



**MANAGED FUNDS  
ASSOCIATION**

MFA DATA | 2019

# **THE LIFECYCLE OF DATA IN CONTEXT:**

## **How Data Proliferation is Shaping Alternatives**

September 2019

FOR INQUIRIES CONTACT:

**Onur Erzan**

Senior Partner  
McKinsey & Company  
3 World Trade Center, New York, NY 10007  
Tel: +1 212 446 7172  
E-mail: [Onur\\_Erzan@mckinsey.com](mailto:Onur_Erzan@mckinsey.com)

**Brooke Harlow**

Chief Commercial Officer  
The Managed Funds Association  
546 Fifth Avenue, New York, NY 10036  
Tel: +1 212 542 8460  
E-mail: [BHarlow@MANAGEDFUNDS.ORG](mailto:BHarlow@MANAGEDFUNDS.ORG)

Prepared by

**McKinsey  
& Company**

Using data to make investment decisions is not a new concept. What is new is the sheer volume and scope of data in existence today. Satellite imagery, credit card transactions, and sensor and real-time geospatial information are a few examples of the proliferation of new sources of data. According to research by IBM, 90 percent of all data existing today was generated in the last two years.

In this context, the question that many investors have naturally been asking is whether and where value exists within the morass of new and changing information. Armed with improvements in technological infrastructure to store and process data more efficiently, and advances in the mathematical and statistical techniques to analyze data more effectively, investors of different stripes are grappling seriously with understanding how they can separate the signal from the noise.

How successful have they been? How have they approached data strategy and spending? Which tools and techniques are delivering the most value, and which are still unproven? What paths have firms pursued in hiring talent, organizing, collaborating effectively, and overcoming cultural challenges?

To find the answers, the Managed Funds Association (MFA) conducted one of the most comprehensive recent research efforts focused on investors' use of data and analytics. The results are the product of interviews with the COOs, CTOs, heads of data management, general counsels, and other leaders at over 25 MFA member firms, collectively representing \$1.5 trillion of assets, as well as a quantitative survey of MFA membership. In many cases, the research surfaced more questions than answers. We feel that is appropriate for a field that is evolving so rapidly and hope this report serves as a foundation for further dialogue within and outside the industry.

## **What we heard**

### *Data: The jury is out on alternative data*

The firms that participated in this research spanned a wide breadth of investment strategies and this, to a large extent, drove meaningful variations in their data acquisition and management practices.

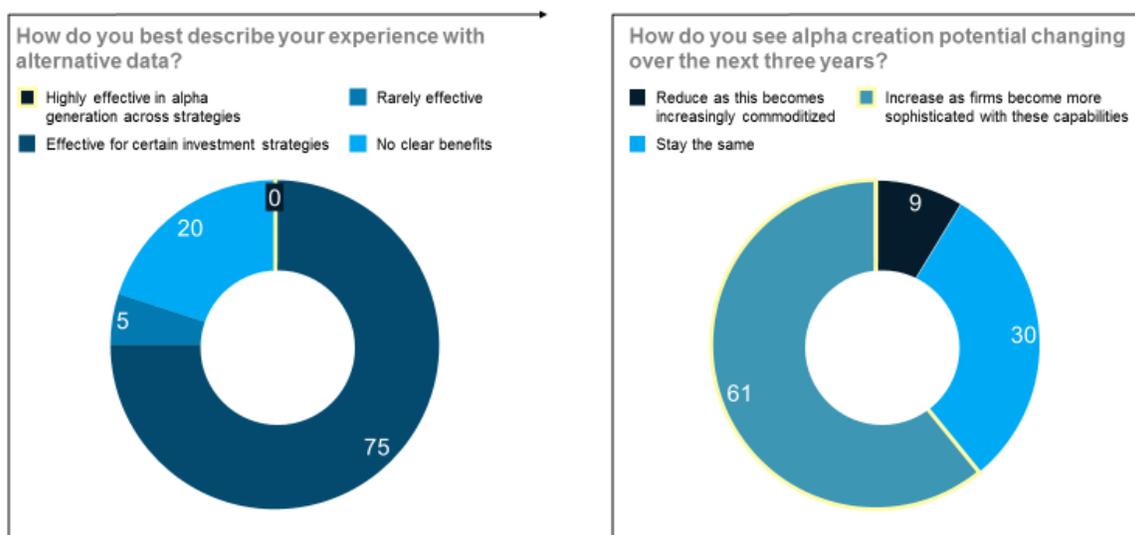
Despite these differences, over half of survey respondents described quantitative research as core to some or all parts of their investment approach, and almost every firm has been experimenting with so-called “alternative data” as part of this. Yet even as they experiment, many firms remain skeptical of the sustainable value that alternative data can deliver. Even at firms that subscribe to hundreds of different data feeds and have a broad aperture and appetite for testing new types of data, the common refrain was that more traditional data still drives the lion share of decision making and investor focus. Accordingly, 40% of respondents said that less than 10% of their firm's expenditure on external data went towards what they would define as “alternative data”.

Of course, the lines between what is “traditional” and what is “alternative” are blurry. Different firms draw these boundaries differently. For some, “alternative” data connotes anything outside

pricing, accounting and financial data. For others, the distinction refers less to the *type* of data (especially as newer types of data become more commonplace) and more to the level of *effort* required to access or create data, or to the level of *uncertainty* that the data will yield something valuable.

## Exhibit 1

### Lukewarm enthusiasm for alternative data, but optimism for the future

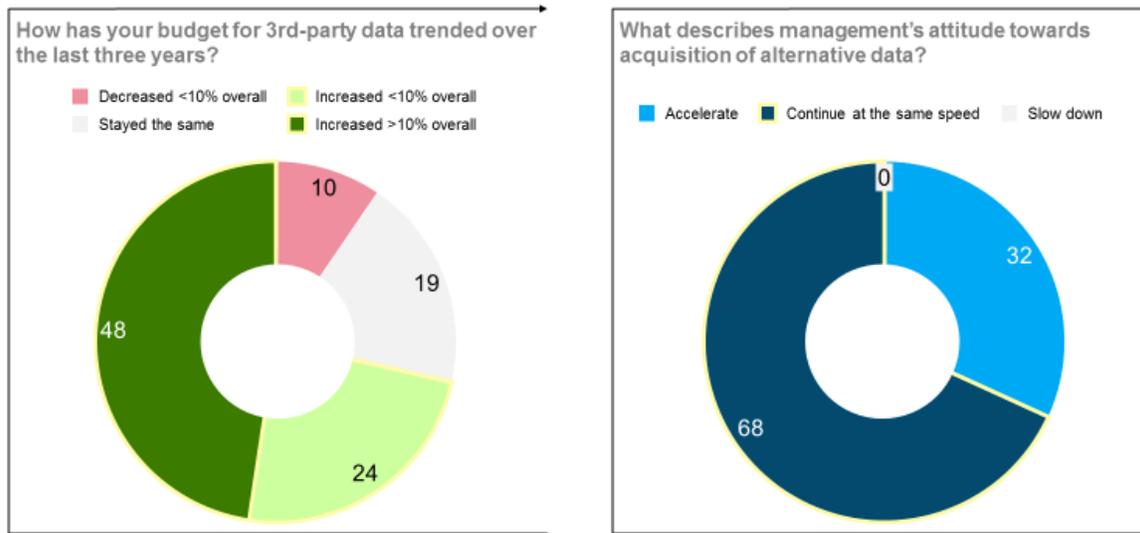


Given this orientation, it is perhaps not surprising that a quarter of survey respondents report seeing “no” or “very limited” benefits from alternative data, with no respondents describing alternative data as highly effective in alpha generation across strategies (Exhibit 1). The skepticism about the sustainable value of alternative data is more pronounced for discretionary investors relative to systematic or multi-strategy firms, and for non-equity-focused investors relative to their equity-focused counterparts. Credit investors, for example, were much more focused on using automation to improve decision-making productivity than on alternative data acquisition.

Given firms’ mixed success in using alternative data, the general optimism respondents expressed about the future is striking. A majority (61%) said that alpha creation potential from alternative data will increase as the quality of data improves, and the corresponding analytical techniques become more powerful. And it appears that firms are putting capital behind this conviction. Almost half of survey respondents reported that their budget for third-party data has increased by more than 10% over the last three years (Exhibit 2). Two-thirds predicted that their firms will continue acquiring alternative data at the same speed, and the remainder predicted an acceleration. A vocal minority disagreed, however. For them, the heyday of deriving alpha from nontraditional data has already passed. They told us that both the data and the data science are becoming commoditized and that only the most sophisticated firms with the orientation and ability to invest significantly in data acquisition, engineering and analysis will retain any sustainable edge.

## Exhibit 2

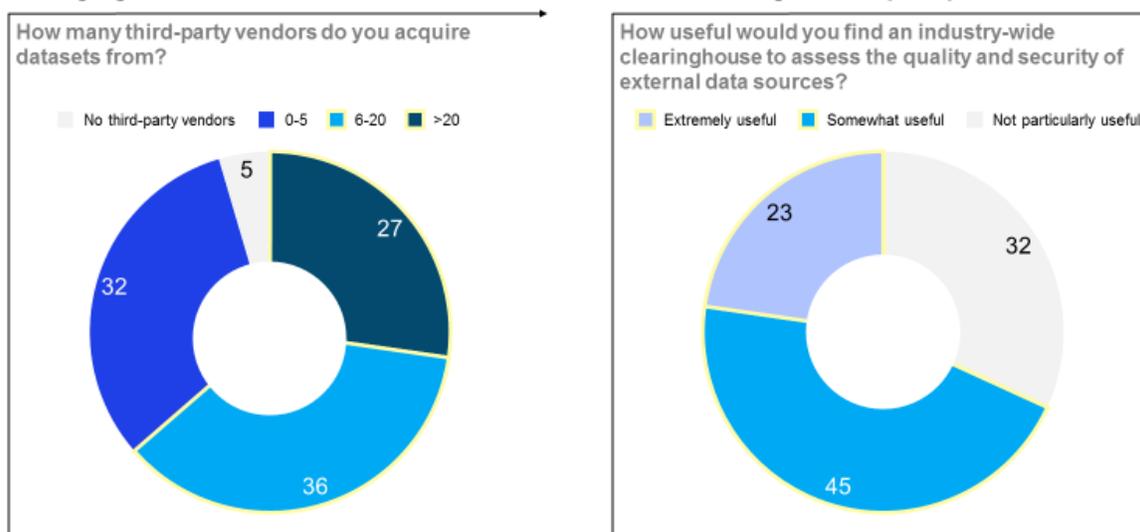
### Rapid growth in third-party data expenditure, with no sign of abating



Despite this divide, most respondents agreed that interfacing with third-party data vendors is onerous in terms of data quality, cost, and—perhaps more importantly—risk and reputation (Exhibit 3). One cause of complexity in interfacing with vendors is the fragmentation of data sources. Nearly 64% of respondents indicated that they utilize six or more third-party vendors for their data, while 27% reported using 20 or more data vendors. While the most sophisticated firms may enjoy privileged relationships with data providers eager to prove themselves and have the internal tools and resources to stay abreast of the dynamic landscape, this is not the case for most.

## Exhibit 3

### Managing external data vendors is resource intensive and a significant pain point



Over two-thirds of survey respondents (68%) said that an industry-wide clearinghouse that assesses the quality and security of external data sources would be very or somewhat useful, especially given the potential regulatory changes that could impact access to data (e.g., Geolocation Privacy and Surveillance Act, the American Data Dissemination Act). The clearinghouse would not be a complete replacement for the internal checks that firms go through, but for many firms, it could reduce the time, effort, and risk involved with this process. Of course, the natural question that remains is whether such a mechanism would only serve to catalyze the commoditization of data that many respondents already perceive. After all, how useful would it be to know that a specific data set is of a high quality if all your competitors have the same information?

### *Analytics and tools: Look beyond machine learning*

Raw data by itself is of limited use, whether alternative or traditional. What is necessary alongside it is the technical and human infrastructure to clean, organize, test, analyze and derive value from this data. Particularly with regard to that last step—analysis—it is difficult to avoid the buzz surrounding machine learning. This class of techniques has come to dominate the discourse in recent years, sometimes to the point that the terms are used interchangeably with advanced analytics.

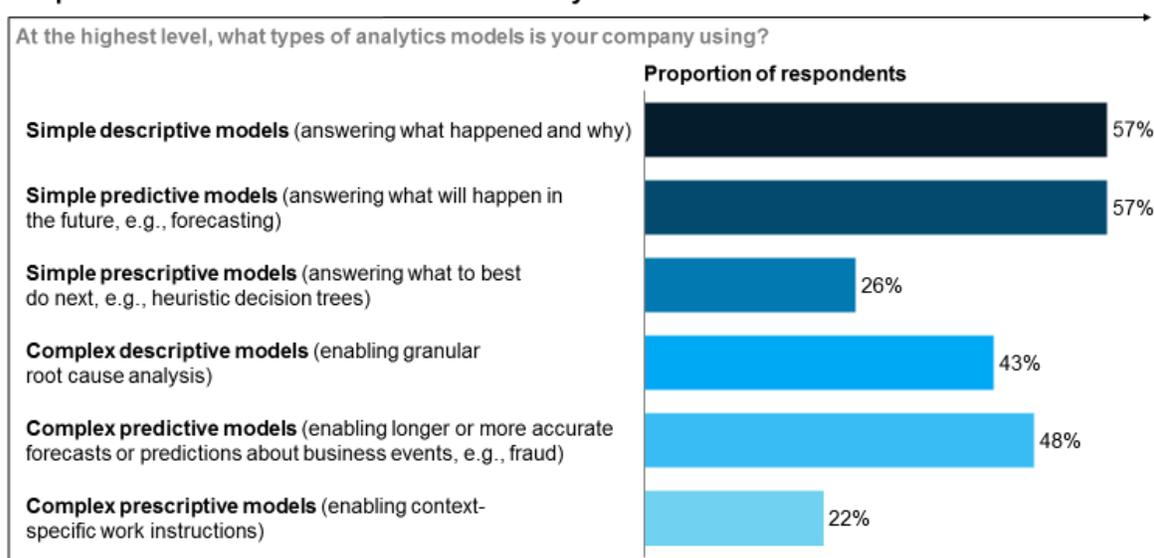
Machine learning indeed has some valuable applications for the investment process, such as identifying signals, helping with portfolio construction, and supporting predictive analysis. However, our interviews reinforced the perspective that machine learning is only one arrow in the quiver. Traditional statistical techniques often are perfectly adequate for the task at hand and are invariably more transparent and easier to understand, which are important considerations for adoption.

Several respondents also highlighted that the risk of overfitting is high if one attempts to force-fit reinforcement learning or deep learning techniques to the investing context. They consider these methods more suitable when the volume of data being grappled with is vastly greater than for most investors. Accordingly, only a minority of respondents reported using complex predictive or prescriptive models in their analytical approaches (Exhibit 4). This is another area where the distinction between discretionary and systematic approaches is important. Many of these advanced techniques are more relevant for systematic investing, especially within high-frequency trading.

By and large, however, respondents reported that their firms have gained less value from these complex analytical techniques than from certain more fundamental process improvements. Specifically, the ability to rapidly onboard and test new data has yielded positive outcomes for many, often catalyzed by the transition from on-premise to cloud environments. In a few instances, respondents described “road testing” new processes with internal data from non-investment functions (e.g., HR, compliance) before deploying a refined process to the investment function.

## Exhibit 4

### Simple can be beautiful when it comes to analytics



Even as these processes and techniques evolve, it was notable that many firms are still in the early days of building sophisticated data management capabilities. While almost 70% of respondents indicated that their firms have a centralized data warehouse, there is no clear consensus on who would have access to this data warehouse. Survey responses are nearly evenly split across the answer choices of the entire firm, all traders, portfolio managers, and quant research/analytics functions. Our conversations indicated that many were grappling with how to effectively catalog data on a continuous basis and provide the right people with access at the right time.

### *Talent and governance: Fiercely competitive and firm-specific*

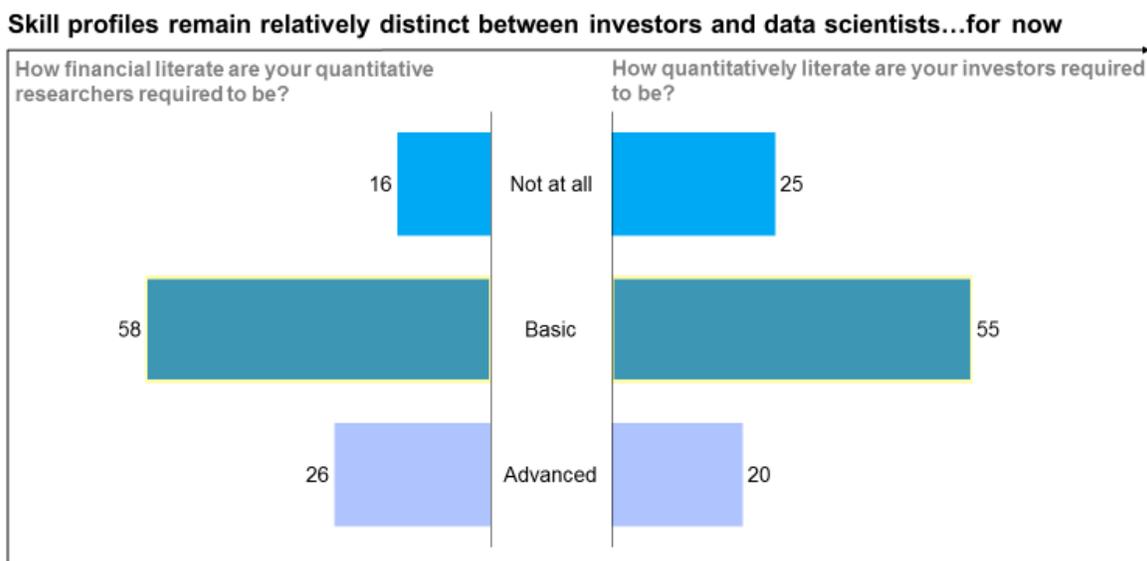
There was a spectrum of approaches towards talent acquisition, organization, and governance reported by our respondents, driven as much by philosophy and culture as anything else. Despite specific differences, there was the general recognition that collaboration between functions is increasingly vital to realizing value from data and analytics in a cost-effective and low-risk manner. In most cases, it takes an orchestra of investors, technologists, data scientists, risk managers, HR professionals, and legal experts to sustain a well-functioning data and analytics operating model.

A critical element in this effort is the ability to hire and retain high-quality data and technology talent. Almost without exception, we heard that competition for this talent is increasingly fierce, and not limited to financial services. Many firms are looking beyond rival firms to Big Tech for the best talent—and for a template for creating the kind of environment where this new talent can flourish.

The firms that are succeeding in attracting and keeping high-quality talent have taken several different steps. Some have built quant and technology communities, creating opportunities for employees to pursue research projects that might not always be directly tied to alpha generation. A second approach has been to offer clear career trajectories, showing diverse development pathways that fit a variety of interests (e.g., roles that marry technical requirements with greater industry context versus those that are purely functional). A few firms have successfully maintained links to the broader academic community (e.g., presenting papers, open-sourcing architecture), which helps with retention of existing talent and access to new talent (e.g., through faculty referrals).

Although investing and data science are becoming more and more intertwined, the talent profiles remain relatively distinct. Most firms still do not require traders to have advanced knowledge of descriptive statistics or predictive analytics, with 55% of respondents indicating that traders only need to know basic descriptive statistics (Exhibit 5). Similarly, most firms do not require their quantitative researchers or data scientists to have advanced financial literacy, with 58% of respondents indicating that quantitative researchers only need basic knowledge of trading and firm strategy. There has been convergence between the two capabilities, most observably within systematic investing, and the prevailing sentiment among those interviewed was that this will extend to discretionary investing, certainly at more junior levels to begin with and perhaps beyond that.

### Exhibit 5



While over two-thirds of respondents have centralized data and analytics groups that perform data acquisition and management, the degree to which specific analytics talent is centralized varies greatly by firm culture, size and strategy. Sixty percent of respondents said their firm houses their quantitative researchers centrally; most of the rest housed them by vehicle or strategy, while only a few placed them in individual desks. Often this choice was the result of the

degree to which investment teams themselves collaborated. In cases where investment team IP is closely guarded and internal competition between investment groups is high, technical talent tends to be embedded with investors. Some respondents acknowledged and were comfortable with the fact that this could lead to duplication or inefficiency (e.g., in data procurement and management).

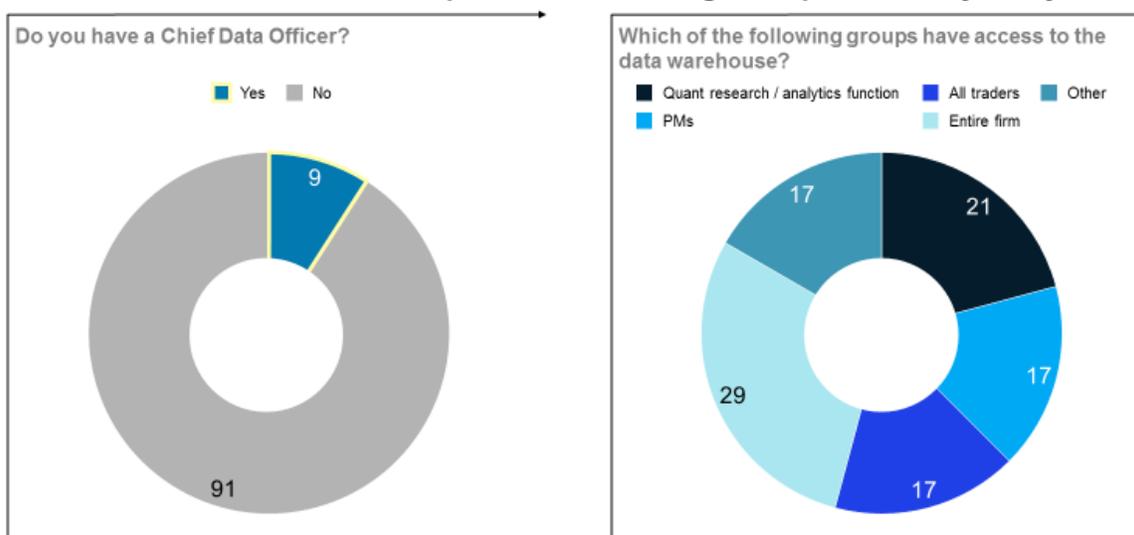
On the flip side, respondents at firms with generally more collaborative cultures extolled both the efficiency, best-practice sharing and community-building virtues of having technical talent located centrally. Unlike some of their federated peers, these firms tended not to focus on or struggle with attribution of credit for alpha between the investment and the technical teams. In these cases, there is a big opportunity for firms to take more advantage of insights that are directly applicable to one part of the firm but may have indirect value to another—for instance, cross-pollination of data and insights relevant across both Real Estate and Commercial Mortgage-Backed Security teams.

Another cause of differences in centralization was investment strategy. Eighty percent of systematic investors reported taking a decentralized approach, with researchers aligned to specific strategies or desks. For discretionary investors, there was a more even split. Firm size is also a determinant of this choice, with most smaller firms opting for a central team for practical reasons.

Many firms approach the governance of this capability by committee—with a cross-functional group of business and technology leaders helping set the use case agenda and budgetary guidelines (e.g., for data acquisition). From respondents where the capability rolled up to technology, we heard that it is crucial for the CTO to be fully involved in decision-making related to the investment function. Notably, 90% of respondents said their firm did not have a formal Chief Data Officer role, again reflecting the nascency of data management relative to other industries (Exhibit 6).

## Exhibit 6

### Chief Data Officers are not commonplace and data management practices vary widely



## The journey from here

Firms participating in this research spanned maturity levels. Some are “quant natives,” for whom advanced data ingestion and analysis is a core part of their DNA; others are just starting on their data and analytics journey. The questions and reflections that they had were therefore appropriately distinct. However, we heard a few common themes from firms at each maturity stage.

**Early-maturity firms are focused on laying foundations.** Respondents told us that the key for firms just starting on their data journey was to lay a set of solid foundations. First, hiring enough experienced data scientists to have a “full quiver” of analytical methods is considered critical for ensuring that data strategies are deployed in line with broader investment strategies. Second, respondents emphasized the importance of building a clear roadmap for data acquisition (e.g., principles for back-testing data), which helps to avoid endless mining of low-value data sets. Finally, they said, building out the architecture for storing and structuring data often starts by “working backward” from specific use cases, focusing on specific datasets or even fields that enable these use cases—an approach that can reduce wasted effort and build organizational buy-in.

**Emerging-maturity firms are focused on democratizing datasets.** Respondents at firms that are starting to mature along the data journey laid out a slightly different set of high-impact actions. Primarily, firms at this stage are focused on defining and implementing robust data storage, structuring, and governance policies to ensure that data is both accessible and discoverable, while complying with security standards. Often, this involves establishing processes to ingest new data into the cloud and creating logical, exploitable catalogs investors can use to find relevant datasets.

**Mature firms begin to diverge.** Respondents at firms with high-functioning data infrastructure diverge in where they see the most value at the next stage of their journey. Some firms told us that they were most excited about identifying proprietary sources of data; they were particularly interested in exploiting privileged data or creating their own data (e.g., data generated by owned assets like real estate or infrastructure; internal HR / employee email data). Others opined that data access will ultimately be commoditized and saw their competitive edge in data *ingestion*, quickly onboarding, storing, cataloging, and structuring datasets at a scale that the rest of the field could not—and thus are investing heavily in data engineering. Still others—though the smallest minority for now—believe that data *science* will be the ultimate differentiator, and thus spend heavily on research and development, building new quantitative approaches to exploiting datasets.

From the energy with which participants engaged in this process, it was clear to us that the topic is an important one for most firms—and that it will likely become even more so over the coming years. As the ideas outlined in this perspective evolve and the external environment changes, we invite you to continue the conversation—both with us and, more importantly, with your peer firms.



*Onur Erzan is a Senior Partner in McKinsey & Company's New York Office, where Sudeep Doshi is an Associate Partner, and Andrew Min and Eddy Lee are consultants.*



The Managed Funds Association is  
the trade association of the global alternative investment industry.  
We represent the interests of our members – professionals in  
managed futures and hedge funds – and their investors.  
MFA provides educational resources for industry regulators,  
policy makers, professionals, investors and the media.

LEARN MORE AT [WWW.MANAGEDFUNDS.ORG](http://WWW.MANAGEDFUNDS.ORG)



**MANAGED FUNDS  
ASSOCIATION**

*546 Fifth Avenue - 12th Floor, New York, NY, 10036  
600 14th Street, N.W. Suite 900, Washington, DC 20005  
[www.managedfunds.org](http://www.managedfunds.org)*