

TECHNOLOGY

## Harnessing Alternative Data for Competitive Advantage

by Anil K. Gupta, Jon Norberg, Evan Schnidman, and Kai Wu



Image Credit | Markus Spiske

*Companies now have access to a tsunami of non-traditional data for creating and capturing value.*

✔ INSIGHT | FRONTIER 07 Nov 2022

---

A data tsunami is headed our way, with new types of data emerging at a staggering pace.<sup>1</sup> Even simple devices such as EZPass transponders used by several U.S. states collect data on over 10 million toll transactions per day recording the time, location, vehicle ID, and its image. According to the National Oceanic and Atmospheric Administration, climate related data will grow from 83 petabytes to over 650 petabytes in ten years. And, on YouTube, users upload over 2000 petabytes of content each year.

#### RELATED CMR ARTICLES

**“How to Appropriate Value from General-Purpose Technology by Applying Open Innovation”** by Jialei Yang, Henry Chesbrough, & Pia Hurmelinna-Laukkanen. (Vol. 64/3) 2022.

**“How to Compete When Industries Digitize and Collide: An Ecosystem Development Framework”** by Michael G. Jacobides. (Vol. 64/3) 2022.

While companies, investors, and governments have used *traditional data* such as market research, financial statements, stock price movements, and census reports for decades, the new data tsunami owes its origins to the rise of *alternative data* and the power of AI technologies such as computer vision and natural language processing to make such data usable. Examples include data from satellite imagery, weather tracking, smartphone geolocation, ratings and reviews, Internet search, job listings, social media, IoT sensors, and lots more.<sup>2</sup>

Quantitative hedge funds were the earliest adopters of alternative data (or alt-data) over 20 years ago. Today, over half of all hedge funds—quant and fundamental—consume various types of alt-data to garner an advantage in the quest for alpha.<sup>3</sup> More recently, alt-data usage has started spreading to private equity and venture capital firms, who find such data useful for identification of investment targets and deeper and more comprehensive due diligence.<sup>4</sup> Since most such targets are unlisted firms, alt-data can be especially potent.

With the growing diffusion of AI technologies, alt-data is now making its way into the non-financial corporate sector. We offer below guidelines regarding how such companies can harness alt-data for competitive advantage.

## How Alt-Data Can Add Strategic Value

The value of alt-data lies in enabling a more comprehensive, more granular, and more timely visibility into the factors that drive strategic decisions. In the process, alt-data can enable analysis of relevant factors from unique and unconventional lenses. Consider some examples.

A food products company wants to develop early insights into emerging consumer preferences to guide new product development. Traditional market research, while useful to assess various aspects of current products, is miserable at uncovering ideas for new products. Analysis of social media posts can be extremely insightful in this regard. At the other end of the value chain, let's say the company wants to predict the prices of various commodities more accurately. Satellite imagery (on acreage under cultivation, where the crop is being cultivated, and crop conditions) combined with weather predictions can yield a more fine-grained and real-time assessment of supply prospects than monthly data from government agencies.

A retail chain wants to assess and improve the performance of its stores. What matters is not just what each store sells but also the store's ability to convert those entering the store into buyers. A store with higher sales but lower conversion may well be cause for greater concern than one with lower sales but higher conversion. Smartphone geolocation data can give the company information on how many people enter each store, vital information to compute conversion rates. Footfall data can also be useful for monitoring the attractiveness of every competitor's every store and more broadly the differing and changing attractiveness of locations.

Alt-data serve as powerful complements to the company's internal data. Consider an auto manufacturer keen to make its supply chain more responsive. Instead of relying solely on internal data, it can work with say Google to build a digital twin of the supply chain. The

digital twin can automatically incorporate real-time external data on looming natural disasters, labor strife, political upheaval, and other phenomena. As a result, the company can sense and respond to supply shocks far more effectively than would be possible through internal data alone.

As these examples suggest, using alt-data to guide decisions offers several advantages. *First*, such data can be collected with much greater frequency and thus have more real-time relevance. *Second*, they can be collected and analyzed at much finer levels of granularity (specific stores, micro customer segments, specific crop plantings, specific supplier locations). *Third*, by integrating complementary datasets, decision-makers can move closer to a 360-degree understanding of targeted phenomena.

## **Business Leaders as Drivers**

Although data professionals can be expected to know what types of alt-data exist, where they can be sourced, and how best to use them, business leaders are the ones responsible for strategic decisions and must therefore play the lead role in converting alt-data into competitive advantage.

Business leaders can be expected to know the most important business questions where deeper data-driven insights could add the most value: Could this data help in discovering new product or market opportunities? Or, in figuring out how best to retain existing customers? Or, in attracting and retaining key talent? Or, to provide insight on production and supply chain uncertainties? Business leaders with P&L responsibilities know these questions intimately and must therefore be the prime movers in assessing the potential value of alt-data. If they are not in the driver's seat, they are less likely to view potential insights from alt-data as trustworthy or meaningful.

For most organizations, a federated structure is likely to be the most appropriate. Under this structure, data scientists are embedded within operating units. Proximity ensures that data scientists will have frequent interaction with their business counterparts and see data needs from the lens of business problems. Yet, they would also be part of a centrally managed data science group. Such quasi-centralization helps ensure critical mass

necessary to recruit, retain, and develop scarce technical talent. It also facilitates the development of a standardized approach for data management, reduces the likelihood of duplication, enables the integration of complementary data across units, and ensures continued buy-in from business leaders.

## **Raw Chocolate or Bonbons**

Any company keen on using alt-data must decide whether to buy such data in raw form from the original data collector or to work with a value-added intermediary. As a marketing executive in a consumer goods company put it more vividly: “Do you want to buy raw chocolate or are you better off buying bonbons?”

Most companies are likely to be better off working with value-added intermediaries. There are already over 2000 different types of alt-data available across dozens of categories.<sup>5</sup> Most companies that collect such data are still young and have neither the skills nor the resources to develop tailored services for specific verticals. This is the forte of consulting firms, who are also more geared to helping clients address their specific needs, instead of just selling more data.

Given the burgeoning supply of alt-datasets, companies should lean towards a needs-based approach for acquiring data and avoid succumbing to a just-in-case syndrome. Buying the data and deploying staff to make sense of it can be costly. And too much data can be just as unhelpful as too little data. In scouting for relevant data, it is also important to factor in marginal utility. Will the new dataset provide additional insights beyond what you can already deduce from existing data sources? Alternatively, will the new dataset be a superior and/or less expensive substitute for one or more existing datasets?

## **Screening for Data Quality**

The adage “garbage in, garbage out” captures well the crucial importance of data quality. Given its newness, this caution is especially relevant for alt-data. The quality of alt-data depends on several factors.

*Provenance.* Was the data collected legally and does the data seller own the rights to license it to others? You should also check that the sharing of data does not violate legal requirements and ethical considerations pertaining to privacy, including regulations such as the EU's General Data Protection Regulation (GDPR) and California's Privacy Rights Act (CPRA).

*Coverage.* Considerations here include whether the dataset adequately covers the entities you want to observe. How many companies in the relevant industry are included in the dataset? What proportion of the stores are covered in the footfall data? What proportion of acreage under cultivation is covered in satellite imagery?

*Accuracy.* Inaccuracies can creep in due to both intentional and unintentional reasons. For example, in a dataset from an employment data aggregator, could there be pollution from resume collection scams? In the data exhaust from a social media site, what proportion of the supposed users are bots rather than humans?

*Consistency.* Three types of consistencies matter. First, is there consistency in formats and labels? Simple inconsistencies such as how time is expressed (time zones, time format, etc.) and how names are written (last name first, or first name first) can dramatically reduce the value of the data. Second, do different cases in the dataset refer to the same phenomenon? For example, parking data from the entrance to a garage are different from data from roadside parking meters. Third, is there compatibility from one time period to another? Such compatibility is crucial if you are interested in changes over time, e.g., changes in consumer sentiment.

*Freshness.* The more current the data, the more useful it is both to understand what happened as well as to make decisions for the future. Freshness also refers to the frequency with which the data is updated. Since fresher data is also more expensive, data buyers need to check that the value from incremental freshness justifies the incremental cost.

*Reference Data.* Companies often find it important to integrate multiple datasets. This may be impossible without an identifier common to each dataset. For data pertaining to companies, the identifier could be the company name or the ticker symbol. For data on

retail footfalls, the identifier could be longitude and latitude coordinates.

## Converting Alt-Data Into Useful Insights

While business leaders must remain the strategic drivers, data scientists play extremely important roles in helping to access and convert alt-data into useful insights. These include:

- Specifying the most effective and efficient mechanism for the company to access the data from the data provider (via application program interfaces or APIs, cloud-based storage such as AWS S3, or file transfer protocols or FTP);
- Using machine learning to convert unstructured into structured data;
- Normalizing the data, e.g., elimination of redundancies, and establishment of relationships across tables; and
- Cleaning the data, which would include making sure that the reference variables critical for integration of multiple datasets are common across the datasets.

Once a dataset is ready for use, it is the business questions that come to the fore and will determine which internal and external datasets need to be integrated and how they should be analyzed to yield the desired insights.

Learning to identify, access, and utilize alt-data can also yield important second-order benefits. Since new types of alt-data keep emerging, companies that start using alt-data today will become smarter at using new types of alt-data tomorrow. As well, a better understanding of the value of alt-data will sensitize the company to potential opportunities to commercialize internal data. Some large retailers are already earning over \$100 million annually from external monetization.

Given today's reality, it is not an exaggeration to suggest that, for most medium to large companies, alternative data can no longer remain an afterthought. It needs to become mainstream.

# Endnotes

1. According to IDC estimates, 64 zettabytes (i.e., 64 billion terabytes) of data was created or replicated in 2020 and, after accounting for deletions, over 1 zettabyte of new data was carried over from 2020 to 2021. See <https://www.idc.com/getdoc.jsp?containerId=prUS47560321>.
  2. For more details on various types of alternative data, see <https://alternativedata.org>.
  3. For more details, see <https://www.eaglealpha.com/>.
  4. For some examples, see <https://firstoinvest.com/alternative-data-for-private-equity-funds/>.
  5. See <https://alternativedata.org>.
- 



Anil K. Gupta [Follow](#)

Anil K. Gupta is Michael Dingman Chair and Professor of Strategy at the Smith School of Business, The University of Maryland; he is also a member of the World Economic Forum's Global Expert Network.



Jon Norberg

Jon Norberg is Chief Strategy Officer of LinkUp, an alternative data company, and has extensive experience in institutional investment organizations.



Evan Schnidman

[Follow](#)

Evan A. Schnidman is founder of EAS Innovation; he is also founder of Prattle, a company that uses natural language processing to provide investment managers with comprehensive metrics evaluating potentially market-moving communications by central banks and corporations.



Kai Wu

[Follow](#)

Kai Wu is founder and CEO at Sparkline Capital, an investment management firm that applies machine learning models to uncover alpha in large, unstructured data sets.