



speed-centric perspective on alt-data's value is pervasive and neatly captured by the tagline of a leading alt-data platform operator: "Alternative data is untapped alpha."<sup>3</sup>

We argue that alt-data's core value proposition is, however, meaningfully different for Investors than that slogan would suggest. Investors (as defined earlier) have a distinct comparative advantage over other market participants: patience. Because of their long operating horizons, Investors can pursue investment strategies unavailable to other market players. This comparative advantage is more aligned with defensive and defensible approaches to alt-data than it is with the exploitative strategies that short-horizon investors tend to pursue. That is, Investors will likely be better off using alt-data in ways that are unharmed by competition over alt-data (i.e., nonrivalrous) or for activities others cannot easily replicate (i.e., excludable).<sup>4</sup> In rethinking how alt-data will be most valuable to long-term strategies, we contend that Investors must also rethink how they evaluate and characterize alt-data, along with whom they should partner in gaining access to alt-datasets.

Rethinking these three issues could guide Investors in selecting alt-datasets, and strategies for analyzing and acting on them, that better fit with their organizational contexts. We seek to help Investors re-examine how alt-data could best serve their needs and offer recommendations that are informed by both formal empirical findings and our own close interactions with Investors. We also explore examples of how alt-data can be creatively used in defensive or defensible strategies.

Although building capacity around alt-data is strategically valuable in its own right, doing so has the added benefit of promoting innovation. Using alt-data demands (almost by definition) that Investors depart from the status quo in their decision making. As such, thoughtful design of an alt-data program can drive innovation in all aspects of an Investor's business (e.g., creative improvements in processes, people's skill sets,

and technology). Finding partnerships that facilitate, rather than forfeit, opportunities to innovate and learn from alt-data is therefore a key issue we address and one that is likely to materially affect Investors' success (with alt-data and beyond).

The rest of this article is organized as follows. We first make the case that Investors are better off designing their alt-data strategies around defensive and defensible approaches to using alt-data than aiming to use it for alpha-oriented, opportunistic purposes. We provide examples of creative uses of alt-datasets under these strategies. These examples emphasize how alt-data can be used for deeper understanding of risk and generating operational alpha. We then cover why existing systems for characterizing alt-datasets do not fit Investors' needs. We consider a replacement system that could improve the appraisal of alt-datasets in terms of how well their characteristics align with an Investor's specific objectives and capabilities. Next, we distill our empirical findings about Investors' organizational attitudes on, and capacities for, alt-data. Our analysis concludes that Investors will generally need to partner for access to alt-data and to realize efficiencies in organizing and (pre-)processing alt-datasets. We detail the benefits and costs of partnering with different types of entities and remark on how opportunities for innovation may be a core consideration in selecting alt-data partners. We then describe how the growing arms race around alt-data could affect Investors. Finally, we close by summarizing our findings and highlight additional facets of alt-data strategies that Investors might wish to rethink in the future.

## RETHINKING ALT-DATA'S VALUE PROPOSITION

Although alt-data have garnered increased attention in recent years, their use in finance is not new. Alt-data have played an integral role in investing ever since humans first began keeping records of trade: They deepen the connections between financial valuations and real-world sources of value. For instance, some enterprising merchants in ancient Babylon used measurements of the Euphrates' depth and flow to gain an informational edge in trading various commodities (because they realized that these variables were correlated with market supply) (Lo and Hasanhodzic 2010).

What has recently changed about alt-data's role in finance is its degree of accessibility. Perhaps the most

---

<sup>3</sup>See: <https://www.quandl.com/alternative-data>.

<sup>4</sup>An example of a nonrivalrous application of alt-data is in screening public equities based on sustainability criteria for the underlying companies. An example of excludable alt-data use would be for due diligence on direct investments in startup companies to which an Investor has privileged access (e.g., a university endowment having first access to funding spinouts from its research laboratories). In this sense, an Investor benefits not from an alt-dataset being excludable but from its own ability to use that data being an excludable (i.e., not easily repeated or imitated) capability.

recited example of alt-data in finance is hedge funds counting cars in retailers' parking lots (which supposedly is a leading indicator of sales performance). In the past, such counts had to be made manually, with analysts physically located in or near cars they tracked. Apart from a small number of well-resourced hedge funds, few financial organizations could devote sufficient resources to such a narrow endeavor. Currently, however, these data are accessible through a subscription service to any investment organization inclined to purchase it (thanks to lower costs of satellite imagery).

More generally, the number and diversity of alt-data sources that are readily accessible to financial entities has mushroomed. The tally of large-scale alt-data vendors who specifically cater to investment organizations has gone from a few dozen to several hundred in less than half a decade.<sup>5</sup> The total alt-data sources potentially relevant to investment decision making that can be cheaply and easily accessed is in the many millions. Furthermore, tools for acquiring and processing these plentiful datasets are increasingly user friendly.<sup>6</sup> Alt-data are steadily becoming mainstream.

As a result, the rate at which any one type or source of alt-data becomes conventional—and therefore ceases to be alt-data—is likely to increase. If the value of alt-data is premised on their conferring advantages in faster exploitation of trading opportunities (as is the case for many financial-market participants), then this means the value of any given alt-dataset will probably deteriorate at an accelerating rate because both alt-data and their value are relatively determined. Notice that data may qualify as alternative at any of three levels: the firm, the industry, and the financial ecosystem as a whole. For example, a dataset may be unconventional for a given hedge fund, but not for other funds in the hedge-fund industry. Likewise, some data may be conventional for a given firm, yet be unconventional for most organizations in the wider financial system. When enough

---

<sup>5</sup> Here, we make a meaningful distinction between providers or sources of alt-data (*point vendors*) and alt-data access providers (*platform vendors*). Later, we discuss why this distinction is relevant. For now, we simply note that the number of alt-data vendors vastly exceeds the number of platforms, and this gap is only likely to widen in the future.

<sup>6</sup> These tools may be standalone (e.g., <http://scikit-learn.org/stable/>) or part of the suite of offerings from an alt-data platform (i.e., an entity that offers not just alt-datasets but also additional support or tools for working with them).

organizations make use of any alt-dataset, it stops being alternative at a system-wide level.

Similarly, two relative dimensions help determine the value of any alt-dataset: *rivalry* and *excludability*.<sup>7</sup> Rivalry is the extent to which one entity's use of a resource diminishes its value for another entity.<sup>8</sup> Excludability is the degree to which one entity can prevent another from using a resource. When alt-data's value is premised on allowing market players to better exploit trading opportunities, then alt-datasets will tend to exhibit high rivalry. Moreover, rising accessibility of many alt-datasets is tending to lead to lower excludability.<sup>9</sup> These trends suggest the shelf lives for alt-datasets may be shortening if their value comes solely from helping to exploit opportunities.<sup>10</sup>

### Defensive and Defensible Value

When an alt-dataset's value is premised on it improving a market participant's ability to speedily seize trading opportunities, there is an embedded assumption that the participant will need to act quicker than others to realize that value. This value proposition for alt-data implies that alt-datasets should be more useful for financial organizations with comparative advantages in rapid execution.

Speed is, in general, not a comparative advantage for Investors, and for sound reasons: They are long-lived organizations whose success is mission critical for their beneficiaries. Building an investment strategy around speed can greatly increase the risk of

---

<sup>7</sup> The dimensions of rivalry and excludability are conventionally used to classify economics goods as private, public, club, or common pool. For such purposes, rivalry and excludability are usually treated as binary categories (i.e., something is either rivalrous or nonrivalrous and excludable or not). We see them here as continuous properties.

<sup>8</sup> Rivalry is a *congestion effect*, which is the opposite of a *network effect* (i.e., a resource's value grows with popularity).

<sup>9</sup> This decreasing excludability may become more prevalent as methods for dataset emulation and replication (e.g., statistically synthesizing better proxy datasets) techniques improve. Likewise, the bigger the market for alt-data becomes, the less incentivized many vendors are likely to be, given that they may be able to maximize revenue by selling their datasets to a wider demand base.

<sup>10</sup> A plausible circularity may exacerbate the shrinking shelf lives of alt-datasets: As the number of alt-datasets grows, more value accrues to those market participants that build alt-data capacity, which makes providing alt-datasets that much more appealing for vendors, who then increase market supply further, and so on.

losing unacceptable amounts of capital. Because most speed-oriented strategies are expensive to implement (e.g., they usually require specialized infrastructure or talent), they are often only efficient to deploy when large amounts of capital can be allocated to them. This risk profile for speed-based investing makes it unpalatable for most long-term Investors to undertake directly. In contrast, many asset-management firms (e.g., hedge funds, active mutual funds, or other organizations that extract management fees) can be relatively short-lived entities (i.e., they may not exist after their founders leave), and their failure would usually be less socioeconomically disastrous than it would be for Investors; thus, their cost of allocating most of their capital to speed-driven strategies is far lower.

Investors are also comparatively disadvantaged in terms of agility. As noted, rising rivalry and declining excludability of many alt-datasets means that market participants who attempt to use alt-data to exploit opportunities must be somewhat flexible to succeed; when some alt-datasets lose value from becoming more conventional, others must be sought. Because alt-datasets are largely heterogeneous, organizations that design investment strategies around them need to be agile. The level of agility required for this purpose would overwhelm the data-management and governance systems of many Investors. Although it can be argued that Investors should strive to improve such systems, in many cases it is more pragmatic to align their use of alt-data with their native strengths.

Perhaps the most powerful comparative strength that Investors have is patience. Their long horizons of operation mean that Investors can reap greater gains than other market participants by being more methodical and disciplined in their investment activities. Accordingly, we assert that the deepest value proposition alt-data has for Investors entails *defensive* and *defensible* strategies.

Defensive strategies prioritize capital preservation and prudent risk-taking over speedily exploiting opportunities. Hence, defensive strategies that incorporate alt-data should be centered on pursuits such as advanced risk analysis and management or improving operating efficiencies. Done correctly, these strategies can substantially decrease the degree of rivalry over an alt-dataset (i.e., one Investor building a defensive strategy around an alt-dataset need not reduce the value to another Investor of doing likewise). Risk management and exclusionary screening in responsible/sustainable investing are

quintessential examples of defensively applying alt-data: Alt-data can be an invaluable source of intelligence on environmental, social, governance, and other factors that are germane to responsible/sustainable investment decisions, and use of an alt-dataset for exploring those factors does not necessarily degrade its value for use in the same type of decisions by others.

Defensible alt-data strategies, meanwhile, can help Investors increase the excludability of an alt-dataset by either restricting access to it (e.g., via making it proprietary) or by developing execution capabilities around it that are not replicable by other market participants (e.g., through having privileged access to infrastructure deals via special relationships with local governments).

In this article, we concentrate on defensive alt-data strategies because we believe these are most broadly applicable across various Investor types and circumstances. We cover defensible strategies briefly in the final section of this article, and we reserve a detailed treatment for a companion article.<sup>11</sup> From what we see, the two clearest categories of defensive alt-data strategies for Investors are deeper understanding of risk (to better allocate and manage it) and driving operational alpha.

## Understanding Risk

Modern efforts in risk management largely emphasize simplifying risk over deeply comprehending its sources. Put differently, such risk-management paradigms are better at detecting that specific risks have materialized in the past than revealing why they have done so. For example, they may uncover how price movements for a given basket of securities correlate when responding to some event, but they deliver scant insight into why the event transpired in the first place. For market participants that operate over short horizons, knowing the correlation may suffice for managing risk, but for Investors to better leverage their capacity for patience, understanding the reasons why can be essential.

This need to more deeply probe causality is due to the fact that correlations in conventional datasets often break down over longer horizons and typically do not reflect the entire spectrum of events that could occur over long periods of time. Alt-data can (partly) mitigate these shortcomings by supplying more context about how events in the wider world drive downside moves

---

<sup>11</sup> See Monk, Prins, and Rook (2018).

in markets. Although it is true that rapid detection of such events might allow Investors to exploit opportunities, a less rivalrous (and more durable) benefit of early detection is that it allows more time for Investors to respond to downside events once they are flagged as likely. Moreover, added context can help warn about unprecedented downside events. When more variables are tracked, there is a higher likelihood of catching anomalous behavior that heralds highly atypical events, even if the precise impacts of such events might not be immediately apparent.<sup>12</sup> The ability to be alerted about unusual events is of prime importance to Investors. Large market crashes practically never play out in the same ways their predecessors did, but a single crash can fully nullify many years of outstanding performance.<sup>13</sup>

The purpose of defensive alt-data strategies is not to totally eliminate risk exposure for Investors but more to distribute it selectively.<sup>14</sup> Selective risk exposure is the chief idea behind smart-beta investment strategies, which seek to control exposures by holding positions in assets that are not necessarily proportional to their respective market capitalizations. Today, many Investors pursue smart-beta investing through purposed exchange-traded funds (ETFs), but smart-beta ETFs often lack fine control over risk exposure. For one, such ETFs are usually only ever composed of public securities and thus are not helpful for controlling private-asset exposure. Second, the asset weightings for the vast majority of ETFs are based on factors derived from conventional data (e.g., company size, dividends, or price momentum). These factors mostly fail to reflect risk in any nuanced way. For finer control over risk exposures through smart-beta ETFs, Investors must purchase shares in niche ETFs that can have high liquidity risk and management fees. Finally, the programmatic rebalancing rules for passive (and many semiactive) smart-beta ETFs

---

<sup>12</sup> Consider a parable example: An island civilization that never has witnessed (or even heard of) a tsunami may nonetheless get advanced warning of an impending anomalous event because of the sudden, dramatic recession of shoreline that characteristically precedes a tsunami.

<sup>13</sup> Long-lived entities are more likely to encounter such crashes, so being able to not do too badly during these crashes is as good as, if not better than, exploitation speed. Investors cannot just shut their doors if they do poorly.

<sup>14</sup> That is, by augmenting information sets with alt-data, Investors may reduce unwanted exposures (e.g., to climate change or reputational risk of investee companies) in a more controlled way, while increasing their desired exposures.

can create unintended—and severely disadvantageous—consequences when abrupt market downturns occur.

Judicious use of alt-data may allow Investors to deploy smart-beta (or similar) strategies in ways that avoid these shortfalls. A suitable supply of alt-data could allow Investors to design index-construction methods for public (or private) assets that create tailored, controlled risk exposures.<sup>15</sup>

The use of alt-data to more deeply understand risk is not confined to portfolio construction. Indeed, alt-data have applications in other areas of risk management, such as in asset oversight and due-diligence processes, especially in private markets. For example, if an Investor directly owns a real-estate development project in an emerging market, it may hire a local manager to oversee that asset's construction. However, this delegation can generate agency problems, such as when the Investor must rely primarily on the local manager's reports about the project's progress. A form of alt-data that might lessen such problems is images of shadow lengths from the project's construction site (e.g., taken from aircraft or satellites). Algorithms such as those developed by Orbital Insight are capable of converting the lengths in such images into calculations of the pace of projects so that an Investor might enjoy greater clarity about whether its local manager is providing valid reports.<sup>16</sup>

An example of alt-data's use for deeper understanding of risk in due diligence involves the analysis of a venture capitalist's networks in determining whether to invest in one of its funds. The relevant networks might be derived from alt-data sources, such as LinkedIn (for general partners' professional and social networks), or

---

<sup>15</sup> In practice, such methods might be similar to those used by Kensho Technologies to construct its *New Economy Indices*, which capture public companies' degrees of involvement in thematic technological trends, such as artificial intelligence, autonomous vehicles, or drones. To derive its indexes, Kensho uses natural-language processing to identify a company's exposure to a given trend by parsing its public filings (e.g., 10-Ks, 20-Fs) for information on (for example) product lines, supply chains, or planned capital expenditures. Although such filings do not qualify as alt-data, this approach could be applied on other, less-conventional text documents to construct indexes (e.g., sustainability reports).

<sup>16</sup> Another example that may materialize in the future could involve Investors using internet-of-things data feeds from their investee companies or assets. Such data could be used in risk management, help in monitoring human work patterns and information flow, give greater clarity on microjudgments, and help make valuation more real time.

built from scraping websites or digital newsfeeds (to capture what other funds were co-investors on specific deals). Because relationships are integral to most venture capitalists' success, understanding the strength or weakness of a fund manager's networks can be a crucial variable for deciding whether an Investor should allocate capital to that manager.<sup>17</sup>

Some other examples of how alt-data may be used defensively for understanding risk include the following:

- harvesting dynamic pricing information from online sources to garner a clearer, more real-time picture of inflation (and draw on wider or more targeted sources of pricing information than are usual in generic consumer-price indexes);
- aggregating label information (e.g., nutrition facts, ingredients lists) from food-product companies' offerings to see how they may be vulnerable to shifting dietary trends or new warnings by health agencies (Investors may then be able to compel company managers to alter their offerings—e.g., through shareholder activism for publicly traded companies);
- assembling online price and ratings histories of possible competitors (e.g., from Airbnb, TripAdvisor, or Yelp) or price series of airfares to that locale when doing due diligence on candidate direct investments in leisure-related properties (e.g., hotels or casinos);
- using microsensors (or other remote sensors) to track fluctuations in soil moisture for determining what plants are best suited to intercropping in a plantation-forestry investment; and
- controlling reputational risk from investee companies by monitoring controversies about them that arise in social-media posts (or other localized or unconventional news outlets).

## Generating Operational Alpha

Alongside deeper understanding of risk, Investors can also use alt-datasets in defensive ways by turning

<sup>17</sup> More specifically, an Investor may have little ex ante clarity about the specific startup companies in which a venture capitalist will invest (and no control over how it does so once capital is pledged). The quality of the venture capitalist's likely co-investors, however, may be easier to discern and serve as an indicator of the ultimate riskiness of its portfolio.

them into sources of *operational alpha*. The chief idea behind operational alpha is to better align operating resources with investment strategies by eliminating internal inefficiencies in how investment processes are executed. This concept is (loosely) related to investment alpha, which is the generation of returns in excess of some benchmark, after adjusting for the riskiness of the assets used to generate the excess returns. Although operational alpha has a secondary benefit of (potentially) improving gross investment returns, its chief aim is to improve net returns by reducing unneeded operating costs. Because such reductions are often risk free, operating alpha can substitute for, and in many instances is superior to, investment alpha.<sup>18</sup> It can also complement investment alpha because it frees up room in the risk budget and thus allows pursuit of strategies with higher upside.

Alt-data can aid Investors in driving operational alpha. Perhaps surprisingly, most Investors already possess large volumes of alt-data within their own organizations. Because alt-data are defined as data not conventionally used in decision making, novel forms of internal data count as alt-data.

Aggregation and disaggregation are key to converting conventional internal data into alt-data. For instance, inventive collation and synthesis of documents (e.g., e-mails, investment memos, and contracts) can uncover precious metadata that is able to provide insights for enhancing communication, culture, negotiation, time allocation, benchmarking, and diligence. Likewise, the disaggregation of collective processes into individual contributions can give a clearer picture of where latent organizational resources—and opportunities to improve them—reside. For example, by tracking how individual internal users query and access documents in organizational databases, an Investor can construct a map of intraorganizational knowledge flows and examine the typical approaches its analysts use in problem solving. More granular visibility of these individual activities can not only expose areas for improvement but also help better identify best practices.<sup>19</sup>

<sup>18</sup> Notably, operational alpha can be (almost or fully) market agnostic.

<sup>19</sup> Such added visibility of internal processes also has a potential risk-management benefit in the form of compliance. Newly legislated requirements for data handling (e.g., the European Union's General Data Protection Regulation) mandate that users be made aware of how their personal data are being treated. In the case of

## Implications of a Changed Value Proposition

In rethinking the value proposition of alt-data, Investors will need to re-examine other views and approaches they have regarding alt-data. Specifically, in pursuing defensive or defensible alt-data strategies, Investors will likely need to alter how they characterize and access alt-datasets. In the next two sections of this article, we discuss pragmatic paths for addressing each of these matters.

## RETHINKING HOW ALT-DATA IS CHARACTERIZED

Because the number and diversity of alt-datasets is enormous, Investors need to be discriminating when selecting which alt-datasets deserve resources (e.g., money to acquire; time to store, prepare, and analyze; and capacity to be governed). Such selectivity requires characterizing alt-datasets to establish which will be most valuable for organizational needs. As the value any dataset has to an Investor lies in the questions it can help answer, there is a need for data-characterization methods that can reflect the question-answering capabilities of datasets (alternative or otherwise).

Alt-data are defined in an exclusionary way—by stating what they are not (conventional). However, unlike alt-data's definition, a characterization system for alt-data should not be constructed around exclusion: It is more reasonable to characterize an alt-dataset by those properties that it verifiably exhibits, rather than those it does not. Problematically, however, few Investors—or, for that matter, financial organizations in general—have any such system for alt-data characterization. In fact (and as we will detail later), Investors rarely have any formal criteria for establishing whether a dataset is indeed alternative (i.e., a threshold that divides conventional from unconventional data on the basis of scarcity, novelty, or another relevant quantitative or qualitative dimension).<sup>20</sup>

Unsurprisingly, because few Investors have any systems for distinguishing or characterizing alt-data,

---

Investors, these users can be their employees. Because the definition of what constitutes personal data is evolving, Investors stand better chances of remaining compliant if they already have developed processes and systems for tracking diverse forms of internal data in their organization.

<sup>20</sup> More generally, many Investors have no formalized models or system for characterizing data or judging data quality.

few use any consistent process for valuing its worth in advancing organizational objectives. Undoubtedly, rigorous valuation of alt-data (or any data, for that matter) is a difficult undertaking and subject to wide error margins.<sup>21</sup> Characterization is a more achievable step: It at least facilitates judgments about whether a given alt-dataset aligns with organizational capabilities and strategic priorities. Lack of characterization systems, however, invites the expenditure of resources on alt-datasets that do not fit with organizational priorities and resources and promotes avoidable waste.

Apart from being wasteful, not having characterization systems can challenge an Investor's fulfillment of its fiduciary duties or regulatory compliance: Investors may be hard-pressed to claim that they are engaging in responsible decision making when decisions are made based on data that are not well understood (e.g., in terms of blind spots it may create). Suitably understanding data (whether alternative or conventional) in any consistent way requires a means of characterizing it.

## Existing Characterization Systems

Existing systems for characterizing alt-datasets are not suitably aligned with the value propositions we have described. These existing systems either ignore the ways in which an alt-dataset is likely to create value for an Investor (and so neglect organizational context) or assume that any dataset's main use will be driving investment alpha (or a similar short-term, opportunistic pursuit).

For example, Kolanovic and Krishnamachari (2017) posited a characterization system for alt-data that focuses on the origins of datasets (Exhibit 1). This system is not ideal for Investors' purposes for several reasons. First, although it encompasses many sources of alt-datasets, it is not necessarily exhaustive. Second, it gives no indication of how valuable a given alt-dataset is to an Investor. Taxonomical schemas such as this are not best suited to help Investors evaluate alt-data.<sup>22</sup>

---

<sup>21</sup> Inarguably, an alt-dataset's value should be positively related to its quality. Yet no quality metrics exist that are universally applicable across datasets or free of restrictive assumptions. We must resort to using properties of data that can serve as context-appropriate proxies for quality. It is on these properties that alt-data should be characterized.

<sup>22</sup> Taxonomical systems are characterization systems that are (or attempt to be) mutually exclusive and collectively exhaustive—that is, the items they characterize must fit into one, and only one, classification category within the system.

## EXHIBIT 1

### Kolanovic and Krishnamachari's Characterization System for Alt-Data

Source Category	Specific Alt-Data Source
Individual Processes	Social media, news and reviews, web searches, personal data
Business Processes	Transaction data, corporate data, government agency data
Sensors	Satellites, geolocation, other sensors

Source: Kolanovic and Krishnamachari (2017).

Kolanovic and Krishnamachari (2017) proposed another taxonomical schema for alt-data characterization that does embed a value proposition and strives to indicate the usefulness of alt-datasets in relation to use cases based on asset class and investing style. Unfortunately, that system is premised on investment-alpha generation, and so it does not cover defensive or defensible uses, which thus undercuts its relevance for Investors (which is further lowered by being taxonomical).

Dannemiller and Kataria (2017) avoided the taxonomical approach and instead suggested that alt-data be characterized on a “continuum ... from structured to unstructured.” For the purposes of indicating the likely value of an alt-dataset, using continuums, and not discrete categories, makes sense, but whether a dataset is structured or unstructured does not immediately reflect its value for an Investor. It is true that more effort may be required to extract insight from unstructured datasets (which makes them more expensive from an organizational-resource perspective), but this does not necessarily reflect the full value that an alt-dataset holds. For example, both unstructured and structured alt-datasets may be relevant (or not) for defensive or defensible approaches by Investors.

Although big data and alt-data are not perfectly identical, there are cases in which alt-data qualify as big data. It may thus be hoped that characterization schemas for big data could sometimes be applicable to alt-data. The most prevalent such schema is the 3 Vs of big data: volume, velocity, and variety. IBM's Big Data unit suggests a further dimension: *veracity* (i.e., the degree of uncertainty around a dataset).<sup>23</sup> These systems are a step

<sup>23</sup> See: <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>.

in the right direction because veracity, velocity (the rate at which new data arrive), and volume (the size of a dataset) could all potentially add to a dataset's value for an Investor.<sup>24</sup> Yet these dimensions by themselves are incomplete, and none seem to squarely encapsulate how specific properties of an alt-dataset should translate into value. For example, velocity may be important for assets that have value-determining properties, which change frequently, but not so important for those without such properties (e.g., many private assets).<sup>25</sup> Thus, freshness—how well a dataset reflects the most recent changes that are material for decision making—might be more appropriate. Likewise, volume seems to be less important for Investors than whether a dataset is comprehensive. That is, a dataset may contain many items (i.e., have high volume) from only a narrow number of categories of interest. In such a case, a dataset that has smaller volume, but encompasses more categories (i.e., is more comprehensive), would likely have higher value. We thus need a different characterization scheme.

The system devised by Kitchin (2015) comes closest to what Investors need. It builds upon the 3-Vs setup (but is still intended for characterizing big data, rather than alt-data) by adding four additional dimensions: *comprehensiveness*, *granularity* (how fine- or coarse-scaled the data are), *relationality* (how many fields a dataset shares with other datasets of interest), and *flexibility* (how easily new fields can be added to a dataset).<sup>26</sup> Comprehensiveness and granularity seem to be apt fits for Investors' purposes, but it is less clear that relationality or flexibility are pertinent concerns. Furthermore, Kitchin's scheme gives no explicit consideration to the known quality (i.e., reliability) of data. Knowing how reliable a dataset is can be essential for Investors to decide how it can be used.

### Six Dimensions of Alt-Data

We adapt Kitchin's (2015) system by replacing relationality, flexibility, variety, and volume with the dimensions of *reliability*, *actionability*, and *scarcity* (and replacing the velocity dimension with the more fitting notion of *freshness*). Reliability (which covers the

<sup>24</sup> Velocity may concern the rate at which new datasets are onboarded or the rate at which existing ones are refreshed.

<sup>25</sup> Velocity may also be valuable (for example) in rapidly detecting reputational risks for Investors in social-media activity.

<sup>26</sup> Kitchin actually uses “exhaustivity” and “resolution” in place of comprehensiveness and granularity, respectively.























Lo, A., and J. Hasanhodzic. *The Evolution of Technical Analysis: Financial Prediction from Babylonian Tablets to Bloomberg Terminals*. New York: Bloomberg Press, 2010.

Monk, A., M. Prins, and D. Rook. 2018. "Playing Data Defense in Institutional Investing." Mimeo.

Monk, A., and D. Rook. 2018. "The Technological Investor: Deeper Innovation through Reorientation." SSRN eLibrary, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3134078](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3134078).

*To order reprints of this article, please contact David Rowe at [d.rowe@pageantmedia.com](mailto:d.rowe@pageantmedia.com) or 646-891-2157.*