by Tadhg Nagle and Dave Sammon, Co-Directors, Master's of Science in Data Business, Irish Management Institute

Fast and Flexible: Exploring Agile Data Analytics

For the purpose of Cutter Consortium's latest *CBR* survey, we define Agile data analytics as "the application of Agile methods to data analytics initiatives in order to increase flexibility, provide faster 'time to value,' and support collaborative relationships between business users and IT developers." Further still, as described by Agile data analytics expert Ken Collier, Agile methods are built on a "simple set of sensible values and principles but require a high degree of discipline and rigor to properly execute" and as a process Agile falls somewhere in the "middle between just enough structure and just enough flexibility."[1] For firms embracing Agile values and principles, which promote early and continuous delivery of business value throughout the development lifecycle, the primary objective of Agile data analytics is to deliver high-quality, high-value, working business intelligence (BI) solutions.

When asking survey respondents to describe their understanding of Agile data analytics, we see that the majority (nearly 70%) also view the key component of Agile data analytics as the use of Agile methodologies in the development of BI/analytics solutions (see Graph 1 in the Survey Data section beginning on page 17). This reinforces Collier's notion of Agile data analytics as being a development "style" rather than a methodology or framework. Consequently, organizations need certain capabilities and behaviors as well as a creative thinking mind-set to successfully grow Agile data analytics within their business. As Cutter Senior Consultant Sebastian Hassinger comments in the companion article to this issue of *CBR*, Agile is "a mindset unencumbered with previous assumptions."

Agile methodologies have been successfully adapted within the software development domain, with customer involvement, transparency, and speed of output being major factors of its success. The fact that Agile methodologies are making an impact in the domain of data analytics suggests that as the area grows and the frequency of data projects increase, there is a higher necessity for business stakeholders to take a more active role in these projects. As Graph 1 shows, the second most popular defining feature of the term "Agile data analytics" is the ability to draw on a wide variety of mechanisms to gain insight in a short time frame. Again, this notes recognition of the wide technological/analytical choice available to practitioners but highlights a need to stay focused on gaining insightful outputs quickly in order to take advantage of emerging opportunities (the third most popular description in the survey). All this indicates that while the area of data analytics is going through a major technological overhaul, organizations are cognizant of the practicality of utilizing Agile methodologies in the deployment of these new technologies since they provide a foundation of risk reduction and increased transparency — key requirements when deploying unstructured projects and new technologies.

GOING AGILE

Very few organizations in today's economic climate can afford to lose customers, mismanage resources, or, put more simply, operate at a loss through making what seem to be straightforward business decisions. However, faster decision cycles are a requirement for achieving business agility and while organizations can deploy enabling technologies to increase "speed to insight" and "time to value," having high-quality data in front of decision makers in a timely fashion and in a usable format is of the utmost importance. To complicate matters, if not appropriately managed, data/information quality deteriorates over time. Unfortunately, this observation is backed up by the survey results, with the bulk of respondents indicating they have either no capability (scoring a 1 on a 1-5 scale) or a very low maturity (scoring a 2) across the capabilities we offered in the survey (see Graph 2). Indeed, as suggested by Richard Wang et al, it may be the case that organizations still struggle to separate data/information (content) and its management from the IT artifacts themselves, where "organizations often focus inappropriately on managing the lifecycle of the hardware and software systems that produce the information, instead of on the information itself."[2] However,

according to Wang et al, most organizations often inappropriately treat information as a "by-product" as opposed to a "product" and do not effectively manage data as a "business resource"[3] or "corporate asset."[4]

The need to up-skill executives in quantitative and analytical techniques has become quite evident in the volume of management/ quantitative-type books released each week.

Benefits

In Graph 3, we report the benefits that organizations expect to come from Agile data analytics. The top benefit — better planning and forecasting selected by 47% of respondents — highlights the forward-looking perspective that organizations have on initiatives. Organizations are beginning to realize that data analytics plays an important role in the decisions they make about their future. The second-highest expected benefit (numerous and accurate business insights — 39%) could indeed support this role. This finding suggests that analytics will indeed provide more options, or at least different perspectives, on future plans of action. At the other end of the scale, it is interesting to see social marketing and Internet behavior (i.e., clickstream data) ranking very poorly in the survey, with 10% and 4% of respondents identifying them as benefits respectively. With experts touting social and unstructured data as a major source of value, the lack of perceived benefits by respondents may suggest a lack of maturity in the area and a resulting lack of understanding on how organizations can realize the value. Furthermore, the lack of perceived benefits from fraud detection (6%) is also surprising. This may suggest that organizations have positioned Agile data analytics as a revenue generator rather than a cost saver. If indeed this is true, it is a very positive indication since recent research has shown that technology projects focused on revenue generation are more successful that those that focus on cost efficiency.[5]

Barriers

When it comes to barriers for Agile data analytics, staffing/skills is top of the list, with 61% of respondents identifying it as such (see Graph 4). As suggested by Sebastian in the next article of this issue, investing in skills and technologies is an important consideration for all organizations wanting to embrace Agile data analytics successfully. In the last number of years, demand for data-savvy hires has increased dramatically, with the role of data scientist being noted as the "sexiest job in the 21st century,"[6] a title past bestowed upon software developers. In fact, a report produced by e-skills UK on behalf of SAS UK revealed that demand for data scientists had grown by 350% during the period Q3 2011 to Q3 2012.[7] This is impressive growth seeing that the role was, according to the report, "near non-existent prior to 2011."

In addition, the need to up-skill executives in quantitative and analytical techniques has become quite evident in the volume of management/quantitative-type books released each week. This need to educate business leaders about data analytics is also reflected in the survey results, which highlights the lack of business sponsorship as the second-biggest barrier, tied with difficulty in architecting a solution. As a comparison, almost twice as many survey respondents view the lack of business sponsorship as a barrier to Agile data analytics (35%) against those that see building business cases as a barrier (18%). This suggests that while the justification of projects is a significant barrier, it is much more straightforward than the actual implementation of a project.

Another key insight from the data shows that technology itself is the lowest barrier. More specifically, respondents don't view already implemented data technologies as major limiting factors to Agile data analytics, with current database software flexibility and speed being highlighted as barriers by relatively low numbers of respondents (26% and 10%, respectively) and current data warehouse rigidity to reports and OLAP identified by 12% as a barrier.

Technologies

Graphs 5a and 5b report survey findings on deployed technologies in data analysis; respondents were asked to rate deployment levels on a 1-5 scale, with 5 being full/widespread deployment. One of the most striking results of the survey highlights the ever-present dependence of organizations on spreadsheets as a data analytics tool, with just under 95% having varying degrees of spreadsheet deployment and 36% having near-to-full/widespread implementation (see Graph 5a). While spreadsheets do play a significant role in an organization's information supply chain, the survey indicates that spreadsheets are significantly overutilized in comparison to newer and more error-resilient technologies, even though spreadsheets have been reported to be flooded with errors a whopping 88% of the time.[8] This won't come as much of a surprise to most, and while it highlights a definite need for enterprise data solutions, it also indicates a significant gap between the data

 ©2013 Cutter Information LLC

products trending in industry discourse and the reality of how comfortable organizations are with these products. This is an interesting challenge, and although organizations must carry the bulk of the responsibility of overcoming it, vendors need to play their part as well. For the most part, vendors such as EMC and SAS have identified the importance of educating the customer with rollouts of university, industry, and executive programs,[9] while new startups such as DataHero are beginning to provide easy-to-use data tools that may lessen organizations dependence on the spreadsheet while at the same time move them up the maturity ladder.

However, what is of concern is the lack of master data technology implementation. In-memory analytics and master data management (MDM) are the technologies least implemented, with about 80% of respondents indicating they have no or very little deployment of in-memory analytics (see Graph 5a) or MDM (see Graph 5b). In-memory analytics is indeed a very new technology that aims to improve analytical performance by storing data in RAM rather than physical storage disks. Given its novelty, we would expect that organizations would not have yet implemented such a technology to a large extent. However, master data management is not a new technology and arguably a key technology that ensures data quality and provides a foundation for other analytic technologies to operate. The cost of leading-class MDM solutions has been noted as a key barrier to implementing such technologies; however, this barrier is present for all technologies. A more plausible explanation is that MDM has more to do with an organization's approach to managing data than any technical- or cost-related factor. Moreover, given the overall lack of data maturity indicated in the survey, it is not surprising that respondents view MDM as low priority in comparison to the other technologies listed.

INFORMATION SUPPLY CHAIN

In analyzing the rest of the results from the survey, we have chosen to use the information supply chain as a framework to position the findings and to better communicate the practical implications of the results (see Figure 1). The information supply chain depicts how data flows through an organization and details the key tasks required to derive value from an organization's data resource. These five key tasks include:

1. **Acquisition** — gathering data from numerous sources
2. **Integration of data sources** — physical and/or logical
3. **Analysis** — analytical process on subsets of data
4. **Delivery** — supplying analysis results in a suitable format

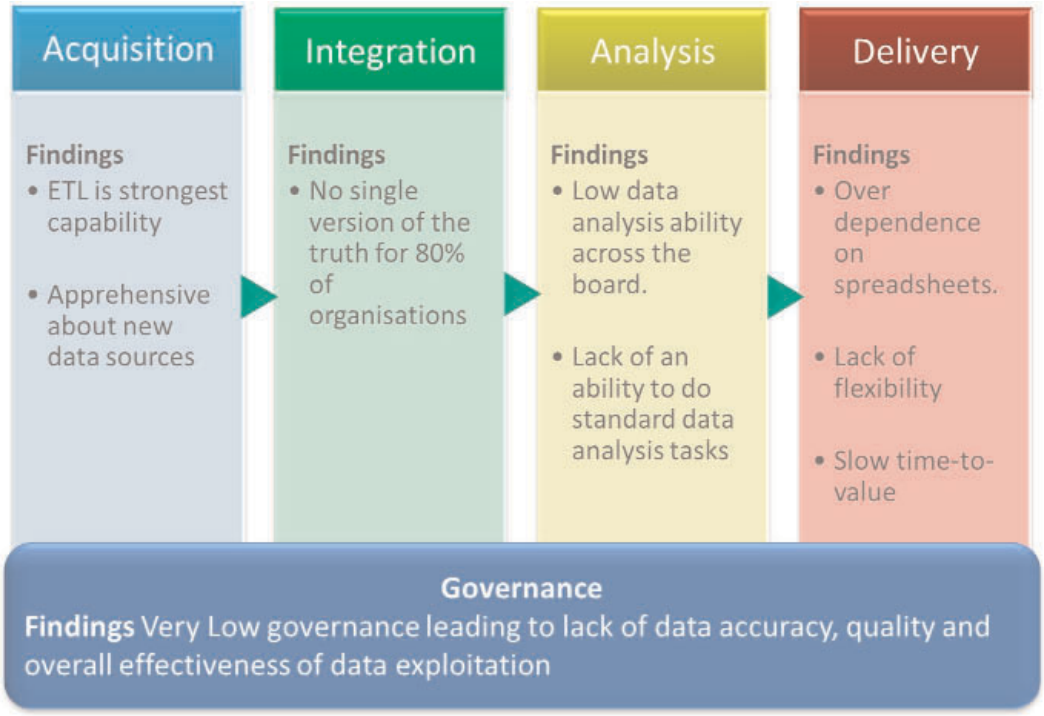


Figure 1 — Key survey findings and how it impacts an organization's information supply chain.

5. **Governance** — promotion of data quality and integrity through structured programs and policies

This information supply chain provides a picture of the difficulties that organizations often face, as highlighted by Sebastian in his article, in data access, integration, normalization, and answering business questions.

Acquisition and Integration

As suggested from the survey, the strongest data capability revolves around ETL and data integration tools with over 60% of respondents indicating a significant maturity, scoring a 3 or higher on a 1-5 scale. (see Graph 2). While only 6% indicate full maturity in ETL, this is still the strongest capability in what we can describe as a capability immaturity profile for all organizations. This result also highlights the significant amount of improvement needed to cope with the barrage of new data sources that organizations now encounter on a daily basis. As it stands, respondents demonstrate an apprehension with new data sources such as clickstreams with only 4% seeing them as a benefit (see Graph 3). Such a low figure may highlight an issue with new data sources, which fits with the maturity profile depicted in the survey results.

Not having a single version of the truth makes it difficult for decision makers to base their decisions on one standard set of facts, consistently and in a timely manner.

Let's now turn to Graph 6, where respondents were asked to rate the extent that various data analysis scenarios apply to their organizations, rating each on a 1-5 scale, with 1 being never and 5 being always. Nearly 60% of organizations rarely or do not have their data residing in one data repository, while 47% also have a pressing need to integrate data across their business functions.

It may come as no surprise then that 64% of organizations only have 21%-60% of their organizational systems integrated (see Graph 7), while 62% do not appear to have a "single version of the truth" across their operations, as shown in Graph 6. Put another way, only about 20% of responding organizations actually have a single version of the truth. This staggering statistic is further supported by the fact, as mentioned earlier, that 80% of organizations apparently do not have, or unlikely have, a master data warehouse deployed in their organization (see Graph 5b). Such an enabling technology is a key feature of delivering a single version of the truth for organizations. However, irrespective of MDM technologies and approaches adopted, not having a single version of the truth makes it difficult for decision makers to base their decisions on one standard set of facts, consistently and in a timely manner. The survey provides further support for this assessment in that 53% of organizations can access information quickly enough to make important decisions in no more than four out of every 10 decisions being made, as illustrated in Graph 7. This suggests that fact-based decision making is not a consistent reality for one in every two organizations.

Analysis and Delivery

The profile of low maturity in general data capabilities is also visible in responding organizations' ability to perform data analysis. For instance, as illustrated in Graph 8, 35% of respondents indicate that they are never able to find patterns and dependencies in multivariant analysis while nearly half report that they can do so only very infrequently. Together, this amounts to nearly 85% of respondents that either never or very infrequently complete an analysis task that falls under a somewhat standard data analysis process. Moreover, the results for this task are not an outlier, 82% and 76% indicate the same level of ability to effectively provide predictive performance of processes and products, respectively. While this further supports a low level of maturity within the organizations, it also highlights the sizeable hurdles that organizations must overcome to fully exploit the potential that organizational data resources have to offer.

In promoting the ideal of Agile analytics, Collier argues that "it takes creative thinking to build the smallest/simplest back-end data solution needed to produce business value on the front end."[10] However, just under half of the respondents believe they can develop a new/ad hoc report within two to seven days (see Graph 9). As a very distinctive data point, this sets a clear benchmark for organizations with low maturity. While two to seven days is not an extraordinary long period, it does indicate a lack of agility in dealing with existing data sets as well as the resulting need for significant effort to provide the type of agility required to prevent data analytics becoming a bottleneck for value realization. Furthermore, the significant lack of ability to do anything in less than one hour highlights severe data

©2013 Cutter Information LLC

restriction organizations are incurring. For instance, all respondents say they can't create a new dashboard/visualization in less than an hour. This is very surprising given the visualization functionality embedded in even the most basic of spreadsheet applications. All respondents also report they are not able to integrate a new data source, and only 4% can mine an existing data set within an hour. This is also quite striking given the fundamental nature of these tasks and the necessity for organizations to complete such tasks in short time frames. These statistics call into question the provision of faster "time to value" and "speed to insight," key components of Agile data analytics.

Governance

In an article a few years ago, Vijay Khatri and Carol Brown reported that data governance is one of the five success "practices" for deriving business value from data assets.[11] Furthermore, according to Khatri and Brown, governance includes "establishing who in the organization holds decision rights for determining standards for data quality."

As Graph 7 illustrates, two-thirds of organizations believe that no more than 60% of their organizational data is accurate. This is coupled with the fact that none of the surveyed organizations "always" has a program in place to verify compliance with data definitions and data standards between operational systems on either a preventative or corrective basis (see Graph 10). Furthermore, while 43% of organizations view data quality as extremely important within their organization (see Graph 6), only 10% have a data governance plan currently implemented with 35% having no plans at present around data governance (see Graph 11). This observation is in line with BI expert Wayne Eckerson who reports that only "1 in 10 organizations have an enterprise data strategy" and as a result are not managing their data as a corporate asset.[12] According to Eckerson, the reality is that the ongoing behavior of these companies leads to "a misplaced perception about the quality of their data" and to the absence of a data quality plan/data governance program. It is also worth mentioning that data quality is the "most unanticipated reason for system failure since organizations began building systems."[13]

While data governance programs have a direct link to data quality, the impact of this lack of data quality can be severe, where the cost of "completing a unit of work" is "10 times more" with "defective" (bad) data than with appropriate "perfect" quality data.[14] Yet, it is interesting to note the equal but scattered distribution of responses for the time frame required to implement a data governance policy across the range of two to four weeks (18%) to greater than a year (18%), as shown in Graph 9. This highlights the varying standards of policies organizations are implementing but also their difficulty in grasping the true meaning of a governance policy. Therefore, the ongoing behavior of an organization to the quality of its data is a key examination point. Organizations can uncover these behaviors by first asking themselves: do we calculate the cost of bad data? Surprisingly, or not, as Graph 6 shows, 83% of surveyed organizations infrequently, if ever, calculate the cost of bad data! A recent TDWI best practices report suggested that "most organizations are moderately satisfied with their ability to address data quality problems."[15] However, the *CBR* survey notes there is a danger of this moderate satisfaction turning into extreme dissatisfaction, because organizations are not investing into the capabilities needed to arrive at an appropriate level of data quality maturity. Overall, the survey reflects a less than desirable level of maturity around data governance. This calls into question the actual importance of data quality to our responding organizations since the behavior and discipline needed to maintain data quality appears to be lacking. However, the importance of data governance cannot be overstated. In the words of Khatri and Brown, "governance is a key element in not only improving economic efficiency and growth, but also enhancing corporate confidence."[16]

Overall, the survey reflects a less than desirable level of maturity around data governance.

CONCLUSION

To recap, the survey respondents for this *CBR* have primarily defined Agile data analytics as the application of Agile methodologies to data analytics and BI. This is important as it provides a pointer to Agile software development as a key reference on how to complete unstructured projects using new technologies with the highest possibility of success. Furthermore, as Agile is not a new concept, most organizations with an IT capability should be in the position to transfer and apply the underlying principles of Agile to their data projects. However, Agile in itself is just a development

style or project management methodology. The survey reveals a much deeper issue in the restricted data capabilities within organizations across the information supply chain. Unfortunately, these capabilities are beyond the scope of what Agile can accomplish, especially in data governance. Simply put: *without good data governance, even well-guided implementations will run a risk of failure, stemming from a lack of data quality and accuracy.*

Since one of the key principles of Agile development is the prioritization of projects, we'll conclude by highlighting what aspects we propose organizations should prioritize:

1. Designing and implementing a data governance program to enforce data standards and maintain data quality.
2. Reducing the dependence on spreadsheets as a primary analytics tool to uncover new ways of looking at unexpected things. (This introduces what Sebastian refers to as "creative flexibility" and breaking the "emotional attachment" to maintaining the status quo.)
3. Designing and implementing a MDM solution to ensure the organization has a common semantic meaning for important business objects (e.g., customer, product).

These three recommended actions are all achievable in an Agile fashion through experimentation. They help ensure that organizations bite off a little piece of each action in a cost-effective manner and evaluate the impacts of the actions in order to appreciate two key issues: (1) their own ability to do the project and their level of maturity around the capability, and (2) the business value, or otherwise, of the initiative (embracing the Agile principle of "failing fast").

ENDNOTES

[1]Collier, Ken. *Agile Analytics: A Value-Driven Approach to Business Intelligence and Data Warehousing*. Addison-Wesley Professional, 2012.

[2]Wang, Richard Y., Yang W. Lee, Leo L. Pipino, and Diane M. Strong. "Manage Your Information as a Product." *Sloan Management Review*, Summer 1998, pp. 95-105.

[3]Levitin, Anany V, and Thomas C. Redman. "Data as a Resource: Properties, Implications, and Prescriptions." *Sloan Management Review*, Fall 1998, pp. 89-101.

[4]Eckerson, Wayne. "Creating Enterprise Data Strategy: Managing Data as a Corporate Asset." BeyeNETWORK, 2011, pp. 1-39.

[5]Mithas Sunil, Ali R. Tafti, Indranil Bardhan, and Jie Mein Goh. "Information Technology and Firm Profitability: Mechanisms and Empirical Evidence." *MIS Quarterly*, Vol. 36, No. 1, 2012, pp. 205-224.

[6]Davenport, Thomas H., and D.J. Patil. "Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*, October 2012.

[7]"Surge in Demand for UK Big Data Specialists." SAS UK/e-skills UK, [year?] (www.sas.com/reg/gen/uk/eskills-big-data-report).

[8]Olshan, Jeremy. "88% of Spreadsheets Have Errors." Market Watch/*The Wall Street Journal*, 20 April 2013 (www.marketwatch.com/story/88-of-spreadsheets-have-errors-2013-04-17).

[9]See, for example, Irish Management Institute's Diploma in Data Business (www.imi.ie/imi-diplomas/imi-diploma-in-data-business).

[10]Collier (see 1).

[11]Khatri, Vijay, and Carol V. Brown. "Designing Data Governance." *Communications of the ACM*, Vol. 53, No. 1, 2010, pp.148-152.

[12]Eckerson (see 4).

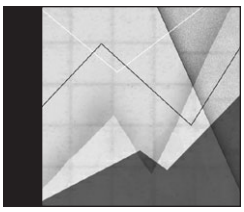
[13]McKnight, William. "Improving Data Quality Through Data Modelling." CA Technologies, 2010, pp. 1-9 (http://odbms.org/download/Data_Quality_McKnight_ERwin.pdf).

[14]Redman, Thomas C. *Data Driven*. Harvard Business School Press, 2008.

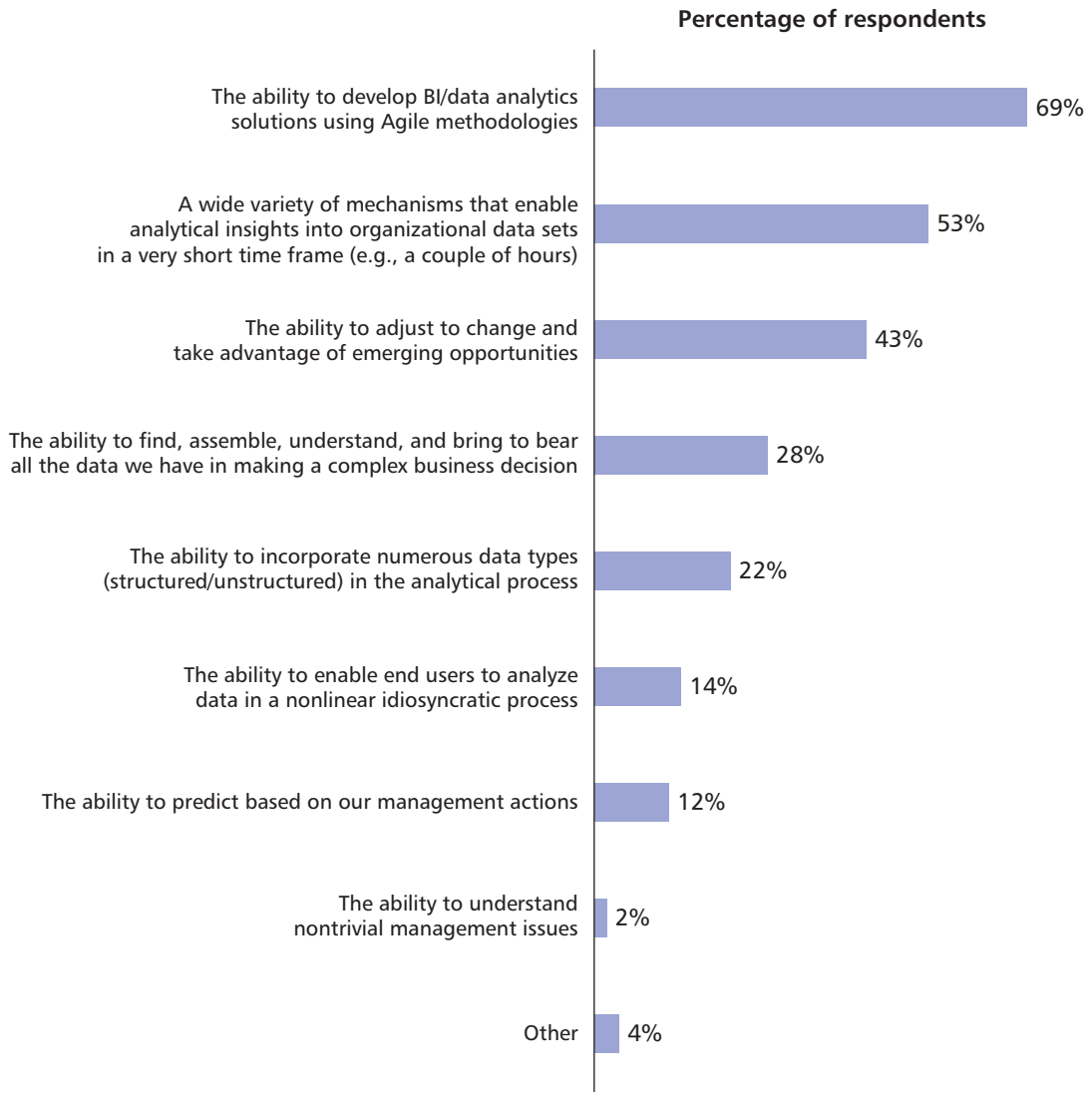
[15]Stodder, David. "Achieving Greater Agility with Business Intelligence." TDWI, 2013, pp. 1-36 (http://tdwi.org/research/2013/01/tdwi-best-practices-report-achieving-greater-agility-with-business-intelligence.aspx).

[16]Khatri and Brown (see 11).

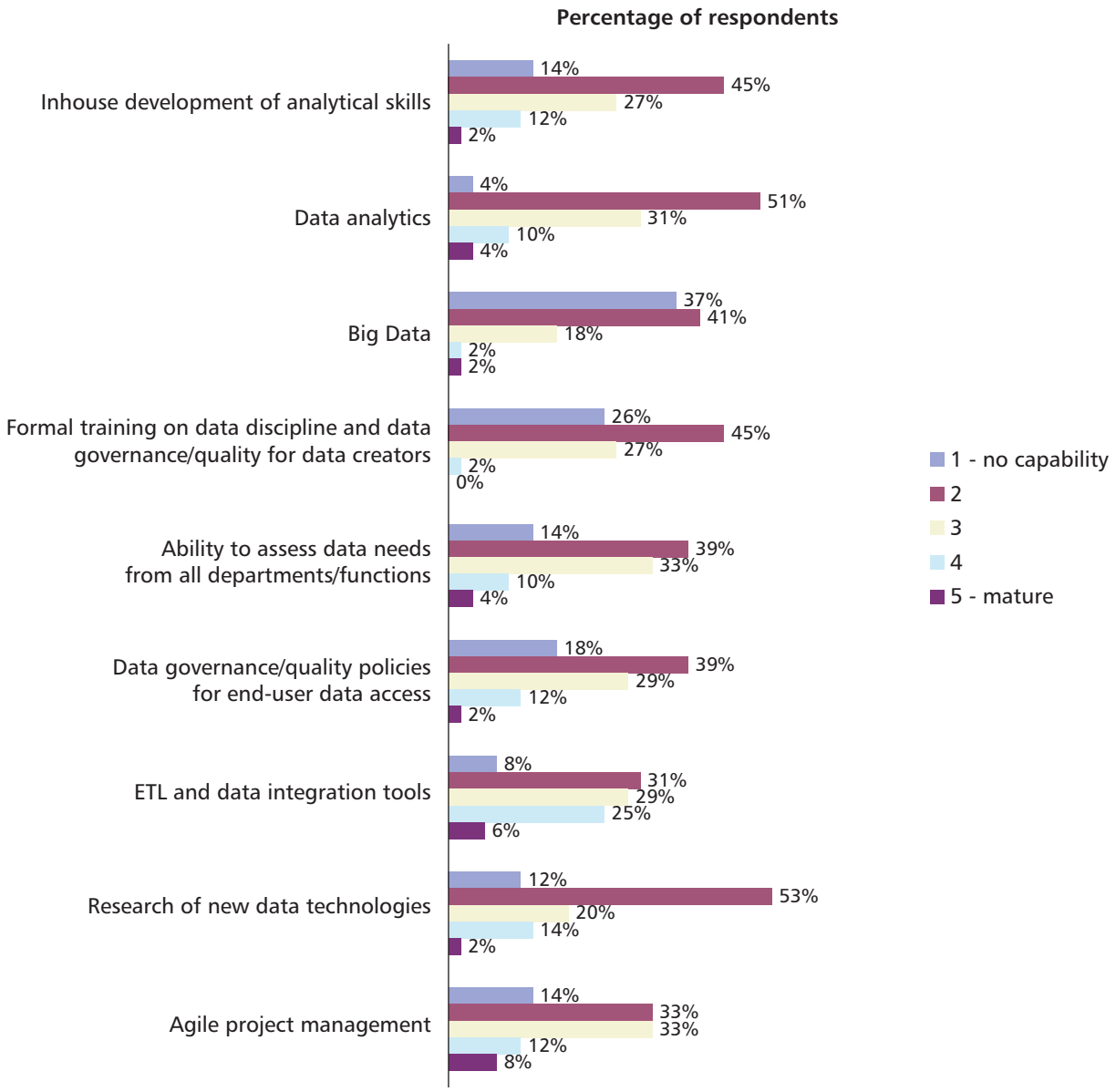
©2013 Cutter Information LLC



Agile Data Analytics



Graph 1 — Which of the following statements accurately describe your understanding of Agile data analytics? (Respondents asked to select a maximum of three responses.)

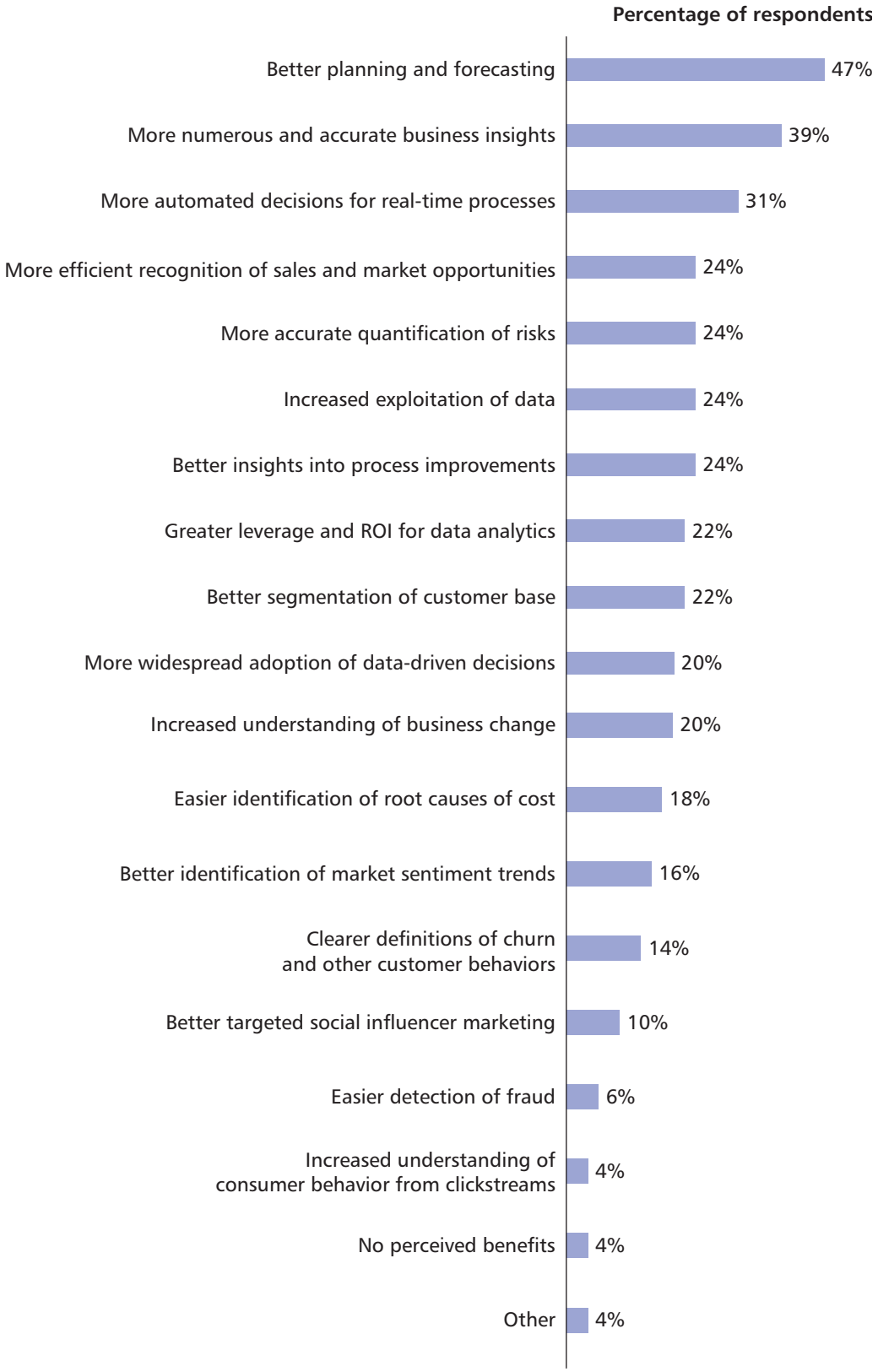


Graph 2 — Rate the maturity of each of the following capabilities in your organization.

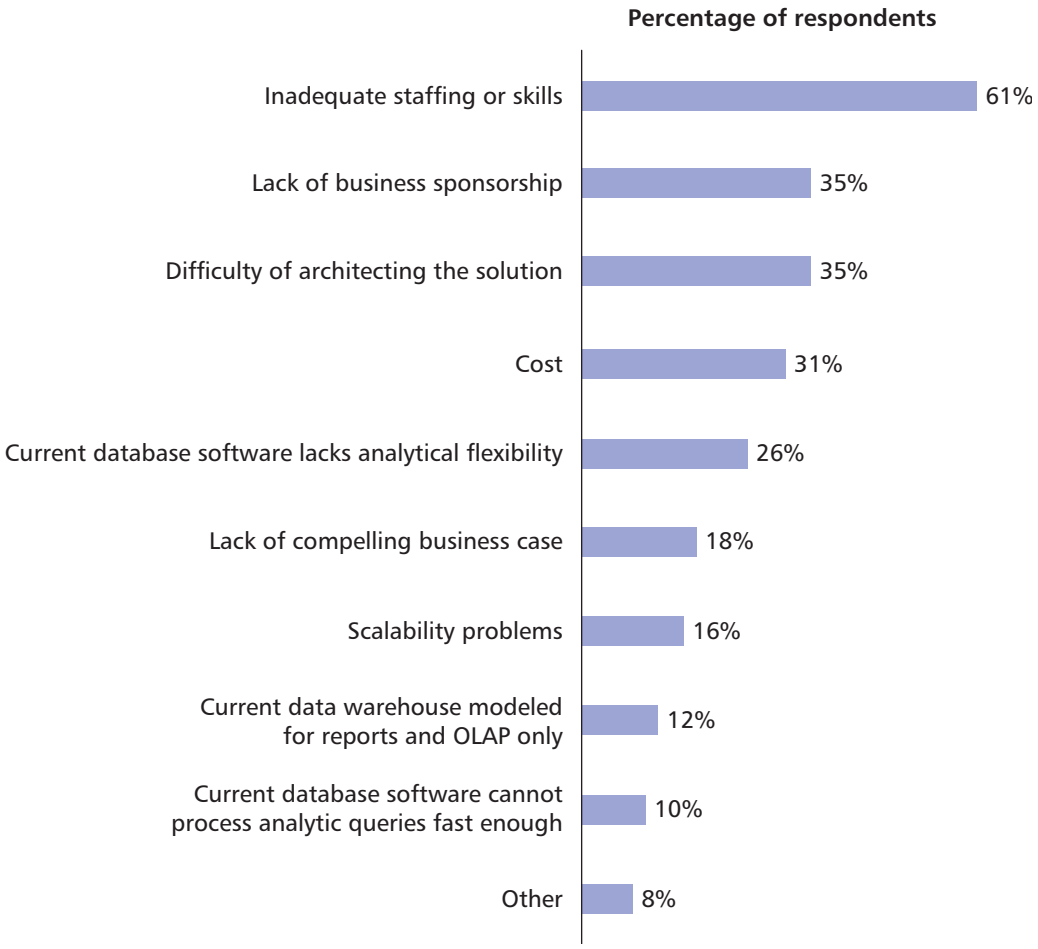
SURVEY DEMOGRAPHICS

This survey examined organizations' understanding of Agile data analytics; the drivers behind, the goals and benefits of, and the barriers to Agile data analytics; the maturity and use of various data capabilities; and whether a data governance program has been implemented. Fifty-five percent of the 51 responding organizations are headquartered in North America, another 22% in Asia/Australia/Pacific, 14% in Europe, and 6% in Africa, with the remainder headquartered in South America and the Middle East. Seventy-three percent report having fewer than 100 IT professionals, with 12% reporting between 100 and 500, 6% between 500 and 1,000, and 10% reporting more than 1,000. Responding organizations' annual IT budgets are also on the smaller side, with 33% having IT budgets under US $500,000, 29% having IT budgets between $500,000 and $10 million, 12% having IT budgets between $10 million and $100 million, and the remaining 12% having IT budgets over $100 million (14% of respondents do not know their annual IT budget). Forty-seven percent of respondents report annual revenues under $10 million, while 16% report annual revenues of over $10 billion. Thirty-five percent hold senior management/policymaking or IS/IT management titles, with software engineering/programming (18%), consulting (12%), and project management (10%) being among the other job titles.

©2013 Cutter Information LLC

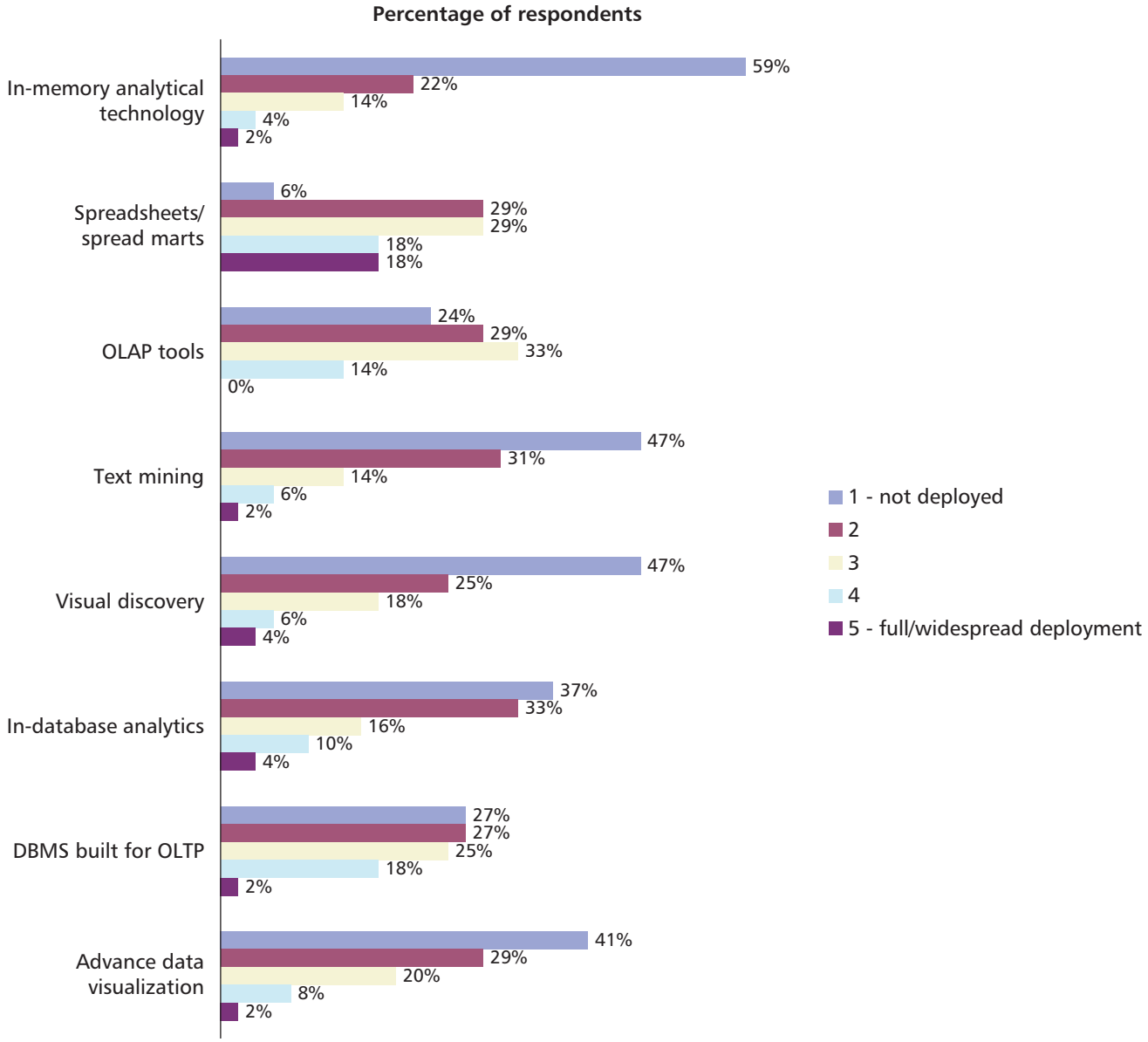


Graph 3 — What are the main benefits your organization has realized or expects to realize from Agile data analytics? (Respondents asked to select a maximum of five responses.)

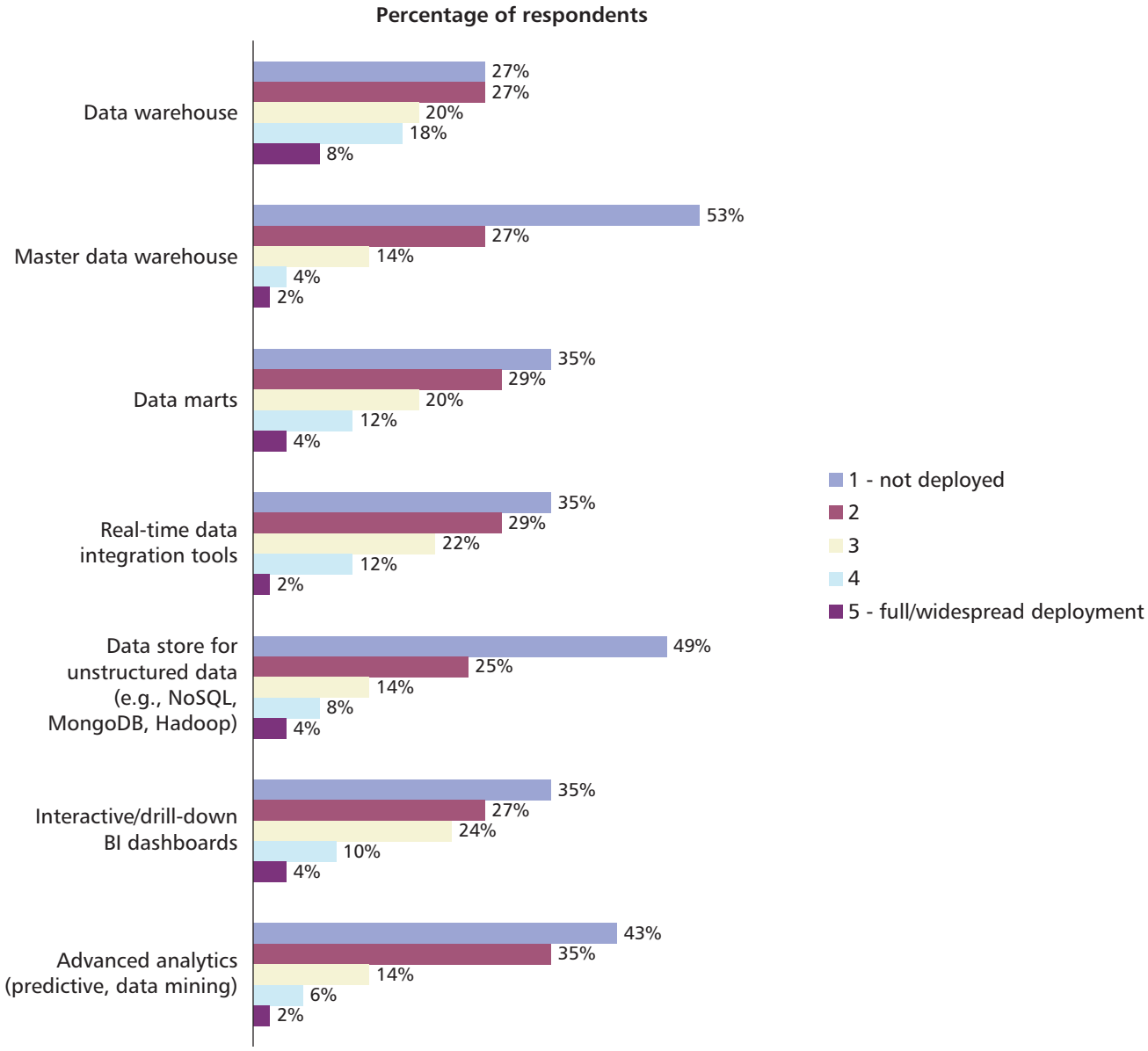


Graph 4 — What are the main barriers to Agile data analytics in your organization? (Respondents asked to select a maximum of five responses.)

©2013 Cutter Information LLC

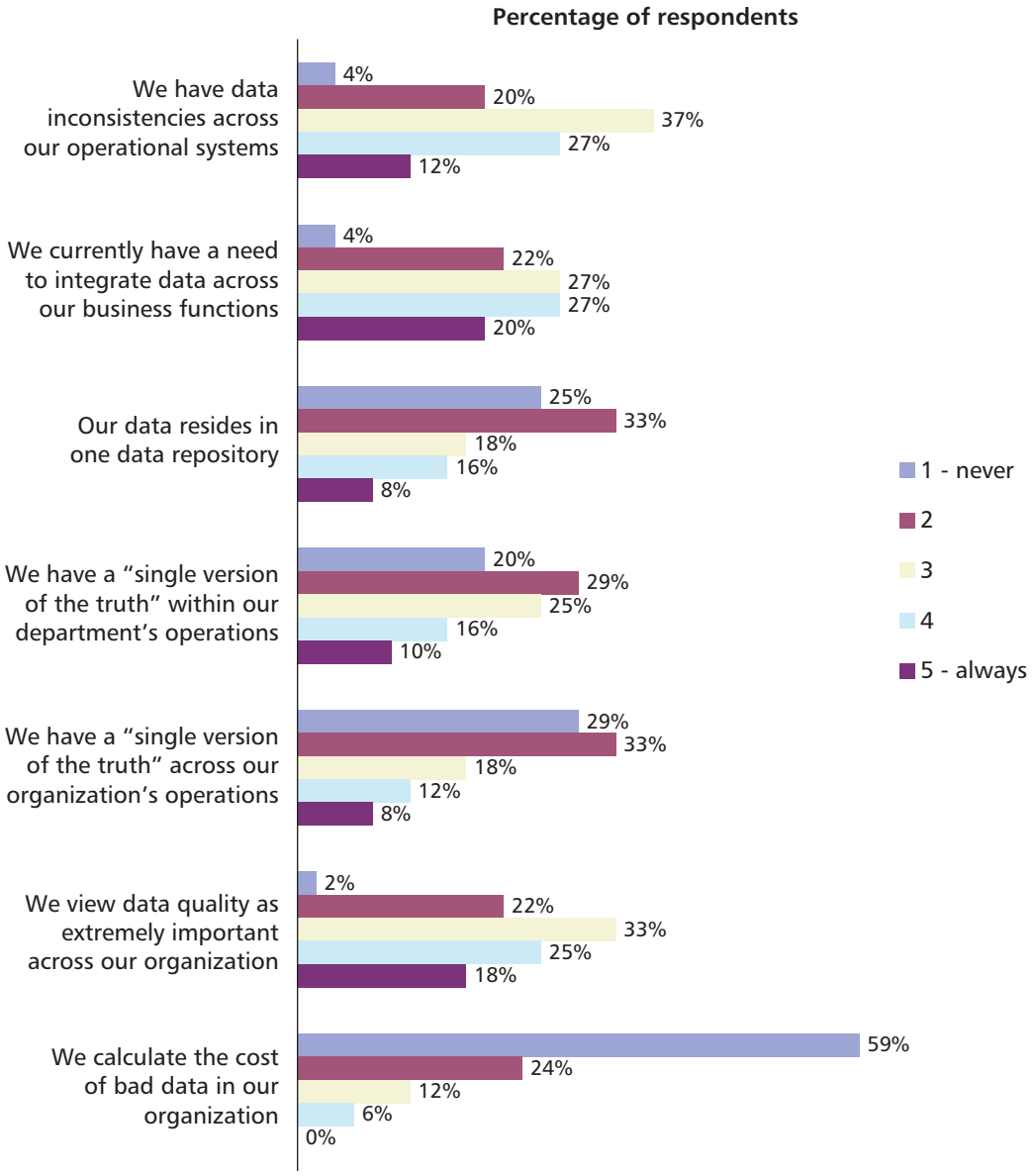


Graph 5a — To what extent is each of the following enabling technologies deployed in your organization?

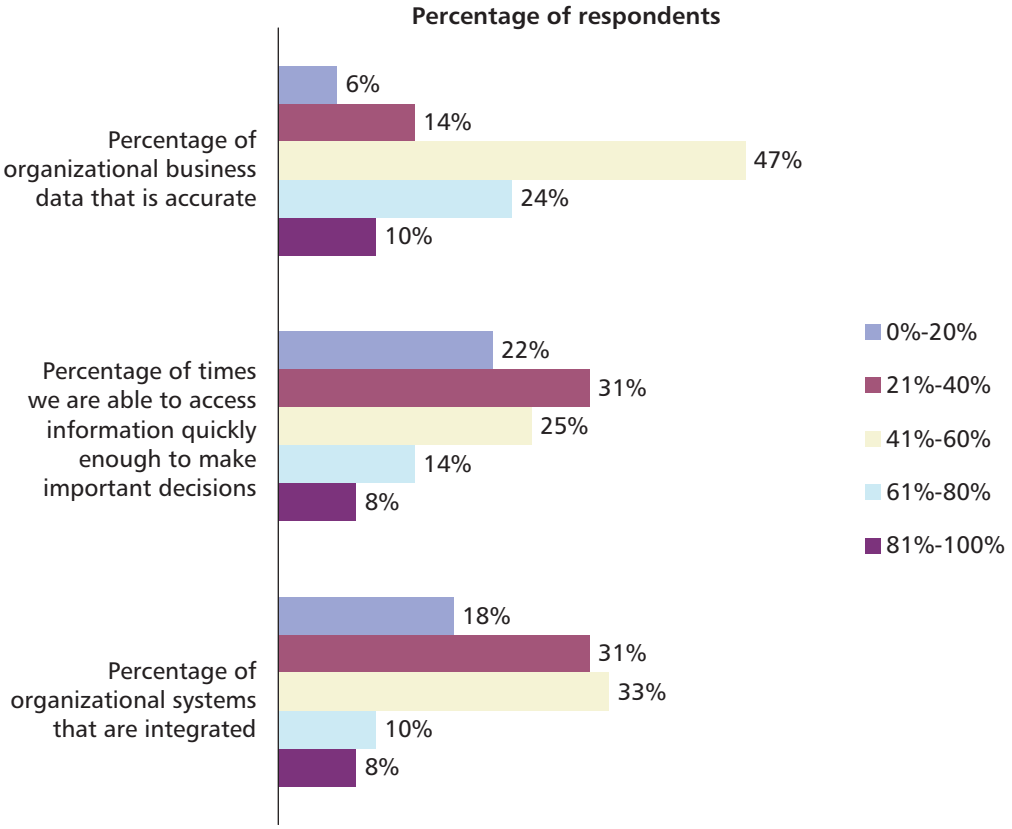


Graph 5b — To what extent is each of the following enabling technologies deployed in your organization?

©2013 Cutter Information LLC

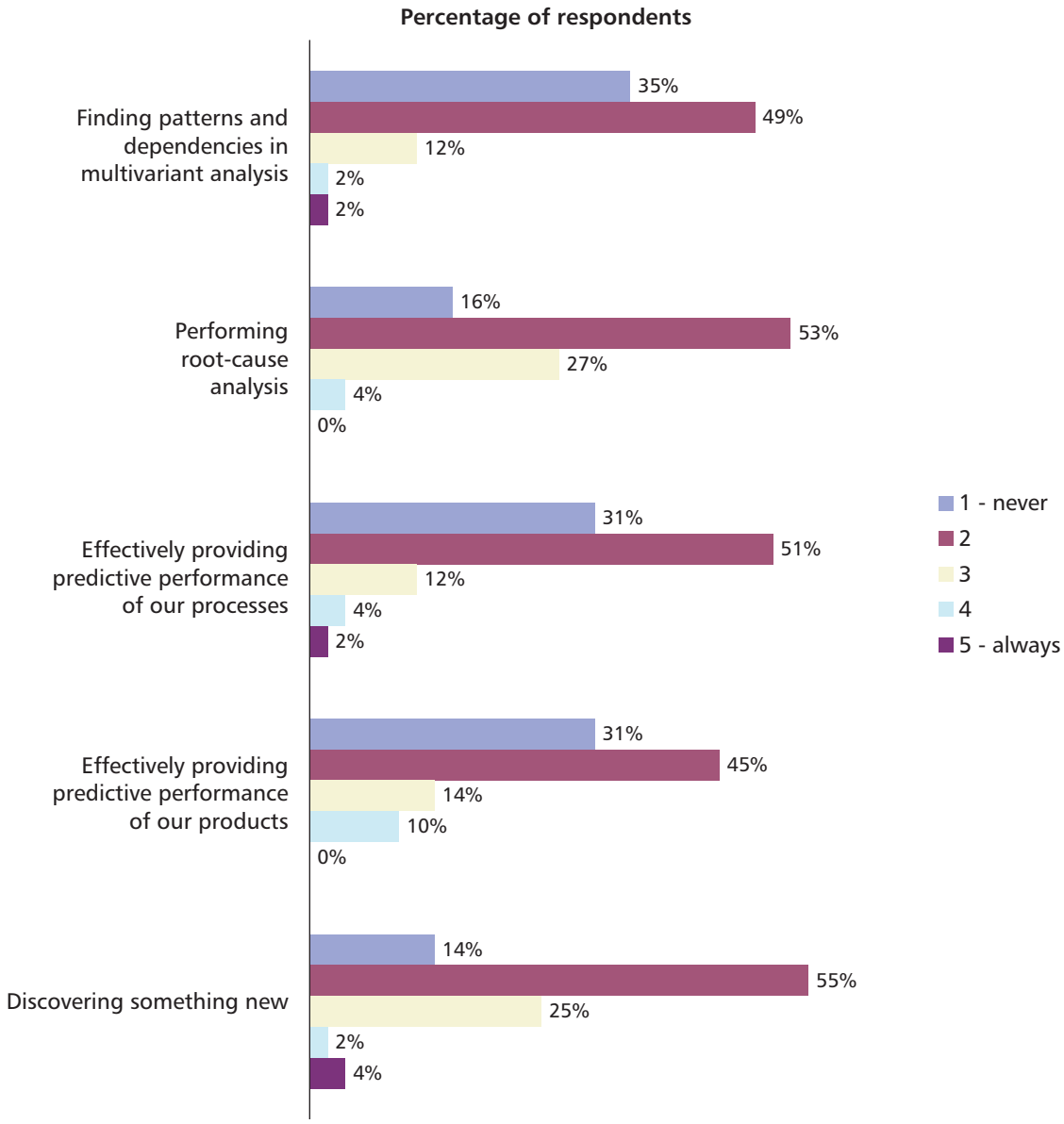


Graph 6 — To what extent does each of the following statements apply to your organization?

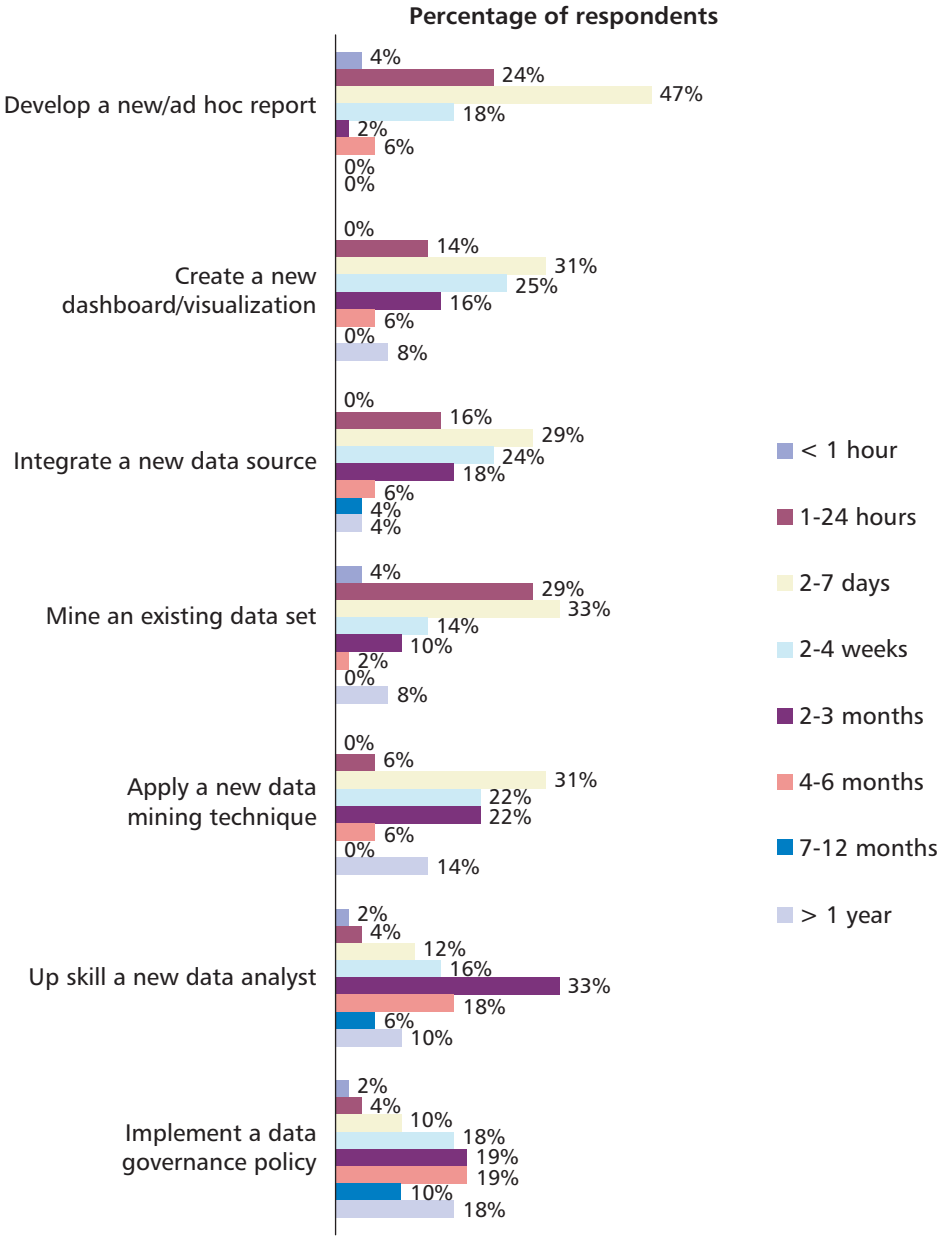


Graph 7 — What is the estimate percentages for each of the following for your organization?

©2013 Cutter Information LLC

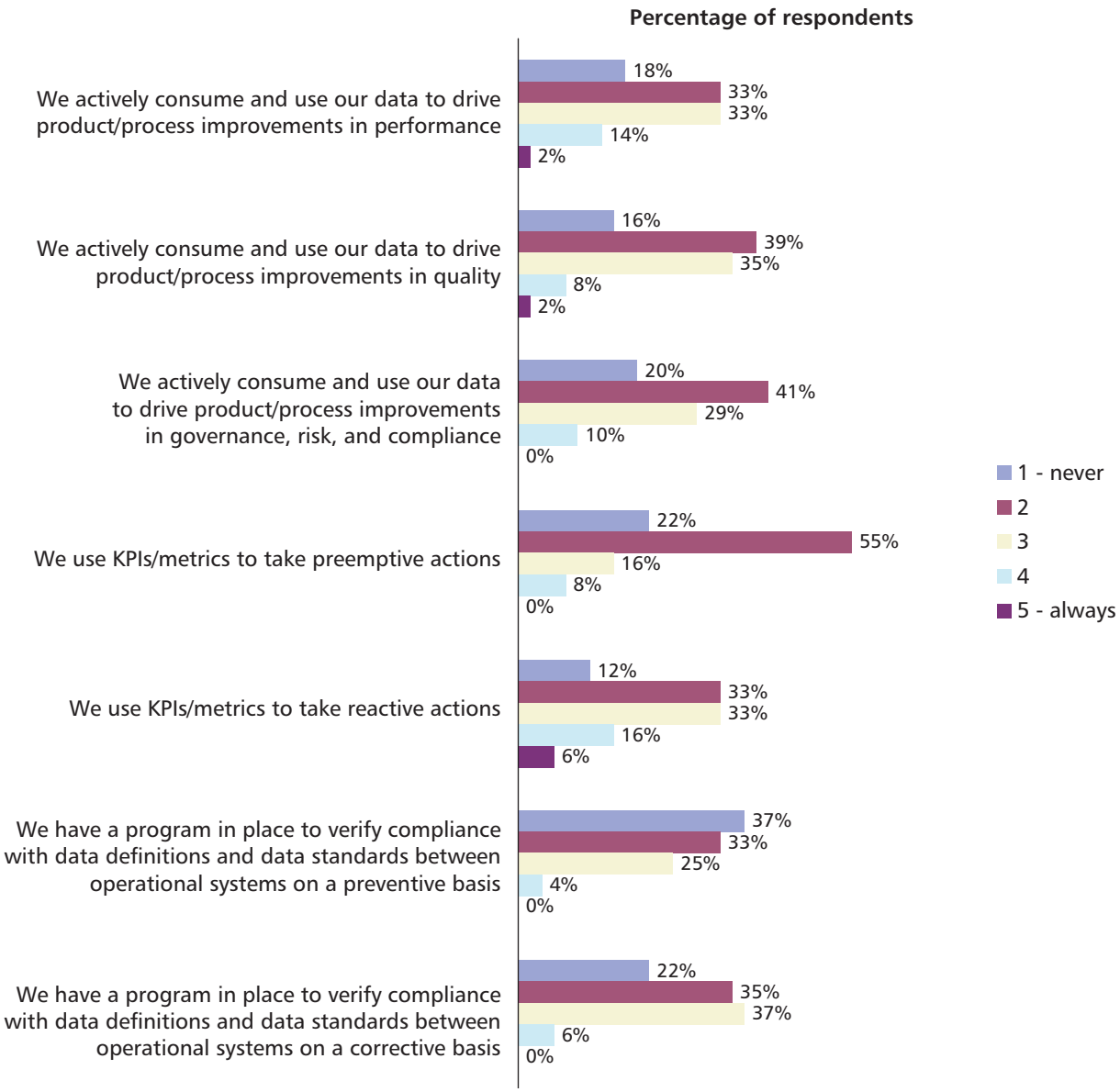


Graph 8 — How easily is your organization able to gather and mine data for each of the following purposes?

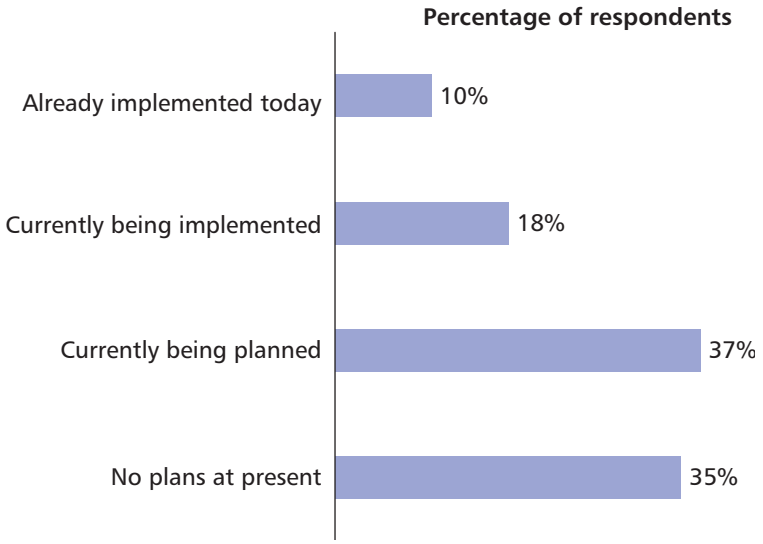


Graph 9 — What is the average time frame your organization requires/would require to complete each of the following tasks?

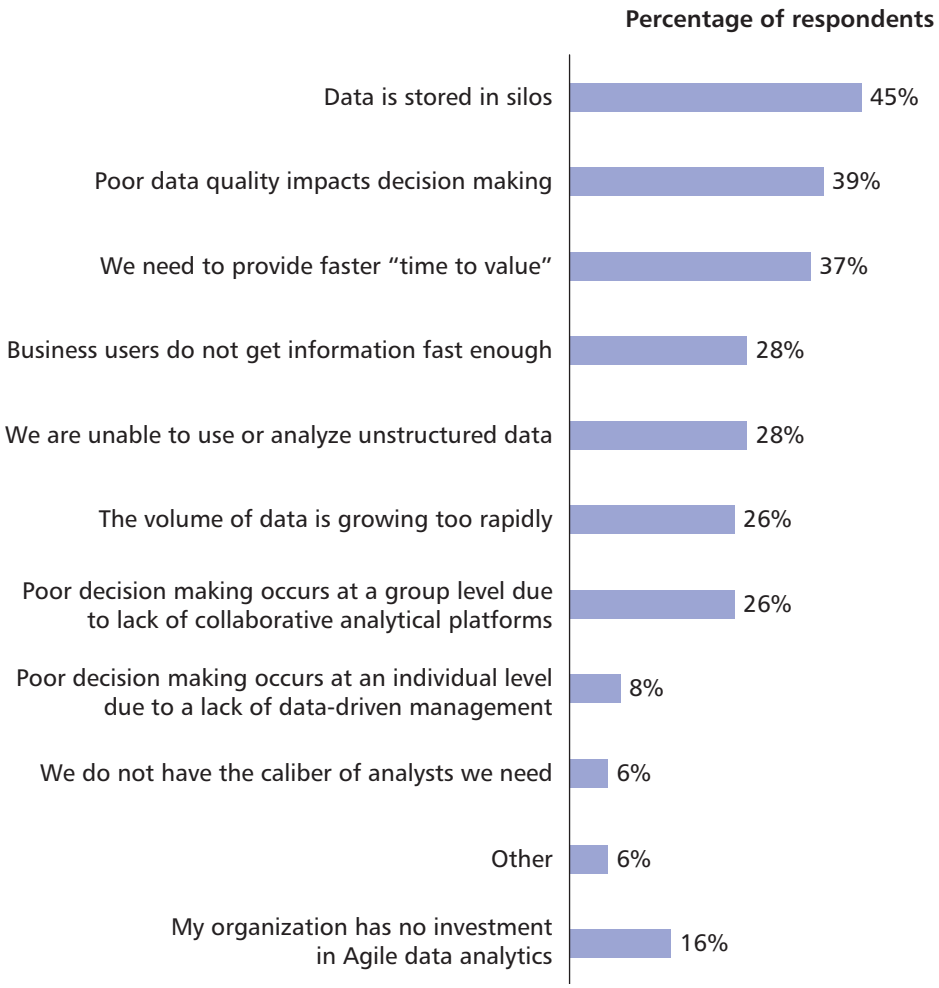
©2013 Cutter Information LLC



Graph 10 — How frequently do each of the following statements apply in your organization?

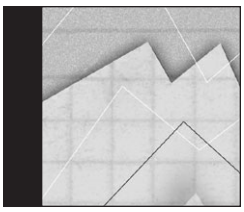


Graph 11 — What is the status of your data governance program?



Graph 12 — What are the pressures driving your Agile data analytics investments? (Respondents asked to select a maximum of three responses.)

©2013 Cutter Information LLC



About the Authors

Joseph Feller is a Senior Consultant with Cutter Consortium's *Data Insight & Social BI* and *Business Technology Strategies* practices. He is a Senior Lecturer at the University College Cork (UCC), where he coordinates the PhD program in business information systems, supervises postgraduate and postdoctoral researchers, and conducts research. Prior to joining UCC, Dr. Feller taught at the Ringling College of Art and Design. His research focuses on open innovation and the peer production phenomena, including open source software, open content, crowdsourcing, innovation marketplaces, citizen science, and social media collaboration platforms. Dr. Feller has published five books, more than 60 academic papers, and more than 100 practitioner articles and reports. His research in open source software has been particularly influential, cited over 1,000 times in academic literature. Dr. Feller has directed various international research projects funded by Irish and EU agencies, private industry, and government partners. He is a frequent presenter and panelist at international conferences and has been invited to deliver keynotes, seminars, workshops, and policy briefings by numerous industry organizations, private firms, NGOs, and Irish and European government bodies. He can be reached at jfeller@cutter.com.

Sebastian Hassinger is a Senior Consultant with Cutter's *Agile Product & Project Management* practice. He has worked in the IT industry for over 20 years both in large corporations and as an entrepreneur and has provided strategic consulting for large and small firms on projects as broad as corporate direction and as focused as product development strategy. Mr. Hassinger founded two ISPs, helped launch several other startups, and held senior product strategy and business development roles with IBM, Oracle, and Apple. He currently leads Tech Propulsion Labs, a Vietnam-based Agile development shop specializing in Web and mobile applications. Mr. Hassinger holds MBAs from Columbia University and London Business School, is a published author, and holds over a dozen software and business model patents. He can be reached at shassinger@cutter.com.

Tadhg Nagle is a researcher/lecturer in the Department of Accountancy Finance and Information Systems at University College Cork, Ireland. He is also co-director of Ireland's first executive program in data business at the Irish Management Institute. With a background in finance services, Dr. Nagel became a Business Analyst Lab Leader at the Digital Enterprise Research Institute from which he holds a PhD. Focusing on emerging technologies, strategic innovation, and disruptive technologies, he has published several papers in leading international IS conferences and journals. Dr. Nagel's current focus is on the implementation of cloud and Big Data technologies with a view toward creating real business value. He can be reached at T.Nagle@ucc.ie.

David Sammon is a researcher/lecturer in business information systems at University College Cork, Ireland. He is also co-director of Ireland's first executive program in data business at the Irish Management Institute. Dr. Sammon's current research interests focus on the areas of conceptual data modeling, master data/information management, theory and theory building, and redesigning organizational routines through mindfulness. He has been published extensively in international journals and conferences. Dr. Sammon is Associate Editor of the *Journal of Decision Systems* and coauthor of *Enterprise Resource Planning Era: Lessons Learned and Issues for the Future*. He can be reached at DSammon@afis.ucc.ie.