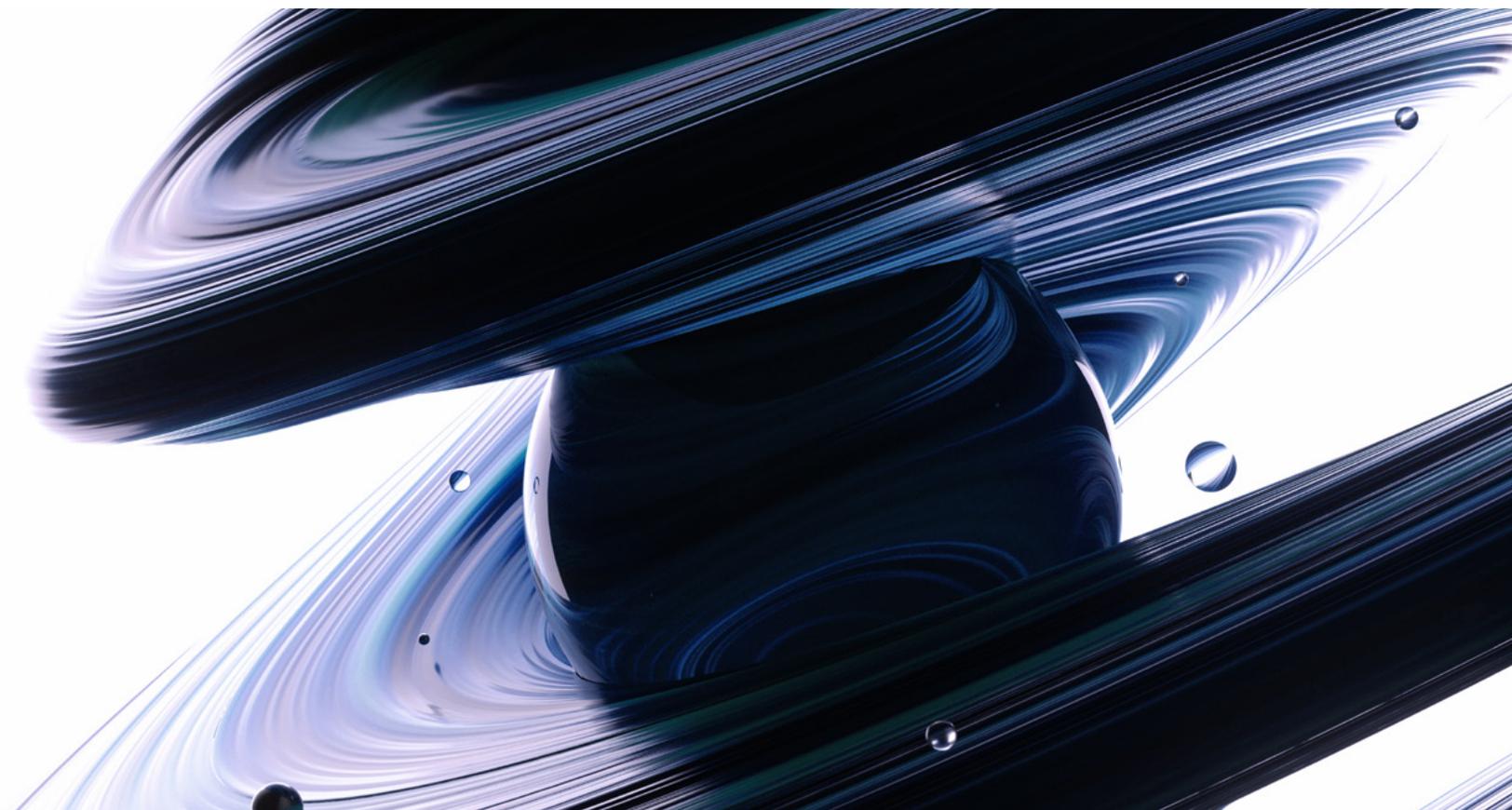


Harnessing the power of external data

Few organizations take full advantage of data generated outside their walls. A well-structured plan for using external data can provide a competitive edge.

By Mohammed Aaser and Doug McElhaney



Many companies have made great strides in collecting and utilizing data from their own activities. So far, though, comparatively few have realized the full potential of linking internal data with data provided by third parties, vendors, or public data sources. Overlooking such external data is a missed opportunity. Organizations that stay abreast of the expanding external-data ecosystem and successfully integrate a broad spectrum of external data into their operations can outperform other companies by unlocking improvements in growth, productivity, and risk management.

The COVID-19 crisis provides an example of just how relevant external data can be. In a few short months, consumer purchasing habits, activities, and digital behavior changed dramatically, making preexisting consumer research, forecasts, and predictive models obsolete. Moreover, as organizations scrambled to understand these changing patterns, they discovered little of use in their internal data. Meanwhile, a wealth of external data could—and still can—help organizations plan and respond at a granular level.

Although external-data sources offer immense potential, they also present several practical challenges. To start, simply gaining a basic understanding of what's available requires considerable effort, given that the external-data environment is fragmented and expanding quickly. Thousands of data products can be obtained through a multitude of channels—including data brokers, data aggregators, and analytics platforms—and the number grows every day. Analyzing the quality and economic value of data products also can be difficult. Moreover, efficient usage and operationalization of external data may require updates to the organization's existing data environment, including changes to systems and infrastructure. Companies also need to remain cognizant of privacy concerns and consumer scrutiny when they use some types of external data.

These challenges are considerable but surmountable. This article discusses the benefits of tapping external-data sources, illustrated through a variety of examples, and lays out best practices

for getting started. These include establishing an external-data strategy team and developing relationships with data brokers and marketplace partners. Company leaders, such as the executive sponsor of a data effort and a chief data and analytics officer, and their data-focused teams should also learn how to rigorously evaluate and test external data before using and operationalizing the data at scale.

External-data success stories

Companies across industries have begun successfully using external data from a variety of sources (Exhibit 1). The investment community is a pioneer in this space. To predict outcomes and generate investment returns, analysts and data scientists in investment firms have gathered “alternative data” from a variety of licensed and public data sources, many of which draw from the “digital exhaust” of a growing number of technology companies and the public web. Investment firms have established teams that assess hundreds of these data sources and providers and then test their effectiveness in investment decisions.

A broad range of data sources are used, and these inform investment decisions in a variety of ways:

- Investors actively gather job postings, company reviews posted by employees, employee-turnover data from professional networking and career websites, and patent filings to understand company strategy and predict financial performance and organizational growth.
- Analysts use aggregated transaction data from card processors and digital-receipt data to understand the volume of purchases by consumers, both online and offline, and to identify which products are increasing in share. This gives them a better understanding of whether traffic is declining or growing, as well as insights into cross-shopping behaviors.

Exhibit 1

Companies can obtain data from many types of external sources.

Geospatial and satellite	Point of interest Footfall	Real estate/ property	Weather	Temperature and precipitation Storms and adverse events Forecasts	
Private business	Revenues Head counts Locations	Industry classifications Technographics	News, IP, and legal	News-feed services Research-journal feeds	Patent filings Legal actions
Consumer	Transactions Consumer panels	Search trends data Census data	Public data	Federal filings State and local filings Macroeconomic indicators	
Web-harvested, online and app	Online reviews Job and product listings	Web-traffic data Digital-app metrics	Industry specific	Trade flows and shipping Hotel and travel bookings	Healthcare claims Agriculture/crops

- Investors study app downloads and digital activity to understand how consumer preferences are changing and how effective an organization’s digital strategy is relative to that of its peers. For instance, app downloads, activity, and rating data can provide a window into the success rates of the myriad of live-streaming exercise offerings that have become available over the last year.

Corporations have also started to explore how they can derive more value from external data (Exhibit 2). For example, a large insurer transformed its core processes, including underwriting, by expanding its use of external-data sources from a handful to more than 40 in the span of two years. The effort involved was considerable; it required prioritization from senior leadership, dedicated resources, and a systematic approach to testing and applying new data sources. The hard work paid off, increasing the predictive power of core models by more than 20 percent and dramatically reducing application complexity by allowing the insurer to eliminate many of the questions it typically included on customer applications.

Three steps to creating value with external data

Use of external data has the potential to be game changing across a variety of business functions and sectors. The journey toward successfully using external data has three key steps.

1. Establish a dedicated team for external-data sourcing

To get started, organizations should establish a dedicated data-sourcing team. In our experience, a key role on this team is a dedicated data scout or strategist who partners with the data-analytics team and business functions to identify operational, cost, and growth improvements that could be powered by external data. This person also would be responsible for building excitement around what can be made possible through the use of external data, planning the use cases to focus on, identifying and prioritizing data sources for investigation, and measuring the value generated through use of external data. Ideal candidates for this role are individuals who have served as analytics translators and who have

Exhibit 2

External data can help companies create value in several key areas.

Customer analytics		Identify ideal B2B prospects and look-a-likes by leveraging firmographics, employment growth, technographics, retirement-plan investments, etc Identify fast-growing consumer trends and marketing opportunities by utilizing search data, social-media analysis, transaction panels, and receipt panels
Strategic analysis		Benchmark organizational talent against peers by analyzing existing talent profiles and job postings Identify product-improvement opportunities by analyzing reviews across social-media and e-commerce platforms
Operations and forecasting		Predict how real-estate prices will change , based on local market characteristics and demographic shifts Forecast which customer segments will grow , using firmographics, technographics, and private-company data
Risk management		Reduce operational risks , based on real-time analysis of news and social-media data for raw-material suppliers Reduce supplier and reputational risks by understanding parent-subsidiary relationships, ownership, news, and legal proceedings

experience in deploying analytics use cases and in working with technology, business, and analytics profiles.

The other team members, who should be drawn from across functions, would include purchasing experts, data engineers, data scientists and analysts, technology experts, and data-review-board members (Exhibit 3). These team members typically spend only part of their time supporting the data-sourcing effort. For example, the data analysts and data scientists may already be supporting data cleaning and modeling for a specific use case and help the sourcing work stream by applying the external data to assess its value. The purchasing expert, already well versed in managing contracts, will build specialization on data-specific licensing approaches to support those efforts.

Throughout the process of finding and using external data, companies must keep in mind privacy concerns and consumer scrutiny, making

data-review roles essential peripheral team members. Data reviewers, who typically include legal, risk, and business leaders, should thoroughly vet new consumer data sets—for example, financial transactions, employment data, and cell-phone data indicating when and where people have entered retail locations. The vetting process should ensure that all data were collected with appropriate permissions and will be used in a way that abides by relevant data-privacy laws and passes muster with consumers.

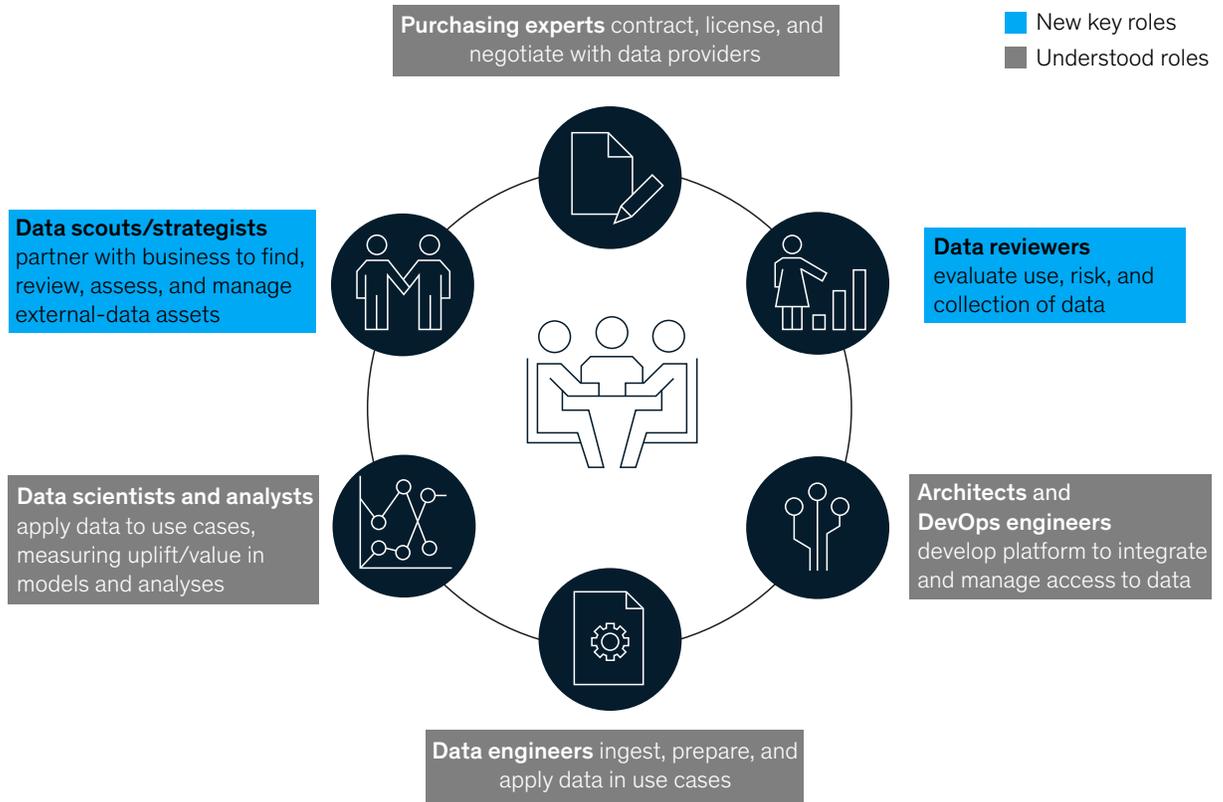
This team will need a budget to procure small exploratory data sets, establish relationships with data marketplaces (such as by purchasing trial licenses), and pay for technology requirements (such as expanded data storage).

2. Develop relationships with data marketplaces and aggregators

While online searches may appear to be an easy way for data-sourcing teams to find individual data sets, that approach is not necessarily the

Exhibit 3

An effective data-sourcing team combines six roles, including data scouts and data reviewers.



most effective. It generally leads to a series of time-consuming vendor-by-vendor discussions and negotiations. The process of developing relationships with a vendor, procuring sample data, and negotiating trial agreements often takes months.

A more effective strategy involves using data-marketplace and -aggregation platforms that specialize in building relationships with hundreds of data sources, often in specific data domains—for example, consumer, real-estate, government, or company data. These relationships can give organizations ready access to the broader data ecosystem through an intuitive search-oriented platform, allowing organizations to rapidly test dozens or even hundreds of data sets under the

auspices of a single contract and negotiation. Since these external-data distributors have already profiled many data sources, they can be valuable thought partners and can often save an external-data team significant time. When needed, these data distributors can also help identify valuable data products and act as the broker to procure the data.

Once the team has identified a potential data set, the team's data engineers should work directly with business stakeholders and data scientists to evaluate the data and determine the degree to which the data will improve business outcomes. To do so, data teams establish evaluation criteria, assessing data across a variety of factors to determine whether the data set has the necessary characteristics for delivering valuable insights (Exhibit 4).

Exhibit 4

A thorough evaluation of external data explores criteria in ten areas.



Depth and breadth of data
 Does the data set have relevant data elements for its main use cases?
 Are there additional fields that could be useful or apply to other areas of the business?



Match/fill rates
 Are the match rates high enough to justify a return on investment?



Data profile
 Does the data profile describe the distribution of the data, frequencies by column, missing variables, and changes in variables and panel over time?
 How does the data set compare with ground truth?



Coverage/panel overview
 Does the data set cover the right geography and population?
 Does the way the data are sourced introduce any bias in the data, relative to the use case?



Timeliness of data
 How far back in time is the data set reliable?
 Is history needed?
 How frequently are the data updated?
 What is the delay?



Data delivery
 How are the data delivered?
 Are application programming interfaces (APIs) available? What is the format?
 How do you know the data have transferred successfully?



Potential impact/lift
 Can the proposed or planned analysis be completed?
 Is there an evaluation data set to model against?
 What impact does the data generate?



Total cost
 What is the total cost for the data in evaluation and production?
 Do you need engineering resources to make the data accessible?



Procurement and contracting
 How quickly will you be able to onboard the data set when needed?
 Who else needs to review the agreement and technology?
 What representations, warranties, etc do you need from the vendor?



Risk
 Is the data set sourced appropriately and ethically?
 Is the data set being used appropriately?
 Is there reputational risk?

Data assessments should include an examination of quality indicators, such as fill rates, coverage, bias, and profiling metrics, within the context of the use case. For example, a transaction data provider may claim to have hundreds of millions of transactions that help illuminate consumer trends. However, if the data include only transactions made by millennial consumers, the data set will not be useful to a company seeking to understand broader, generation-agnostic consumer trends.

3. Prepare the data architecture for new external-data streams

Generating a positive return on investment from external data calls for up-front planning, a flexible data architecture, and ongoing quality-assurance testing.

Up-front planning starts with an assessment of the existing data environment to determine how it can support ingestion, storage, integration,

governance, and use of the data. The assessment covers issues such as how frequently the data come in, the amount of data, how data must be secured, and how external data will be integrated with internal data. This will provide insights about any necessary modifications to the data architecture.¹

Modifications should be designed to ensure that the data architecture is flexible enough to support the integration of a continuous “conveyor belt” of incoming data from a variety of data sources—for example, by enabling application-programming-interface (API) calls from external sources along with entity-resolution capabilities to intelligently link the external data to internal data. In other cases, it may require tooling to support large-scale data ingestion, querying, and analysis. Data architecture and underlying systems can be updated over time as needs mature and evolve.

The final process in this step is ensuring an appropriate and consistent level of quality by constantly monitoring the data used. This involves examining data regularly against the established

quality framework to identify whether the source data have changed and to understand the drivers of any changes (for example, schema updates, expansion of data products, change in underlying data sources). If the changes are significant, algorithmic models leveraging the data may need to be retrained or even rebuilt.

Minimizing risk and creating value with external data will require a unique mix of creative problem solving, organizational capability building, and laser-focused execution. That said, business leaders who demonstrate the achievements possible with external data can capture the imagination of the broader leadership team and build excitement for scaling beyond early pilots and tests. An effective route is to begin with a small team that is focused on using external data to solve a well-defined problem and then use that success to generate momentum for expanding external-data efforts across the organization.

Mohammed Aaser is McKinsey’s chief data officer, based in the Minneapolis office, and **Doug McElhaney** is a partner in the Washington, DC, office.

The authors wish to thank Kelly Brennan and Aqil Dattoo for their contributions to this article.

Copyright © 2021 McKinsey & Company. All rights reserved.

¹ For more, see Antonio Castro, Jorge Machado, Matthias Roggendorf, and Henning Soller, “How to build a data architecture to drive innovation—today and tomorrow,” June 2020, McKinsey.com.