analysts, vendors and end users alike that the amount of data flowing into organizations—from end users, partners, applications, sensors, and machines—is unprecedented, and estimates put the total amount of data in the world at 295 exabytes. In fact, some analysts estimate that the amount of digital data in the world surpassed that of analog data, and at this time, more than 94% of our memory is now in digital form. The problem is, a large portion of it exists across cyberspace in an unorganized fashion.

Much of this explosive growth consists of unstructured information, which is creating new types of challenges in terms of governance, management and security practices. Unstructured data typically consists of files such as word processing documents, spreadsheets, PDFs, social media and messaging system content, video, multimedia, and graphics. Unlike relational data, unstructured information lacks defined data types and rules to enforce where that data is stored. In addition, unstructured data is created and generated across a broad cross-section of end users, and is stored in user-defined directories, beyond the reach of enterprise rules. However, much of this unstructured data may be crucial to the organization, as it could contain everything from product design documents in businesses to training manuals within military organizations and electronic health records in medical institutions.

A new survey of data and IT managers finds that companies are only beginning to grasp the complexities created by all this new unstructured data. Even the most mature organizations that acknowledge they depend on unstructured data still do not have effective governance or best practices in place. In many cases, managers and professionals within organizations are not aware of the extent of their unstructured information stores. The survey gathered input from 446 data managers and professionals who are subscribers to *Database Trends and Applications*, and was conducted by Unisphere Research, a division of Information Today, Inc., in partnership with MarkLogic in March and April of 2011.

The survey results imply that companies are missing the opportunity to leverage the full value of this unstructured data—that many view it as a storage or technology challenge to be

dealt with accordingly. However, companies that are merely salting away this data into storage—without fully enabling it to be indexed and searchable—may miss the business opportunities that greater insight may bring. In addition, a segment of companies are even leveraging unstructured data, embedding or building it into information products and services.

Survey respondents have a variety of job roles both within IT and business, and represent a wide range of company types and sizes. One-fifth of respondents have more than 500 database instances at their establishments. The largest segment of respondents consists of professionals who are on the front lines of database administration and development, and includes database administrators, followed by programmers and developers. About one-fourth of respondents are in management.

More than one-fourth of respondents come from very large organizations with more than 10,000 employees. But there is also a sizable contingent of small-to-medium-size businesses represented in the survey as well. In terms of industry groups, the largest segment seen in this survey consists of organizations involved in IT services and consulting, identified by 12% of respondents. The government sector is also heavily represented, with 11% of respondents representing federal, local or state agencies and organizations. Another 10% are either with financial services firms, or healthcare facilities. (See Figures 38–41 at the end of this report.)

The survey uncovered these key findings:

- Unstructured data, as defined by respondents, runs the gamut from web content to word processing documents. Eight out of 10 respondents say this type of data is increasing, and about half expect it to surpass traditional relational data in the near future.
- While two out of five respondents say upper management is barely aware of it, those on the front lines have serious concerns about the volume of unstructured data in their organizations. Respondents on the business and management side are worried that the growth of unstructured data is resulting in data clutter, as well as challenges with indexing and tagging the information appropriately. More technical respondents such as DBAs, on the other hand, are more anxious about increased storage and security requirements. Organizations with higher concentrations of unstructured data are leading the way in terms of data governance—but even these companies face issues with corporate awareness.



-
- Most respondents report they either do not have or are not aware of governance procedures in place for unstructured data, suggesting that much of this kind of data is haphazardly being distributed and managed across enterprises. In addition, a majority of companies—even those with large concentrations of unstructured data—still devote most resources to managing relational data. A large segment of respondents' organizations are now including unstructured data under a common governance structure. Most respondents do not feel their current data infrastructures are capable of handling the growth of unstructured data. Among companies that already have high levels of unstructured data, only 18% say their current infrastructure will be sufficient.
 - Email and messaging systems are the primary vehicles for creating unstructured data, but much of it also ends up in content management systems. Most companies with large concentrations of unstructured data are moving such information into non-relational databases.

WHERE UNSTRUCTURED DATA LIVES

Unstructured data, as defined by respondents, runs the gamut from web content to word processing documents. Eight out of 10 respondents say this type of data is increasing, and about half expect it to surpass traditional relational data in the near future.

Respondents identify a variety of formats and file types as being “unstructured data” (see Figure 1), as well as “Big Data.” (See Figure 2.) In a majority of cases, unstructured data includes business documents such as spreadsheets, PDF files, social networking content, digitized articles, and video or audio files. Unstructured data is a significant part of the Big Data explosion, and 91% of the respondents in the survey say they are aware of unstructured data files within their enterprise systems. Only 2% say they have no unstructured data at all. More than one-fourth of the survey respondents now say that the majority of their enterprise data is unstructured. (See Figure 3.)

The degree of known unstructured data on premises varies by industry sector. Leading the way in the leveraging of unstructured data are organizations within government, education, or non-profits, where 36% report that the majority of their data assets are unstructured files. The other leading group, also at 36%, is the information technology sector, comprised of software development, integration and consulting firms. (See Figure 4.)

Which types of data do respondents find the most difficult to manage in their day-to-day work? A majority say they encounter difficulties with standard business documents, which can include presentations, spreadsheets, and other types of files. More than one-third are having difficulties with managing and storing PDF files, and close to one-third are concerned about what to do with social content such as tweets, posts, blog, forums, or wikis. Such social media and social networking data—coming both from

within and from outside the enterprise—is increasingly playing a role in managing brand awareness, or reaching out to and engaging with key customer segments. Interestingly, only 12% report difficulty in managing their relational data, suggesting that the highly structured RDBMS platforms offer a well-contained process model that has well-defined data types and rules to enforce where that data is stored. (See Figure 5.)

For most respondents, the amount of unstructured data across their enterprises has changed dramatically over the past 3 years. At least 78% say there has been an increase in this type of data, and 30% regard this growth as significant. (See Figure 6.) A similar number anticipate continued growth over the next 3 years. (See Figure 7.)

Ultimately, about half of the respondents agree that the amount of unstructured data in their enterprises will eventually surpass the amount of structured or relational data—at least within the next 10 years or so. At this time, about one out of five say unstructured data already tops their total relational data stores. Much of the growth behind “Big Data”—hundreds of terabytes and even petabytes worth of data—consists of unstructured data flowing into organizations from end users, partner systems, sensors and the web at large. (See Figure 8.)

The industry group that anticipates the most growth of unstructured data over the long run is manufacturers. Two-fifths of respondents in this segment predict the amount of unstructured data within their enterprises will exceed structured or relational data within the next 3 to 10 years. This is a surprising finding within an industry sector traditionally not associated with new data types, and may be the result of the increasing adoption of digital product design and computer-aided manufacturing. (See Figure 9.)

Figure 1: Respondents' Definitions of "Unstructured Data"

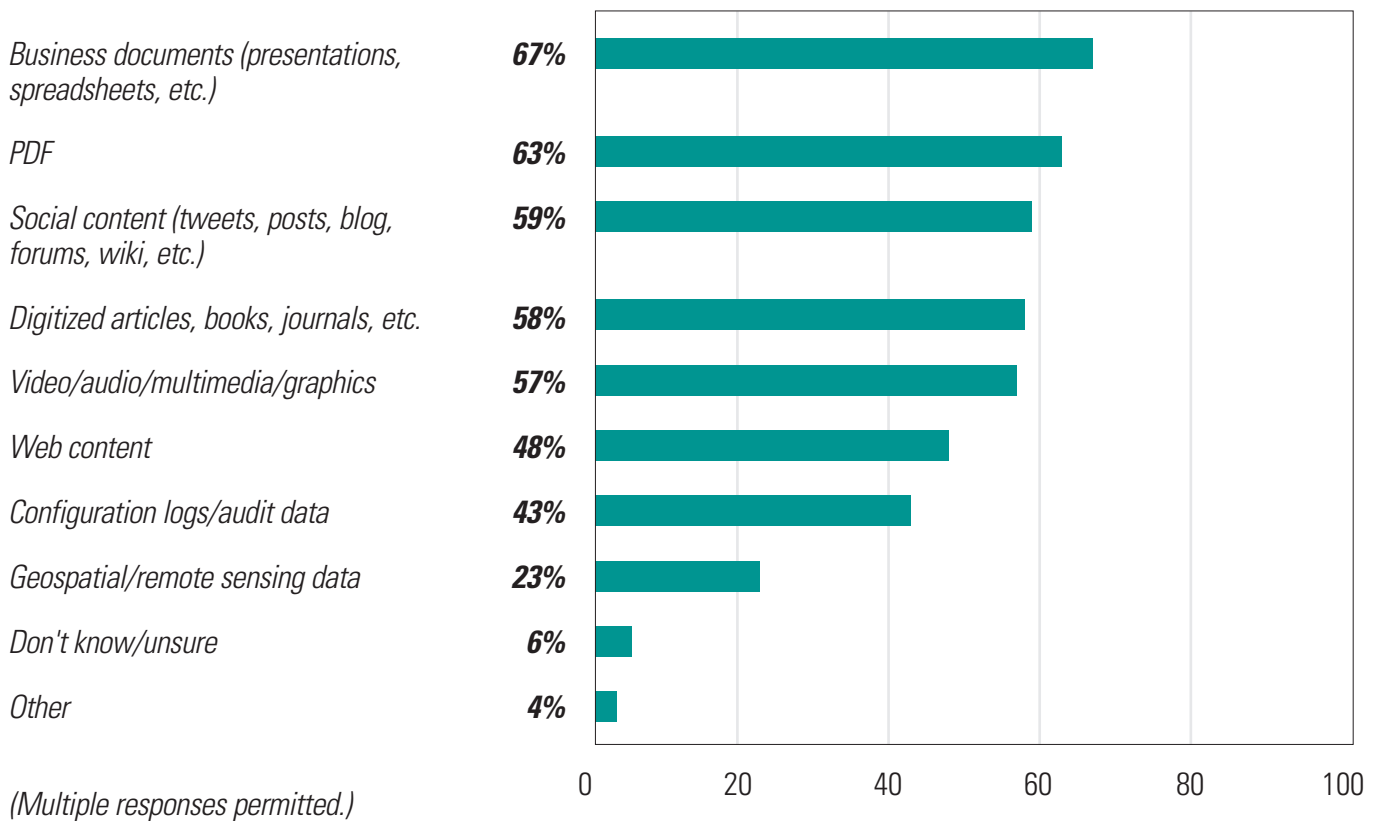
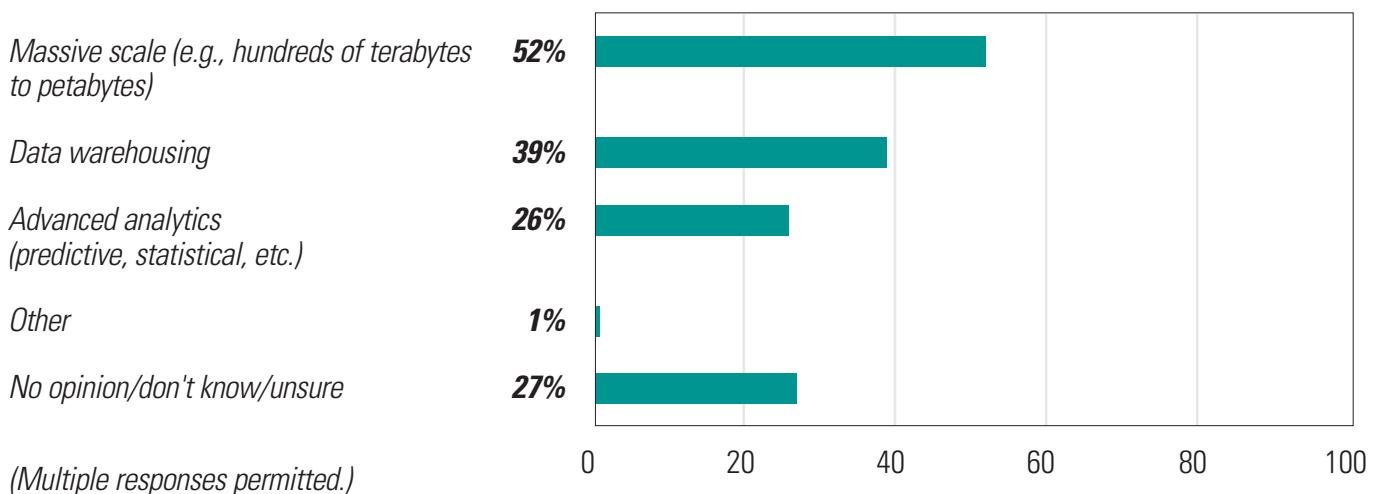


Figure 2: Respondents' Definitions of "Big Data"



The Post-Relational Reality Sets In: 2011 Survey on Unstructured Data is sponsored by MarkLogic Corporation and produced by Unisphere Research, A Division of Information Today, Inc. Unisphere Research is the market research division of Unisphere Media, publishers of Database Trends and Applications magazine, www.dbta.com and the 5 Minute Briefing email newsletters. To review abstracts of past reports or to obtain full final reports prepared by Unisphere Research, contact publisher@dbta.com.

Data collection and analysis performed with SurveyMethods.

Figure 3: Percentage Known Unstructured Data Within Respondents' Enterprises

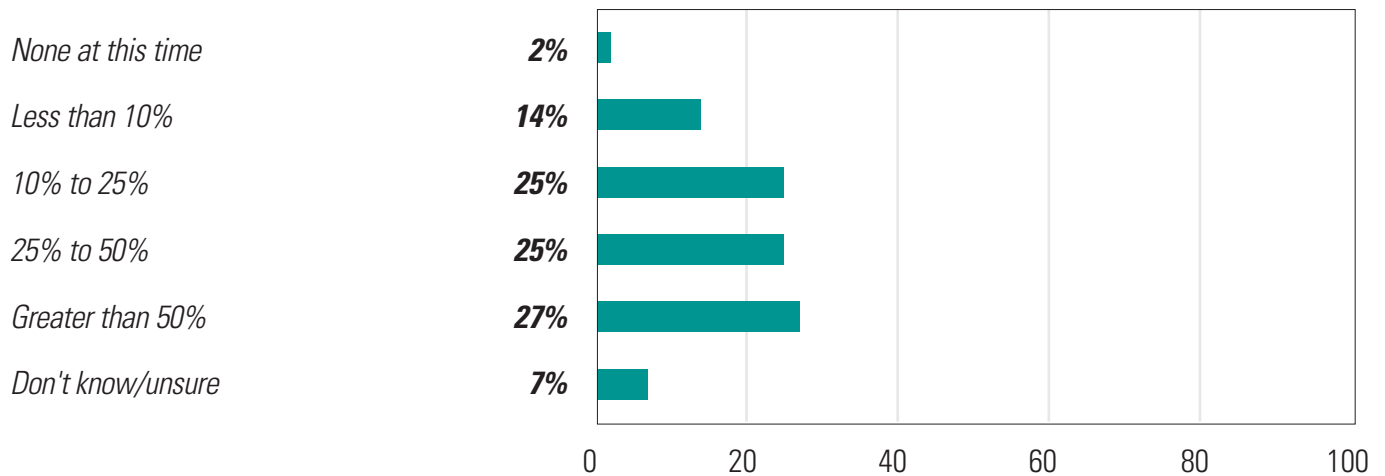


Figure 4: Percentage of Enterprises in Which Majority of Data is Known to be Unstructured (Greater than 50%)—By Industry Group

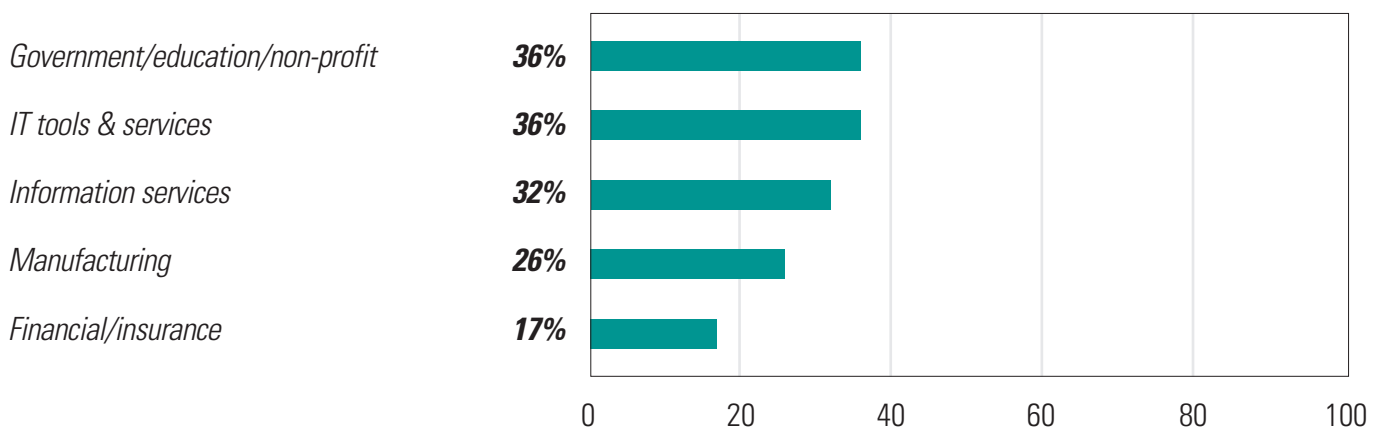


Figure 5: Data Most Difficult in Day-to-Day Work

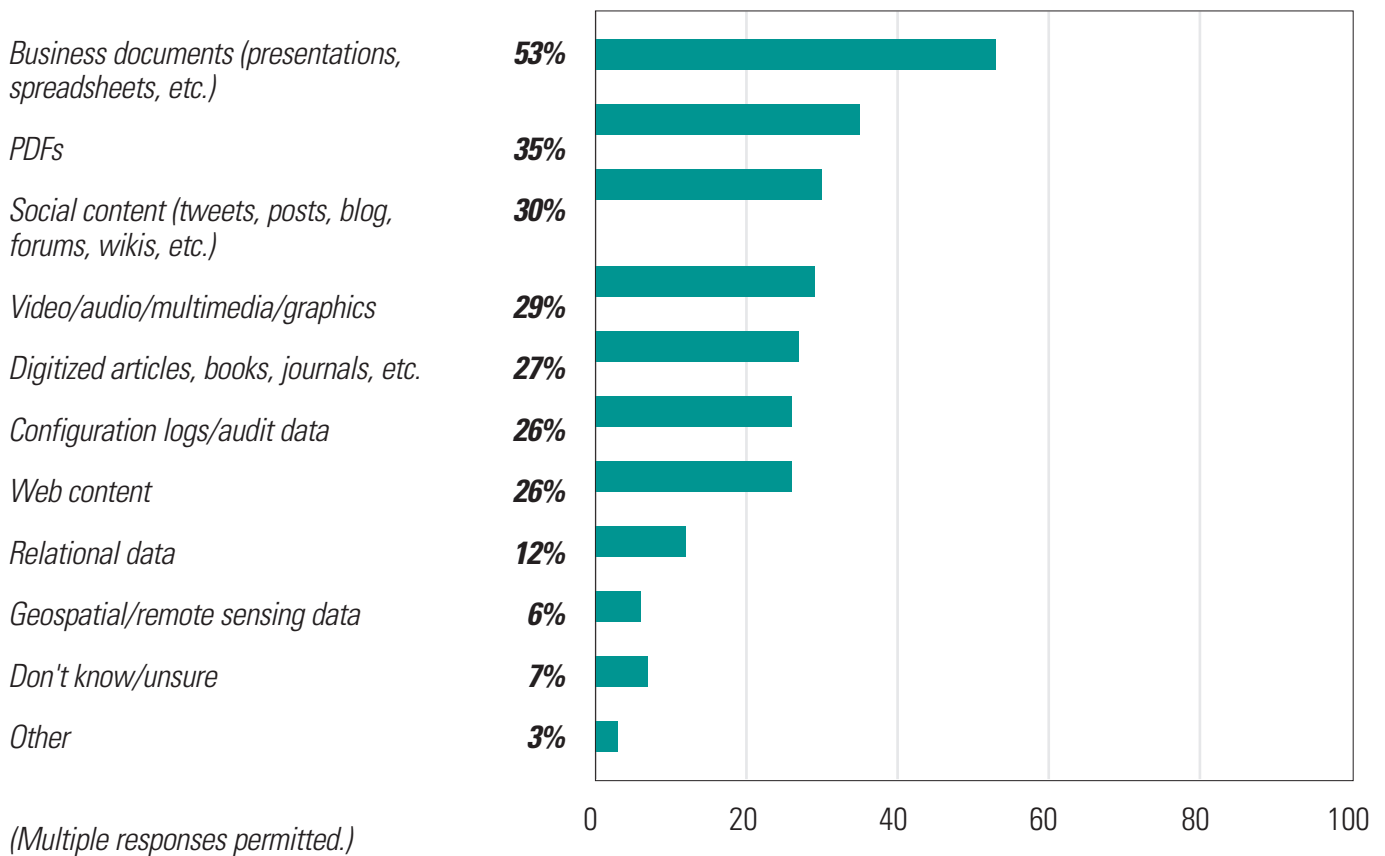
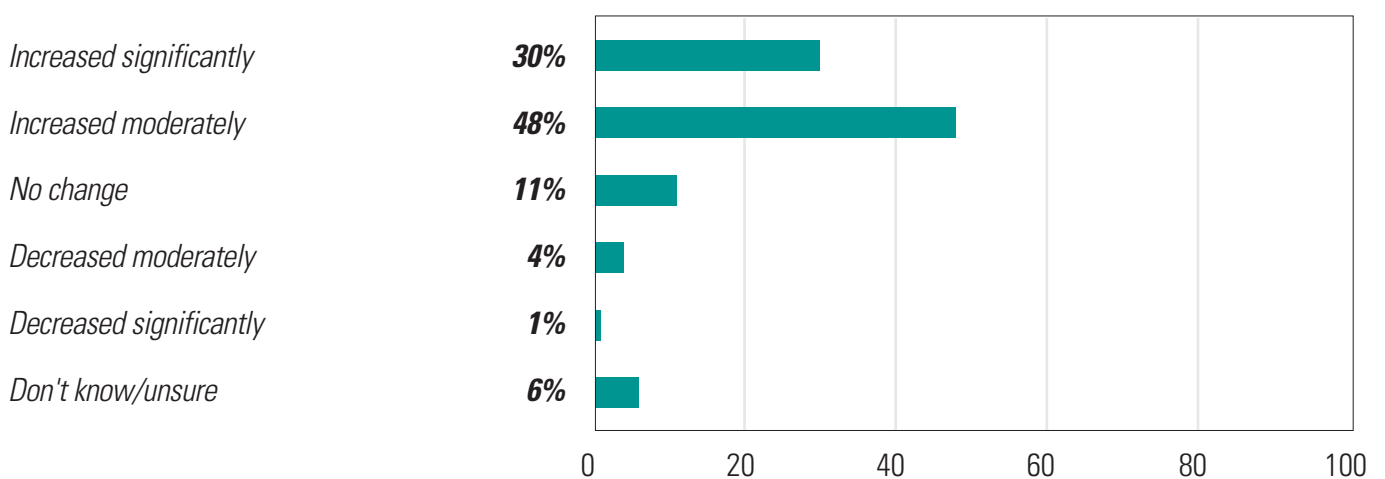


Figure 6: How Amount of Enterprise Unstructured Data Has Changed Over Past 3 Years



The Post-Relational Reality Sets In: 2011 Survey on Unstructured Data is sponsored by MarkLogic Corporation and produced by Unisphere Research, A Division of Information Today, Inc. Unisphere Research is the market research division of Unisphere Media, publishers of Database Trends and Applications magazine, www.dbta.com and the 5 Minute Briefing email newsletters. To review abstracts of past reports or to obtain full final reports prepared by Unisphere Research, contact publisher@dbta.com.

Data collection and analysis performed with SurveyMethods.

Figure 7: How Amount of Enterprise Unstructured Data Will Change Over Next 3 Years

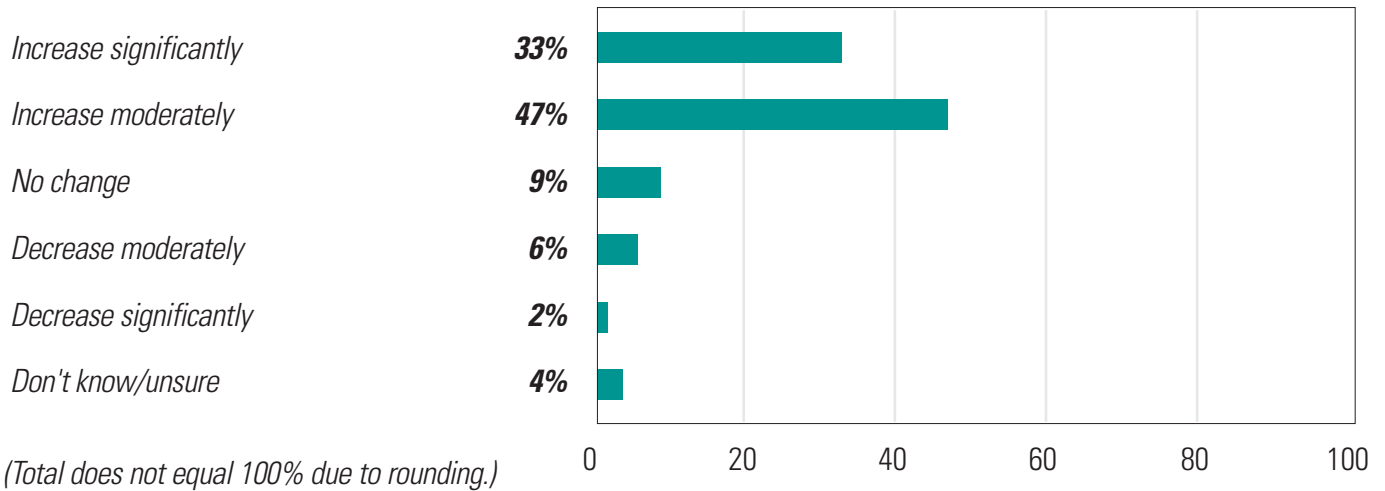


Figure 8: Will Amount of Unstructured Data Surpass Structured/Relational Data?

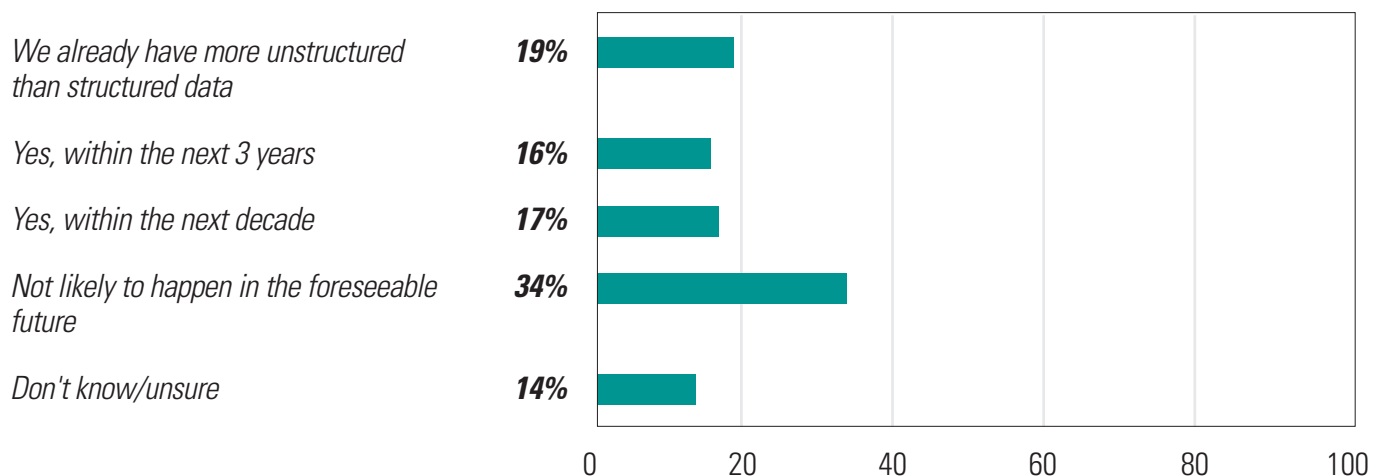
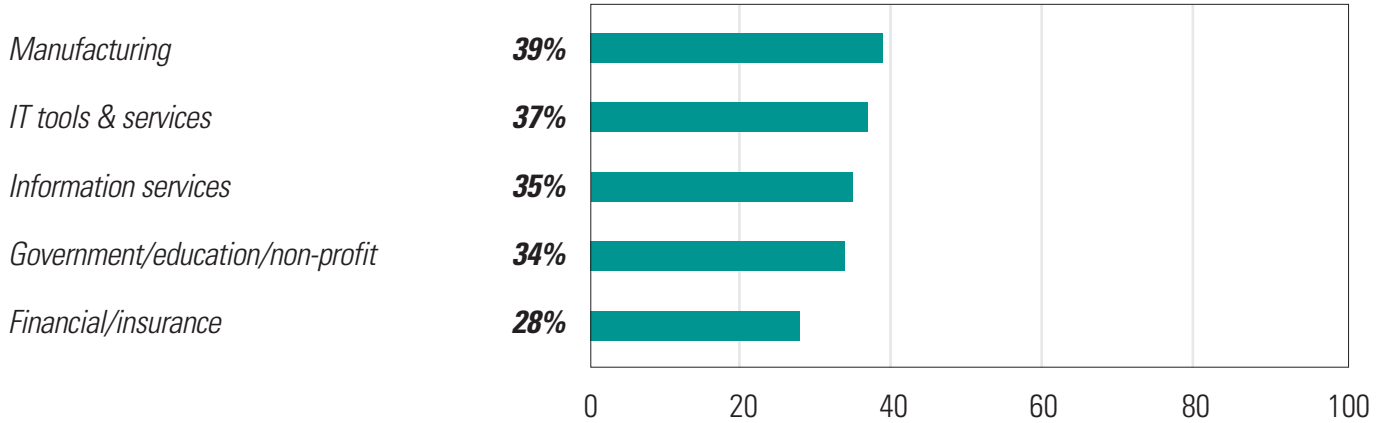


Figure 9: Where Unstructured Data Will Surpass Structured Data —By Industry Group

(Percentage reporting volume of unstructured data in their enterprises will surpass structured data within next 3–10 years)



ORGANIZATIONAL ISSUES WITH UNSTRUCTURED DATA

While two out of five respondents say upper management is barely aware of it, those on the front lines have serious concerns about the volume of unstructured data in their organizations. Respondents on the business and management side are worried that the growth of unstructured data is resulting in data clutter, as well as challenges with indexing and tagging the information appropriately. More technical respondents such as DBAs, on the other hand, are more anxious about increased storage and security requirements. Organizations with higher concentrations of unstructured data are leading the way in terms of data governance—but even these companies face issues with corporate awareness.

With the relentless growth of unstructured data comes the issue of data “clutter,” in which files of varying formats and serving various purposes may be piling into data centers or across user-defined directories and client systems. This is a great concern among a majority of survey respondents, along with the added storage requirements such data brings. Close to half are also concerned about better managing these multiple formats.

Additionally, about half have focused on the challenges of being able to index or tag unstructured files, to enable the information to be quickly and intuitively accessed on demand, as well as cross-linked to other pieces of information. A preponderance of text mining and text analytics tools have emerged on the market to address these concerns, as well as increased support for tagging and creating metadata, which enhances the searchability of unstructured files. (See Figure 10.)

Often, these issues are driven by differing perceptions across organizations. For example, when sorted by job titles, respondents on the business side are more concerned about data clutter than more technical staff members such as DBAs (70% versus 56%, respectively). However, DBAs tend to be much more concerned with managing additional storage than business users (57% versus 41%), as well as any security risks that come with unstructured data files (49% versus 30%, respectively). (See Figure 11.)

How aware are managers and professionals in respondents’ companies of the extent to which unstructured data exists in their enterprises? About half indicate there is a degree of awareness on the part of upper management (“highly” or “moderately” aware). Less than one-fifth (17%) could say there is a high level of awareness, however. Notably, a substantial percentage (40%) of respondents say that management at their organizations is either only “slightly” aware or not aware at all. This lack of management awareness means it may be difficult for data managers to secure the funding and resources needed to properly secure, store, and fully leverage the large volumes of unstructured data coming into their organizations. (See Figure 12.)

There is a natural increase in awareness evident among those companies in the survey with high levels of unstructured data (those reporting more than half of their enterprise information is unstructured). However, even among these companies, only 28% say their business-side managers are highly cognizant or engaged with such data. This suggests that the disconnect between executive management and data managers is wide among the companies with the most to gain from effectively leveraging their unstructured data. As a result, business users may only be accessing a fraction of the data available within the enterprise. (See Figure 13.)

With only about half of the companies covered in this survey fully aware of the extent of unstructured data within their enterprises, how committed are companies to addressing the challenges associated with managing unstructured data? Again, about half (45%) report their companies are “highly” or “moderately” committed to addressing the unstructured data challenge. (See Figure 14.)

Respondents are taking some steps to help address unstructured data challenges. Close to half report they are actually re-examining their data management processes, which may include measures to include unstructured data within existing databases, or deploying new types of databases. Along these lines, 43% also report they are looking at new technologies to help manage these processes, and three out of 10 are at the point of investing in new technologies or infrastructure. It’s notable that few companies at this time see increasing staff as a way to address these issues—only 8% are hiring new staff or consultants. This may indicate a reluctance to invest larger amounts of resources to handle increasing data levels at this time. As discussed earlier, there is an extreme lack of awareness, or disconnect, on the part of managers and executives about the extent of unstructured data within enterprises. (See Figure 15.)

A majority of respondents report that unstructured data is an essential part of their business, meaning that it may be a component of services or products offered to customers or constituents. At least 57% indicate that unstructured data plays an “extremely” or “very” important role in their businesses. About one out of five, or 18%, consider unstructured data to be at the core of their business. (See Figure 16.)

The percentage of respondents citing unstructured data as extremely important varies by industry, led by information services firms (25%), as well as manufacturing (22%). As reported in the previous section, manufacturing companies appear to be taking the lead with unstructured data deployments, suggesting that non-relational data is becoming a key part of product development processes. (See Figure 17.)

Figure 10: Challenges in Unstructured Data Growth

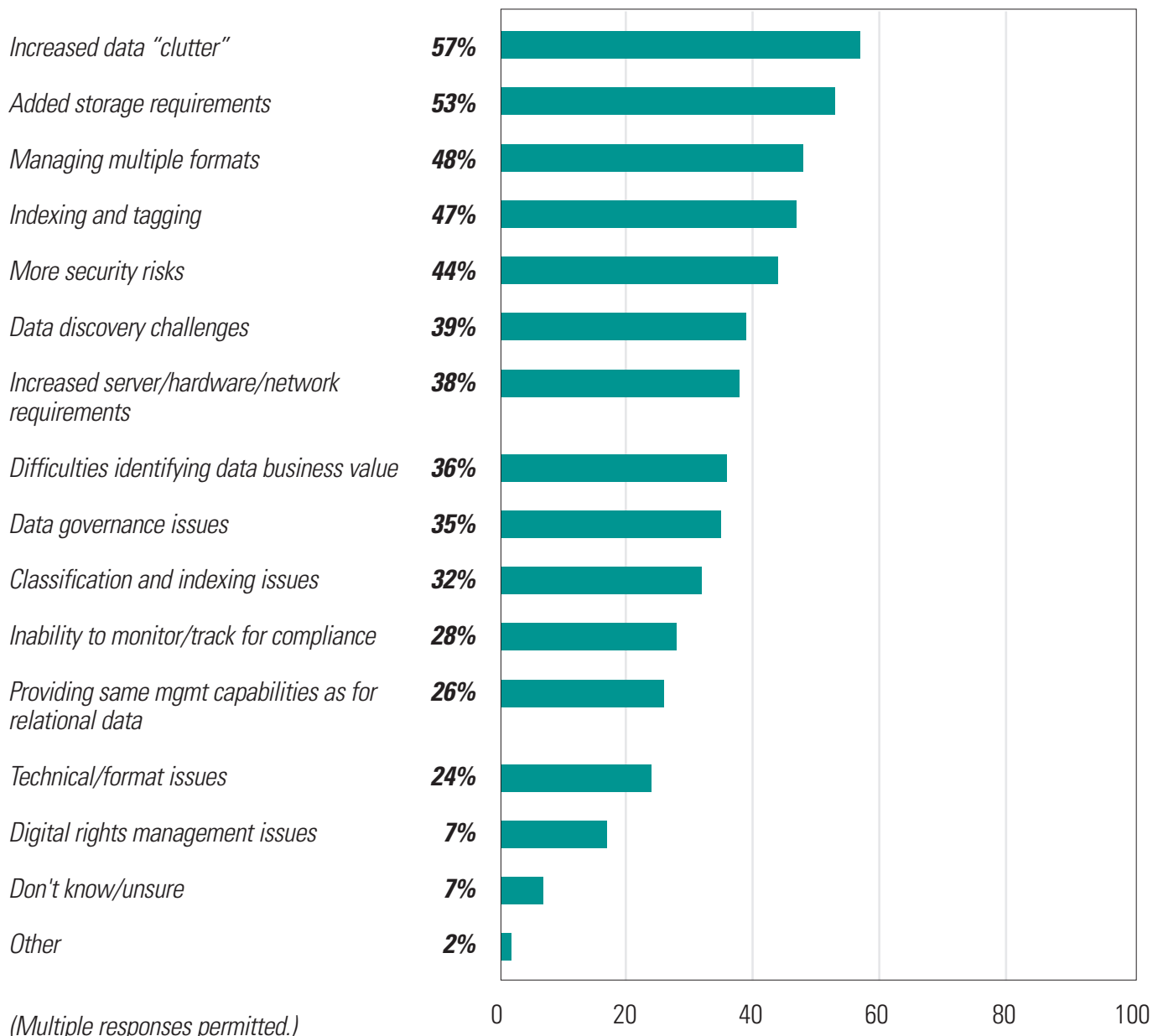
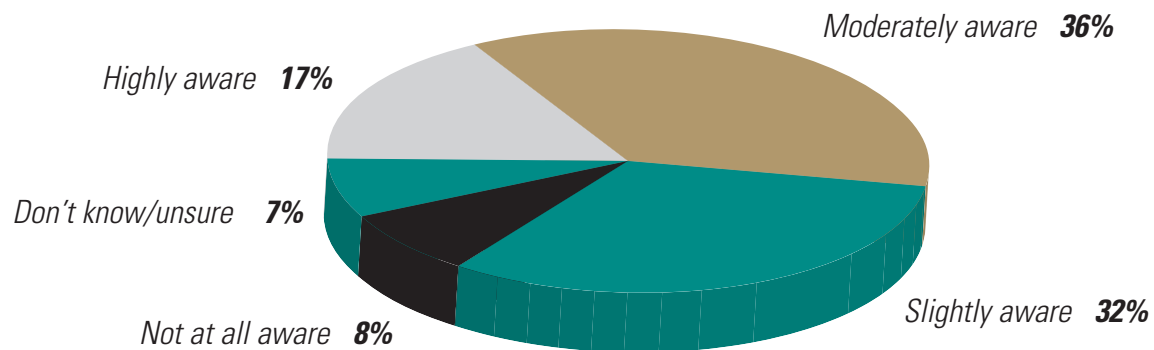


Figure 11: Top Challenges in Unstructured Data Growth—By Job Title

(Multiple responses permitted.)

	<i>DBAs</i>	<i>Devs/analysts</i>	<i>IT Mgrs</i>	<i>Business Mgrs</i>
<i>Increased data “clutter”</i>	56%	57%	62%	70%
<i>Added storage requirements</i>	57%	55%	57%	41%
<i>Managing multiple formats</i>	54%	50%	44%	59%
<i>Indexing and tagging</i>	44%	54%	58%	56%
<i>More security risks</i>	49%	43%	44%	30%
<i>Data discovery challenges</i>	35%	42%	50%	26%
<i>Increased server/hardware/ network requirements</i>	45%	35%	35%	26%

Figure 12: Corporate Awareness of Existing Unstructured Data



The Post-Relational Reality Sets In: 2011 Survey on Unstructured Data is sponsored by MarkLogic Corporation and produced by Unisphere Research, A Division of Information Today, Inc. Unisphere Research is the market research division of Unisphere Media, publishers of Database Trends and Applications magazine, www.dbta.com and the 5 Minute Briefing email newsletters. To review abstracts of past reports or to obtain full final reports prepared by Unisphere Research, contact publisher@dbta.com.

Data collection and analysis performed with SurveyMethods.

Figure 13: Corporate Awareness of Existing Unstructured Data —By Concentration of Unstructured Data

(Percent reporting “high awareness”)

Among those with high levels (>50%)
of unstructured data

28%

Among those with low levels
(0%-10% or unknown) of unstructured data

17%

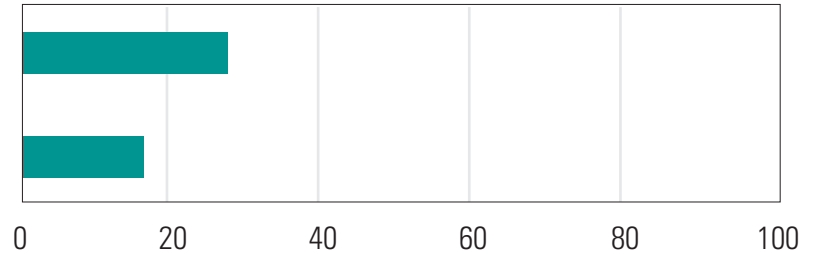


Figure 14: Corporate Commitment to Addressing Unstructured Data Challenges

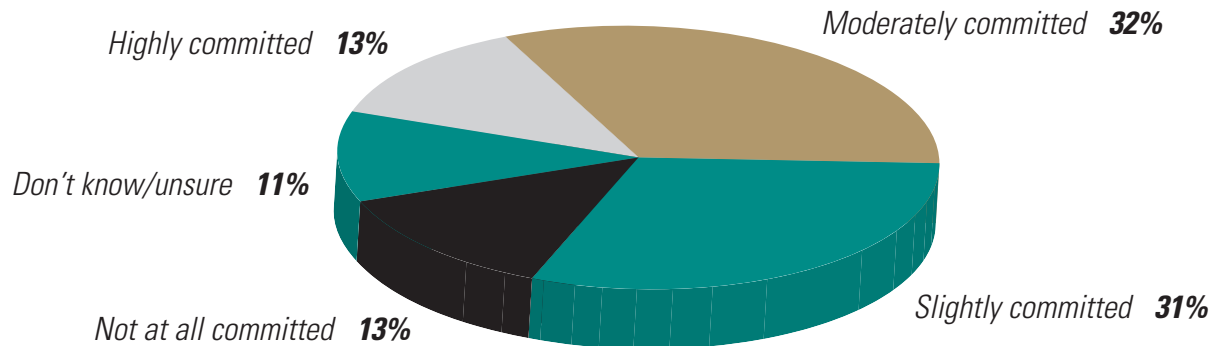


Figure 15: How Companies are Addressing Unstructured Data Challenges

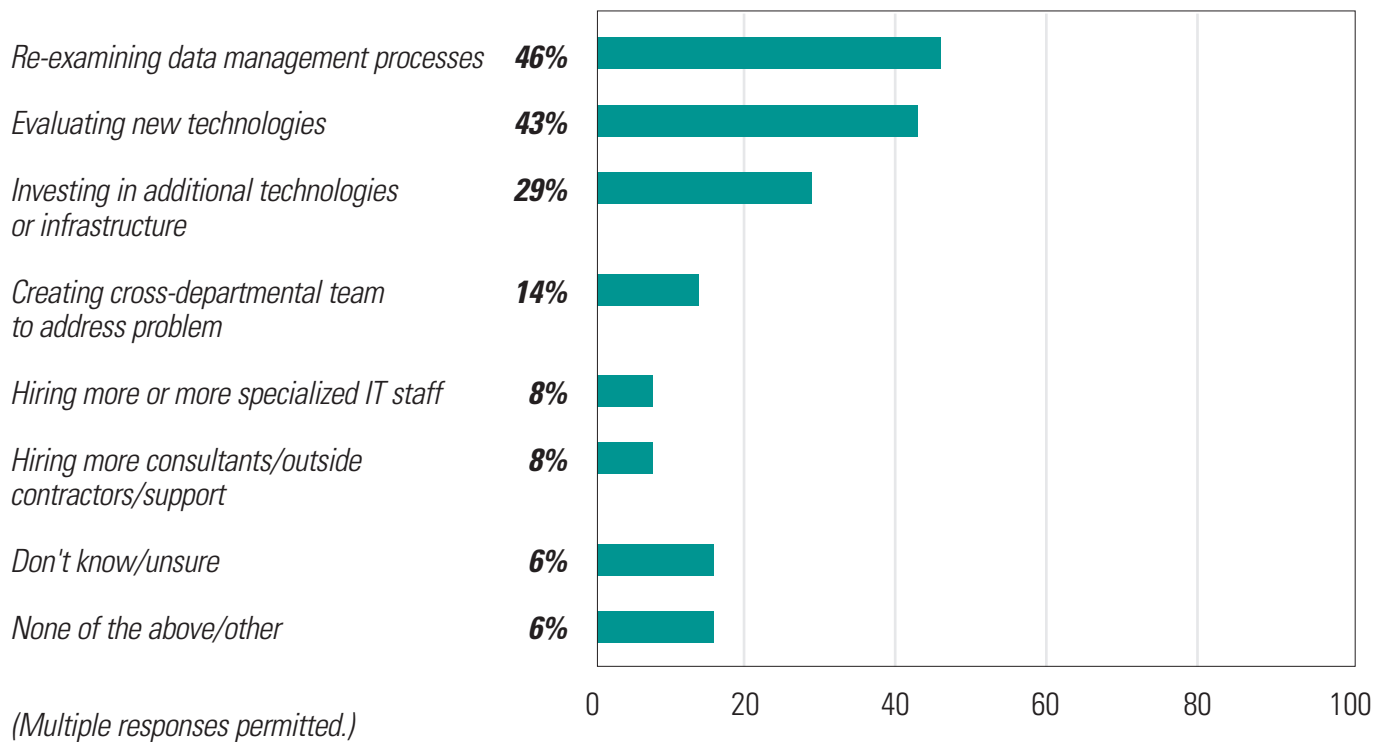


Figure 16: Importance of Unstructured Data to Respondents' Businesses

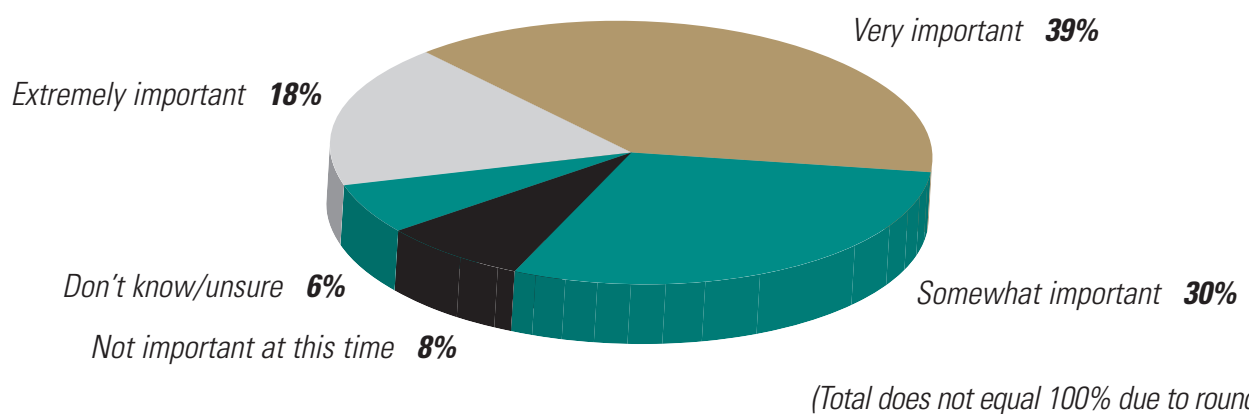
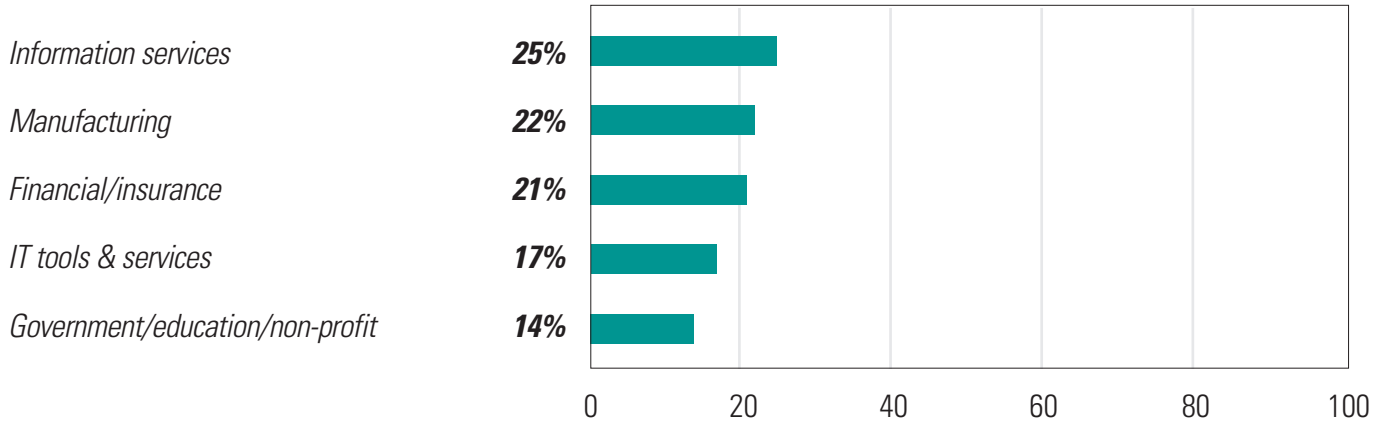


Figure 17: Where Unstructured Data is a Primary Business Driver —By Industry Group

(Percent of respondents indicating that unstructured data is “extremely important” to their business)



MANAGING UNSTRUCTURED DATA

Most respondents report they either do not have or are not aware of governance procedures in place for unstructured data, suggesting that much of this kind of data is haphazardly being distributed and managed across enterprises. In addition, a majority of companies—even those with large concentrations of unstructured data—still devote most resources to managing relational data. A large segment of respondents’ organizations are now including unstructured data under a common governance structure. Most respondents do not feel their current data infrastructures are capable of handling the growth of unstructured data. Among companies that already have high levels of unstructured data, only 18% say their current infrastructure will be sufficient.

More than half the organizations in the survey still provide more of their IT resources to structured relational data, compared to 27% either providing the same level or more to unstructured data. (See Figure 18.) This may include services and capabilities such as storage space, data security, monitoring and oversight, and dedicated administrators.

Among organizations with large concentrations of unstructured information, there is a greater tendency to devote the same or more resources to this kind of data, though this is still a minority even in this group. One-third of companies with a majority of their data now consisting of unstructured information say they devote the same or more resources to this kind of data versus that devoted to relational data. By contrast, 18% of companies with low levels of known unstructured data files will devote resources in this direction. (See Figure 19.)

Among industry groups, information services businesses show the greatest propensity to devote an equal amount or greater resources to managing their unstructured data stores, as indicated by 35%. However, it’s notable that this is still in the minority. Manufacturers follow with 31%, and IT tools and services providers at 30%. (See Figure 20.)

In most cases, unstructured data currently does not fall under the same data governance (policies and procedures) as structured or relational data, respondents report. However, this may change soon. While 27% currently have a comprehensive governance approach to both types of data, another 32% say they are still considering how to bring unstructured data under the same governance umbrella. (See Figure 21.)

Those organizations with high concentrations of unstructured data are more inclined to have addressed policies and procedures for this type of data within a common governance framework, the survey finds. Close to two-fifths report a common governance approach, versus 29% of companies with low or no levels of known unstructured data stores. (See Figure 22.)

Given, then, that most companies do not extend their current structured data governance policies, procedures or solutions to unstructured data, are there cases of "separate" governance structures for this data? The answer is not likely at this time. Most do not have separate governance procedures in place for unstructured data, suggesting that much of this kind of data is haphazardly being distributed and managed across enterprises. At this time, only 11% of respondents can confirm they have applied governance to manage the intake and leveraging of unstructured data. A majority, 58% either have no governance or don’t know what they have in place. (See Figure 23.)

At the same time, there is considerable doubt that current systems have the capability or capacity to manage all the unstructured data that is moving into enterprises. In fact, less than one-third could say with any certainty that the IT infrastructure and processes they maintain for relational data are also adequate for managing unstructured data. (See Figure 24.) In the case of companies with high concentrations of unstructured data, there is less confidence in existing systems. (See Figure 25.) Respondents are divided as to whether their current systems will even be enough in the coming years as more unstructured data flows into their organizations. (See Figure 26.) The gap grows especially wide between companies that already have high levels of unstructured data—only 18% say their current infrastructure will be sufficient, versus one-third of companies with low levels of unstructured data. (See Figure 27.)

What would be organizations’ primary motivations for improving the management of unstructured data within their companies? Topping the list is the ability to improve business processes that engage these types of data, cited by half the respondents. Close to half also point out that better managing unstructured data leads to improved decision making, as well as improved employee productivity. (See Figure 28.)

What are the main barriers to effectively improving the management of unstructured data within respondents’ companies? For a majority, cost is a major challenge (56%), followed by the need for more staff and more awareness about the amount and role unstructured data is playing within their organizations (40%). To some degree, this suggests companies continue to take a short-sighted view on unstructured data, viewing it as a “technology” problem that needs to be fixed, versus an opportunity to move the organization in a new direction. Often, organizations attempt to tackle these new challenges with older approaches. Relational database management tools are not a fit for these new environments. New technology that more effectively leverages new types of data offers orders of magnitude improvement in efficiency. (See Figure 29.)

Figure 18: Resources Committed to Unstructured Versus Structured Data

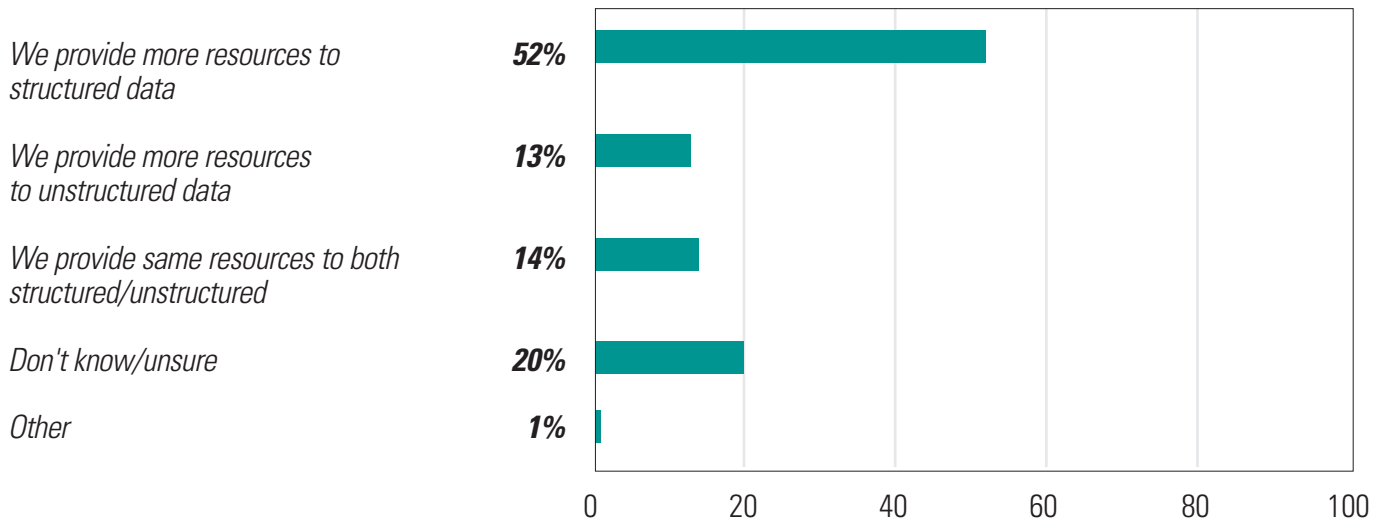


Figure 19: Resources Committed to Unstructured Versus Structured Data—By Existing Levels of Unstructured Data

(Percent reporting their companies provide same levels or more resources than to relational data)

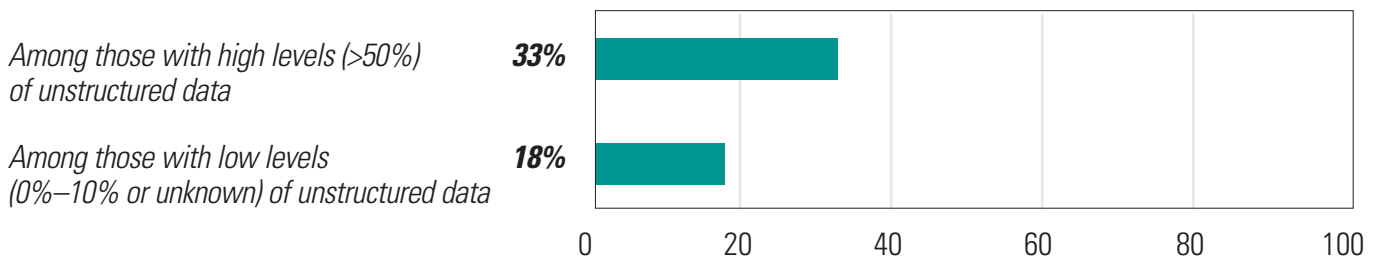


Figure 20: Resources Committed to Unstructured Versus Structured Data—By Industry Group

(Percent reporting their companies provide same levels or more resources than to relational data)

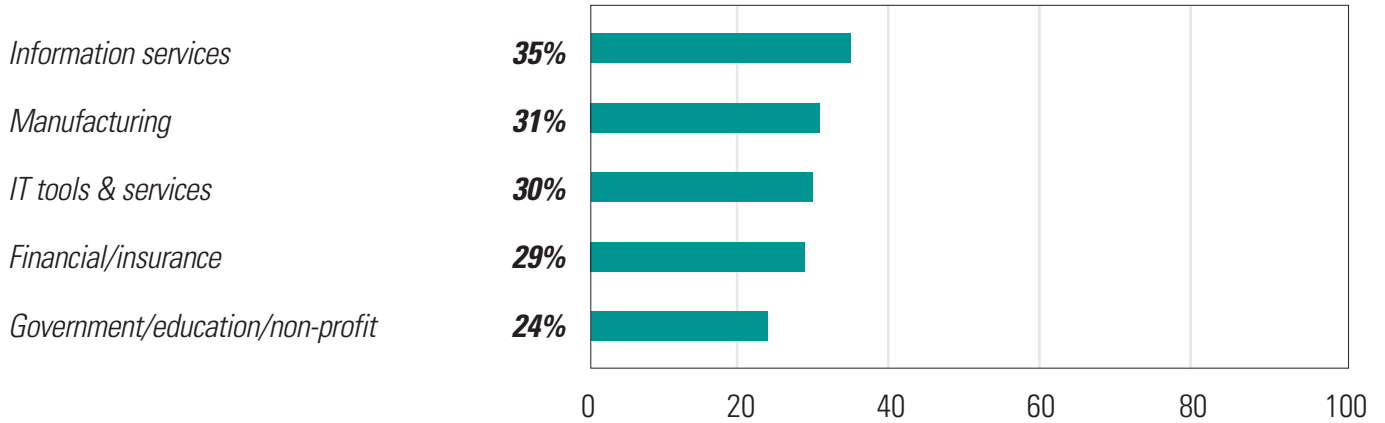
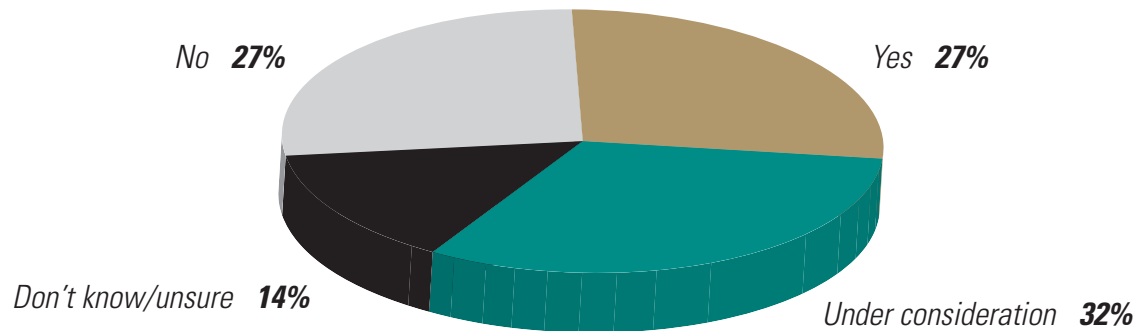


Figure 21: Data Governance Policies Extend to Unstructured Data?



The Post-Relational Reality Sets In: 2011 Survey on Unstructured Data is sponsored by MarkLogic Corporation and produced by Unisphere Research, A Division of Information Today, Inc. Unisphere Research is the market research division of Unisphere Media, publishers of Database Trends and Applications magazine, www.dbta.com and the 5 Minute Briefing email newsletters. To review abstracts of past reports or to obtain full final reports prepared by Unisphere Research, contact publisher@dbta.com.

Data collection and analysis performed with SurveyMethods.

Figure 22: Data Governance Policies Extend to Unstructured Data —By Level of Unstructured Data

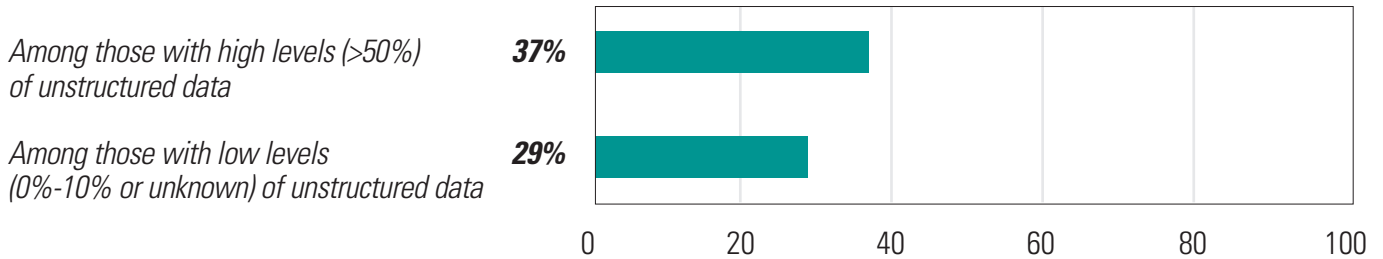


Figure 23: Governance Procedures for Unstructured Data Apart from Structured Data?

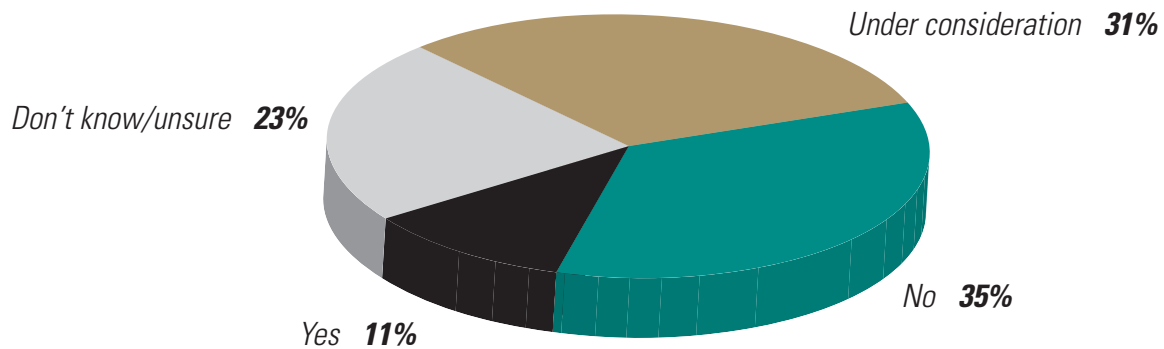


Figure 24: IT Infrastructure and Processes Adequate for Managing Unstructured Data?

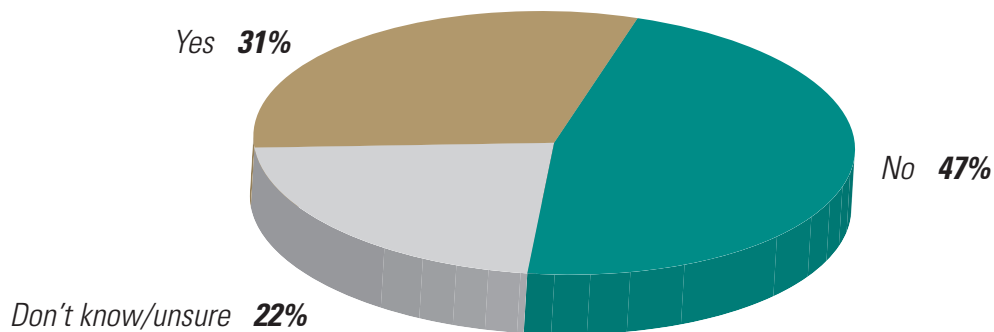


Figure 25: IT Infrastructure and Processes Adequate for Managing Unstructured Data—By Level of Unstructured Data

(Percent reporting adequate levels)

Among those with high levels (>50%) of unstructured data

28%

Among those with low levels (0%–10% or unknown) of unstructured data

36%

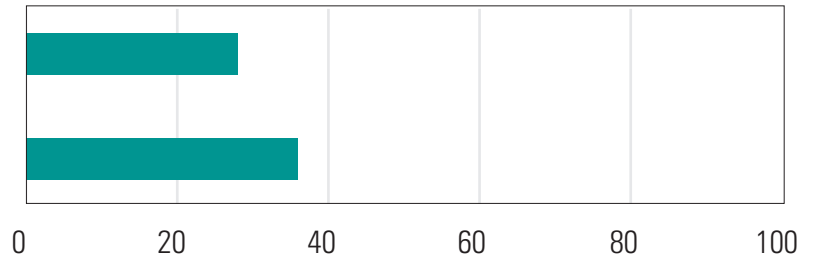
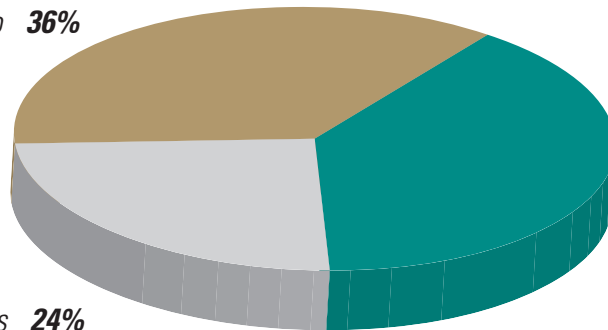


Figure 26: IT Infrastructure and Processes Adequate for Managing Unstructured Data 3 Years From Now?

No **36%**

Yes **24%**

Don't know/unsure **39%**



(Total does not equal 100% due to rounding.)

Figure 27: IT Infrastructure and Processes Adequate for Managing Unstructured Data 3 Years From Now—By Level of Unstructured Data

(Percent reporting levels will be adequate)

Among those with high levels (>50%) of unstructured data

18%

Among those with low levels (0%–10% or unknown) of unstructured data

34%

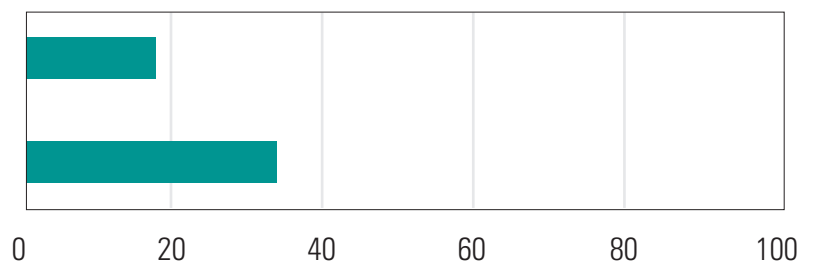


Figure 28: Primary Motivations for Improving Management of Unstructured Data

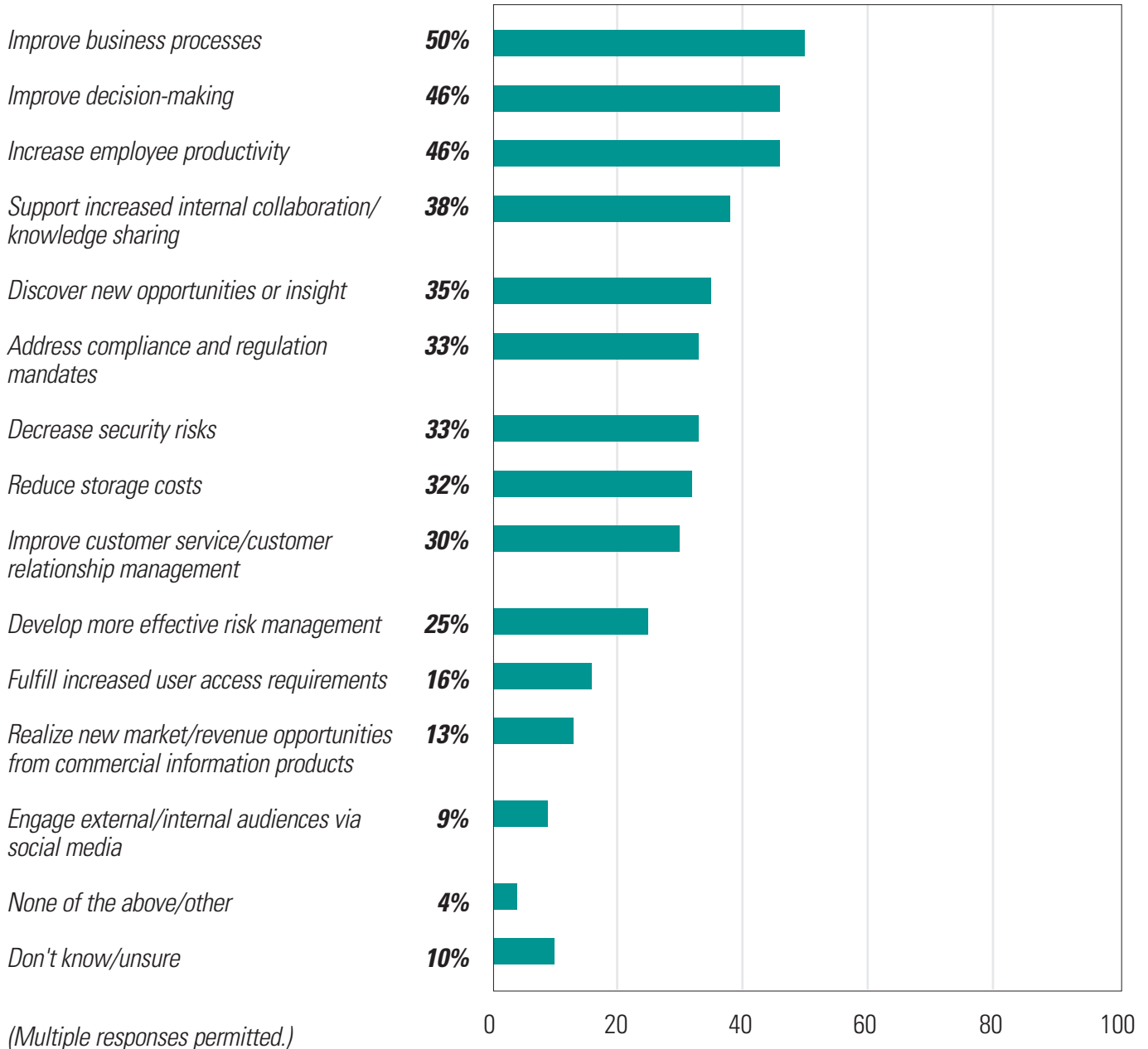
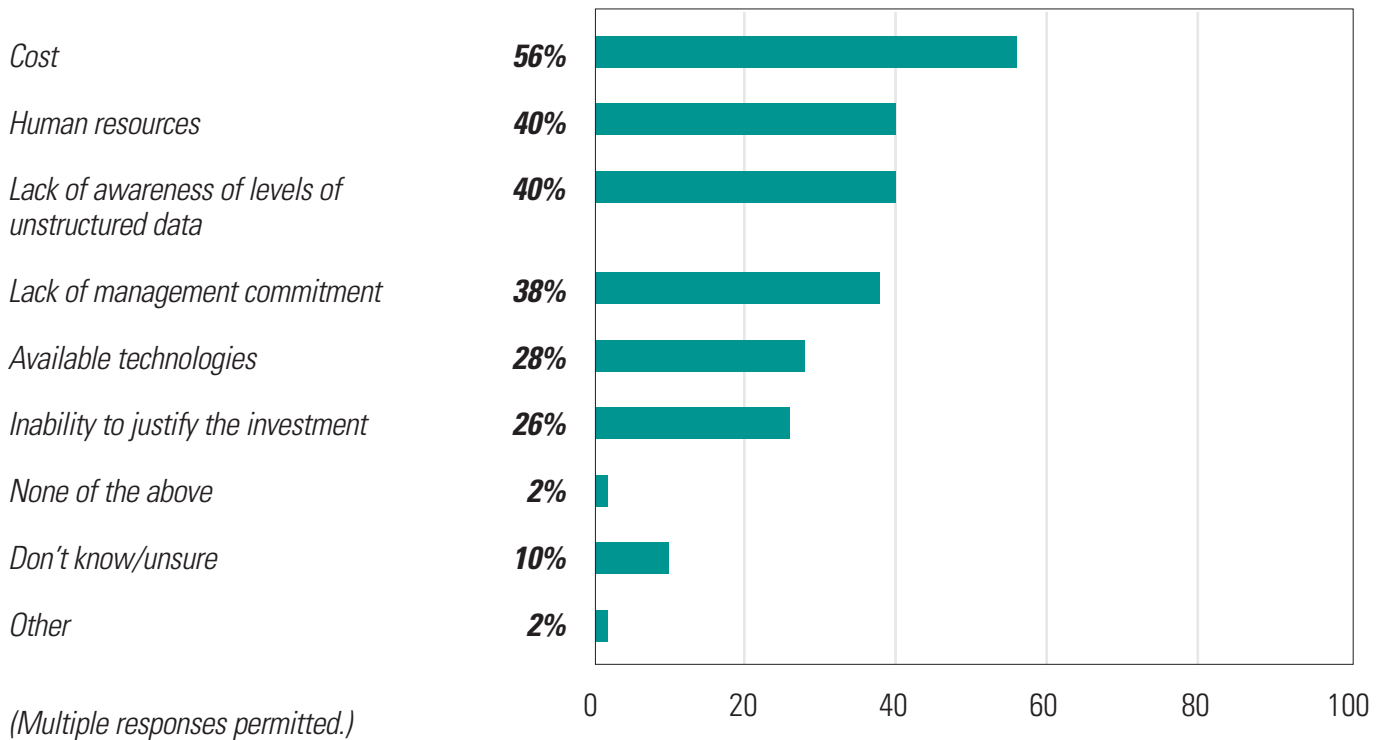


Figure 29: Main Barriers to Improving Unstructured Data Management



TECHNOLOGY FOR UNSTRUCTURED ENVIRONMENTS

Email and messaging systems are the primary vehicles for creating unstructured data, but much of it ends up in content management systems. Most companies with large concentrations of unstructured data are moving such information into non-relational databases.

What are the primary systems in respondents' organizations that create unstructured data? Three-fifths report that much of this data is coming out of email or instant messaging systems. This suggests that at this time, the vast majority of business information is being created and exchanged via email, versus business documents such as spreadsheets and word processing documents. Finding a way to effectively leverage email repositories and the information within them would provide vast productivity gains and business improvements. (See Figure 30.)

What are the primary systems in respondents' organizations that consume unstructured data? The leading types of systems where these documents end up are file systems or document repositories, cited by more than four-tenths. A similar number say their unstructured data is contained within email or instant messaging systems, while more than one-third say it ends up within content management systems. Both email and content management systems are notoriously difficult to search or leverage for information discovery. (See Figure 31.)

One of the challenges of unstructured data is indexing it and making it easily searchable by end users. A majority of respondents (77%) feel that it is either "very" or "somewhat" important to automate processes such as tagging and creating metadata for unstructured data at this time. More than one-third place a high priority on these types of activities. As discussed in the previous section, half of the respondents—and an even larger ratio of business users—are concerned about the ability to adequately capture unstructured data files and render them searchable to end users. An organization with a large trove of video files, for example, may end up assigning employees to manually review each and every file to properly index and tag the files—a costly and time-consuming process. (See Figure 32.)

In most cases, unstructured data is stored in non-relational databases (e.g., file systems, content management systems, special-purpose databases)—as cited by two-thirds of respondents. Another 42% say they deposit such data into

relational databases, in many cases right alongside relational data. This indicates the dysfunction and dominance of the status quo, as people will default to what they know—in this case, the traditional relational database management systems that are already onsite. Unstructured data does not belong in a relational structure, and attempting to fit this round peg into a square hole will ultimately cost the organization in the long run, in misappropriated and difficult-to-find files. (See Figure 33.)

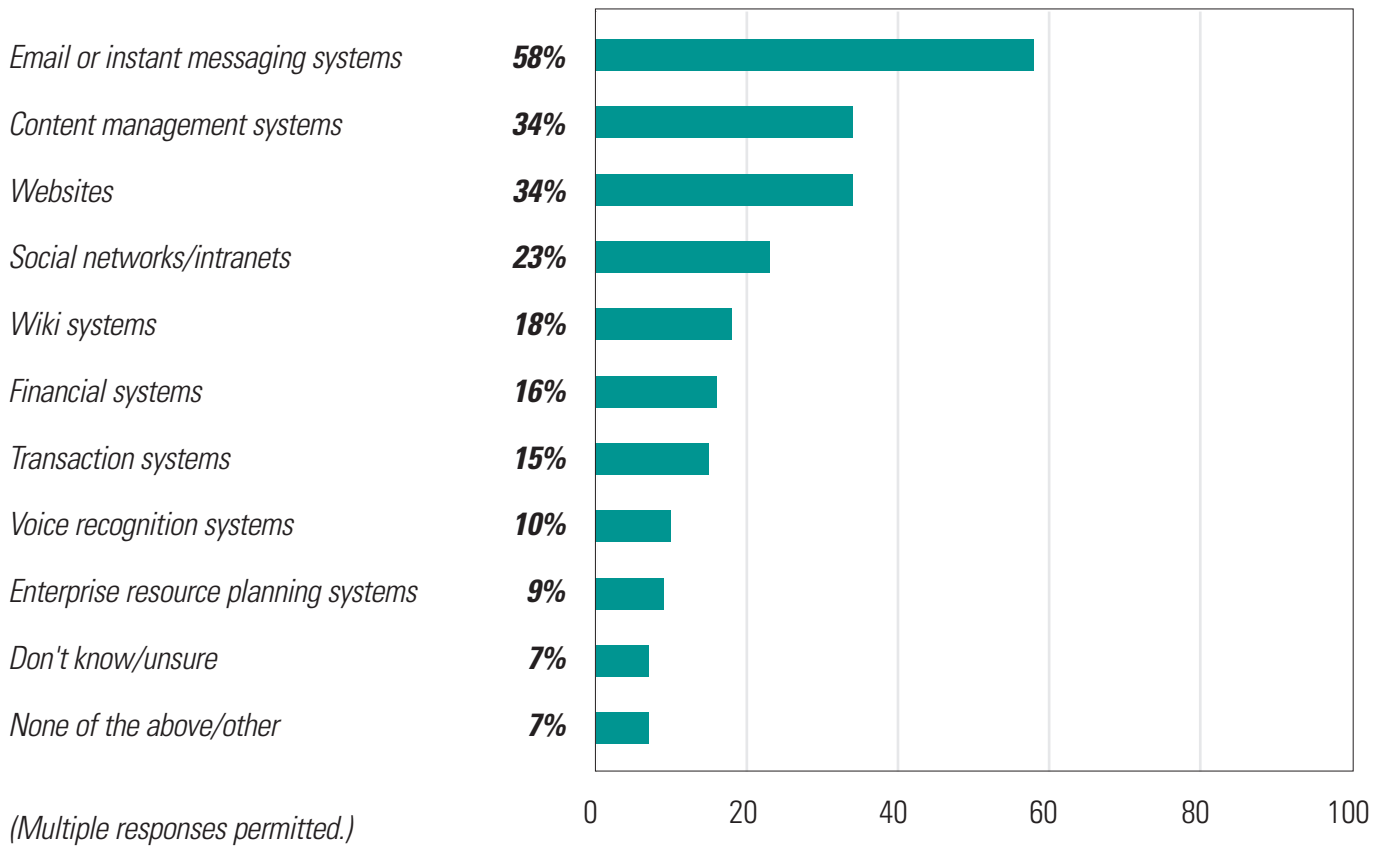
Among companies with high concentrations of unstructured data, however, a large majority have moved to non-relational databases. About 82% of the unstructured data-intensive enterprises report they store such data in non-relational databases. Companies with low levels of known unstructured data are just as likely to put whatever unstructured data they have into a relational database as they are into a non-relational database. (See Figure 34.)

In most cases, respondents report they employ file systems to manage, integrate and/or store unstructured data. This is a defective, sub-optimal process, and may point to a perception on the part of some managers that storage and technology is the main challenge posed by unstructured data—versus viewing this data as presenting an opportunity to provide greater insights and resources to decision-makers, or as a new business opportunity. (See Figure 35.)

What technologies do respondents employ to search, analyze and/or deliver unstructured data? The largest percentage of respondents, 37%, employ search and access solutions against their content management systems for these tasks, while just under one-third do so against relational databases. Data mining and text analytics are employed at one-fourth of the respondents' companies in the survey. (See Figure 36.)

There are also a number of emerging solutions that respondents are exploring to better manage the data explosion, including log monitoring and reporting tools (being considered by 19%, in-memory and NoSQL databases (18% and 17%, respectively), and Apache Hadoop, a software framework that supports data-intensive distributed applications, enabling applications to work with thousands of nodes and petabytes of data. (11%). (See Figure 37.)

Figure 30: Primary Systems That Create Unstructured Data



The Post-Relational Reality Sets In: 2011 Survey on Unstructured Data is sponsored by MarkLogic Corporation and produced by Unisphere Research, A Division of Information Today, Inc. Unisphere Research is the market research division of Unisphere Media, publishers of Database Trends and Applications magazine, www.dbta.com and the 5 Minute Briefing email newsletters. To review abstracts of past reports or to obtain full final reports prepared by Unisphere Research, contact publisher@dbta.com.

Data collection and analysis performed with SurveyMethods.

Figure 31: Primary Systems That Consume Unstructured Data

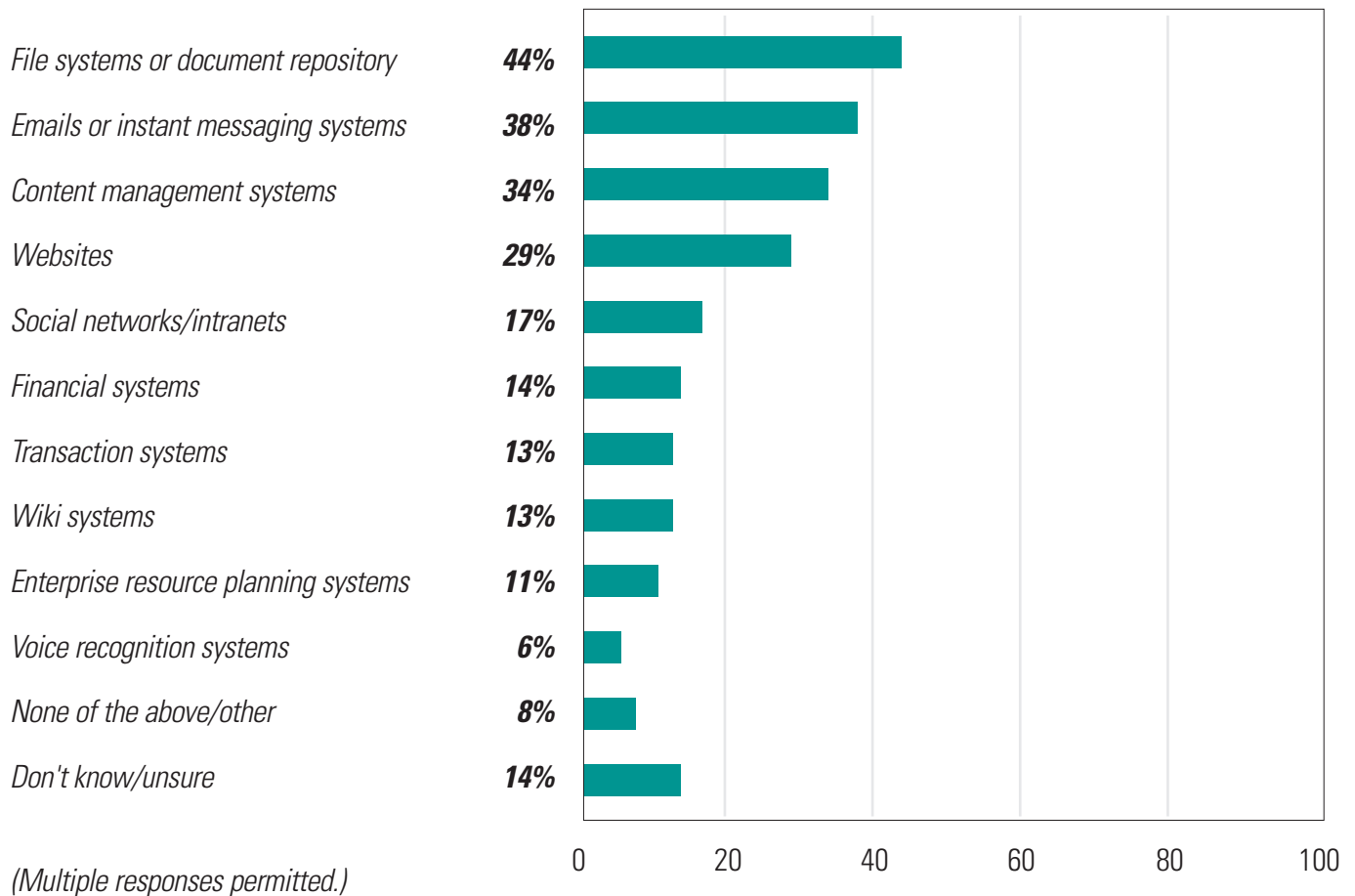
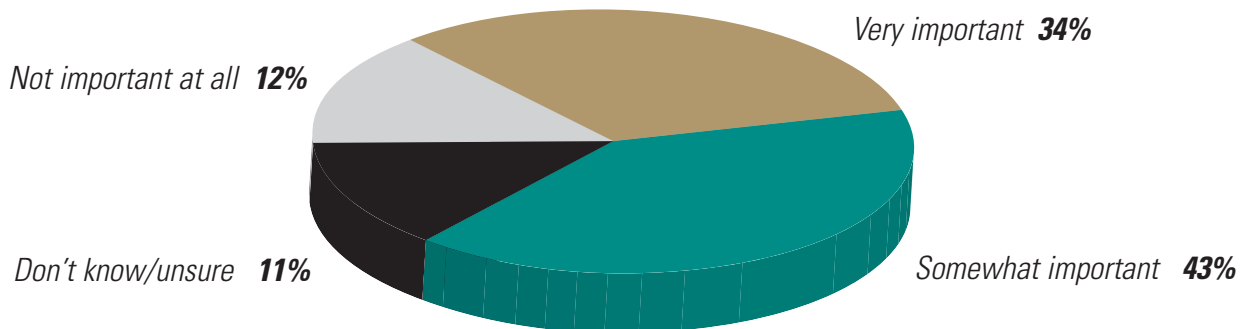


Figure 32: Importance of Automating Processes (Such as Tagging and Creating Metadata) for Unstructured Data



The Post-Relational Reality Sets In: 2011 Survey on Unstructured Data is sponsored by MarkLogic Corporation and produced by Unisphere Research, A Division of Information Today, Inc. Unisphere Research is the market research division of Unisphere Media, publishers of Database Trends and Applications magazine, www.dbta.com and the 5 Minute Briefing email newsletters. To review abstracts of past reports or to obtain full final reports prepared by Unisphere Research, contact publisher@dbta.com.

Data collection and analysis performed with SurveyMethods.

Figure 33: Where Majority of Unstructured Data is Stored

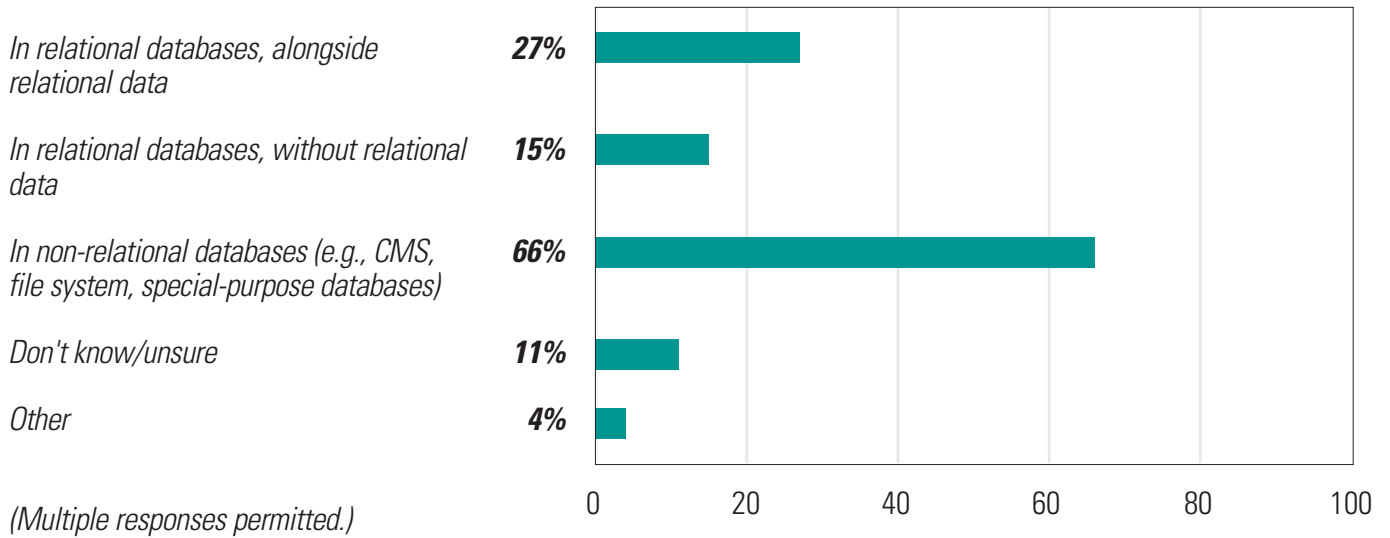


Figure 34: Where Majority of Unstructured Data is Stored —By Existing Levels of Unstructured Data

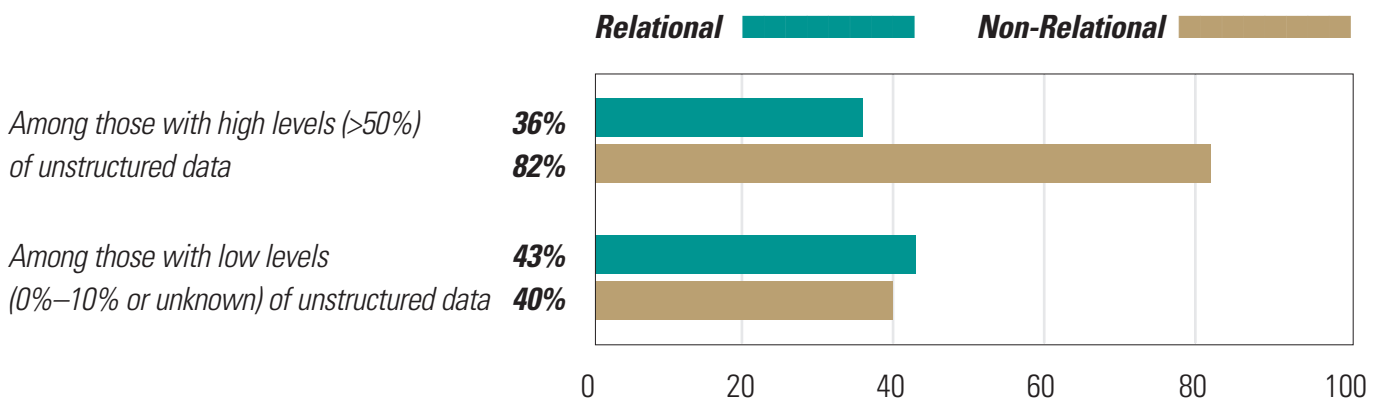


Figure 35: Technologies Employed to Manage, Integrate and/or Store Unstructured Data

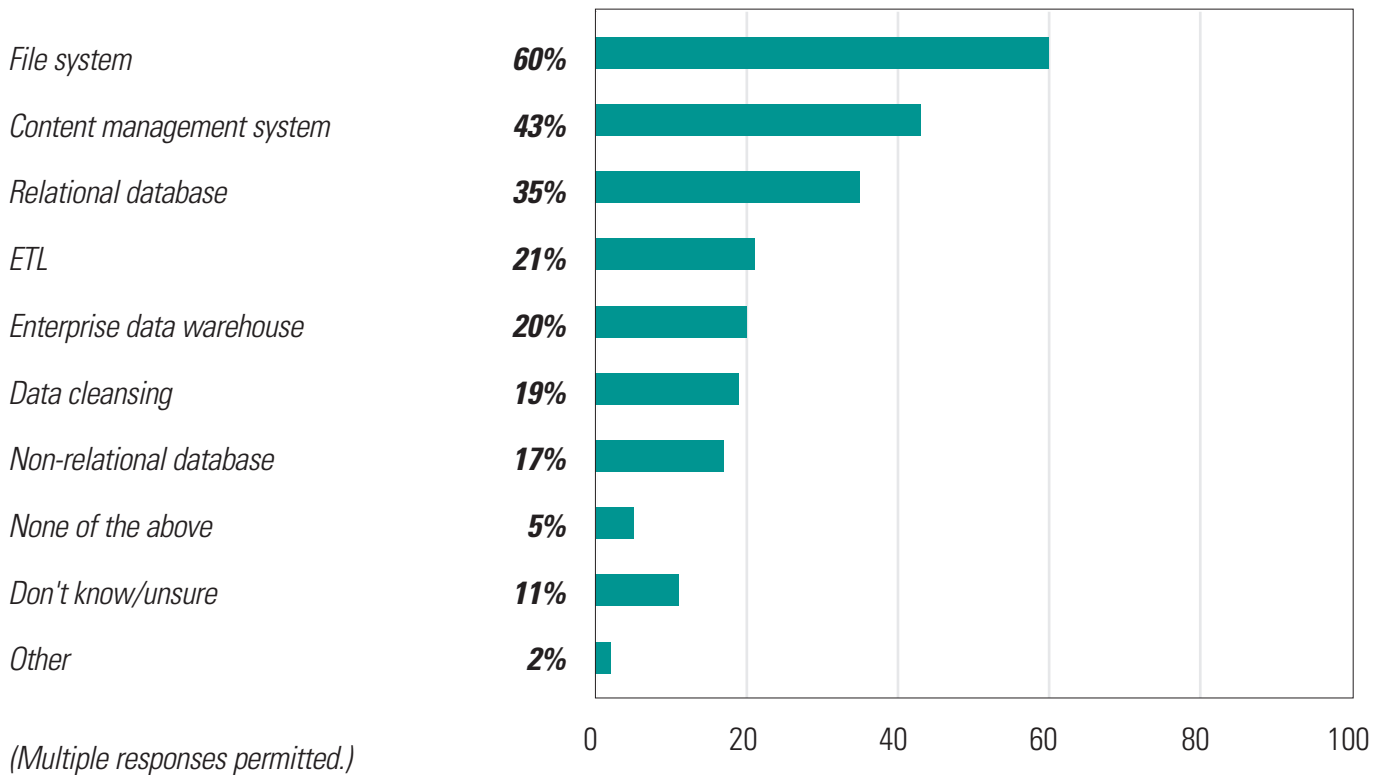


Figure 36: Technologies Employed to Search, Analyze and/or Deliver Unstructured Data

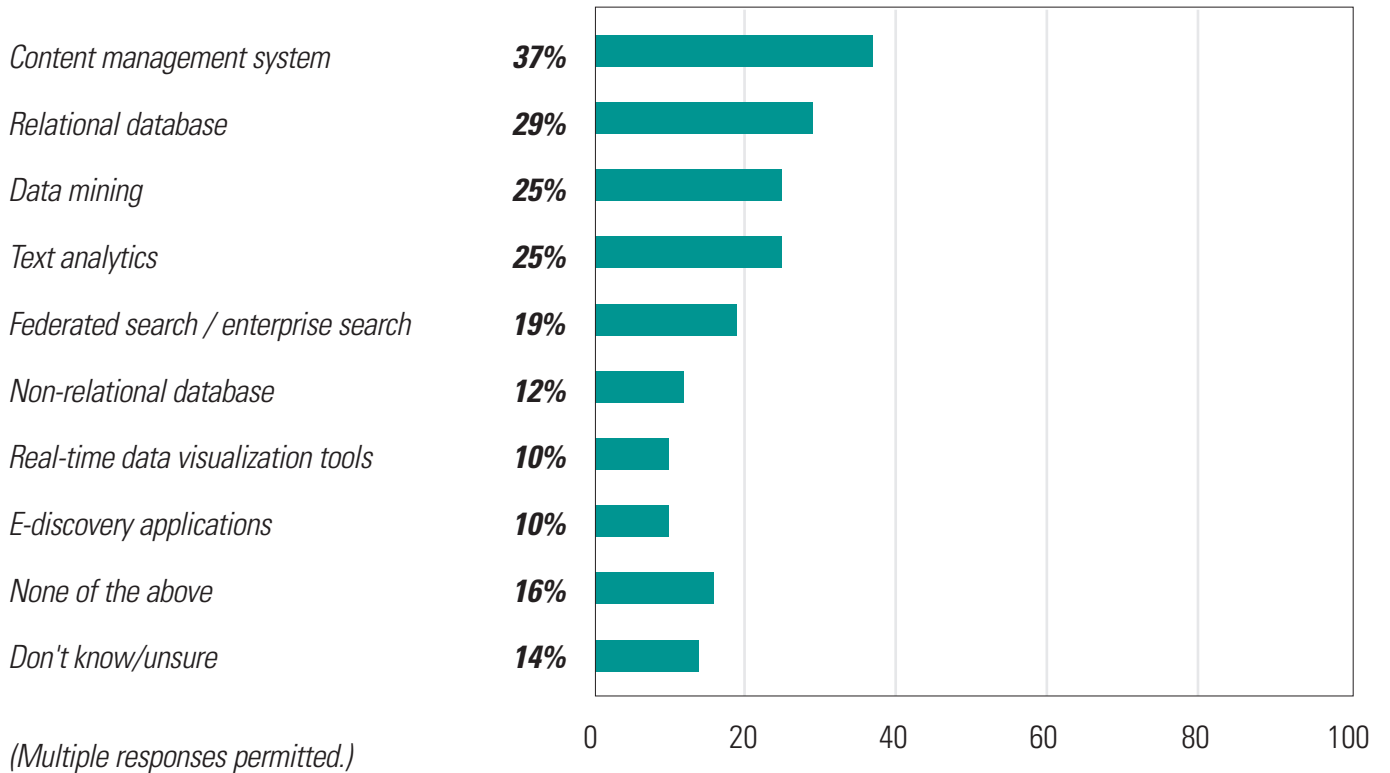
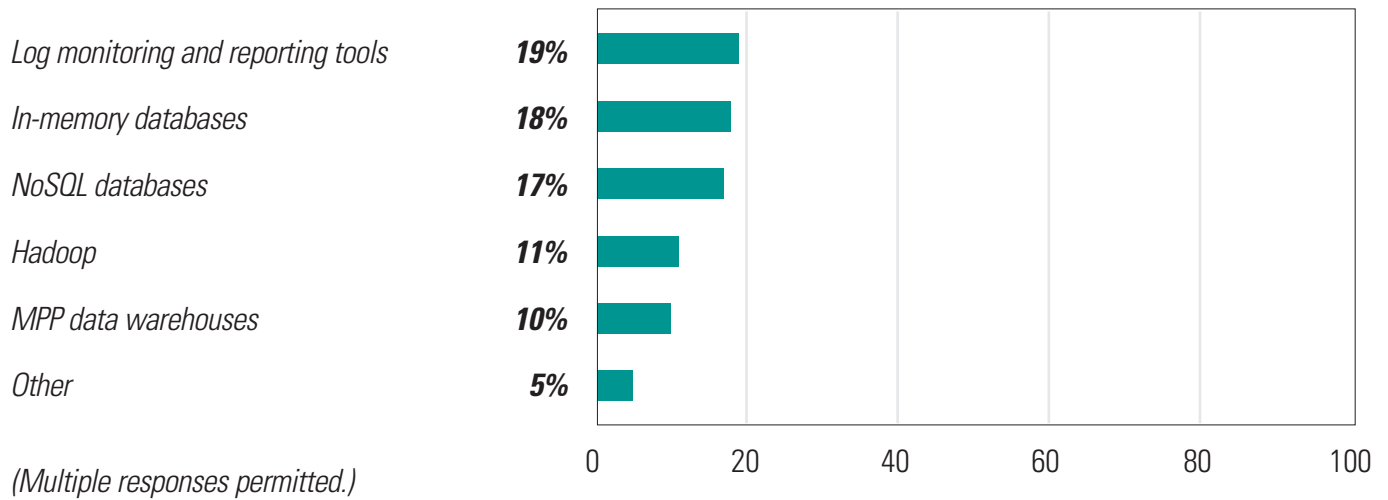


Figure 37: New Technologies for the Unstructured Data Era



CONCLUSIONS AND RECOMMENDATIONS

As far as most of the managers and professionals in this survey are concerned, there's no question that the era of Big Data is here in full force, and is led by a rising tide of unstructured data—which includes non-relational files such as word processing documents, spreadsheets, PDFs, multimedia, social media, messaging, and graphics. Almost half of the respondents to this survey agree that they will soon be managing more unstructured data than relational file structures within their enterprises.

Organizations are only beginning to take their first steps toward grasping the complexities and opportunities created by unstructured data, however. In fact, at this point, most companies are likely missing out on the full opportunities they have to parlay this data into new insights and new business.

The following are key recommendations for tackling the challenges of unstructured data and turning this information into real business value:

Think of unstructured data as a business asset, not a liability: As this survey finds, unstructured data often falls under the domain of database administrators, whose job it is to view such data as a problem that needs to be addressed with the right hardware and systems. The prevailing viewpoint is that unstructured data is little more than an expensive nuisance, when it, in fact, is probably one of the greatest business opportunities to arise in decades. Unstructured data can be the cornerstone of knowledge management, new products, and enhanced decision-making—if recognized as such by business leaders.

Conduct an assessment: Close to half of the survey respondents say there is very little understanding in their organizations about the extent or types of unstructured data now used within their enterprises. It's important to understand how much of this data may exist across the organization, and what role it plays in both formal and informal decisions. There may be a wealth of social networking taking place with customer groups, for example, outside of formal sales and marketing channels. There may be new sales taking place due to informal channels. This kind of information needs to be understood.

Build awareness, and educate: Even among organizations where it is known that high levels of unstructured data exist, only about one in four could say that there was high degree of corporate awareness. In order to effectively manage and leverage unstructured data, there needs to be greater awareness, and perhaps even levels of training for employees and managers working with the data. Increasing management awareness will help data managers secure the funding and resources needed to properly secure, store, and fully leverage the large volumes of unstructured data coming into and being created by their organizations.

Establish good governance practices: As the survey reveals, most unstructured data moves through organizations informally, with little oversight. In some cases, organizations have applied the same governance principles as is applied to structured relational data. To effectively leverage the business opportunities unstructured data provides, as well as develop a lifecycle management approach to storing, archiving and retiring the data, organizations need an approach to oversee the way the data is managed. The business can only benefit from unstructured data if they have a say in how it is managed.

Consider alternative technologies: Current relational database management systems are not up to the task of managing the terabytes and, soon, petabytes worth of digital information that will sweep through enterprises. As the survey shows, there is considerable doubt that current systems have the capability or capacity to manage all the unstructured data that is moving into enterprises—less than one-third of managers and professionals could say with any certainty that the IT infrastructure and processes they maintain for relational data are also adequate for managing unstructured data. The good news is that there is a new generation of technologies available that can address the requirements of unstructured data, from NoSQL and in-memory databases to large data management tools such as Hadoop and MPP data warehouses.

DEMOGRAPHICS

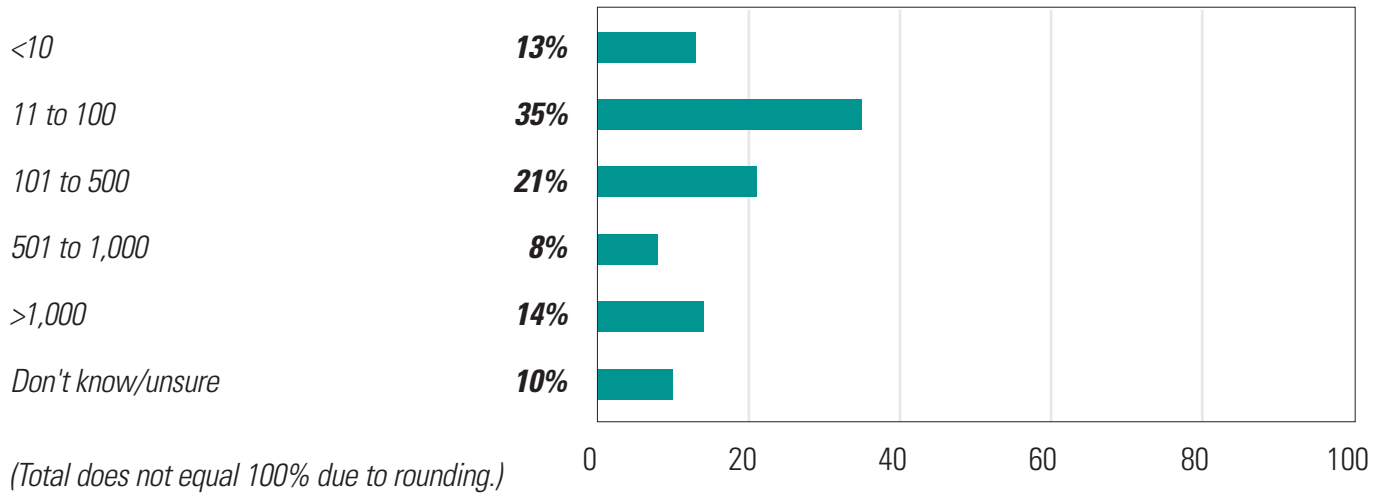
Figure 38: Number of Distinct Database Instances Within Respondents' Organizations

Figure 39: Respondents' Primary Job Titles

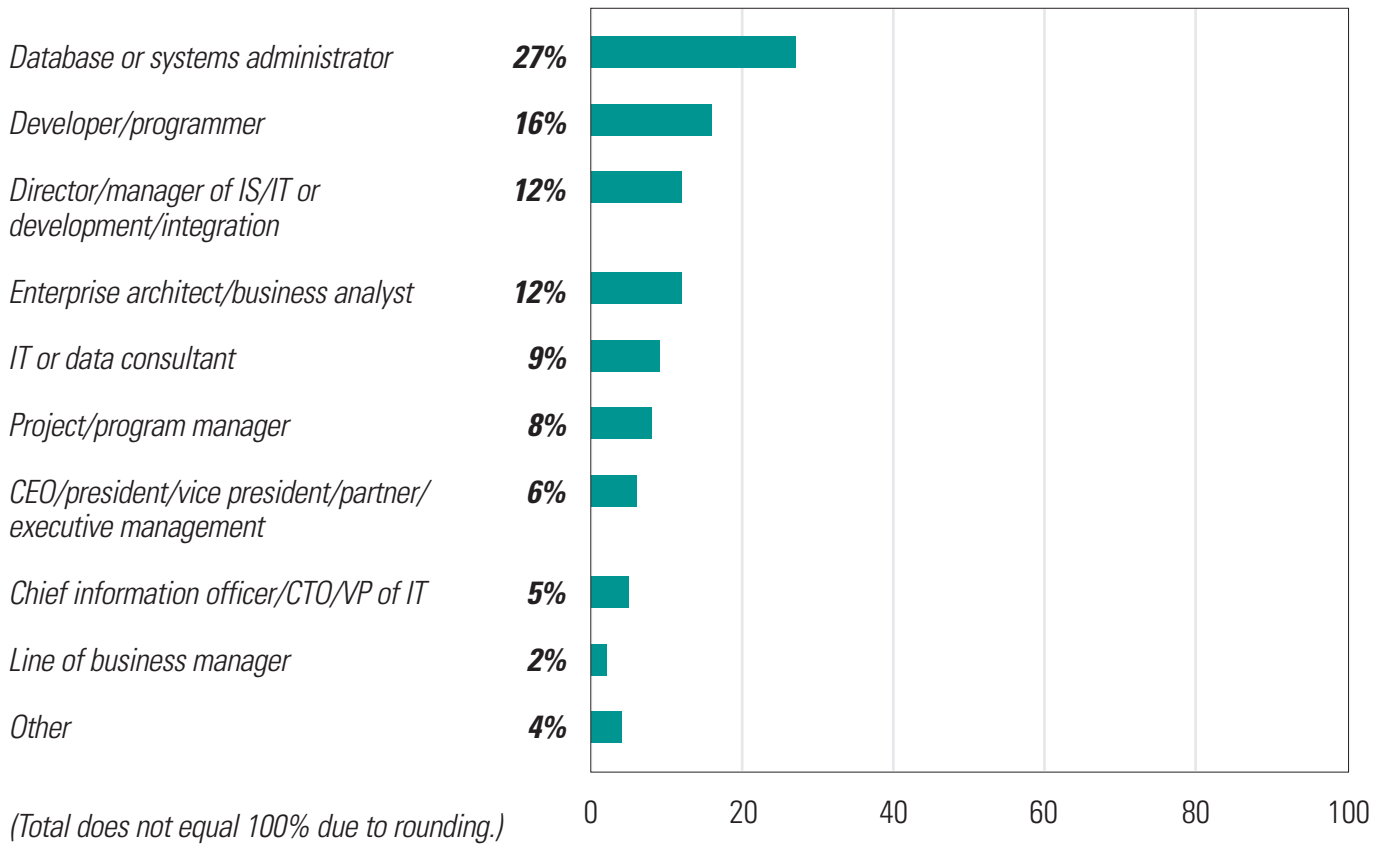


Figure 40: Organization Size—By Number of Employees

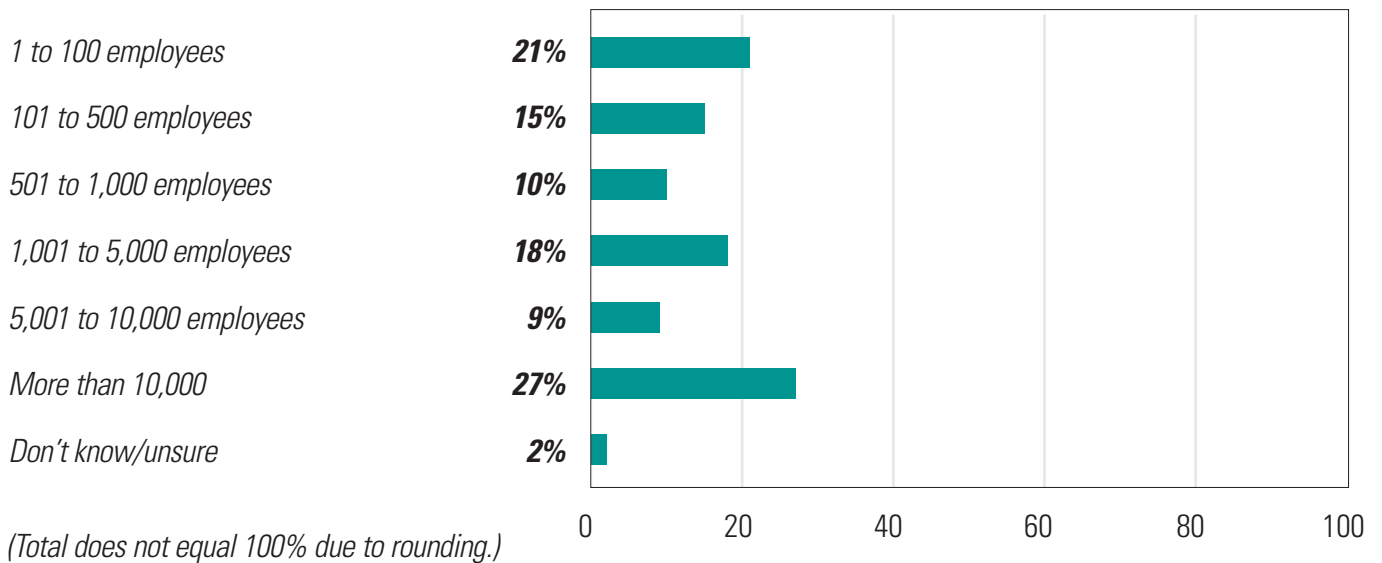


Figure 41: Respondents' Primary Industries

