



DATA MONETIZATION

Edition 3

www.sellalldata.com

Table Of Contents

Section 1: Introduction	2
Section 2: The Alternative Data Revolution	3
Section 3: The Data Monetization Opportunity	9
Section 4: Advice From Vendors	15
Section 5: Advice From Buyside Firms	22
Section 6: Solutions For Vendors Of Alternative Data.....	28
Appendix	31
A: Interviews With Data Vendors That Have Sold To Buyside.....	31
B: 18 Buyside Firms Answer The 10 Most Common Vendor Questions	47
Glossary	69

Section 1: Introduction

Purpose Of This Report

The purpose of this report is to provide owners of alternative data with advice regarding how to sell their data to buy-side firms¹. This report is the third in a series of reports published by Eagle Alpha. We welcome all comments, suggestions and ideas to improve the next edition of this report.

The Alternative Data Revolution

According to Goldman Sachs Asset Management in 2010 the world entered the 'Data Revolution'. The Data Revolution has led to an increasing amount of data companies and exhaust data providers. This non-traditional "alternative" data is being leveraged by buy-side firms. Spend by the buy-side on alternative datasets is forecast to rise to over US\$2bn, according to the consulting firm Opimas.

The Data Monetization Opportunity

There are 300 buy-side firms purchasing alternative data today. This number will increase to thousands of firms over the coming years.

In order to be able to sell a dataset to all buy-side firms it needs to be relevant to all types of firms (e.g. quant, hedge, long-only) and asset classes (equity, macro, credit). In order to maximize the entire addressable market, vendors need to deliver data in 4 ways: i) raw data; ii) aggregated/tickerized; iii) democratizing data through dashboards; and iv) reports.

In Section 3 we outline the factors that affect sales volume and pricing. The four key items to consider regarding sales volume are: 1) delivery format; 2) relevance; 3) exclusivity; and 4) sales to the sell-side. There are eight key factors that vendors need to take into account when pricing a dataset. These are: 1) ROI; 2) uniqueness; 3) quant vs. discretionary; 4) micro vs. macro; 5) coverage; 6) comps; 7) license structure; and 8) feedback.

Advice From Vendors

We interviewed 15 successful vendors. Their advice can be broken down into seven topics (see Section 4). These are: 1) understanding the needs of buy-side firms; 2) difficulty of standing out; 3) price discovery; 4) lengthy onboarding process; 5) Co-operation among vendors and platforms; 6) hire data scientists; and 7) be open to feedback.

Advice From Buy-side Firms

We interviewed 18 buy-side firms. Section 5 includes their advice for vendors split into the following topics: 1) dataset development; 2) legal and compliance; 3) initial pitch; 4) Q&A with vendors; 5) samples; 6) trials; and 7) dataset delivery.

Solutions For Vendors Of Alternative Data

Eagle Alpha works with data owners in six ways: 1) Freemium: we offer free marketing to our global client base for vendors who create and manage their profile on our platform; 2) Sales Referral: we consistently provide fully qualified sales leads to over 100 Eagle Alpha partners; 3) Premium Add-Ons: we have created solutions to help vendors overcome marketing, legal and compliance and sales obstacles; 4) Sales Referrals (exclusive): where Eagle Alpha is the exclusive distributor to the buy-side we provide all the premium add-ons for free to boost sales; 5) Dashboards: we build tailored dashboards for partner datasets broadening the addressable market of end users; and 6) Joint Ventures: Eagle Alpha partners with the owners of unique data to build new datasets for the buy-side.

¹ See the Glossary on page 69 for a definition of the different types of buy-side firms.

Section 2: The Alternative Data Revolution

Introduction

In this section we introduce the concept of the data revolution, provide examples of industries that are already monetizing data, define alternative data, outline the benefits and use cases and finally discuss adoption and dollar spend.

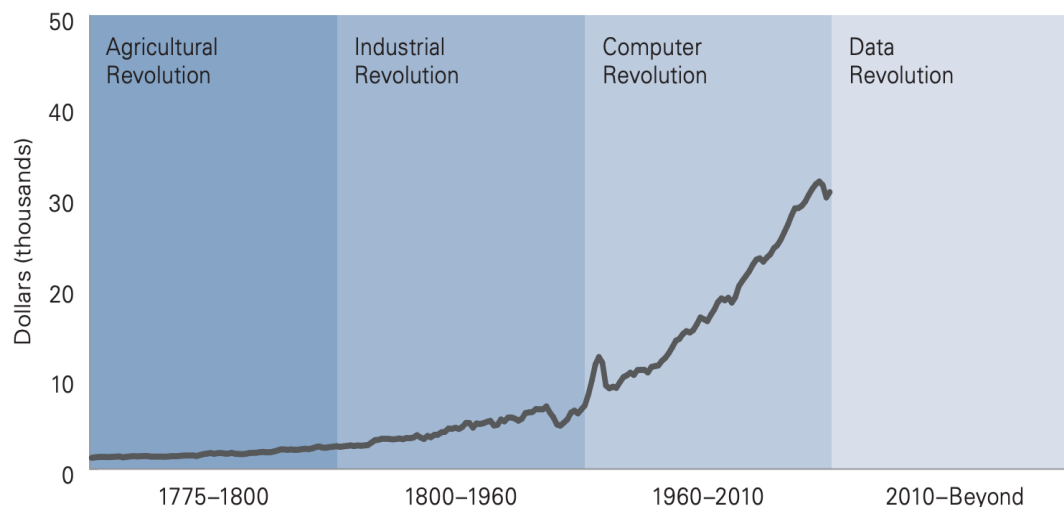
The Data Revolution

In 2010 the world entered the “Data Revolution”, according to Goldman Sachs Asset Management. The quote below describes the diagram in Figure 1.

“The Data Revolution is not the first-time technology promises to transform our world. It may, however, be the fastest. In the long view of economic history, the first major technological shift forward was the Agricultural Revolution, which benefited humankind with more efficient farming methods and a higher standard of living. Next, in the Industrial Revolution of the late eighteenth and early nineteenth centuries, economic productivity doubled in about 150–200 years. Productivity skyrocketed in the Computer Revolution, with a three-fold gain in the half century starting in the 1960s. Now, at the onset of what we believe to be a fourth revolution, the world may be entering a period of even more rapid productivity gains thanks to the remarkable power which can be gained from data”.

Figure 1: The Four Revolutions

United States GDP Per Capita: 1775 – 2010



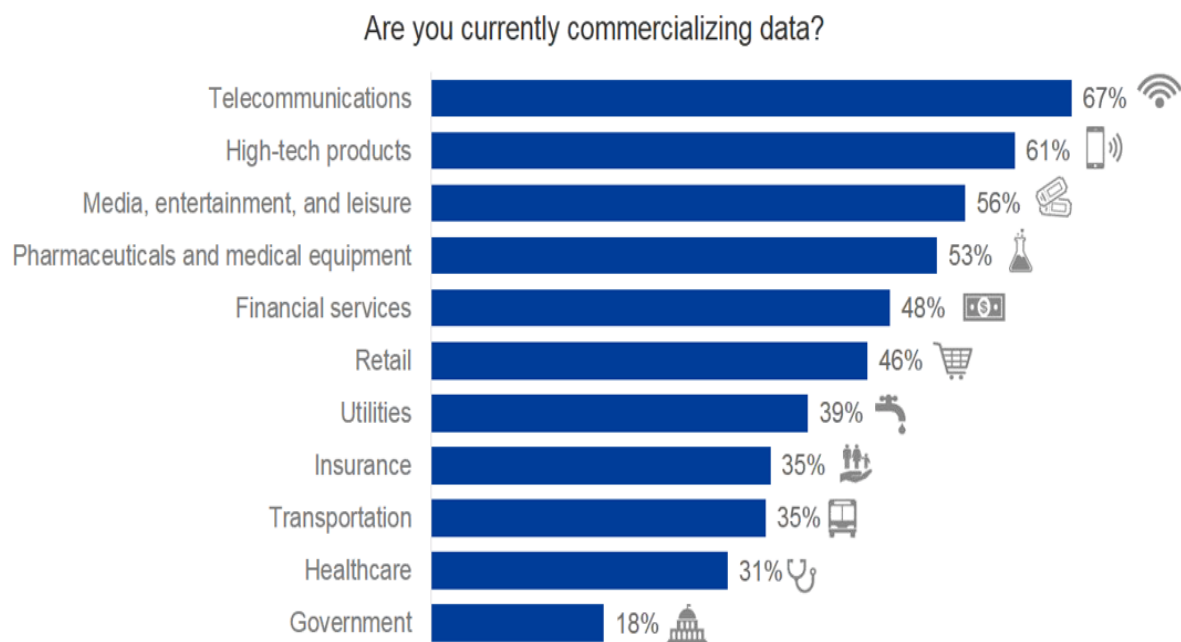
Source: Goldman Sachs Asset Management - The Data Revolution – From Volume to Value².

² Source: ‘[The Data Revolution: From Volume to Value](#)’, September 2015.

Data Monetization

The Data Revolution enables owners of alternative data to increase revenue through commercialization opportunities. In a 2017 Forrester survey³, it was highlighted that 48% of companies were already commercializing their data, up from 32% in 2016. This statistic is echoed by a McKinsey survey⁴ in 2017 that stated 41% of respondent companies have begun to monetize data.

Figure 2: Industries That Are Monetizing Data



Source: Forrester Data Global Business Technographics Data And Analytics Survey (2017) of global data and analytics business decision makers. Note: not all industries are shown.

What Is Alternative Data?

Buyside firms⁵ around the world all have access to the same traditional information and data from traditional sources. They are increasingly looking to 'alternative data' to give them an edge. Alternative data is non-traditional data that can be used in the investment process. Because alternative data refers to a wide variety of data, it is frequently defined by what it isn't, rather than what it is.

Traditional data includes market price data, financial data, and government and industry statistics. Eagle Alpha identifies 24 categories of alternative data, defined by either the collection method or the nature of the data (see Figure 3).

As of March 6th 2019, there were 898 datasets in our database that were spread across these 24 categories. It is interesting to note that some of our clients have already identified over 2,000

³ Source: ['The Age of Alt: Data Commercialization Brings Alternative Data To Market'](#), June 2017.

⁴ Source: ['Fueling Growth Through Data Monetization'](#), December 2017.

⁵ The term "buyside" refers to all types of investment managers including: quantitative funds, discretionary hedge funds, traditional long-only funds, sovereign wealth funds and private equity funds.

alternative datasets that are relevant to the buy-side. We forecast there being 5,000 datasets by the end of 2020, which several of our clients believe is conservative.

Figure 3: Eagle Alpha's 24 Categories of Alternative Data



Source: Eagle Alpha

Alternative data sometimes qualifies as big data, and sometimes it is small enough to analyse in a spreadsheet. It may be structured or unstructured, or it may be structured data that's derived from unstructured data, such as a sentiment score for text. In our view the word "alternative" is currently in vogue but will fall away. The term the industry will use is simply "data", it will not matter if its traditional or non-traditional (alternative).

This was echoed in an interview we conducted for [Altdata.tv](https://altdata.tv)⁶ with Mani Mahjouri, CEO and CIO of Blueshift Asset Management:



"Ultimately, the sort of mega trend is this curve technology that's enabling us to measure information more effectively and more comprehensively. We are hitting that inflection point where it actually is feasible to put this together. You have a confluence of consumers of this data, folks on the buy-side recognizing its value, with the right entrepreneurs looking to develop the products."

[Link to video.](#)

⁶ Our online TV station for the alternative data space.

Why Buyside Firms Use Alternative Data

There are five key benefits for buyside firms using alternative data: 1) Greater volume of data and information; 2) Unforeseen insight; 3) Competitive edge; 4) Fiduciary duty; and 5) Efficiency. Figure 4 describes each benefit.

It is important that vendors understand how buyside firms use alternative data. We produce a 125-page [white paper](#) dedicated to fifty case studies that are based on third party datasets as well as Eagle Alpha proprietary datasets and analysis. Many of the case studies have been backtested and/or verified by third parties such as JPMorgan, Citi Research, and academic researchers. The case studies are classified by category of alternative data, asset class and manager type.

Figure 4: Key Benefits For Asset Managers

	Alternative Data	Traditional Data	Description
Greater Volume of Data and Information	<ul style="list-style-type: none"> Frequently updated Shorter history Wide breadth 	<ul style="list-style-type: none"> Sparsely updated Longer history Narrow breadth 	<ul style="list-style-type: none"> Large amounts of data are being generated constantly, leading to more available information for analysis Frequent updates mean managers can carry out analysis sooner, enhancing portfolio construction
Unforeseen Insight	<ul style="list-style-type: none"> Wide breadth Profound implications to multiple assets classes, or even industries. 	<ul style="list-style-type: none"> Narrow breadth Only provides particular information, normally only directly relevant to few assets. 	<ul style="list-style-type: none"> New data provides information on more than just financial performance, leading to better trend and performance predictions Discovery of hidden relationships can help devise investment strategies
Competitive Edge	<ul style="list-style-type: none"> Requires investment and data capabilities 	<ul style="list-style-type: none"> Available to all Analysis can be conducted relatively easily 	<ul style="list-style-type: none"> Talent and technologies are needed to gain value from alternative data, giving asset managers with the right resources higher alpha-generation
Fiduciary Duty	<ul style="list-style-type: none"> Make use of available information and data 	<ul style="list-style-type: none"> Make use of financial data and few conversations only 	<ul style="list-style-type: none"> Talent and technologies are needed to gain value from alternative data, giving asset managers with the right resources higher alpha-generation
Efficiency	<ul style="list-style-type: none"> Rapid and efficient research and analysis Wide coverage 	<ul style="list-style-type: none"> Highly manual research and analysis Narrow coverage 	<ul style="list-style-type: none"> Alternative data provides insights into multiple assets, and can replace certain aspects of the current research process. Research analysts spend longer time on better modeling or investment strategies

 Enhance alpha generation

Source: Quinlan & Associates



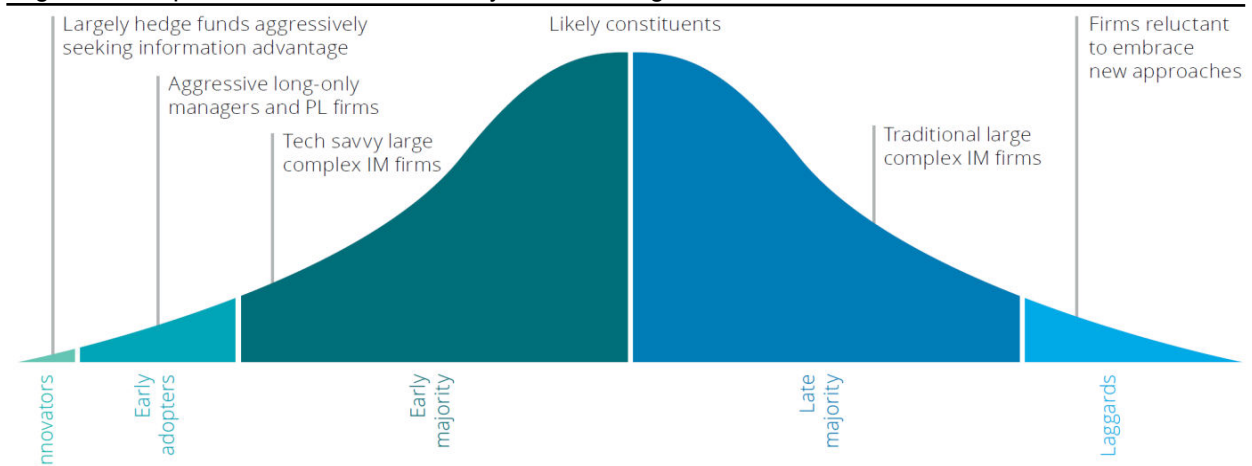
[Link to download.](#)

Table of Contents	
SECTION 1: EXECUTIVE SUMMARY	4
SECTION 2: ALTERNATIVE DATA	6
SECTION 2.1: DEFINITION AND BENEFITS	6
SECTION 2.2: DRIVERS OF ADOPTION	8
SECTION 3: DATASETS IN DEMAND (PROPRIETARY ANALYTICS)	14
SECTION 3.1: EAGLE ALPHA DATABASE CLICK-THROUGH ANALYTICS	14
SECTION 3.2: 1-ON-1 MEETING ANALYTICS	16
SECTION 3.3: SELECTED PROFILES OF DATASETS IN DEMAND	16
SECTION 4: USE CASES	18
SECTION 4.1: OVERVIEW	18
SECTION 4.2: BY CATEGORY OF ALTERNATIVE DATA	19
SECTION 4.3: BY ASSET CLASS	24
SECTION 4.4: APPLICATIONS BY TYPE OF ASSET MANAGEMENT FIRM	29
SECTION 4.5: BUYSIDE USE CASES MENTIONED ON ALTDATA.TV	33
SECTION 5: FIFTY CASE STUDIES	38
SECTION 6: QUANTITATIVE APPROACHES TO ONLINE SEARCH DATA	117
SECTION 7: OVERVIEW OF EAGLE ALPHA	120

The Tipping Point Is Finally Upon Us

Alternative data is not new. More than 50 innovative buy-side firms have been working with alternative data for over 5 years. These firms are primarily the large quantitative and 'quantamental' firms. According to an EY survey back in 2017, 78% of US hedge funds were already using, or said they use, alternative data. We predict that the tipping point for the global asset management industry will be during 2019. See Figure 5.

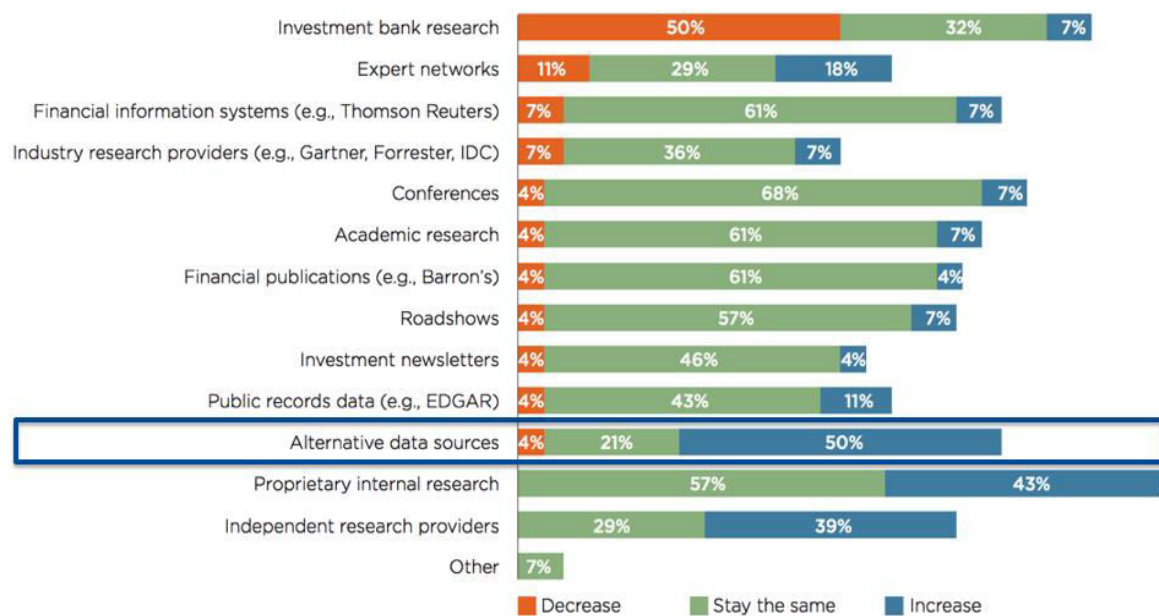
Figure 5: Adoption Of Alternative Data By Asset Managers



Source: Deloitte

According to the Greenwich Associates 2018 Future of Investment Research Study, alternative data is the source of investment research that is forecasted to grow the most. See Figure 6.

Figure 6: Alternative Data's Projected Growth



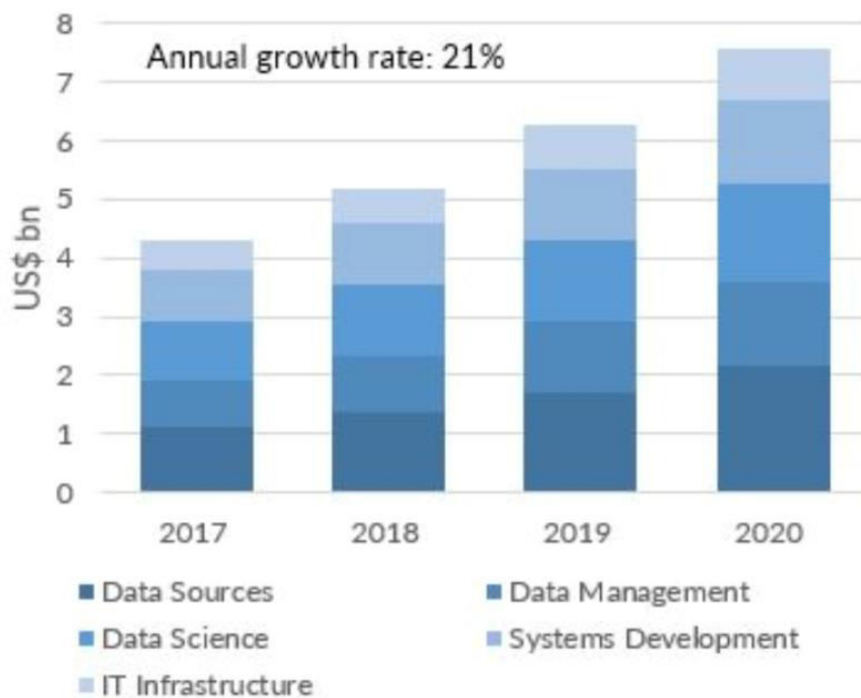
Source: Greenwich Associates

Dollar Spend Is Increasing Substantially

More and more buy-side firms are procuring alternative datasets and the overall spend is forecasted to rise significantly – see Figure 7. A research and consulting firm Opimas estimates that buy-side firms will be spending up to \$2 billion by 2020. Furthermore, they expect alternative data to “contribute significantly to a further shrinkage in the hedge fund population, as firms unable to exploit the information needed to compete effectively in the new world of intelligent investing will fall behind.”

Other firms such as JPMorgan, TABB Group and Greenwich Associates have also given forecasts. In our view the bottom-up analysis by Greenwich is the most accurate. Clearly the spend will increase substantially in addition to the number of firms purchasing alternative data.

Figure 7: Spending On Alternative Sources/Datasets



Source: Opimas

J.P.Morgan

TABB
GROUP

GREENWICH
ASSOCIATES

“Our estimate of the investment management industry’s spend on Big Data is in the \$2-3bn range, and the number is expected to have double digit annual growth (e.g. 10-20%), in line with Big Data growth in other industries. This spend includes acquiring datasets, building Big Data technology, and hiring appropriate talent”

“TABB Group estimates that the current [2016] market for alternative datasets in the US is at \$200 million and expects the market to grow to \$400 million in the next 5 years. The drivers of this growth is both the expansion of users into smaller quant funds, fundamental funds, the sellside and venture capitalists”

“The average investment firm spent about \$900,000 yearly on alternative data...extrapolating this data across the whole market, we estimate that annual industry budgets for alternative data now stand at \$300 million — almost double from one year ago”.

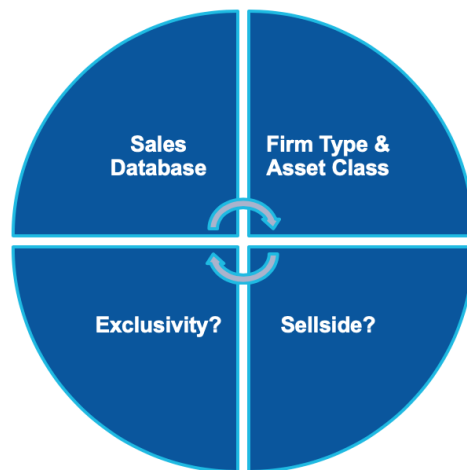
Section 3: The Data Monetization Opportunity

Introduction

In this section we outline the factors that affect sales volume and pricing.

Sales Volume Considerations

The four key items to consider regarding sales volume are: 1) delivery format; 2) relevance; 3) exclusivity; and 4) sales to the sellside.



1. Delivery Format.

In order to maximize the addressable market, vendors need to deliver data in 4 ways: i) raw data; ii) aggregated/tickerized; iii) democratizing data through dashboards; and iv) reports.

Olga Kokareva, Director of Business Development at Quantstellation, highlights the difference between monetization, commercialization and productization:

“Data monetization can be a one-time deal. You can just sell the data exhaust to an intermediary or a hedge fund and forget about it. You don’t have to deal with data preparation, data engineering, or anything like that. But commercialization is different, it is a long-term business of selling data. To commercialize a dataset, a data owner should invest up-front in building a high-quality data product in the technical team and salesforce. This will help you in building a scalable business where a data vendor can raise capital and eventually make a lucrative exit over the course of time.

Moving to productization, there is a big difference between data and a data product. Although there are still some unique data sets in the market, given that the market is saturated now and in most categories of alternative data, funds can find a few competing vendors. So, when vendors are pressured by price competition and data commoditization, data quality will become a major competitive edge. In my personal experience, vendors who have invested in productization and standardizing

2. Firm Type & Asset Class.

There are 300 buy-side firms purchasing alternative data today. This number will increase to thousands of firms over the coming years. In order to be able to sell a dataset to all buy-side firms it needs to be relevant to all types of firms (e.g. quant, hedge, long-only) and asset classes (equity, macro, credit).

3. Exclusivity?

Should you offer exclusivity when selling your data? Eagle Alpha does not recommend that data owners offer exclusive deals to the buy-side. The majority of our buy-side clients agree:

"We avoid exclusive datasets: exclusive offerings are not desirable".

"Non-exclusive makes compliance easier".

"We avoid exclusive relationships given MNPI concerns".

"Exclusivity isn't required (or desired)".

"Complete exclusivity may be a compliance concern, but limited distribution is always an interesting
"We are generally not interested in exclusive agreements".

"We aren't interested in exclusivity and I would argue that selling exclusively is myopic and won't allow for much revenue growth".

"We don't look for exclusivity. If a vendor has several financial clients, we see this as a positive. The vendor will be able to invest more in improving the dataset. All things being equal we would much rather pay half the price (improving the net profitability of the data substantially) at the cost of having another firm accessing the same data".

4. Sellside?

Most of our buy-side clients advise vendors not to sell their datasets to the sellside because it will dilute the alpha opportunity. If a vendor wants to sell to a sellside firm, Eagle Alpha recommends selling data at an aggregate level. Below are the views of asset managers regarding selling to the sellside:

"Should sell solely to buy-side firms".

"Exclusivity to the buy-side is vital – if the data is sold to the sellside we will see the conclusions through them so are very unlikely to pay for the data".

"Selling only to buy-side is more ideal, but if they work with sellside it is not a non-starter".

"No to sellside. Aggregated numbers to third party resellers or strategic partners".

"No to selling to sellside."

"We are not against [selling to the] sellside".

"Seems reasonable that they should sell broadly. We would employ a dataset in a very specific way in our process, so we are quite relaxed about others having access to it".

"We actually don't mind sales to the sellside or others. We consider data as the starting and not the ending point. Everyone will start with the most obvious questions (e.g. summing email receipts for sales estimation), but the diversity of questions explodes once you get into 2nd and 3rd order thinking. So as long as there is sufficient degrees of freedom in a dataset (those are what we are interested in anyways), we are not that concerned".

Pricing Considerations

There are eight key factors that vendors need to take into account when pricing a dataset. These are: 1) ROI; 2) Uniqueness; 3) Quant vs. Discretionary; 4) Micro vs. Macro; 5) Coverage; 6) Comps; 7) License Structure; and 8) Feedback.

1. ROI

Vendors need to consider what return on data (“ROD”) can a buy-side firm make by integrating their dataset into an investment process. Quants and systematics typically seek for more than 10x returns from a single dataset as highlighted in this comment by a quant fund:

“Methods of paying for data vary but most funds will typically be paying out of money they earn on performance fees. At an (optimistic) 10% performance fee, a \$100k dataset has got to improve overall performance by \$1M a year – every year – just to break even. Given the uncertainty around future performance, most funds would need to make several times more than that to be happy to pay for a dataset”.

Clearly measuring ROI/ROD on alternative data spend is not always straightforward. For example, some of our discretionary clients measure it based on the number of interactions with a centralized ‘Data Insights’ team. The information derived from a dataset is part of a mosaic that leads to an investment decision. While the results of that decision are measurable, it can be difficult to determine how much credit should be attributed to each piece of the information mosaic. Frequency of consulting a dataset can be a good proxy for measuring its value to the asset management firm. A discretionary fund manager highlighted this point saying that no dataset provides a strong enough signal on its own for them to buy a stock.

2. Uniqueness

The more unique a dataset, the more valuable it can be. Buy-side clients will ask how widely a dataset is disseminated. As one client put it, “3rd vs. 30th is a price driver”.

Given the increasing number of alternative datasets, vendors need to ask whether buy-side firms are able to get similar data elsewhere. Sometimes there are proxies that achieve the same outcome.

In an [Altdata.tv](https://www.altdata.tv) interview with Qaisar Hasan, Portfolio Manager at Lombard Odier, it was highlighted that the buy-side now has more data options than they know what to do with it. Mr. Hasan noted:

“The challenge now both for the buy-side and the vendor community is how to stand out amongst that noise. Too many times we come across datasets that haven’t even been mapped to companies. A buy-sider can do it, but it’s highly time consuming and it’s increasing the overall cost of the dataset and ultimately impacts the ROI”.



[Link to interview.](#)

3. Quant vs. Discretionary

It is important for data vendors to note that quants and discretionary investors have very different needs and value data differently. Quants typically want an all-in API price. Advice from a quant fund:

“Different firms have different objectives and different uses in mind for the data, so depending on how a vendor sets their pricing structure, they may limit who is interested in their data. For example, we have seen some firms that charge by the ticker, so as a quant shop interested in the entire data set the price becomes unreasonably high”.

Discretionary firms, on the other hand, want options that could be threefold: 1) entire datasets; 2) subset of the dataset; and 3) per ticker. The data structure and delivery mechanism are also key for discretionary firms. Typically, these firms are not experienced with dealing with large datasets, and don't have the data science and engineering skills to work with raw data. Providing a more refined version of the data or an intuitive user interface is critical. A fund manager explained their strategy:

“We tend to spend little with suppliers initially, until the concept is proven internally. Then we can significantly increase our spend. For us, willingness to sell a proportion of the data for a relatively small fee is important and can lead to larger purchases later”.

Another discretionary fund manager highlighted that a per ticker pricing is a good starting point for a vendor to get their foot in the door as it is hard to get internal traction and buy the whole dataset to just look at one stock. One manager clarified this view:

“We explore a dataset on a per ticker pricing but would need a massive discount then upselling to the whole dataset. There is a need for a tiered structure so it's not insane to buy everything.”

Discretionary firms can work with less history and don't always require coverage universes that are as large.

4. Asset Class

There are three different asset classes: 1) equity; 2) macro; and 3) credit. Equity funds are the most frequent buyers of alternative datasets and therefore offer the largest addressable market. However, recently we have witnessed a steady increase in the number of enquiries from both macro and credit funds.

Company level data is typically more actionable and therefore valuable than macro level data. Macro data which is highly relevant to an asset class, such as global currencies or agricultural commodities, may be the exception.

It is important to adjust the data structure and delivery. This point was highlighted in our recent interview with a data vendor:

“Our data is structured in a way that allows investors to roll up datasets across any specific industry or geography if they want to get more macroeconomic insights or alternatively dissect the company’s own operations on the most granular level.”

5. Coverage

Assuming the same level of predictiveness, a dataset that covers more tickers/companies is typically more valuable. A fundamental manager said that they would pay up when a vendor is very good at one specific sector. A quant fund mentioned what they usually ask vendors relating to coverage:

“What length of history is available and how has coverage changed through time which is key to determining how suitable the dataset will be for back-testing”.

Most vendors don’t know the typical price point for alternative datasets in general and datasets they are in direct competition with. Pricing needs to be in-line with comparable datasets that are already available to the buy-side. One of our buy-side clients advised:

“Vendors need to price it towards the market and realize that they aren’t the only dataset out there. Better to hook a few whales at a lower price, gain traction, learn what is going right and going wrong, and build new products”.

As a leading aggregator of alternative data we know the pricing of hundreds of datasets and are therefore in a unique position to advise vendors on pricing.

7. License Structure

The majority of licenses are sold on annual contracts on a per ‘team’ basis. Some of our buy-side clients consider enterprise pricing as more valuable and won’t engage with vendors offering usage-based structures. A quantitative manager mentioned that their preference is to be on the meter and avoid paying for something they don’t need but would prefer an a la carte pricing per team or enterprise if there is enough demand. Another quantitative fund manager said:

“Avoid per usage models where possible. Don’t even consider asking to link the fee to AUM/investment performance. Per user or per team pricing models are usually not compatible with quant firms, where a common research environment is available to all researchers. Embrace enterprise licensing models (for larger firms) as we’re not going to pay for the same dataset multiple times within our group company”.

A discretionary fund offered this advice:

“Our main request would be for vendors to offer an option wherein the customer in principle has access to all the data but only incurs a cost for the portions analysed, which recognises that the value of data comes from its use. SimilarWeb is a good example - they used to only offer a (steep) enterprise option at which time we were not a customer. Then they added more options including API access with different price points for different ‘API call’ budgets. This worked much better for us - we know we’ll do 100 web-traffic queries per quarter, we just don’t know which companies they’ll be about until the time arrives. This reflects our main use case being as an assist to research analysts who do a deep dive investigation on companies occasionally”.

8. Feedback

Dataset vendors should continually analyze feedback from asset managers. The data market is undergoing rapid change as are the techniques for analyzing data. For example, recent competition in

the consumer transaction category in the US is resulting in increased pricing pressure for vendors. Or, a newly released academic study on similar data may spur new interest from potential clients. For new datasets that are coming to market, it is critical to obtain feedback during the “pilot fishing” stage to best understand how to market and price their data.

In an interview that we conducted with LinkUp, an employment data provider, their CEO (Toby Dayton) discussed the process of price discovery:

“The pricing strategy was a process of discovering the right price for the right audience. Price discovering methodology consisted primarily trying to figure out how could the dataset be priced. We started looking for relevant benchmarks, other kinds of data, what kinds of datasets asset managers were buying in the market. We also looked at pricing for other market data subscriptions, such as Bloomberg, FactSet, and Capital IQ. The challenge back in 2015 was to filter reference points for pricing alternative data relative to broader-based, much more established and traditional market data.

But the conversations helped to articulate what really could be a reasonable kind of pricing for the scheme of data. Over time, a pricing framework emerged around an annual subscription price of roughly \$100,000 per year for large systematic, quant funds subscribing to LinkUp Raw. This price was tested out in conversations with who at that time we thought were potential buyers for data, which included asset managers, hedge funds, and other firms in and around the capital markets.

The conversation was a key element in defining price and LinkUp was pretty open to talking to anybody and getting feedback. Over time, combined with the positive feedback we were receiving about the product, we were able to establish conviction around their pricing for large quant user and settled on \$120,000 per year subscription fee.”

Section 4: Advice From Vendors

Eagle Alpha has been working with alternative data vendors since 2012 gaining valuable experience as to what has and hasn't worked for them throughout the process of data monetization. The advice presented in this section is based on 15 interviews and 7 seminars we conducted in various cities (Amsterdam, Frankfurt, London, Paris, San Francisco, Boston, NYC). Please refer to Appendix A for full transcripts (page 31).

The advice from successful vendors can be broken down into seven topics. These are: 1) Understanding The Needs Of Buyside Firms; 2) Difficulty of Standing Out; 3) Price Discovery; 4) Lengthy Onboarding Process; 5) Co-operation Among Vendors and Platforms; 6) Hire Data Scientists; and 7) Be Open To Feedback.

1. Understanding The Needs Of Buyside Firms

Successful vendors understand that buyside firms have different needs, so they adjust their products depending on the profile of a client.

Audit Analytics, a vendor specializing in audit and compliance data, pointed out that they pay a lot of attention to preparation prior to the initial meeting. They advise vendors to research funds, research meeting attendees, prepare all relevant material to showcase a dataset and set expectations.

Orbital Insight, an alternative data vendor specializing in geospatial data analytics, advised to understand the data gaps that buyside firms have:

"Work with your customers to understand the data gaps in their trading strategies. We've found that the buyside is willing to use our data science and engineering expertise to conduct joint R&D. If there is an opportunity to generate alpha, there are many customers in the financial services industry that are willing to work with alternative data firms to discover novel, uncommon solutions."

Alternative data vendors will go through a learning curve associated when working with buyside firms. Accern, a data design company that helps automate research and data analysis processes, stated:

"There is a steep learning curve when playing with big data and we, at Accern, carefully work with our clients to help them structure the data and tailor it to their use case. We begin this process by explaining the complexity and power of our rich analytics and history, and after we understand their requirements and finish testing, we get them on board. We learned a lot about the learning curve considering our 4-year history providing data on equity asset class, and it has helped us tremendously as we expanded into cryptocurrency, commodities and FX."

Adoption of alternative datasets is becoming more widespread, so different types of buyside firms enter the market and the vendors will need to adjust. Here is the strategy used at Advan, a geospatial data provider:

“We have seen the idea of adopting alternative data mature amongst market participants. There is already wide adoption of alternative data from the largest quantitative investors but more recently we observe more fundamental investors entering the space. As a result, we are adjusting our products to be more relevant to this community. Items such as simpler representation of the data, graphical interfaces, easier to digest signals, intuitive normalizations are all designed for these types of investors.”

Consumer Edge, an alternative data vendor providing credit/debit card transaction data, focuses on having a clear segmentation of target buy-side firms:

“Investor client base is diverse and evolving rapidly. Clients need handholding at every level of sophistication. We have four categories of client base: 1) discretionary portfolio manager/analyst; 2) discretionary data analyst/coordinator; 3) discretionary data scientist; 4) central data team.”

Asset managers pay attention to the vendors that can clearly show where the data is coming from and that there are no privacy concerns. Edison Trends, a market intelligence data provider, explains this:

“Vendors can maximize their revenue opportunity by being extremely responsive to the needs of the client. They need to make sure that the data they deliver is well constructed with low noise and privacy compliant. Vendors should create a detailed document that delivers the information customers need to showcase how their data is valuable and unique.”

It might take a considerable amount of time for vendors to understand the buy-side world. Here is how Terageance, a telecom data provider, dealt with this issue:

“The biggest challenge for us was to bridge the semantic gap between our operational world and the buy-side world. What matters in an operational context does not necessarily matter in a financial context and vice versa. So we had to take a “buy-side for dummies” crash course to understand what people were looking for (I can highly recommend Eagle Alpha’s work in that regard).”

2. Difficulty of Standing Out

Alternative data vendors need to stand out as the market is getting increasingly crowded. Advan gives a sound advice to newcomers:

“Newcomers to the alternative data arena should do proper due diligence on the markets they would like to approach. We often see vendors that don’t have the market expertise needed to succeed in today’s competitive environment. Steps such as understanding the market’s needs, create a proof of concept product, provide backtests that showcase the data are all items newcomers should consider.”

It is essential to build trust with buy-side firms. Orbital Insight explains this point:

“Selling to the buy-side requires patience, and there is one chance to make a good impression. Bringing competitive products to the table and surpassing your customers’ expectations is fundamental to building trust. Our disciplined product delivery approach allows us to continuously surprise and delight clients.”

Demonstrate how your data can help buy-side firms improve their investment process. You will also need to prove the worth of your data in different ways as Distill Analytics, a vendor specializing in employment data, highlights:

“Demonstrations to potential buyers that show the value your data has. This is different for different types of buy-side firms. What works as value demonstration to a pension fund will look very different to a demonstration to purely quantitative driven fund.”

You will also need to demonstrate that you are a reputable provider of alternative data as ScrapingHub, a vendor specializing in providing web scraped data, points out:

“For web data the biggest challenge selling to buy-side firms is demonstrating that you are a reputable provider of web data. Web scraping can be a polarizing topic and there are numerous web data providers and freelancers that have a checkered history. As a result, your reputation is your #1 concern when providing buy-side firms with web data. Firms need to trust that you are an ethical web scraper and actively minimize any legal risk associated with extracting data from the web.

Accern advises vendors to be patient versus aggressive monetization strategies:

“Data providers first have to be very responsible for what kind of data they are providing. We, at Accern, use all publicly available information and use the power of our machine learning models to learn and understand what the internet is implying. Vendors also have to be aware of the different use cases of their dataset as we have learned that aggressive data monetization can lead to unethical or illegal use of the data.”

3. Price Discovery

Price discovery can be problematic in the alternative data industry as buy-side firms are usually quite secretive and usually don't reveal their budgets. LinkUp highlights this issue:

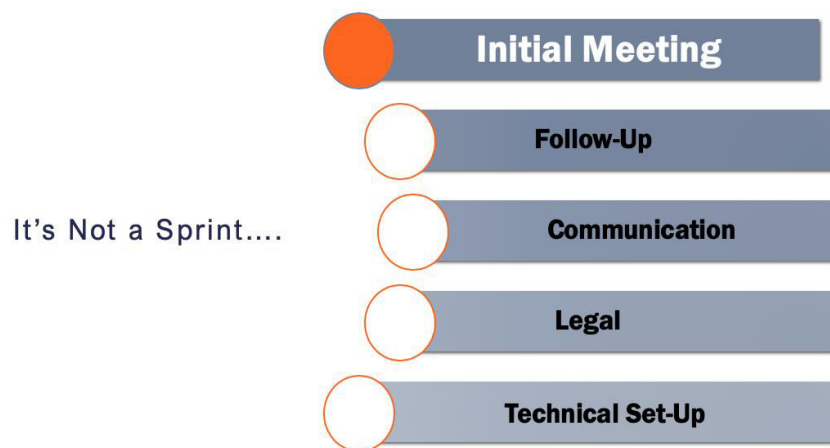
“I would say first and foremost that there is less price discovery on both sides of the equation than what we experienced early on and even just a few years ago. The market has grown, the alternative data industry is obviously more mature than it was early on, and the industry has evolved significantly in just a few short years. With a more active market for data, there are more buyers and sellers and everyone has more data points to benchmark against which creates a more established structure or framework for pricing discussions.

To be sure, though, price discovery will continue to be a fundamental aspect of the industry for the foreseeable future. The alternative data industry and the entire ecosystem around it continue to be incredibly dynamic and as things evolve, new types of data providers, new business models, new products, services, and solutions, and new distribution platforms are emerging that create new areas for price discovery and create the need for a more open conversation around pricing.”

4. Lengthy Onboarding Process

One of the biggest challenges facing alternative data vendors is the lengthy onboarding process. Audit Analytics highlights that it is not a sprint (see Figure 8) and can take 6-9 months:

Figure 8: Process



Source: Audit Analytics

Successful vendors like LinkUp experience this issue as well:

“The biggest challenge for the past two years or so has been the backlog that buyers have in onboarding new datasets for trial and evaluation. With the massive growth in all kinds of alternative datasets, it’s definitely a more crowded environment. On top of that, the current labor market is incredibly tight and the shortage of data talent is particularly acute. Putting those together means that queues to get our data into buyer’s hands are longer and longer. We’re fortunate, however, to have a unique dataset with broad coverage and strong, wide ranging alpha signals with exposure

“There are two main challenges we see in the geolocation data space. One has to do with the industry as a whole. Buyside firms have been approached from numerous alternative data vendors, some reputable, others of questionable quality. This results in elongated and sometimes inconclusive decision making processes. The other one has to do with the education and wider adoption of geolocation data. The first movers are already consumers of geolocation data and have found ways to integrate the data in their daily processes, but the great majority of the buyside is still not aware of this type of data. Geolocation being the biggest of any alternative datasets with vast applications across multiple asset classes makes it also the most complex one to work with.”

Testing takes a considerable amount of time and buyside firms don’t have unlimited resources. Edison Trends share their experience:

“The alternative data landscape is flooded with valuable new data sets and pretenders to the space. Even valuable data offerings are often minutely specific in their application. Buyside firms are inundated by these providers and struggle with resources necessary to understand, on-board, and test the variety of data sets available to them. Buysides would like to test every data set but they simply do not have the resources – the common theme expressed by panel members at events the last six months is, “be patient, we will get to you.”

Vendors need to make sure they help explain their datasets and guide buyside firms throughout the process. SESAMm, a sentiment data provider, shares its strategy:

“The main challenge is mostly the time that these firms have. Analyzing alternative data usually takes weeks to months. We aim at providing as many use cases, examples, and we also perform projects with clients who want to avoid this bottleneck.”

BizQualify, a vendor specializing in employment data, explains how they guide buyside firms during the onboarding process:

“Buyside firms are extremely rigorous and thorough in their testing, which is exactly as it should be. Due to this high level of due diligence, the testing process is sometimes lengthy and involves extensive Q&A. BQ helps explain how the data is collected and processed, what the data means from an accounting standpoint, and how the data relates to other sources of information. Furthermore, advanced statistical analysis is typically performed, which sometimes requires guidance from BQ in terms of what factors to focus on to investigate specific investment hypotheses.”

5. Co-operation Among Vendors And Platforms

Several vendors realize there is benefit in co-operating with other market players. Accern explains this point:

“Data providers have to be able to identify the value of their data and partner with other providers to enrich their data. We have learned that our clients are using more than one data set to design their strategies. This kind of overlay of data has been very profitable for Wall Street, and we are also slowly seeing a rise in vendor partnerships. For instance, a satellite imagery provider could partner with news and credit card provider so that the consumer doesn't have to look elsewhere. This kind of partnership will help vendors maximize their revenue as the overall market opportunity will expand.”

Co-operation is particularly important when vendors rely on downstream data providers as highlighted by Edison Trends:

“Vet your data source carefully as not all data is useful data. Pay attention to data source accuracy in particular and ensure that information is representative and applicable to mass audiences and behaviors.”

Teragence, a telecom data provider, discusses new trends observed in the telco industry:

“In our particular market, we see an increasing trend towards open or semi-open datasets. Whether you are talking about telco, smart cities or autonomous vehicles, these ecosystems are creating large datasets. These datasets have a primary use case (e.g. traffic planning), but their owners are often looking to explore other use cases as well and gain some extra income on the back of it. We believe that if you can articulate what kind of dataset you want, you can find it (or at least a good proxy) in the market. At the same time, we see an increasing role for middlemen who can bridge between the different spheres and can scrub, translate and format data in such a way that it becomes useful for the buy-side firms.”

Make sure you become part of the ecosystem. Finding a right partner for this market can be a good way forward as highlighted by Distill Analytics:

“Working with a partner to understand how the buy-side will use your data to make investment decisions is key. Partners are critical to gain access to the highly specialized world if you're a vendor who is unfamiliar with the area.”

BizQualify also advocates for this strategy:

“First, attending financial services industry events in major financial hubs such as New York and London is helpful to evangelize the vendors' data and learn about how analysts leverage alternative data. Second, working with a partner that has expertise and connections in the industry is an excellent way to accelerate sales efforts.”

It can also be beneficial to work with firms that complement your offering as explained by SESAMm:

“Work with firms that have a modeling capability (ideally quantitative and data science) to build examples of investment strategies that can be showcased to buy-side firms. This can help them find new clients and maximize their revenue opportunity.”

6. Hire Data Scientists and Engineers

Data science and engineering capabilities are essential as the marketplace becomes sophisticated. Consumer Edge advises vendors to embrace flexibility and simplicity (Figure 9) – clearly significant data science and engineering capabilities are needed to successfully deliver on all of the below.

Figure 9: Key Learnings from Consumer Edge

Embrace flexibility and simplicity

Data formats

Standard templates, machine readable files pre-aggregated, or granular transaction data

Traditional data delivery in every flavor

Email, S3, Web-Portal

Next-gen cloud delivery

Secure, in-browser cloud-hosted real-time analysis on granular data

Data + machine learning for investment professionals

Source: Consumer Edge

Buy-side firms are building their own data science teams but they might not know the intricacies of data collection when it comes to a specific type of alternative data. For example, data extraction and scraping isn't a core competency for buy-side firms as pointed out by ScrapingHub:

"One of the biggest data monetization trends we have observed over the last few years is that investors are increasingly using large scale web crawling to extract huge quantities of product data to track the underlying performance of e-commerce stores and CPG companies. The scope and frequency of data these firms now require has pushed many web scraping projects into the enterprise level, where significant infrastructure, experience and resources are required to build and maintain these web scraping projects. As data extraction isn't a core competency or focus of most buy-side firms, they are increasingly outsourcing this data extraction to dedicated web scraping companies or buying the datasets straight from data vendors."

Buy-side firms prefer datasets with longer history and a clear collection methodology. BizQualify provides sound advice on how to manage and present datasets:

"First, it is critical to preserve or collect the longest possible history of data and do so in a manner that ensures the historical data is point in time. This ensures no bias is introduced to the statistical testing procedure. Second, it is typically necessary to map the data to tickers or securities to make it easier for analysts to test the data. Third, it is helpful to have a technical paper explaining the methodological process being utilized to collect, parse, and analyse the data."

Data science and engineering talent is needed to fully understand the dataset's potential. Buy-side firms might not value internal backtesting results but it is beneficial to be on the same page and speak the same language as pointed out by SimilarWeb:

"Understand the full depth of what your data can provide and not just the top level values and benefits. Complex data sets can be incredibly insightful and investors want to see a full and comprehensive picture. Our experience has shown us that value begins with understanding fully the way you can provide value. Do you truly understand your own data set, the full capabilities, have you tested it internally, and productized it effectively in the way investors need to digest information. Ultimately, your goal is to empower these investors do more, and that means really understanding where the value is and how to provide it in an accessible way."

Some buy-side firms might need handholding when working with specific datasets and it might be valuable to use the data science expertise in order to create a user-friendly platform as noted by Thinknum:

“By their nature, buy-side firms seek investment insights that have not been publicized broadly and that their competitors have not taken advantage of yet. The desire to be the first movers often requires buy-side firms to rely on data that has not been interpreted for signals or recommendations. Yet, many buy-side firms do not have extensive data science capabilities that allow them to work with large complex datasets. Therefore, the biggest challenge has been creating a platform so intuitive and user-friendly that even the least technically adept investors can use to generate their own insights, test their own theses and visualize data in a manner that allows communication of the most complex data trends with their colleagues. Make sure your data is mapped using industry standards such as CUSIPs and ISINs which significantly simplify the testing process for the buy-side. Mapping to securities outside equities gives an additional edge as many buy-side firms invest in multiple asset classes now.”

SESAMm highlights the importance of use cases when speaking with buy-side firms:

“A good starting point could be to make sure that your communication is aligned with that of buy-side firms. Real quantitative and fundamental investment use cases very much help firms pre-evaluate the quality of the data acquisition and analysis process, and its value.”

7. Be Open To Feedback

Embrace feedback from buy-side firms, both positive and negative, in order to adjust your offering and develop better products and keep networking with other vendors to learn from their feedback. Audit Analytics noted how important it is to collect feedback (Figure 10).

Figure 10: Key Learnings from Audit Analytics

All Questions Are Good Questions

Pros

- Data is being looked at and reviewed
- Provides opportunities to clean up data and address any issues that are raised
- Opportunity to engage with the potential client and learn more about their data needs

Cons

- The 80/20 Rule
- If greater than 80%, then indications are that your data is not a good fit

Source: Audit Analytics

Orbital Insight advises vendors to use feedback in order to build a scalable product:

“Take the insights you’ve garnered from working with existing buy-side clients to see what opportunities exist to build a scalable and repeatable product. From our own domain expertise and numerous joint venture R&D projects we were able to quantify market demand. The buy-side ultimately wants edge. You’ll always have their attention if you are able to marry novel data with reliable correlations.”

Advan highlights that it is important to have an open eye:

“Understand the clients’ needs and adjust for these. Always keep an open eye to clients’ feedback and incorporate it to design new products. You can always price the right product at a premium.”

Teragence advises vendors to navigate the market with experienced aggregators and brokers:

“The buy-side alternative data market can be attractive, but be aware that it is also a very specific niche market with specific players, a specific language, and a specific way of doing business. Trying to “door knock” your way into this market will be hard work. Instead work with seasoned aggregators and brokers who will help you navigate this world, position it with the right players and maximise your revenue opportunity.”

Section 5: Advice From Buyside Firms

We interviewed 18 buy-side firms and asked them what advice they have for vendors of alternative data. Section 3 includes advice on sales and pricing while this section covers other topics presented below. Please refer to Appendix B for full interview transcripts (page 47).

1. Dataset Development: What are the most important attributes of the best alternative datasets?
2. Legal and Compliance: What are the key items your legal counsel requires?
3. Initial Pitch: What are the key items you want in an initial pitch? What is the best way for vendors to demonstrate the value in their dataset?
4. Q&A with Vendors: What are the most common questions that you ask data vendors and why?
5. Samples: What should vendors provide within samples?
6. Trials: What does the ideal trial look like?
7. Dataset Delivery: Which mode(s) of delivery should vendors provide?

1. Dataset Development: What are the most important attributes of the best alternative datasets?

In an [Altdata.tv](https://altdata.tv) interview with Olga Kokareva, Director of Business Development at Quantstellation, Olga advised vendors to do their homework when approaching buy-side firms:

“First of all, many vendors still think that it is so hard to reach hedge fund managers, specifically quants because they are so secretive. It is not exactly the case. Hedge funds and quants who are serious about using alternative data, they are actually employing teams of people who are going out trying to find new datasets. So it’s easy to reach out to us but it doesn’t necessarily mean that your pitch is going to be successful and it will get to at least a testing stage. I would say that the most important thing for a vendor to do is to do its homework. Try to find out about a strategy and data needs and if your dataset has nothing to do with this strategy, you’d be much better off not wasting time trying to close someone who’s unclosable.”



[Link to video.](#)

Buyside firms prefer datasets with long history and granularity in order to properly backtest the data. As mentioned earlier in the report, quantitative and discretionary firms have slightly different preferences. Quantitative firms prefer access to raw data with wide coverage and long history as explained by this EMEA-based fund with more than \$10 billion AUM:

“The most important attributes are: 1) Novelty, but within the realm of what can be reasonably applied to trading; 2) History [lots of it]; 3) Breadth of coverage/application (i.e. not a dataset that provides insight to just 10 stocks); and 4) Access to the underlying data, not pre-packaged signals.”

A more detailed answer was given by another quantitative manager:

“As a systematic fund - the best alternative datasets have:

- Broad coverage – will potentially relate to a large number of the markets we trade.
- Long history – going back to 1990-2000 is optimal. Less than five years of history is difficult for us to work with currently.
- Market impact – the market cares about the quantity that the dataset is measuring e.g. we see large moves in related news announcements.
- Distinctiveness – the data has low correlation to price trends or to common equity factors.
- Frequency – data that frequently changes over time is more valuable than slowly changing data.”

Discretionary firms are looking for datasets that will enable them to analyze a company in different ways:

“Being a long-term focused fund, we are not as interested in short term trading signals. Rather, we want to better understand the real world. Our questions can be wide ranging – from pricing to customer uptake, etc. Hence a dataset that is sufficiently granular that can enable us to answer multiple questions about a company (reliably) and consistently is worth a lot. It is more important to be deep than broad for us”.

Discretionary firms prefer datasets that would help them understand intricacies of particular businesses as highlighted by an EMEA-based fund with \$500 billion AUM:

“Ability to access insights from granular-level data (i.e. not just mass aggregates like total sales but details like customer switching, demographics, etc.) without excessive cost. If it’s the kind of data that the companies, we invest in would find insightful to make their own tactical / strategic business decisions then we’d be interested too.”

Discretionary funds who are just starting their alternative data journey would prefer simple solutions as highlighted by an EMEA-based manager with \$1 trillion AUM:

“The most important dataset attributes are: 1) Novelty, but within the realm of what can be reasonably applied to trading; 2) History (lots of it); 3) Breadth of coverage/application (i.e. not a dataset that provides insight to just 10 stocks); and 4) Access to the underlying data, not pre-packaged signals.”

When developing your datasets for buy-side firms, data provenance and methods used are extremely important as mentioned by this quantitative fund manager with \$20 billion AUM:

The most important dataset attributes are: 1) Transparency regarding the data sourcing and collection process; 2) Transparency regarding the methodology and cleansing/normalization process; 3) A robust legal and compliance process.”

2. Legal and Compliance: What are the key items your legal counsel requires?

Reputational risk is a big concern for buy-side firms as none of them would like to be featured in news outlets for negative reasons. Therefore, alternative data vendors need to pass strict compliance procedures. Some of the most important points highlighted during interviews are presented below:

“Indemnity – the knowledge that this data is gathered in a compliant manner and ownership (or at least an ability to resell) is clear.”

“Some items they require: (1) that the correct rights to license/productize the data are in place between all third-party providers involved. Compliance needs to understand in great detail where the data is sourced from and how the vendor has obtained rights to it and to relicense it. (2) That the vendor is comfortable providing a rep and warrant that the data is free of MNPI and PII. (3) That the vendor has not breached a duty to provide this information. (4) Compliance seems to prefer when a vendor has a compliance officer on staff as well as a Code of Ethics that employees must follow.”

“Often it’s just a case of getting the basics correct: 1) Legal entity is correct; 2) No indemnifying against 3rd party claims brought against the vendor; 3) Trial period is correct; 4) Definition of proprietary data are not overly restrictive; 5) Limit to number of clauses and consistency between clauses; 6) Clauses read correctly.”

“Key items are: 1) A representation that the vendor owns and/or has the right to sell the data for the intended purpose; 2) A representation that the vendor has complied with all applicable privacy policies; 3) A representation that the data does not include any MNPI or PII.”

“For multi-year deals, it’s good to have an out-clause at the end of every year.”

“Indemnity is probably the one non-starter. If we don’t have that, can’t sign the agreement.”

“No PII, no MNPI, explicit consent from originators of data to sell to third parties.”

“Right to coordinate any disputes through one of the firm’s entities where affiliates have been party to the agreement.”

3. Initial Pitch: What are the key items you want in initial pitch? What is the best way for vendors to demonstrate the value of their dataset?

It is important for buy-side firms to understand how the data is collected and what methods vendors are using to deliver a higher quality product. Case studies might not even be looked at. Advice from buy-side firms:

“Methodology and methodology. There is really no need to show us the inevitably (selection biased) backtesting results. Rather, convince us that the data capture methodology is robust and be up front about limitations. Don’t let us discover them first.”

“Demonstrate a deep understanding of the data and provide details about how the data is obtained (for example, server logs, web scraping, etc.). Explain how the data is filtered if it is incomplete.”

“A clear description of the ‘shape’ of the dataset and a clear summary of the identifiers used.”

“Proper documentation of fields is very important. How they are derived, collected.”

“Not a backtest. Good points are: validation against some hard ‘ground truth’ facts; honest description of the weak points of the data (i.e. biases, gaps, etc); examples of answering a fundamental investment thesis question.”

“Where does it come from, how does it compare to competition, and be honest about any shortcomings of the data to save us time.”

“Have detailed documentation ready to provide: methodology, schema, data dictionary, white papers, etc.”

“I’d like to know what your edge is – why do you have access to this data and how is it distinct/similar to competitor offerings? You can demonstrate value by telling me that you already have financial clients using your data profitably and by demonstrating your length and breadth of history.”

Buyside firms like to see how vendors think their data could be used, but won’t trust backtesting results presented by vendors:

“Intuition behind why they think it is useful for alpha generation (a back-test is of very limited use here, just describe the relationships the data is seeking to model).”

“Backtests are helpful – but we have to be sceptical as we don’t have access to the original process you went through so I can’t tell whether the vendor overfitted. I’m nervous if I feel that the vendor has iterated through the backtesting process multiple times and this has fed back into the creation of the dataset. Be transparent and willing to talk through your data generation approach – try not to hide behind phrases such as “machine learning”. It’s okay if the vendor is using relatively simple techniques – they are often more robust and the IP is in the engineering of getting this to work rather than the specific techniques the vendors are using.”

Discretionary managers are more likely to value your own backtesting results:

“When possible I would like to see all data sets back tested against publicly available data. A range of examples of data use is also useful.”

“Some sort of whitepaper or use case is always a good start. Doesn’t make or break anything but it helps.”

4. Q&A with Vendors: What are the most common questions that you ask data vendors and why?

Expect to be scrutinized about your data provenance and your methodology. Buyside firms will also pay a lot of attention whether your methodology changed through time which would make it hard for them to compare different time periods. Examples of questions buyside firms ask vendors:

“What are the weaknesses, and where does the data come from? We care a lot about provenance as it lies at the root of truly understanding the weaknesses, and hence knowing which questions it will be good at answering and which not.”

“How has your process changed over time? If so, can we see old revisions of the modeled data to compare (we really need to have a strong understanding of the vendor process – backtesting robustly is hard, overfitting is easy!).”

“Is the history point in time or do you have delta/change files if the history is restated?”

“Length of history, granularity, latency, delivery frequency, wide universe coverage, security tagging, lack of historical revisions, uniqueness (not distributed by resellers, no productized derivative data circulated to the market).”

“What length of history is available and how has coverage changed through time. This is key to determining how suitable the dataset will be for back-testing.”

“What tickers (if relevant) have been used? Are they corporate action aware through time?”

“We always ask compliance related questions from the get go since the answers can be deal breakers. For instance, if a vendor it is licensing credit card data we will ask them how they control for PII so that we don’t receive it. If they don’t have a satisfactory answer that vendor is not a good fit for us (or the investment community in general).”

Buy-side firms will also ask company-specific questions to understand your motivations and your history working with other funds in the space:

“What is the history of your company and customer acquisition?”

“What differentiates you from the other vendors in your space?”

“What is the size/history of the company? (Part of due diligence for us – it’s often difficult for us to gauge this from a website).”

5. Samples: What should vendors provide within samples?

It is essential to provide high quality sample data with a detailed data dictionary:

“Documentation (data dictionary, methodology, schema) describing how the data was collected, whether there have been changes in methodology over time, coverage and any caveats for working with the data is more important than a sample but a sample allows us to understand the format of the data and what is available. Each dataset can be different, but we usually ask to see all fields and history available for 2-3 financial instruments.”

“Longer is more important than wider (we look for at least 2 years minimum generally). It is better to do a bit really well than to do a lot poorly. Also, please make sure methodology and panel changes are clearly documented!”

“A detailed data dictionary explaining the sample data. i.e. it is not always easy to infer the meaning of every field and datatype on spec.”

6. Trials: What does the ideal trial look like?

Buyside firms will generally ask for trial periods of at least 3 months long and are not willing to pay for the trial data. Historical data is of no value to fund managers, they only need it for backtesting purposes:

“Free historical data which excludes the latest 6 months of data.”

“3 months of free or low-fee trial with production data.”

“We are rarely, if never, willing to pay for trials. Trials should be free of charge. Given increasing scrutiny regarding PII and MNPI, the more assurance a vendor can provide that all data is anonymized and aggregated – and that no PII will be provided – the better.”

“As close to live data as possible i.e. full breadth and history. The more data the vendor gives us, the easier it is for us to find “proof of value”. The value of historic data is entirely limited to discovering if there is value in a live subscription, we cannot make any ongoing investment decisions off of the back of a historic dump of the data. Data comes pre-mapped to some common identifiers – we don’t mind which. This makes the initial research process faster.

Time limited is fine, we try to maintain an ongoing dialog through the trial process with questions being asked on both sides. If the trial doesn’t work out we’d like to keep a dump of trial data in our systems. This is in vendors interest. Our circumstances may change - we may hire a specialist data scientist in the area, or indeed we find a complimentary dataset that unlocks the value in the existing dump of data. Keeping the data around makes it much more likely that we will get back in contact with the vendor six months down the line. In these circumstances the trial agreement can be essentially the same but without the delete clause and limits the user support available. We won’t pay for trials. We interpret charging for trials as a lack of confidence in the production data having value.”

“We actually are rather low maintenance. Give us the data and we will analyze and pepper you with questions about methodology and robustness. Best to provide us with access to the data scientists directly instead of just talking to the relationship people.”

“3 months with responsive support.”

“It depends on the dataset but generally a good trial takes time and requires support from the vendor. So, a trial that is at least three months in length with a direct contact person to answer questions and help with support issues helps make for a successful trial.”

7. Dataset Delivery: Which mode(s) of delivery should vendors provide?

Most firms are willing to work with data delivered via an API and/or FTP while some discretionary firms prefer working with dashboards:

“FTP and API, but if they provide API they should provide proper documentation and an easy query to pull all the history with all fields. API’s are ideal and more secure, but they also require more tech resources to set up.”

“API (versioned) is preferred, but we can work with s3 drops as well. This isn’t that important at the end of the day – we are willing to jump hoops for good data.”

“Most of our vendors provide data via either an FTP or API. As a systematic firm there is little we can do with distribution via email commentary. Email in general is something that is difficult for us to make robust. We are also seeing a gradual rise of use of common platforms such as Nasdaq Analytics hub or Eikon for distribution of data. We have a preference for this as it allows us to reduce the number of distinct data systems we have to depend on and makes our process more reliable.”

“Dashboards for discretionary; API/FTP for quants.”

“Direct AWS transfer, API, CSV, FTP.”

“API, FTP, S3 all work for us.”

Keep in mind that some firms might be blocked from using certain systems:

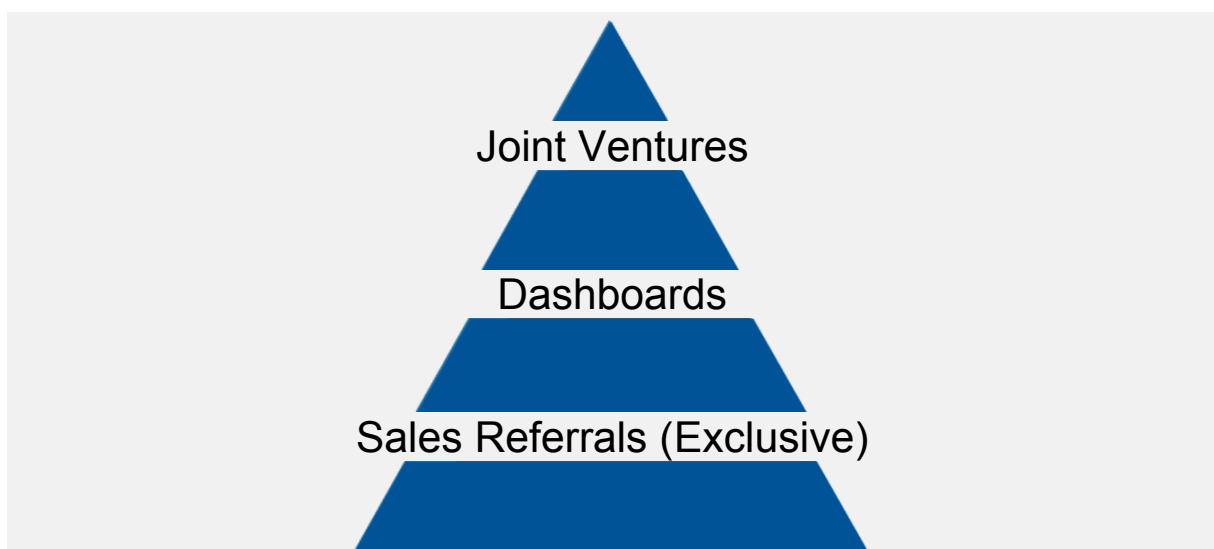
“Cloud-based downloads can be annoying if our firewall blocks access (Google for example; AWS can be a timewaster as well unless vendor is very clear about how to give access to S3 buckets etc).”

Section 6: Solutions For Vendors Of Alternative Data

Since 2012, over 900 vendors have partnered with Eagle Alpha to monetize their alternative data.

Eagle Alpha is the most flexible partner for alternative data vendors. We work with data owners in six ways (Figure 11).

Figure 11: Eagle Alpha’s Vendor Solutions





Source: Eagle Alpha

1. Freemium: we offer free marketing to our global client base for vendors who create and manage their profile on our platform.
2. Sales Referral: we consistently provide fully qualified sales leads to over 100 Eagle Alpha partners.
3. Premium Add-Ons: we have created solutions to help vendors overcome marketing, legal & compliance and sales obstacles.
4. Sales Referrals (exclusive): where Eagle Alpha is the exclusive distributor to the buy-side we provide all the premium add-ons for free to boost sales.
5. Dashboards: we build tailored dashboards for partner datasets broadening the addressable market of end users.
6. Joint Ventures: Eagle Alpha partners with the owners of unique data to build new datasets for the buy-side.

Eagle Alpha is a recognized leader in the alternative data space. For example, white papers by Citi, JPMorgan, BAML and Deloitte have all profiled the company. More importantly, customers view us as a leader as demonstrated by the following testimonials:

Referral Partnerships

“Eagle Alpha has been fantastic to work with. Not only did the business development team immediately begin making valuable introductions to investment firms and asset managers, but the Data Insights team dug into our data and began producing reports and case studies” – LinkUp.

“Eagle Alpha has helped us connect with relevant funds and assisted us with advice and business development resources. Eagle Alpha’s service is fantastic end to end. The team is a pleasure to work with and we are very excited about the future of the partnership” – Cognism.

Premium Add-on: Roadshows

“Eagle Alpha organized a roadshow with a number of investment managers that has clearly demonstrated not just the reach of Eagle Alpha’s network but also the strength of the firm’s relationships with its clients. I cannot speak highly enough about how impressed we’ve been with Eagle Alpha and how quickly they’ve had a significant impact on our data business” – Linkup.

“Eagle Alpha has been organizing roadshows with different investment funds and asset managers across Europe. These roadshows have been extremely successful for us in generating leads, as they are well connected within the industry. We are happy to be partnered with Eagle Alpha and highly recommend them” – App Annie.

"Eagle Alpha has been organizing roadshows with asset managers and different investment funds across Europe. These roadshows have been extremely successful for us in generating leads, clearly, reflecting that Eagle Alpha is well connected across the industry. We are happy to be partnered with Eagle Alpha and highly recommend them" –SESAMm.

Premium Add-on: Events

"Eagle Alpha's alternative data event was exceptional and beat our expectations. Their hard work was evident. The team truly delivered with top tier 1 client introductions, an efficient speed dating approach, and a full calendar of 1:1s! We're excited for the future and what's ahead with our Eagle Alpha partnership" – Omney Data.

"Eagle Alpha has done a super job organizing their London data showcase event. A fabulous team that has taken great strides to support vendors throughout the process, and to provide useful programming and insight from the buy-side perspective. For vendors, not to be missed!" – IPQwery.

"The Eagle Alpha conference was by far the single best one I've attended. It had the right duration, the right people and the right structure" – Semantics3.

"The size and attention was perfect for the vendors and there were some very high quality buyers in the room" – Shareablee.

"Congratulations on a terrific event! We were very impressed with how organized, efficient and effective the day was on Thursday. From the information presented by the speakers/panels (before and after the meetings), materials provided, the meetings themselves and managed expectations, we left with a better understanding of the buy-side needs" – Audit Analytics.
fund attendees and these should be moving forward – Weather Source.

"I thoroughly enjoyed the NYC event and learned immensely. I honestly believe you have created a great system for connecting asset managers with data providers" – BizQualify.

"You guys did a great job organizing the event. Looking forward to formalizing a partnership and building up business together" – PortingXS.

Premium Add-on: Altdata.tv

Dmitri Pekker (Ex OchZiff)



“Alternative data is the hot new area in investment management. However, sometimes it's difficult to separate sober analysis from the hype and wishful thinking. Altdata.tv brings you analysis and commentary by recognized experts whose opinions are based on real-world experience.”

Olga Kokareva (Quantstellation)



“Only a couple of years ago alternative data analysis was perceived as an esoteric method only used by a small handful of quants. Now it is being taken seriously as the majority of market participants from hedge funds to traditional long-only managers recognize that unstructured big data is today's true information advantage. The ability to augment traditional research processes with alternative data analysis is becoming a “must-have” to remain competitive in asset management. Altdata.tv, a new online video series, aims to provide a broader audience of market participants with access to first-hand expertise and updates on the latest developments in the space. I think Altdata.tv will become a great resource of valuable content for both data vendors looking to break into investment industry and asset managers looking to improve their research and investment process.”

Appendix

A: Interviews With Data Vendors That Have Sold To Buyside

Data Vendor #1 – LinkUp

Dataset Overview

“LinkUp's vision is centered around being the global leader in delivering perfect job market information. We do that with a very unique approach to the online job space in that we only index job openings directly from company websites. So every night, we index, right now, about 5 million job openings directly from 50,000 company websites around the world. As a result, our dataset of jobs is always current, there are no duplicate or expired listings, and because we do not aggregate jobs from job boards, we completely eliminate job board pollution.”

Question 1: How do buyside firms use your dataset in their investment process?

“The historical dataset goes back to 2007 and includes about 130 million job openings. We cover about 3,750 tickers, index jobs from roughly 200 countries around the world, and we are continuously and aggressively expanding that coverage. In addition to the data inherent in the job listings themselves, things like company, job title, job description, location, and duration, we also include a wide variety of additional data elements including company identifiers, industry classification, and Standard Occupational Code (SOC) at the individual job level, for example.

And lastly, our job market data solutions include two primary products - LinkUp Raw and Market Reports. LinkUp Raw is an XML data feed that includes the entirety of our dataset along with daily updates and a large collection of analytics and reference files. Market Reports is a feature-set within our web-based SaaS application called Job Data Engine. Market Reports allows users to create and access both standard and customizable analytics files in CSV or Excel file format across essentially every conceivable dimension of our data.

We’ve been leveraging the unique and predictive attributes in our data to generate alpha signals internally for nearly 10 years, and we began developing and selling our data solutions to buy-side firms more than 5 years ago. Our global client base includes essentially the entire spectrum of the industry – quant, fundamental, systematic, discretionary, large and small, and firms are using our data as a unique source of uncorrelated alpha at a macro, sector, and individual company level.”

Question 2: What data monetization trends have you observed over the last few years?

“Having been selling to buy-side firms for quite some time, we have seen a few significant trends over the years.

I would say first and foremost that there is less price discovery on both sides of the equation than what we experienced early on and even just a few years ago. The market has grown, the alternative data industry is obviously more mature than it was early on, and the industry has evolved significantly in just a few short years. With a more active market for data, there are more buyers and sellers and everyone has more data points to benchmark against which creates a more established structure or framework for pricing discussions.

To be sure, though, price discovery will continue to be a fundamental aspect of the industry for the foreseeable future. The alternative data industry and the entire ecosystem around it continue to be incredibly dynamic and as things evolve, new types of data providers, new business models, new products, services, and solutions, and new distribution platforms are emerging that create new areas for price discovery and create the need for a more open conversation around pricing.”

Question 3: What has been the biggest challenge selling to buy-side firms?

“The biggest challenge for the past two years or so has been the backlog that buyers have in onboarding new datasets for trial and evaluation. With the massive growth in all kinds of alternative datasets, it’s definitely a more crowded environment. On top of that, the current labor market is incredibly tight and the shortage of data talent is particularly acute. Putting those together means that queues to get our data into buyer’s hands are longer and longer. We’re fortunate, however, to have a unique dataset with broad coverage and strong, wide ranging alpha signals with exposure correlations near zero, all of which helps us shorten those queues and increase conversions.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“My advice would be to talk to as many people around the entire alternative data space as possible - buyers, sellers and vendors, partners, platforms, advisers, consultants, pretty much anyone who can share information and perspective. Through those conversations, you’d want to learn and absorb everything you can related to pretty much every aspect of the industry – different business models, various ways to think about data solutions, pricing, the challenges in bringing data to market, technical issues, sales cycles, different kinds of buyers and precisely how they differ from one another, etc.

The other thing I’d say is to make sure you don’t cherry-pick the positive feedback and data points. It’s really key to listen to as broad a set of perspectives as possible and be prepared for all kinds of

feedback, positive and negative. It's also important to be realistic and objective and find people whose perspectives are well-founded and weight their feedback appropriately.

That's why firms like Eagle Alpha are so valuable because they've been around a long time and have worked with so many firms on both sides of the equation. That experience, having seen pretty much everything in the industry from its infancy, provides Eagle Alpha with a tremendously informed perspective about the industry and what it takes to be successful. It's really hard to overestimate the value of that kind of perspective."

Question 5: What can vendors do to maximize their revenue opportunity?

"I'd pretty much go back to my earlier answer – talk to everyone you can about your data to figure out what you need to do to maximize the opportunity. The first thing to figure out is whether or not your data truly has value, but assuming it does, it's key to talk to everyone, and not just buyers but sellers, too, and all the emerging players in the space. It's really tough to get people's attention, but it's critical to get a really solid understanding of the industry, the market itself, and how to develop a really compelling value proposition around your data."

Data Vendor #2 – Shareablee

Dataset Overview

"Shareablee is the leader in audience-based social media measurement for financial services firms that want to understand the impact and effectiveness of marketing and social media campaigns on their portfolio investment firms. Shareablee captures the social media actions of more than 2 billion consumers and parses them across 700+ metrics into its global taxonomy of over 475,000 brands, publishers, TV shows, influencers, and celebrities going back to January 2013."

Question 1: How do buy-side firms use your dataset in their investment process?

"We have a couple of user personas on the buy-side. Our quant buyers want a programmatic ingestion of all of our data, their systems trade based on signals that they find in the data. Further down the spectrum, we have 'quantamental' people that are doing some quantitative work and may also be analysts that cover a vertical. And, of course, we have fundamental analysts who are pulling lots of data sources using most likely our platform in that case and not our raw data feeds. They would also

use our data to build the background information to be better prepared for the meeting with management teams.

Our platform is a SaaS offering that allows a fundamental analyst to log in and pull data by company, by brand, by individual brand entity. They can look all the way from the ticker level to the operating units, brands and sub-brands to see what is driving engagement. On the private equity side, we would also have similar use cases and they would also be platform-based rather than API. They are doing similar analysis but most of the time looking at non-public companies for investment opportunities.”

Question 2: What data monetization trends have you observed over the last few years?

“All of our early adopters were quants. They don’t tell us what they do and we almost never talk to them until renewal time. Now we are starting to get into different areas of the buy-side and we spend more time explaining what we do, and often explaining what is social. There are more meetings with more people and it’s a heavier lift for sales.

We are also glad to see that a lot of people now understand the word altdata. We are alternative data providers that sell social data. Things are becoming more institutionalized, people are looking around for alternative data sources and it’s often due to work that companies like Eagle Alpha have been doing in this space.”

Question 3: What has been the biggest challenge selling to buy-side firms?

“A significant number of companies that we go to are not allowed to use social data. They do not know what we are measuring. We can’t tell them ‘this is what we see on Twitter’ or ‘here’s what we are doing on YouTube’ as they literally can’t see it, they’re blocked. Now that we got past super smart and early adopters, the challenge of explaining is bigger. The likelihood of clients having experience with what we are measuring is surprisingly low. People are in tune in the last couple of years as they hear about Cambridge Analytics and Facebook, for example. Vendors need to clearly articulate the provenance of the data, the rights that they had to collect and to distribute the data.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“Do not expect your clients to figure it out. You need to do some lift ahead of time, create some studies and do your own analysis. You need to create a presentation that draws a clear link between your data and the dollar signs. You need to be buttoned up and have a clear story. I would also recommend people get some data scientists. Either have a contract with a data science firm that’s repeatable or with a university. You can’t expect somebody to figure it all out for you. Also, do enough research to understand how the buy-side expects the data to be organized and in tune with their workflow.”

Question 5: What can vendors do to maximize their revenue opportunity?

“The KISS model. Find a unique proposition and keep it simple. Don’t try to sell them everything you got, give them something that they can clearly understand in their evaluation process.”

Data Vendor #3 – Orbital Insight

Dataset Overview

“Orbital Insight develops geospatial data analytics to help its clients unlock societal and economic trends at a global scale. We source petabytes of satellite, drone, balloon and other unmanned aerial vehicle (UAV) data. Using computer vision and machine learning technologies, we process and interpret this data to create intelligence enabling businesses, governments, and NGOs to make better decisions.”

Question 1: How do buy-side firms use your dataset in their investment process?

“Our most popular use cases are in Consumer and Energy. Orbital Insight Consumer is a geospatial analytics platform used by quantitative and fundamental asset managers to access daily satellite and geolocation monitoring data across 260,000 retail locations. The product allows buy-side customers to

understand revenue decomposition and identify consumer trends months ahead of sell-side analysis and public reporting.

Our energy product measures crude oil storage through satellite imagery to monitor 5 billion barrels of oil storage capacity across 25,000 float roof tanks. The data gives 48 months of historical site, country, and regional level coverage – including OPEC and China.”

Question 2: What data monetization trends have you observed over the last few years?

“As a first-mover in geospatial analytics, we’ve experienced incredible interest. Today, the proliferation of alternative data conferences signal growth and customer willingness to try innovative new ways to generate edge.

For example, we’ve recently added mobile phone geolocation data to our geospatial analytics products. By aggregating anonymized mobile phone telemetry data, Orbital Insight helps the buy-side access timely accurate data while meeting their compliance needs. We saw the ingestion of geolocation data as an opportunity to build upon our existing privacy and ethics standards to become an alternative data leader consumer privacy.

As “alternative data” becomes mainstream, Orbital Insight will continue to add new datasets into a combined insight engine serving our target markets.”

Question 3: What has been the biggest challenge selling to buy-side firms?

“Selling to the buy-side requires patience, and there is one chance to make a good impression. Bringing competitive products to the table and surpassing your customers’ expectations is fundamental to building trust. Our disciplined product delivery approach allows us to continuously surprise and delight clients.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“Work with your customers to understand the data gaps in their trading strategies. We’ve found that the buy-side is willing to use our data science and engineering expertise to conduct joint R&D. If there is an opportunity to generate alpha, there are many customers in the financial services industry that are willing to work with alternative data firms to discover novel, uncommon solutions.”

Question 5: What can vendors do to maximize their revenue opportunity?

“Take the insights you’ve garnered from working with existing buy-side clients to see what opportunities exist to build a scalable and repeatable product. From our own domain expertise and numerous joint venture R&D projects we were able to quantify market demand. The buy-side ultimately wants edge. You’ll always have their attention if you are able to marry novel data with reliable correlations.”

Data Vendor #4 – Accern

Dataset Overview

“Accern’s newly launched BlackScope DS2 dataset generates daily predictive signals for S&P 500 constituents. This dataset is generated by the XYME framework using a machine learning model ensemble containing more than 200,000 models (approximately 500 models for each ticker). Our firehose API powers this dataset that has built a reputation over years of having industry-leading analytics such as news sentiment, event sentiment, source ranking, source timeliness. We have also developed an application for fundamental managers using our powerful underlying data.”

Question 1: How do buy-side firms use your dataset in their investment process?

“Buyside firms use our data to follow industry trends, as Accern's capability to track over 1 billion websites, gives them the ability to have access to a rich historical data to rely on and see what trends took place in the past to predict the future. Firms also use our signals to live trade, and the signals range from article sentiment, saturation, author rank and much more.”

Question 2: What data monetization trends have you observed over the last few years?

“The past six months have been particularly exciting and interesting as we have seen a wave of clients who want to learn more about data mining and how to structure big data. Like many other data providers, we have recognized the tailwind from the sympathy of the Cambridge Analytica scandal. The marketplace for big data is growing exponentially and is validating our CEO's vision.”

Question 3: What has been the biggest challenge selling to buyside firms?

“There is a steep learning curve when playing with big data and we, at Accern, carefully work with our clients to help them structure the data and tailor it to their use case. We begin this process by explaining the complexity and power of our rich analytics and history, and after we understand their requirements and finish testing, we get them on board. We learned a lot about the learning curve considering our 4-year history providing data on equity asset class, and it has helped us tremendously as we expanded into cryptocurrency, commodities and FX.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“Data providers first have to be very responsible for what kind of data they are providing. We, at Accern, use all publicly available information and use the power of our machine learning models to learn and understand what the internet is implying. Vendors also have to be aware of the different use cases of their dataset as we have learned that aggressive data monetization can lead to unethical or illegal use of the data.”

Question 5: What can vendors do to maximize their revenue opportunity?

“Data providers have to be able to identify the value of their data and partner with other providers to enrich their data. We have learned that our clients are using more than one data set to design their strategies. This kind of overlay of data has been very profitable for Wall Street, and we are also slowly seeing a rise in vendor partnerships. For instance, a satellite imagery provider could partner with news and credit card provider so that the consumer doesn't have to look elsewhere. This kind of partnership will help vendors maximize their revenue as the overall market opportunity will expand.”

Data Vendor #5 – Advan

Dataset Overview

“Advan is the leading financial technology company for geolocation data. The company offers exact location geofencing (polygon mapping) for quantitative, fundamental and quantamental funds. This data can provide insight into sales and sales by square foot estimates (including traffic by individual location), separate store and parking lot traffic, dwell times, and allows for forecasting. The company was established by seasoned Fintech entrepreneurs to bridge the gap between Big Data and investment decisions. Industries covered include basic industries, capital goods, consumer durables and non-durables, consumer services, energy, healthcare, technology, and others.”

Question 1: How do buyside firms use your dataset in their investment process?

“Top tier institutional investors spanning from quantitative hedge funds to fundamental asset managers have been the main consumers of Advan’s products. Each of these firms have their own unique use cases from alpha seeking functions to single name monitoring.”

Question 2: What data monetization trends have you observed over the last few years?

“We have seen the idea of adopting alternative data mature amongst market participants. There is already wide adoption of alternative data from the largest quantitative investors but more recently we observe more fundamental investors entering the space. As a result, we are adjusting our products to be more relevant to this community. Items such as simpler representation of the data, graphical interfaces, easier to digest signals, intuitive normalizations are all designed for these types of investors.”

Question 3: What has been the biggest challenge selling to buy-side firms?

“There are two main challenges we see in the geolocation data space. One has to do with the industry as a whole. Buy-side firms have been approached from numerous alternative data vendors, some reputable, others of questionable quality. This results in elongated and sometimes inconclusive decision making processes.

The other one has to do with the education and wider adoption of geolocation data. The first movers are already consumers of geolocation data and have found ways to integrate the data in their daily processes, but the great majority of the buy-side is still not aware of this type of data. Geolocation being the biggest of any alternative datasets with vast applications across multiple asset classes makes it also the most complex one to work with.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“Newcomers to the alternative data arena should do proper due diligence on the markets they would like to approach. We often see vendors that don’t have the market expertise needed to succeed in today’s competitive environment. Steps such as understanding the market’s needs, create a proof of concept product, provide backtests that showcase the data are all items newcomers should consider.”

Question 5: What can vendors do to maximize their revenue opportunity?

“Understand the clients’ needs and adjust for these. Always keep an open eye to clients’ feedback and incorporate it to design new products. You can always price the right product at a premium.

Evaluate all strategic partnerships. Do not just work with a partner for the sake of it, make sure they can bring true value to the table and that both parties can benefit from this relationship.

Alternative data consultants, data conferences and data catalogues are all important parts of the ecosystem, provided they have a positive ROI.”

Data Vendor #6 – Edison Trends

Dataset Overview

“Edison Trends delivers quality market intelligence -- accurate, detailed and real-time data updated daily from more than 11,000 brands. Data and analytics are based on actual purchase behavior from a panel of 2.5 million+ US shoppers. The data is gathered from users who have agreed to share anonymized and aggregated data about their purchasing and travel behavior through analysis of their email accounts. Through maintaining the direct relationship with users, Edison Software has clear, compliant, data access rights.”

Question 1: How do buy-side firms use your dataset in their investment process?

“Edison Trends transforms multiform unstructured email receipt data into a normalized digestible consumer transaction dataset. Edison tracks over 1 billion dollars in monthly consumer purchase data, covering online transactions from 0.5% of the overall US online population. Use cases:

- Forecast quarterly revenue. Use panel data as a model input.
- Track key purchase metrics. Customer loyalty, average order value, billing zip codes, new companies, new categories.
- Supplement fundamental data. Forecast sales and revenue, predict earnings. Identify margin trends based on product mix.
- Visualize market segment growth and competitive changes.

Examples of trends which have been identified using Edison can be found here: <https://trends.edison.tech/research>.”

Question 2: What data monetization trends have you observed over the last few years?

“Data monetization has changed the way business gets done today. For example, the use of data analytics has fundamentally changed business practices in sales and marketing functions, and R&D. Across industries, research shows that high tech and energy industries report the greatest number of functions being transformed by analytics.

Among hedge and equity funds, an increasing need for data driven and quantitative methods of predicting market trends has created a growing demand for alternative data. To meet that need, businesses are leveraging cutting edge technologies like artificial intelligence to develop abilities to analyze a variety of data inputs (i.e. crowdsourced data, historical data) and detect patterns across large audiences.

Data monetization will ultimately grow into a major source of revenues, and this year will see consumers become more aware and educated about the power of big data and the impact that data monetization has on them.”

Question 3: What has been the biggest challenge selling to buy-side firms?

“The alternative data landscape is flooded with valuable new data sets and pretenders to the space. Even valuable data offerings are often minutely specific in their application. Buy-side firms are inundated by these providers and struggle with resources necessary to understand, on-board, and test the variety of data sets available to them. Buysides would like to test every data set but they simply do not have the resources - the common theme expressed by panel members at events the last six months is, “be patient, we will get to you.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“Vet your data source carefully as not all data is useful data. Pay attention to data source accuracy in particular and ensure that information is representative and applicable to mass audiences and behaviors.”

Question 5: What can vendors do to maximize their revenue opportunity?

“Vendors should create a detailed document that delivers the information customers need to showcase how their data is valuable and unique.

Vendors can maximize their revenue opportunity by being extremely responsive to the needs of the client. They need to make sure that the data they deliver is well constructed with low noise and privacy compliant.”

Data Vendor #7 – BizQualify

Dataset Overview

“BizQualify (“BQ”) leverages filings with the IRS and Dept. of Labor that pertain to employee benefit plans. BQ tracks every company operating in the U.S. and their locations, accounting for over 95% of U.S. employment. The source filings are audited, cash based, traceable, and updated monthly, spanning 2010 - present. For approximately 4,000 tickers, BQ provides point in time factors mapped to tickers relating to employment, cash contributions, benefits, growth, profitability, and credit; and all the U.S. locations of these tickers. In addition, BQ covers in excess of 12 million U.S. private companies with over 20 million locations, providing core firmographics, growth trends, and credit indicators to enable detailed industry and macro analysis.”

Question 1: How do buy-side firms use your dataset in their investment process?

“First, client backtests demonstrate that BQ factors relating to employment and cash contributions generate superior Sharpe ratios for residual returns in Fama-French style models.

Second, clients leverage BQ locations to assist with their analysis of geolocation data and identify patterns with store locations.

Third, clients utilize BQ filings to track the market share of insurance carriers (e.g., Aetna) in the market for employer health insurance plans; and asset managers (e.g., Blackrock) in the 401k market.

Fourth, because BQ covers the entire U.S. economy, clients perform macro analysis utilizing employment, wage inflation, and corporate credit metrics.”

Question 2: What data monetization trends have you observed over the last few years?

“Numerous datasets are being created based on snippets of code that companies include on their websites (typically in a hidden fashion) that enable technological tracking tools. These code snippets empower web scrapers to deduce what technologies a company is using, e.g. marketing, cloud infrastructure, CRM, SEO optimization.”

Question 3: What has been the biggest challenge selling to buy-side firms?

“Buy-side firms are extremely rigorous and thorough in their testing, which is exactly as it should be. Due to this high level of due diligence, the testing process is sometimes lengthy and involves extensive Q&A. BQ helps explain how the data is collected and processed, what the data means from an accounting standpoint, and how the data relates to other sources of information. Furthermore, advanced statistical analysis is typically performed, which sometimes requires guidance from BQ in terms of what factors to focus on to investigate specific investment hypotheses.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“First, it is critical to preserve or collect the longest possible history of data and do so in a manner that ensures the historical data is point in time. This ensures no bias is introduced to the statistical testing procedure. Second, it is typically necessary to map the data to tickers or securities to make it easier for analysts to test the data. Third, it is helpful to have a technical paper explaining the methodological process being utilized to collect, parse, and analyse the data.”

Question 5: What can vendors do to maximize their revenue opportunity?

“First, attending financial services industry events in major financial hubs such as New York and London is helpful to evangelize the vendors’ data and learn about how analysts leverage alternative data. Second, working with a partner that has expertise and connections in the industry is an excellent way to accelerate sales efforts.”

Data Vendor #8 – SimilarWeb

Dataset Overview

“SimilarWeb is the pioneer of market intelligence and the standard for understanding the digital world. SimilarWeb’s data, sourced via SDK in partner apps tracking app while active in the foreground, provide customers with insights to help them understand, track and grow their digital market share. With our insights businesses, agencies, marketers and analysts can benchmark performance against competitors, reveal competitors online strategy, discover new opportunities, identify emerging trends and understand consumer intent and journey.”

Question 1: How do buyside firms use your dataset in their investment process?

“We are seeing a wide range a of uses for our data in the investment process. Quant funds are looking for signals through web traffic, engagement, conversions, app usage and more. Fundamental investors use it to understand strategic shifts in online and web channel performance, assess international growth, analyse developments of partnerships and more.

Finally, we’re seeing very interesting uses for distressed debt investors, as they look for the strengths and weaknesses of both a private and public company’s digital offering. For example, how successfully can a traditional brick and mortar retailers shift online?”

Question 2: What data monetization trends have you observed over the last few years?

“We have only been involved in the sector for a fairly short time, but have seen impressive demand for high quality data as well as good retention indicating that investors are seeing good returns.”

Question 3: What has been the biggest challenge selling to buyside firms?

“Data sources like these are fairly new, and education is a huge component of building an effective partnership. We have put in tremendous effort to help customers and even prospects really dig into to the potential uses and value they can gain.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“Understand the full depth of what your data can provide and not just the top level values and benefits. Complex data sets can be incredibly insightful and investors want to see a full and comprehensive picture.”

Question 5: What can vendors do to maximize their revenue opportunity?

“Our experience has shown us that value begins with understanding fully the way you can provide value. Do you truly understand your own data set, the full capabilities, have you tested it internally, and productized it effectively in the way investors need to digest information. Ultimately, your goal is to empower these investors do more, and that means really understanding where the value is and how to provide it in an accessible way.”

Data Vendor #9 – Teragence

Dataset Overview

“Teragence set out to create a data product for the mobile telecoms industry: helping operators understand how their own and their competitor’s networks perform for their end-customers. We do this through our Crowdsourcing 2.0 approach – collecting the information straight from third-party apps. This enables us to create a hyper-scale footprint and hyper-detailed network performance insights.”

Question 1: How do buyside firms use your dataset in their investment process?

“Footprint and network performance insights can be very useful for investment professionals, especially those who want to understand what is happening with a specific operator in a specific market (e.g. around network capex). But more generally, we also found that buy-side firms had more interest in other information we collected.

We realised that we have a good view of operator market share and its evolution over time as well as handset market share (down to make and model), and this on a country-by-country basis. As our data is near real-time, it gives investors a forward glimpse into what’s happening in a given telco market.”

Question 2: What data monetization trends have you observed over the last few years?

“In our particular market, we see an increasing trend towards open or semi-open datasets. Whether you are talking about telco, smart cities or autonomous vehicles, these ecosystems are creating large datasets. These datasets have a primary use case (e.g. traffic planning), but their owners are often looking to explore other use cases as well and gain some extra income on the back of it.

We believe that if you can articulate what kind of dataset you want, you can find it (or at least a good proxy) in the market. At the same time, we see an increasing role for middlemen who can bridge between the different spheres and can scrub, translate and format data in such a way that it becomes useful for the buy-side firms.”

Question 3: What has been the biggest challenge selling to buy-side firms?

“The biggest challenge for us was to bridge the semantic gap between our operational world and the buy-side world. What matters in an operational context does not necessarily matter in a financial context and vice versa. So we had to take a “buy-side for dummies” crash course to understand what people were looking for (I can highly recommend Eagle Alpha’s work in that regard). One of the basic differences is between “fundamental” players and “quantitative” players.

If you are an investment professional, this is probably an obvious distinction, but for someone outside of the sector it would be less easily understood. At the same time, it has a major bearing on how you position your data product. “Fundamentals” live and breathe a sector or a company. They have a deep appreciation of a specific industry or company and are looking for data which helps them to confirm or disconfirm specific hypotheses. In our experience, “Fundamentals” love to get their hands on raw datasets to play around with and see if they can confirm whatever hypothesis they are working on.

“Quants”, on the other hand, are essentially sector-agnostic, and are looking for datasets which have a strong, proven correlation with stock prices or other economic indicators. They care less about the actual content of the data than about the fact that it correlates. They look for historical data (preferably going back at least 2 years) so they can backtest the correlations. They care less about raw data and look for a more polished product.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“If you are looking to monetise your data in the alternative data market, look at the data you have and find a friendly market operator to discuss how it might be used. Make sure that you can access your data in an efficient way and understand the key parameters (how big is it – a sample of the “entire universe”, how much history do you have, how much polishing does it require, what companies or sectors does it cover, etc.?).

Also have a sample ready to be shared and reviewed. No one buys data without first road-testing it. And finally, as with any other emerging opportunity, you need to talk to as many people as possible so you can iterate and refine your product taking into account the feedback you receive.”

Question 5: What can vendors do to maximize their revenue opportunity?

“The buy-side alternative data market can be attractive, but be aware that it is also a very specific niche market with specific players, a specific language, and a specific way of doing business. Trying to “door knock” your way into this market will be hard work. Instead work with seasoned aggregators and brokers who will help you navigate this world, position it with the right players and maximise your revenue opportunity.”

Data Vendor #10 – Thinknum

Dataset Overview

“Thinknum creates datasets from a broad array of public online sources, capturing ephemeral information on the products, operating markets and labor markets of over 400,000 global companies across sectors. Datasets include pricing trends for individual products at specific retailers to hiring activity across macro industries and particular companies down to the location. The datasets are updated daily. The indexing and mapping methodology allows users extreme flexibility in testing an unlimited number of hypotheses with ease while the user-friendly platform enables investors to manipulate the data in a simple intuitive way.”

Question 1: How do buy-side firms use your dataset in their investment process?

“Buy-side firms use Thinknum’s alternative data to enhance their investment decisions by understanding companies from the most holistic perspective possible and in real time. Since we combine ALL public information from the web on individual companies, an investor has an opportunity to look at a company’s own strategy such as hiring plans, and geographic expansion; compare the company’s strategy to its competitors; analyze how successful the strategy execution is from a perspective of company’s customers, employees and all other ecosystem participants.

The data is structured in a way that allows investors to roll up datasets across any specific industry or geography if they want to get more macroeconomic insights or alternatively dissect the company’s own operations on the most granular level.”

Question 2: What data monetization trends have you observed over the last few years?

“Data monetization has expanded in tandem with the amount of data being created online as companies across all sectors are move their business operations online. We are seeing very traditional industries such as industrials and energy provide more and more information which allows institutional investors to gain more timely and predictive insights into businesses.”

Question 3: What has been the biggest challenge selling to buy-side firms?

“By their nature, buy-side firms seek investment insights that have not been publicized broadly and that their competitors have not taken advantage of yet. The desire to be the first movers often requires buy-side firms to rely on data that has not been interpreted for signals or recommendations. Yet, many buy-side firms do not have extensive data science capabilities that allow them to work with large complex datasets. Therefore, the biggest challenge has been creating a platform so intuitive and user-friendly that even the least technically adept investors can use to generate their own insights, test their own theses and visualize data in a manner that allows communication of the most complex data trends with their colleagues.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“Make sure your data is mapped using industry standards such as CUSIPs and ISINs which significantly simplify the testing process for the buy-side. Mapping to securities outside equities gives an additional edge as many buy-side firms invest in multiple asset classes now.”

Question 5: What can vendors do to maximize their revenue opportunity?

“Make the datasets investment strategy agnostic. It increases the pool of potential clients and opens an opportunity for unexpected applications.”

Data Vendor #11 – ScrapingHub

Dataset Overview

“Scrapinghub specializes in ad hoc / one-off web scraping projects as well as data subscriptions and large-scale broad crawls. We provide a full suite of professional service resources for the most complex of projects (data scientists, project managers, crawl engineers) as well as data subscription packages for clients looking for a regular stream of web data. We can acquire data for any need - blog and news articles, sentiment analysis, residential and commercial real estate, pricing data, etc.”

Question 1: How do buyside firms use your datasets in their investment process?

“Firms use web data as an alternative data source in their investment processes to give themselves an informational edge over the market. Examples include:

- Firms leverage product pricing data and bestsellers list data to predict ASP, revenue and market share for CPG companies ahead of reported earnings;
- Firms leverage gross reservation data of hotels and online travel websites to predict room-night growth amongst travel websites;
- Firms leverage online property listings data to predict residential and commercial real estate trends (analysing property deal flow and volume, tracking liquidity of the property market, average time from listing to completion, etc.) to give an advanced indicator of strength or weakness in housing related stocks.”

Question 2: What data monetization trends have you observed over the last few years?

“One of the biggest data monetization trends we have observed over the last few years is that investors are increasingly using large scale web crawling to extract huge quantities of product data to track the underlying performance of e-commerce stores and CPG companies.

The scope and frequency of data these firms now require has pushed many web scraping projects into the enterprise level, where significant infrastructure, experience and resources are required to build and maintain these web scraping projects. As data extraction isn’t a core competency or focus of most buyside firms, they are increasingly outsourcing this data extraction to dedicated web scraping companies or buying the datasets straight from data vendors.”

Question 3” What has been the biggest challenge selling to buyside firms?

“For web data the biggest challenge selling to buyside firms is demonstrating that you are a reputable provider of web data. Web scraping can be a polarizing topic and there are numerous web data providers and freelancers that have a checkered history. As a result, your reputation is your #1 concern when providing buyside firms with web data. Firms need to trust that you are an ethical web scraper and actively minimize any legal risk associated with extracting data from the web.

Scrapinghub achieves this with a significant focus on ethical web scraping best practices and due to the fact that we were one of the pioneering web scraping companies.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“When selling to buyside firms legal compliance and having the ability to provide historical datasets will be key to your success. You will need to demonstrate that how you extract web data is legally compliant and poses a low risk to their business. We’ve also found the ability to provide historical web data datasets is crucial if you want to gain traction with buyside firms.”

Question 5: What can vendors do to maximize their revenue opportunity?

“Buyside firms ultimately want alpha from the data you can provide. However, to incorporate alternative data into their investment process they first need to demonstrate return on data (ROD).

This requires them to benchmark and backtest their investment thesis with historical data (3 years as a minimum). If you want to maximize your revenue opportunity you need to be able to give buyside

firms the historical data they need to demonstrate ROD on your data. If you are able to show that you can provide novel data with reliable correlations you will always have the attention of buy-side firms.”

Data Vendor #12 – SESAMm

Dataset Overview

This vendor provides data and quantitative trading strategies based on emotions in web text data. The company aggregates articles and messages (social media, social trading, blogs, forums, news) in 8 languages. Sentiments, emotions and opinions are computed using proprietary Natural Language Processing algorithms. Each dataset (hourly, daily, minute frequency) is composed of 4,300 tickers for 3 to 5 years of historical data.

Question 1: How do buy-side firms use your datasets in their investment process?

“Quantitative investment teams use our sentiment and emotions indicators to develop differentiating and performing investment strategies. We also perform proof of concept projects by solving quantitative and data science use cases for long short, long only, stock picking, asset allocation, portfolio rebalancing and credit signals for clients, using both market/internal factors and alternative data.

Fundamental investment teams visualize our indicators via our web platform and we conduct custom-made studies with a focus on brands /geographic filters / media type / message subject. Custom indicators can be built and integrated directly within our clients’ systems.

We have implemented use cases for clients on different asset classes and obtained good results including on indices, stocks, Forex, commodities and crypto-currencies (intraday to monthly investment periods depending on use case).”

Question 2: What data monetization trends have you observed over the last few years?

“In most cases, asset managers not only use our sentiment data but also collaborate with us on custom POC projects to help them integrate our data into profitable open-source investment strategies (quantitative and machine learning strategies as well as fundamental studies and indicators).”

Question 3: What has been the biggest challenge selling to buy-side firms?

“The main challenge is mostly the time that these firms have. Analyzing alternative data usually takes weeks to months. We aim at providing as many use cases, examples, and we also perform projects with clients who want to avoid this bottleneck.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“Work with firms that have a modeling capability (ideally quantitative and data science) to build examples of investment strategies that can be showcased to buy-side firms. This can help them find new clients and maximize their revenue opportunity.”

Question 5: What can vendors do to maximize their revenue opportunity?

“A good starting point could be to make sure that your communication is aligned with that of buy-side firms. Real quantitative and fundamental investment use cases very much help firms pre-evaluate the quality of the data acquisition and analysis process, and its value.”

Data Vendor #13 – Distill Analytics

Dataset Overview

“Distill Analytics is a Vancouver based startup which utilizes tools, the military and intelligence community uses to profile terrorists and applies them to create profiles of business leaders. Our data tracks executives and alerts Asset Managers to changes in the C-suite’s potential to deliver positive business outcomes and maintain a highly functional, stable, and performance driven executive management team. It reveals how company individuals and team as a whole navigate business cycle transitions and systemic volatility. We can help leading Asset Managers refine their ability to identify management teams that can deliver growth and those that won’t.”

Question 1: How do buy-side firms use your dataset in their investment process?

“To gain a data-driven perspective into the mindset and motivational characteristics of a public company management team. This helps them corroborate evidence about a management team. Or receive a contrarian perspective to challenge the internal narrative of a company of interest.”

Question 2: What data monetization trends have you observed over the last few years?

“There has been a significant increase in interest and attention directed at data monetization in the last couple of years. Increased interest hasn’t translated into increased spending as many buy-side firms have worked first to orient towards the possibilities of alpha in Alternative Financial Data. However, interest does appear to be translating into spending. Data monetization will face pressure in the form of GDPR specifically, but also increased emphasis on the ways data are being sold and packaged generally. As a result, buy-side firms have increasingly become concerned with risk around the data they buy.”

Question 3: What has been the biggest challenge selling to buy-side firms?

“While our firm has the ability to help a buyer make sense of the data, we need some understanding of how a buy-side firm might use it to support their evaluation of it. Building an understanding with a potential buy-side user has been one of the largest challenges faced for two key reasons: (1) in some instances we see extreme secrecy about how a buy-side firm may use our data which accompanies an expectation that our firm have no secrets about our own processes, methods, and data generation approaches; in effect, we’re asked to disclose our underlying intellectual property. The expectation of informational asymmetry does not lead to successful commercial relationships. (2) Buy-side firms have expressed significant interest in our methods and analysis but have on occasion requested a significant amount of unpaid work prior to committing to purchase. This makes initial monetization efforts a risky proposition.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“Working with a partner to understand how the buy-side will use your data to make investment decisions is key. Partners are critical to gain access to the highly specialized world if you’re a vendor who is unfamiliar with the area.”

Question 5: What can vendors do to maximize their revenue opportunity?

“Demonstrations to potential buyers that show the value your data has. This is different for different types of buy-side firms. What works as value demonstration to a pension fund will look very different to a demonstration to purely quantitative driven fund.

Data Vendor #14 – Audit Analytics

Dataset Overview

“We specialize in normalizing non-quantitative aspects of financial reporting that have a demonstrated impact on company performance. We provide detailed research on over 150,000 active audits and more than 10,000 accounting firms. This data can be used as unique, accounting-related variables in your risk and investment models. Also, certain data sets like our financial restatement and late filing databases can be used for testing models that seek to predict a negative outcome or accounting

failure. Some variables, such as late filings, could be used as an early indicator of increased accounting risk.”

Question 1: How do buyside firms use your dataset in their investment process?

“Buy-side firms, by their nature and construct, are discrete as to how they apply datasets to their investment process. As part of our trial touch-point protocol, we ask for insight, feedback and any challenges in order to try to get a glimpse into their process.”

Question 2: What data monetization trends have you observed over the last few years?

“Fundamental analysts incorporating data into their analyses.”

Question 3: What has been the biggest challenge selling to buyside firms?

“First off, there are a lot of good challenges to selling to buyside firms. If we don’t, initially, hit the nail right on the head, we have the opportunity to learn the needs of the market and try to align our product offerings to the demand. On the other hand, we have come up against challenges that we cannot easily overcome without significant R&D.

Our data is event based in nature and might not be a fit for all funds and all strategies. Many of our datasets were developed 15+ years ago and were not designed to include point-in-time stamp, which could present a challenge.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“Be transparent. Trying to fit a square peg in a round hole won’t work. Acknowledge your strengths and be honest about your weaknesses. The funds will find out what they are anyway.

They will respect you for saving them time if you discuss the good and the not so good. Focus on what you do best.”

Question 5: What can vendors do to maximize their revenue opportunity?

“There are 3 lessons that we have learned over the past year.

1. The first is to get in front of the “right” buyers, early and often. Making sure the match is a good match, is critical. Then understand that multiple meetings with the same funds, but perhaps different attendees, is also a good thing. Many funds have several pods so as many touchpoints as you can get, the better.
2. The second thing is to be open to trials. Be generous with your data and the trial period. Nothing moves fast in this sector, so provide the opportunity for exposure – don’t let doors get closed on you.
3. The third lesson we learned is to price low for year one. The funds need to get a better sense of your datasets and how they can work with their long/short portfolio strategies. See section 2! Keep in mind if you price too high in the first year, you will be the first on the list when budgets are assessed.”

Data Vendor #15 – Consumer Edge Insight

Dataset Overview

“Consumer Edge Insight (CEI) offers exclusive street-ready big-data products engineered for alpha generation by Data Science teams and portfolio managers alike. CEI offers a number of alternative data products including CE Transact, our leading U.S. credit/debit card insight product (15M cards).”

Question 1: How do buyside firms use your dataset in their investment process?

“CE Transact products provide aggregated revenue signal, deep fundamental cohort analysis and/or granular transaction-level data feeds with unparalleled metadata. Advantages: shortened latency (T+4), better panel representativeness (9M+ daily panel), and metadata integration such as demographics (at cardholder level). All data is cleaned with CE's Advanced Tagging System.

Ground Truth service highlights company-reported results (KPIs), maintains fiscal alignments, M&A adjustments, and flags key news events that can impact the correlation between transaction data trends and reported metrics.”

Question 2: What data monetization trends have you observed over the last few years?

“Major observations of past two decades include:

- Investor client base is diverse and evolving rapidly.
- Clients need handholding at every level of sophistication.
- Paths to value-add are evolving for everyone.”

Question 3: What has been the biggest challenge selling to buy-side firms?

“We see increased use of data across our entire client base. Clients need handholding at every level of sophistication. We have four categories of client base: 1) discretionary portfolio manager/analyst; 2) discretionary data analyst/coordinator; 3) discretionary data scientist; 4) central data team.”

Question 4: What is your top tip to vendors that are exploring data monetization?

“Embrace flexibility and simplicity. Deliver data in different formats: 1) standard templates; 2) machine readable files pre-aggregated; or 3) granular transaction data. Traditional data delivery in every flavor: 1) email; 2) S3; 3) web portal.”

Question 5: What can vendors do to maximize their revenue opportunity?

“Level set on your ground truth data with careful measurements of KPIs that matter. Identify what the data is supposed to measure, catalogue structural drivers of differences (hunt for signals that can help out) and stay on top of new drivers of differences.”

B: 18 Buy-side Firms Answer The 10 Most Common Vendor Questions

Asset Manager #1

An investment management firm with a fundamental, long-term, contrarian approach.

- AUM: \$40 billion.
- HQ Location: EMEA.
- Firm Type: Traditional fundamental.
- Asset Class: Multi-Asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

“Degrees of freedom for analysis – being a long term focused fund, we are not as interested in short term trading signals. Rather, we want to better understand the real world. Our questions can be wide ranging – from pricing to customer uptake, etc. Hence a dataset that is sufficiently granular that can enable us to answer multiple questions about a company (reliably) and consistently is worth a lot. It is more important to be deep than broad for us.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

“Indemnity – the knowledge that this data is gathered in a compliant manner and ownership (or at least an ability to resell) is clear.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

“Cheaper is obviously better but we are willing to pay for value – sometimes we don’t always need a dataset over time (e.g., we just want to check a single hypothesis). Being willing to price per analysis could be interesting – obviously this would involve trust issues.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

“We actually don’t mind sales to the sell side or others. We consider data as the starting and not the ending point. Everyone will start with the most obvious questions (e.g., summing email receipts for sales estimation), but the diversity of questions explodes once you get into 2nd and 3rd order thinking. So as long as there is sufficient degrees of freedom in a dataset (those are what we are interested in anyways), we are not that concerned. In fact, non-exclusive makes compliance easier.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

“Methodology and methodology. There is really no need to show us the inevitably (selection biased) backtesting results. Rather, convince us that the data capture methodology is robust and be up front about limitations. Don’t let us discover them first.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

“Methodology and limitations as they directly affect data accuracy.”

Question 7 – Samples: what should vendors provide within samples?

“Longer is more important than wider (we look for at least 2 years minimum generally). It is better to do a bit really well than to do a lot poorly. Also, please make sure methodology and panel changes are clearly documented!”

Question 8 – Trials: what does the ideal trial look like?

“We actually are rather low maintenance. Give us the data and we will analyze and pepper you with questions about methodology and robustness. Best to provide us with access to the data scientists directly instead of just talking to the relationship people.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“API (versioned) is preferred, but we can work with s3 drops as well. This isn’t that important at the end of the day – we are willing to jump hoops for good data.”

Question 10 – Other: what other advice do you have for vendors of alternative data?

“Think about the relationship as a partnership – if the data has potential continuously (with the right interrogation), there will be a continual willingness to pay. Make sure expectations are set from the get-go and let us know about material data changes throughout our trial and post-purchase. We have terminated vendors previously for deteriorating data quality as our back-testing is continuous, not just when we are looking to buy.”

Asset Manager #2

A global asset manager with an active, long term and responsible approach to investment.

- AUM: \$20 billion.
- HQ Location: EMEA.
- Firm Type: Quantitative.
- Asset Class: Equity.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

- “Original primary data with clear ownership of the data.
- Alternatively, a really original model.
- Transparency.
- Full coverage.
- Relevance across all (or most) industries.
- Long history (>10 years).”

Question 2 – no response

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

“Realistic pricing with a scale that reflects the importance of the data in the particular investment strategy where it is employed.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

“Seems reasonable that they should sell broadly. We would employ a dataset in a very specific way in our process, so we are quite relaxed about others having access to it.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

- “Clarity.
- A clear description of the ‘shape’ of the dataset.
- A clear summary of the identifiers used.
- Some kind of backtest based on a very simple investment ‘strategy’.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

- “Where did you get the data?
- Who is using the data currently?”

Question 7 – Samples: what should vendors provide within samples?

- “Full history.
- Some data about the components of the modelled score (for modelled data).”

Question 8 – Trials: what does the ideal trial look like?

“3 months with responsive support.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

- “Probably depends on the dataset size.
- Cloud-based downloads can be annoying if our firewall blocks access (Google for example; AWS can be a timewaster as well unless vendor is very clear about how to give access to S3 buckets etc).”

Asset Manager #3

A global investment management firm with capital allocated across a diverse set of strategies pursued by a large number of specialized trading teams.

- AUM: \$35 billion.
- HQ Location: US.
- Firm Type: Discretionary hedge fund.
- Asset Class: Multi-Asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

- “Conform to a normalized metadata schema.
- Well-written, up-to-date data dictionary.
- Deep understanding of data.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

- “No PII or MNPI.
- Non-exclusive datasets.
- Vendor obtains the data legally and with consent, as necessary.
- Vendor has rights to license the data.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

- “Reasonable.
- Offer an enterprise license, which should not be much more expensive than an individual license.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

- “We avoid exclusive datasets: exclusive offerings are not desirable.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

- “Demonstrate a deep understanding of the data and provide details about how the data is obtained (for example, server logs, web scraping, etc.). Explain how the data is filtered if it is incomplete.
- While accuracy is important, consistency is crucial.
- Having an SLA in place is important.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

- “Data dictionary, 10 rows of sample data, history, latency and how data is obtained.

- If there is any processing on the raw data, list the processing logic (for example, if outliers are removed, what is the definition of the outlier and how is it removed?).”

Question 7 – Samples: what should vendors provide within samples?

- “10 rows of sample data at the beginning with data dictionary.
- During trial, full history of the dataset (with or without the latest 6 months of data).
- CSV is good, but parquet on S3 is better, especially when the size of dataset is more than 10G.”

Question 8 – Trials: what does the ideal trial look like?

- “Free historical data which excludes the latest 6 months of data.
- 3 months of free or low-fee trial with production data.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

- “SFTP/FTP.
- S3 (preferred for large datasets).
- API (preferred for low latency datasets).”

Question 10 – Other: what other advice do you have for vendors of alternative data?

- “Any sample code that offers insight is a big plus.
- Any knowledge of related datasets is a plus. Knowledge of joining related datasets is a big plus.”

Asset Manager #4

A quantitative investment unit of an asset management firm.

- AUM: \$100 billion.
- HQ Location: US.
- Firm Type: Asset Manager.
- Asset Class: Multi-Asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

“Methodology/panel stability over time; history; universe depth and breadth; solid identifier/security mapping.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

- “Representation that no PII or MNPI will be provided.
- Indemnification against third-party claims.
- Representation that the vendor has the rights to license the data.
- Representation that the data is anonymized and aggregated (i.e. not at the individual level).”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

- “Coverage – geographical and universe of securities.
- Level of history available.
- Mapping and cleanliness of panel.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

“No issue with sell-side sales. We avoid exclusive relationships given MNPI concerns.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

“An overview of collection/aggregation methodology; description of coverage and history; a list of available fields and potentially a data dictionary; confirmation that no PII would be provided and that the data is already commercially available.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

“What is their data collection methodology, what is the universe coverage, how many years of history are available, what is their client base, what sort of security mapping/identifiers are provided, what is their current client base.”

Question 7 – Samples: what should vendors provide within samples?

- “A one-time historical data dump covering at minimum three years of history.”
- “A data dictionary.”

Question 8 – Trials: what does the ideal trial look like?

“We typically ask for a minimum of three months to evaluate the data following an initial data dump.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“FTP or API delivery is preferred.”

Question 10 – Other: what other advice do you have for vendors of alternative data?

“We are rarely, if never, willing to pay for trials. Trials should be free of charge. Given increasing scrutiny regarding PII and MNPI, the more assurance a vendor can provide that all data is anonymized and aggregated – and that no PII will be provided – the better.

It is extremely helpful if a vendor is willing to utilize our paperwork, both for trials and full-fledged subscriptions. This greatly reduces the contracting process, which can be long and painful.”

Asset Manager #5

An asset manager employing multi-strategy trading approach across a broad range of securities.

- AUM: \$20 billion+.
- HQ Location: US.
- Firm Type: Discretionary hedge fund.
- Asset Class: Multi-Asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

“Auditability and provenance. We like transparency around the source of the data and we also like to see evidence of the rights to license the data.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

“Some items they require: (1) that the correct rights to license/productize the data are in place between all third-party providers involved. Compliance needs to understand in great detail where the data is sourced from and how the vendor has obtained rights to it and to relicense it. (2) That the vendor is comfortable providing a rep and warrant that the data is free of MNPI and PII. (3) That the vendor has not breached a duty to provide this information. (4) Compliance seems to prefer when a vendor has a compliance officer on staff as well as a Code of Ethics that employees must follow.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

Not disclosed.

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

Not disclosed.

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

“We find it helpful to see actual live data. In person meetings tend to be the most productive.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

“We always ask compliance related questions from the get go since the answers can be deal breakers. For instance, if a vendor is licensing credit card data we will ask them how they control for PII so that we don't receive it. If they don't have a satisfactory answer that vendor is not a good fit for us (or the investment community in general).”

Question 7 – Samples: what should vendors provide within samples?

Not disclosed.

Question 8 – Trials: what does the ideal trial look like?

“It depends on the dataset but generally a good trial takes time and requires support from the vendor. So, a trial that is at least three months in length with a direct contact person to answer questions and help with support issues helps make for a successful trial.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“CSV or platform preferred at this point but that may change if we become more sophisticated.”

Question 10 – Other: what other advice do you have for vendors of alternative data?

“Test and re-test your data. Even if there are holes in it it can still be valuable data but be prepared to disclose the weaknesses of your data set. Transparency is appreciated as is being up to speed on financial regulations.”

Asset Manager #6

A diversified quantitative investment manager employing multi-asset and multi-strategy approaches.

- AUM: \$10 billion+.
- HQ Location: EMEA.

- Firm Type: Quantitative.
- Asset Class: Multi-Asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

- “Novelty, but within the realm of what can be reasonably applied to trading.
- History (lots of it).
- Breadth of coverage/application (i.e. not a dataset that provides insight to just 10 stocks).
- Access to the underlying data, not pre-packaged signals.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

- “Confidentiality provisions.
- Representation/Warranty on MNPI and PII controls.
- Clear term and termination rights.
- Right to coordinate any disputes through one of the firm’s entities where affiliates have been party to the agreement.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

“Quants are different to discretionary managers. Historical data is both absolutely required for back testing but also of zero commercial value, so charging for historical data (either in trial or on-going license) isn’t helpful. Avoid per usage models where possible. Don’t even consider asking to link the fee to AUM/investment performance. Per user or per team pricing models are usually not compatible with quant firms, where a common research environment is available to all researchers. Embrace enterprise licensing models (for larger firms) as we’re not going to pay for the same dataset multiple times within our group company.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

“Exclusivity isn’t required (or desired). With regard to selling to strategics, this has some upside for us as most large firms have existing plumbing with the strategics and the ease of delivery likely outweighs concerns about ‘mass availability’ of the data as many firms don’t make use of the more advanced data feeds from the likes of Bloomberg anyway.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

- “Overview of dataset (coverage, history, mapping etc.).
- Intuition behind why they think it is useful for alpha generation (a back-test is of very limited use here, just describe the relationships the data is seeking to model).”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

- “Background on the company, client base, revenues to determine service continuity risk.
- What length of history is available and how has coverage changed through time which is key to determining how suitable the dataset will be for back-testing.
- What tickers (if relevant) have been used? Are they corporate action aware through time?
- What compliance controls are in place? Do you have the right to collect/distribute the data?”

Question 7 – Samples: what should vendors provide within samples?

“A detailed data dictionary explaining the sample data. i.e. it is not always easy to infer the meaning of every field and datatype on spec.”

Question 8 – Trials: what does the ideal trial look like?

“Full history for at least three months, free of charge.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“SFTP/AWS is bucket is easiest for initial cut. Ongoing delivery via RESTful API would be preferred.”

Question 10 – Other: what other advice do you have for vendors of alternative data?

“Not disclosed.”

Asset Manager #7

A multi-strategy quantitative asset manager with offices in North America, Europe, and Asia.

- AUM: \$50 billion.
- HQ Location: US.
- Firm Type: Quantitative.
- Asset Class: Multi-Asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

“Length of history, granularity, latency, delivery frequency, wide universe coverage, security tagging, lack of historical revisions, uniqueness (not distributed by resellers, no productized derivative data circulated to the market).”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

“No PII, no MNPI, explicit consent from originators of data to sell to third parties.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

“All the factors mentioned in item 1.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

“No to selling to sell side. Complete exclusivity may be a compliance concern, but limited distribution is always an interesting option.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

“Highlight how the data set stacks up on dimensions listed in 1.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

“Questions on dimensions listed in item 1.”

Question 7 – Samples: what should vendors provide within samples?

“All historical data with no live information.”

Question 8 – Trials: what does the ideal trial look like?

“60 days, free of charge, all access.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“FTP or AWS.”

Asset Manager #8

This fund manages equities globally using both fundamental and quantitative analysis.

- AUM: \$58 billion.
- HQ Location: US.
- Firm Type: Discretionary hedge fund.
- Asset Class: Equity.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

- “Legally compliant sourcing and history.
- Organized data dictionary and structure.
- Uniqueness and differentiation.
- Understanding of potential applications.
- Organizational stability and likelihood to be around in 3-5 years.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

“NDAs and legal review.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

- “Understanding of potential applications and where it can add value.
- Quants and Discretionary investors have different needs and value data differently.”

Question 4 – Sales: should vendors sell to the sellside and strategics? Should vendors offer exclusivity to a buy-side firm(s)?

“Should sell solely to buy-side firms.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

- “Summary tear sheet: coverage, history, ticker mapping, frequency, lags, pricing, trial policy, potential applications, delivery method.
- Data dictionary and/or dashboard demo.
- Sense of company and customer history.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

- “Why are you selling this data and how did you decide to sell it?
- What is the source of the data?
- What is the history of your company and customer acquisition?
- What differentiates you from the other vendors in your space?”

Question 7 – Samples: what should vendors provide within samples?

“Data dictionary.”

Question 8 – Trials: what does the ideal trial look like?

“2-3 months.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“Dashboards for discretionary; API/FTP for quants.”

Question 10 – Other: what other advice do you have for vendors of alternative data?

“In an evolving market, it would be more comforting to know that you are an expert in the space and understand the different vendors and consumers of the data.”

Asset Manager #9

A data-driven hedge fund that extracts unique trading insights from alternative data, using machine learning.

- AUM: Not disclosed.
- HQ Location: EMEA.
- Firm Type: Quantamental.
- Asset Class: Equity.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

“Length of history, breath of stock coverage, point in time, correlation with returns/revenues.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

“No PII, Vendor’s liability, chain of consent.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

“Buyer’s size (start up, small/big HF) AUM; length of history, panel size, data uniqueness.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

“We believe the real value comes from the selection and combination of data sources and the uniqueness of the analysis.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

“Panel size, length of history, use cases, sources.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

“Same as 5 plus how do they save the data point in time, if there are any updates.”

Question 7 – Samples: what should vendors provide within samples?

“API- real example API queries that work, CSV files with documentations that explain format and content of each field.”

Question 8 – Trials: what does the ideal trial look like?

“90 days, same day support, full history, self-explanatory documentations.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“Direct AWS transfer, API, CSV, FTP.”

Question 10 – Other: what other advice do you have for vendors of alternative data?

“Provide full history for free as well as provide the data for 90 days. Documentation should include real working examples and should be easy to deal with to save vendor's support time.”

Asset Manager #10

A global asset manager trading across most markets, countries and currencies. The fund has a long-term investment horizon.

- AUM: \$1 trillion.
- HQ Location: EMEA.
- Firm Type: Traditional fundamental.
- Asset Class: Multi-asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

“Simplicity – we do not have an in-house data science team, so we need most of the work to have been done for us already. When we buy the data, we are looking for the data in a format which PMs can work with – ideally we want to distil the conclusions to a simple Excel file.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

“Flexibility – we have specific terms which have to be included in all contracts, such as anti-money laundering requirements.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

“We tend to spend little with suppliers initially, until the concept is proven internally. Then we can significantly increase our spend. For us, willingness to sell a proportion of the data for a relatively small fee is important and can lead to larger purchases later.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

“Exclusivity to the buy-side is vital – if the data is sold to the sell-side we will see the conclusions through them so are very unlikely to pay for the data.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

“When possible I would like to see all data sets back tested against publicly available data. A range of examples of data use is also useful.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

“We usually ask for help with processing the data – we outsource as much as possible, so if the vendor can provide the data in a bespoke format, that would help a lot.”

Question 7 – Samples: what should vendors provide within samples?

“Examples of how the data helps long-term fundamental investors (obviously not suitable for all data sets).”

Question 8 – Trials: what does the ideal trial look like?

Time series data is more important for us. We can pay for data without a trial if the back data is good.

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

Excel (as discussed before we like the vendor to do the analysis if possible, then pass on a small data set)

Asset Manager #11

A global fundamental asset management firm.

- AUM: \$12 billion.
- HQ Location: US.
- Firm Type: Discretionary hedge fund.
- Asset Class: Multi-asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

- “Need some level of transparency how the data is collected or how the model is generated.
- If it’s a dataset with some sort of panel, that panel needs to be stable/growing/normalized.
- Point in time history.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

- “Indemnity is probably the one non-starter. If we don’t have that, can’t sign the agreement.
- For multi-year deals, it’s good to have an out-clause at the end of every year.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

- “They need to price it towards the market and realize that they aren’t the only dataset out there. Better to hook a few whales at a lower price, gain traction, learn what is going right and going wrong, and build new products.
- Multi-year deals add stability for both parties.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

- “Selling only to buy-side is more ideal, but if they work with sellside it is not a non-starter.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

- “Some sort of whitepaper or use case is always a good start. Doesn’t make or break anything but it helps.
- Proper documentation of fields is very important. How they are derived, collected.
- Don’t need to spend half of a 30 minute call on your background. Spend the time more on the firm, the data, the vision, maybe a product roadmap.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

- “How do you map tickers (if they map).
- Is the history point in time or do you have delta/change files if the history is restated?
- What made them think to sell this data to hedge fund space.
- How are they different from competitors (most of them aren’t even aware of their own competitive landscape)?”

Question 7 – Samples: what should vendors provide within samples?

- “Data dictionary detailing all fields.
- 3+ years history or max.”

Question 8 – Trials: what does the ideal trial look like?

- “Free.
- 3+ years (or maximum) history or the entire dataset (not just a sample of tickers).
- 90 days minimum time allotment.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“FTP and API, but if they provide API they should provide proper documentation and an easy query to pull all the history with all fields. API’s are ideal and more secure, but they also require more tech resources to set up.”

Asset Manager #12

An international asset management firm with offices in North America, Asia Pacific, Europe, Middle East and Africa.

- AUM: \$500 billion.
- HQ Location: EMEA.
- Firm Type: Traditional fundamental.
- Asset Class: Multi-asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

“Ability to access insights from granular-level data (I.e. not just mass aggregates like total sales but details like customer switching, demographics, etc.) without excessive cost. If it’s the kind of data that the companies, we invest in would find insightful to make their own tactical / strategic business decisions then we’d be interested too.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

“Compliance with UK / European data protection law.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

“Our main request would be for them to offer an option wherein the customer in principle has access to all the data but only incurs a cost for the portions analyzed, which recognizes that the value of data comes from its use. SimilarWeb is a good example - they used to only off a (steep) enterprise option at which time we were not a customer. Then they added more options including API access with different price points for different ‘API call’ budgets. This worked much better for us - we know we’ll do 100 web-traffic queries per quarter, we just don’t know which companies they’ll be about until the time

arrives. This reflects our main use case being as an assist to research analysts who do a deep dive investigation on companies occasionally.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

“Don’t mind. We don’t use any of the typical datasets for obviously crowdable trades so we’d rather sell to more people as the cost for us is likely to be lower.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

“Not a backtest. Good points are: validation against some hard ‘ground truth’ facts; honest description of the weak points of the data (i.e. biases, gaps, etc); examples of answering a fundamental investment thesis question.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

“See above. What are the weaknesses, and where does the data come from? We care a lot about provenance as it lies at the root of truly understanding the weaknesses, and hence knowing which questions it will be good at answering and which not.”

Question 7 – Samples: what should vendors provide within samples?

“Would rather have lots of data aged to no longer be relevant so we can survey for understanding the weaknesses than a small current sample. Or both is excellent!”

Question 8 – Trials: what does the ideal trial look like?

“Minimum 4 weeks, ideally 3 months. We’re very busy, and sometimes the best person to do the trial isn’t available. Definitely freedom for us to choose when the clock starts ticking.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“API, FTP, S3 all work for us.”

Asset Manager #13

An asset manager specializing in systematic investment strategies.

- AUM: \$85 billion.
- HQ Location: US.
- Firm Type: Quantitative.
- Asset Class: Multi-asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

- “For stock specific data, that it is mapped to financial instruments.
- Broad coverage and long history where available.
- Point in time.
- Transparency of the data source, as well as the methodology around any collection, cleaning, aggregating that has been done.
- Economic intuition for why the data may be valuable.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

Not disclosed.

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

“Different firms have different objectives and different uses in mind for the data, so depending on how a vendor sets their pricing structure they may limit who is interested in the data. For example, we have seen some firms that charge by the ticker, so as a quant shop interested in the entire data set the price becomes unreasonably high.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

Not disclosed.

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

- “Communicate what the data is that they are offering (see Q1 & Q6).
- Sometimes a case study can be helpful (but this depends).”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

- “We ask about the attributes mentioned in Q1 above.
- We ask how long they’ve been around for, how long they’ve been collecting this data, when did they start offering it to buy-side firms?
- It’s also helpful to learn a bit more about their business (ie: will they be around next year).”

Question 7 – Samples: what should vendors provide within samples?

“For a trial, we would ideally like to see a full history which helps us evaluate the dataset based on the full breadth of coverage and has the most statistical power. The most current data is less important to us for testing. Earlier in the process, we’d generally like to see a data dictionary where a few sample rows can sometimes be helpful.”

Question 8 – Trials: what does the ideal trial look like?

Not disclosed.

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“API or at least a CSV, GUI is less relevant to us as a quant firm.”

Asset Manager #14

An asset manager employing both quant and fundamental equity strategies.

- AUM: \$20 billion.
- HQ Location: US.
- Firm Type: Quantitative/Discretionary hedge fund.
- Asset Class: Multi-asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

- “History.
- Frequency of update.
- Latency/lag.
- Breadth/how many companies does it cover.
- Is it point in time.
- Reasonable price.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

“Doesn’t contain MNPI or PII and can make ensure they have the right to sell the data.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

“It would be nice for vendors to start skipping the phase where they ask for 7 figures and then bring prices down substantially in the next 3-6 months. On top of this, reasonably consider the points in 1 and recognize that there is a serious opportunity cost associated with a very expensive dataset, so we are fine not purchasing something because of price in order to allocate money to other datasets.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

“We are not against sell side and strategics. We aren’t interested in exclusivity and I would argue that selling exclusively is myopic and won’t allow for much revenue growth.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

“Where does it come from, how does it compare to competition, and be honest about any shortcomings of the data to save us time.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

“The points in part 1 mainly. I also want to know exactly when we would have been able to act on the data (including when during the day it is delivered).”

Question 7 – Samples: what should vendors provide within samples?

“We just need something that is sufficient to give us an understanding of the fields. A data dictionary is often enough but it can be helpful to have a day’s worth of data as well. We would prefer to just receive complete historical data and skip a sample though.”

Question 8 – Trials: what does the ideal trial look like?

“Provide us with the full history and it should take 3 months max. We will be in touch with questions and any issues that might elongate the process. Hopefully its done before that though. Limiting the time frame doesn’t help us and can actually make things more difficult.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“FTP or S3 for bulk/historical data. FTP, S3, or API for ongoing updates.”

Asset Manager #15

A quantitative investment management firm.

- AUM: \$20 billion+.

- HQ Location: EMEA.
- Firm Type: Quantitative.
- Asset Class: Multi-asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

- “Transparency regarding the data sourcing and collection process.
- Transparency regarding the methodology and cleansing/normalization process.
- A robust legal and compliance process.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

- “A representation that the vendor owns and/or has the right to sell the data for the intended purpose.
- A representation that the vendor has complied with all applicable privacy policies.
- A representation that the data does not include any MNPI or PII.
- Indemnity language.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

- “Historical data has no commercial value and should be made available at no charge.
- Similarly, if the data is going to be used for research purposes then there is no commercial value to us and it’s harder to justify a large spend.
- The more work our team needs to clean the data (e.g. mapping identifiers, filling gaps, ensuring a consistent methodology) the less we will expect to pay for it.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

- “We are generally not interested in exclusive agreements.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

- “Be able to describe the data sourcing and collection process in detail.
- Key characteristics of the dataset: universe / coverage, mapping, history, point-in-time, latency, delivery method/schedule.
- Trial details.
- Terms of the pricing structure.
- Have detailed documentation ready to provide: methodology, schema, data dictionary, white papers, etc.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

- “See #5 – this is how we initially determine fit and relevance for a particular project.”

Question 7 – Samples: what should vendors provide within samples?

- “Documentation (data dictionary, methodology, schema) describing how the data was collected, whether there have been changes in methodology over time, coverage and any caveats for working with the data is more important than a sample but a sample allows us to understand the format of the data and what is available. Each dataset can be different, but we usually ask to see all fields and history available for 2-3 financial instruments.”

Question 8 – Trials: what does the ideal trial look like?

- “Ideally, it should be 90 days with access to all data; although, it can exclude the most recent data (e.g. last 3-6 months) if a vendor is uncomfortable providing current data.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

“We accept:

- Google Cloud Storage / Amazon S3.
- API.
- File format: CSV for small data, Parquet for data >10GB, JSON.
- HTTP, FTP, SFTP – checksum should be provided, and revised files should be sent as new files.”

Question 10 – Other: what other advice do you have for vendors of alternative data?

- “The due diligence performed on these datasets is much more extensive than traditional data sources. Vendors should familiarize themselves with the relevant laws and regulations and be prepared for thorough reviews.
- The vendor should take the time to understand a client’s use case. There is a misconception that these datasets are immediately used for trading purposes, but the back testing and research processes can be very time-consuming and it’s not easy to find value in these datasets.
- Time stamped data is essential for quantitative funds that require point-in-time data. The time stamps should be well defined and clear what it represents (e.g. when the event happened, when the data was collected, and/or when the data was made publicly available). “

Asset Manager #16

A quantitative hedge fund.

- AUM: \$5 billion.
- HQ Location: EMEA.
- Firm Type: Quantitative.
- Asset Class: Multi-asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

“As a systematic fund - the best alternative datasets have:

- Broad coverage – will potentially relate to a large number of the markets we trade.
- Long history – going back to 1990-2000 is optimal. Less than five years of history is difficult for us to work with currently.
- Market impact – the market cares about the quantity that the dataset is measuring e.g. we see large moves in related news announcements.
- Distinctiveness – the data has low correlation to price trends or to common equity factors.
- Frequency – data that frequently changes over time is more valuable than slowly changing data.”

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

“Often it’s just a case of getting the basics correct:

- Legal entity is correct.
- No indemnifying against 3rd party claims brought against the vendor.
- Trial period is correct.
- Definition of proprietary data are not overly restrictive.
- Limit to number of clauses and consistency between clauses.
- Clauses read correctly.”

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

“Be realistic – the data is probably less valuable than the vendor anticipates. The benefit to us is what the dataset can offer in addition to other sources we can already access. A backtest that makes money – but is highly correlated to momentum trading - isn’t hugely valuable for us.

Methods of paying for data vary but most funds will typically be paying out of money they earn on performance fees. At an (optimistic) 10% performance fee, a \$100k dataset has got to improve overall performance by \$1M a year – every year – just to break even. Given the uncertainty around future performance, most funds would need to make several times more than that to be happy to pay for a dataset.”

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

“We don’t look for exclusivity. If a vendor has several financial clients, we see this as a positive. The vendor will be able to invest more in improving the dataset. All things being equal we’d much rather pay half the price (improving the net profitability of the data substantially) at the cost of having another firm accessing the same data.”

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

“I’d like to know what your edge is – why do you have access to this data and how is it distinct/similar to competitor offerings?

You can demonstrate value by telling me that you already have financial clients using your data profitably and by demonstrating your length and breadth of history.

Backtests are helpful – but we have to be sceptical as we don’t have access to the original process you went through so I can’t tell whether the vendor overfitted. I’m nervous if I feel that the vendor has iterated through the backtesting process multiple times and this has fed back into the creation of the dataset.

Be transparent and willing to talk through your data generation approach – try not to hide behind phrases such as “machine learning”. It’s okay if the vendor is using relatively simple techniques – they are often more robust and the IP is in the engineering of getting this to work rather than the specific techniques the vendors are using.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

- “What is the length/ breadth of your dataset?
- How do you distribute it?
- How has your process changed over time? If so, can we see old revisions of the models data to compare (we really need to have a strong understanding of the vendor process – backtesting robustly is hard, overfitting is easy!).
- What is the size/history of the company? (Part of due diligence for us – it’s often difficult for us to gauge this from a website).”

Question 7 – Samples: what should vendors provide within samples?

“We don’t make much use of samples. Just enough to demonstrate the data format is fine.”

Question 8 – Trials: what does the ideal trial look like?

“As close to live data as possible i.e. full breadth and history. The more data the vendor gives us, the easier it is for us to find “proof of value”. The value of historic data is entirely limited to discovering if there is value in a live subscription, we cannot make any ongoing investment decisions off of the back of a historic dump of the data.

Data comes pre-mapped to some common identifiers – we don't mind which. This makes the initial research process faster.

Time limited is fine, we try to maintain an ongoing dialog through the trial process with questions being asked on both sides.

If the trial doesn't work out we'd like to keep a dump of trial data in our systems. This is in vendors interest. Our circumstances may change - we may hire a specialist data scientist in the area, or indeed we find a complimentary dataset that unlocks the value in the existing dump of data. Keeping the data around makes it much more likely that we will get back in contact with the vendor six months down the line. In these circumstances the trial agreement can be essentially the same but without the delete clause and limits the user support available.

We won't pay for trials. We interpret charging for trials as a lack of confidence in the production data having value."

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

"Most of our vendors provide data via either an FTP or API. As a systematic firm there is little we can do with distribution via email commentary. Email in general is something that is difficult for us to make robust.

We are also seeing a gradual rise of use of common platforms such as Nasdaq Analytics hub or Eikon for distribution of data. We have a preference for this as it allows us to reduce the number of distinct data systems we have to depend on and makes our process more reliable."

Asset Manager #17

A discretionary asset manager.

- AUM: Not disclosed.
- HQ Location: US.
- Firm Type: Discretionary hedge fund.
- Asset Class: Multi-asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

- "Panel Stability in terms of history and methodology.
- Granular data that allows us to answer multiple in-depth questions.
- Transparency of methodology, and customer support when issues occur."

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

- "How the data is obtained, and proof of their rights to sell the data."

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

- "# of tickers/market cap that can be covered.
- A-la-carte options for discretionary funds.
- # of use cases investors have for the data."

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

- "No to sellside. Aggregated numbers to third party resellers or strategic partners."

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

- “Panel size (if applicable), history, update frequency, lag, skew they observed, delivery method.
- Methodology of collection/ modelling on their data.
- Best way to showcase - Use cases they have explored, show that they have tried evaluating their own data.”

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

- “Source of data – how are you getting it, and do you have the rights to sell it.
- Methodology – is this processed or raw? How is the feed being generated?”

Question 7 – Samples: what should vendors provide within samples?

- “Data dictionary.”

Question 8 – Trials: what does the ideal trial look like?

- “Free trial of around 3 months with 2years+ history.”

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

- “FTP, S3 (GZIP CSV or parquet), API – should have good documentation on data structures.”

Question 10 – Other: what other advice do you have for vendors of alternative data?

- “It would be great if vendors can show that they have done basic level of evaluation on their own data and have a good handle of their methodology.”

Asset Manager #18

A newly launched quantitative hedge fund.

- AUM: Less than \$1 billion
- HQ Location: US.
- Firm Type: Quantitative.
- Asset Class: Multi-asset.

Question 1 – Dataset development: what are the most important attributes of the best alternative datasets?

Coverage, accuracy, and whether the dataset is relevant. A lot of these datasets are not really relevant, and they have to be combined with other datasets. For example, web traffic data is almost worthless on its own, but combined with transaction data it gives you granularity.

Question 2 – Legal and compliance: what are the key items your legal counsel requires?

- Abiding by all current laws.
- Accepting indemnifications and responsibilities for any loopholes or problems with the data.
- No prior record of problems.

Question 3 – Pricing: what factors should a vendor consider when developing its pricing strategy?

The bottom line is – how is your customer going to make money with it. I think vendors have been completely removed from understanding that. It does not matter how unique your dataset is if your customer cannot use it to make money.

Question 4 – Sales: should vendors sell to the sellside and strategics? should vendors offer exclusivity to a buy-side firm(s)?

It matters a lot to me whether or not it is available to the sellside, but from their perspective they have to sell it to the market with most dollars.

Question 5 – Initial pitch: what are the key items you want in an initial pitch? what is the best way for vendors to demonstrate the value of their dataset?

- What KPIs have been developed with your product.
- What the backtested error rate is.

Question 6 – Q&A with vendors: what are the most common questions that you ask data vendors and why?

Coverage, backtested error rate, and KPIs.

Question 7 – Samples: what should vendors provide within samples?

Full history is preferred, but at least 2 years of history.

Question 8 – Trials: what does the ideal trial look like?

At least 6 months' worth of access. A lot of people try to do like 1 month, but there are so many moving parts that you need to get right. You are trying to figure out where the data plugs in and how to use it. It's a long sales cycle.

Question 9 – Dataset delivery: which mode(s) of delivery should vendors provide?

Flat file, S3, API.

Glossary

Quantitative Fund

Quant funds are funds that use quantitative analysis using computer-based models to inform their decisions on whether to buy or sell securities. Quant strategies are often used by hedge funds as well as mutual funds. They offer de-correlation benefits in that managers are not stock picking and instead are relying on a computer's dispassionate assessment of price and value criteria (source: Financial Times Lexicon). They are sometimes referred to as "systematic" funds.

Discretionary Fund

Discretionary investment management is a form of investment management in which buy and sell decisions are made by a portfolio manager or investment counselor for the client's account. The term "discretionary" refers to the fact that investment decisions are made at the portfolio manager's discretion. This means that the client must have the utmost trust in the investment manager's capabilities (source: Investopedia).

Hedge Fund

Investment fund that aims to meet high targeted returns (absolute returns) using aggressive trading strategies, often across several asset classes. Hedge fund managers can take both long and short positions to maximize returns (long/short funds) or use other methods, sometimes purely speculative. The need to achieve high returns can involve the movement of large amounts of funds from one market, asset class or currency to another, involving a high degree of risk to the investor (source: Financial Times Lexicon).

Fundamental Long Only

Fundamental investors select securities to invest in based on the business outlook and the valuation of securities in the marketplace. The term "long only" refers to restrictions that prohibit short selling. Many mutual funds apply a fundamental long only approach as the desire of retail investors is to grow their savings over time while limiting their risk. Pension funds, trusts, endowments and sovereign wealth funds also have assets invested in fundamental long only strategies to balance out the risk/reward trade-off of their investment portfolio.

Private Equity

Injection of funds by specialized investors into private companies with the aim of achieving high rates of return. This is equity or quasi equity funding provided by professional investors to mature private companies with the aim of gaining a financial return through interests, dividends and capital gains at exit (source: Financial Times Lexicon).

Macro fund

Global macro hedge funds are actively managed funds that attempt to profit from broad market swings caused by political or economic events. Global macro hedge funds are market bets around economic events. Investors use financial instruments to create short or long positions based on the outcomes they predict as a result of their research. A market bet on an event can cover a wide variety of assets and instruments including options, futures, currencies, index funds, bonds, and commodities. The goal is to find the right mix of assets to maximize returns if the predicted outcome occurs (source: Investopedia).