



Sell Your Data to Wall Street

The comprehensive guide to monetizing your data assets for professional investors



TABLE OF CONTENTS

Executive Summary	1	Pricing Your Data	16
Why Wall Street Wants Your Data	2	Data Attributes	16
How to Seize the Opportunity	4	Assets Under Management (AUM)	16
Building a Data Product	5	Return on Data Investment	16
Data Quality	5	Tiering	17
Predictive Power	6	Exclusivity	17
Commercial Value	8	Case Studies	18
Packaging Data for Consumption	10	Auto Insurance	18
Delivering Your Data	11	Commodities	19
Marketing and Selling Your Data	12	Business Health Metrics	19
Strategic Partnerships	12	Next Steps	20
Marketing Channels	12	Productization	20
Data Collateral	12	Infrastructure and Delivery	20
The Sales Process	13	Marketing and Sales	21
Protecting Yourself and Your Data	14	About Quandl	22
		Contact Us	23



EXECUTIVE SUMMARY

Professional investors who fail to beat their benchmarks quickly go extinct. Beating benchmarks, known as “generating alpha” in Wall Street’s parlance, is extremely difficult to do. In fact, it keeps getting harder, a reality illustrated by deteriorating hedge fund returns. This is why professional investors are eager for new sources of alpha. **Find alpha or die** is the stark reality for every portfolio manager.

At the same time, we are living through a data explosion. In the world’s 2.5 billion gigabytes of data, Wall Street sees its savior. The prevailing belief is that this data — and the predictive power it promises — is the most powerful alpha source to emerge in the last quarter century. The \$3 trillion hedge fund industry is about to bet its future on it.

This is what we call “**alternative data**” and it refers to unique information content not previously known to financial markets but nonetheless powerful to professional investors. It is not available from traditional market data providers. By 2020, spending on alternative data and its associated infrastructure is expected to exceed USD 7 billion.¹

That you’re reading this white paper indicates you or your firm may own an alternative data asset that could be sold to a Wall Street audience and generate a significant new stream of revenue for your business that goes directly to your bottom line.

¹ <http://www.opimas.com/research/217/detail>

If you expect to sell your data to investors, there are key considerations you must take into account:



Data productization — you need to understand how Wall Street evaluates, transforms, and uses data



Infrastructure and delivery — you need to have the infrastructure to reliably deliver your data at the same time every day, 365 days a year, in a format consumable by your client



Marketing and sales — you need a strong and personalized distribution channel into this audience



Privacy — you need to equip yourself with knowledge about global data privacy regulations and ensure your data assets do not include any personally identifiable information



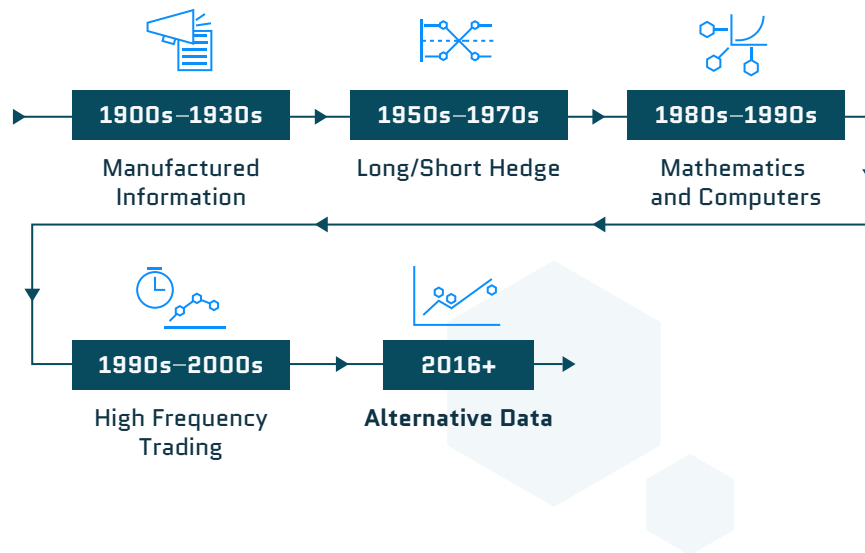
Pricing — you need insight into how investors will value your data so you can develop the right pricing strategy

This white paper will provide you with an in-depth review of all of these considerations, to help you develop a revenue stream from your data.

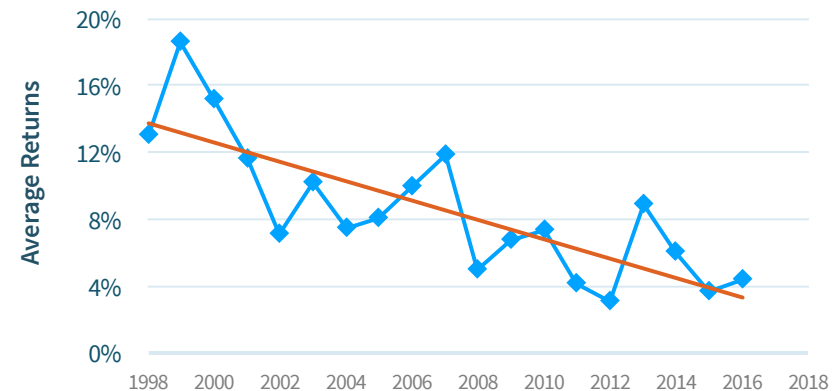
WHY WALL STREET WANTS YOUR DATA

Investors are always looking for an edge: a source that affords the opportunity to earn excess returns on an investment against a benchmark. This is known as generating “alpha” and it is extremely difficult to do.

Over the past hundred and fifty years, many alpha sources have come and gone. In the 1950s, it was the invention of long/short equity strategies by the very first hedge funds. In the 1980s, mathematics and computers held an edge over handheld calculators. In the early 2000s, it was high-frequency trading. These were all strategies or tools that, for a time, gave the people who had access to them an advantage over everyone else. But as they became more and more widespread, their advantage dissipated and investors had to move on to the next thing.



Alpha sources have been lacking in recent years, a fact that is illustrated by lower hedge fund returns over the same period of time. Investors are eagerly searching for new alpha sources.



Five-year rolling hedge fund returns. Source: Credit Suisse Hedge Fund Index

At the same time, the data economy in which we find ourselves is exploding and the prevailing belief is that the predictive signals buried within the mountains of data we are producing represent the next source of alpha. The future of investing will rely heavily on mining the information that is produced as a result of today’s increasingly digital and connected way of living and doing business.



35 zetabytes of data
in the world by 2020



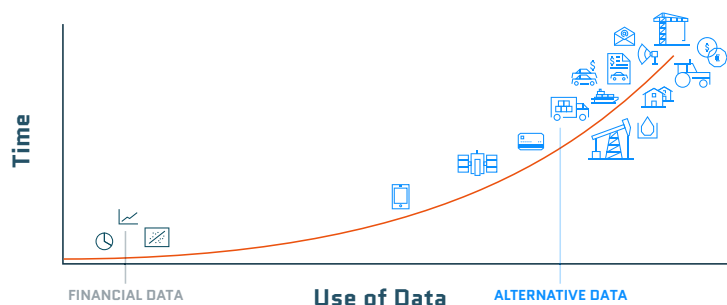
50 billion connected
devices by 2020



90% of all alternative
data is noise

This is what we call “**alternative data**” and it refers to unique information content not previously known to financial markets but nonetheless powerful to professional investors. It is not available from traditional market data providers. By 2020, spending on alternative data and its associated infrastructure is expected to exceed USD 7 billion.

An ever growing number of hedge funds are going “all in” on alternative data, betting that the information advantages it promises will be the primary driver of alpha over the next decade. The data that you produce as a result of your business operations — be it related to logistics, insurance, healthcare, telecommunications, software usage, human resources or other germane topics — is alternative data.



Alternative data is quickly becoming the key driver of investment strategies.

Examples of datasets currently giving investors an edge in the markets include:



Automotive — GPS and telematics data from cars



Big Ag — Sensors placed in agricultural crops, fields, and equipment



Business to Business — Business health metrics from credit platforms



Customer Insights — App downloads and usage from telecommunications providers



Human Resources — Job listing data from digital job boards



Logistics — Cargo transport data recorded through device-equipped fleets



Retail — Consumer footfall and transaction data

The landscape continues to evolve and some early sources of alternative data, such as sentiment data and satellite data, have already become almost as commonplace as stock price history and fundamentals. It is no surprise, then, that banks and asset managers are ravenous for new data assets. They must keep trying to find that edge.

Organizations are now queuing up to pay tens of thousands of dollars for untapped data, including data you might already have.

HOW TO SEIZE THE OPPORTUNITY

If you believe you're sitting on valuable data and you now understand that investors are potentially interested in buying it, what's next? To start, you need to understand the key considerations for getting your data into their hands.

When we talk about "Wall Street", what we mean is anyone on the "buy-side" of capital markets — those who invest in stocks and bonds on behalf of others. They could be on Wall Street or elsewhere — but we find the vast majority of early adopters in this space are in New York.

On the buy-side, hedge funds are the most aggressive players in terms of their willingness to seek excess returns, or alpha. Furthermore, among the hedge fund ecosystem lives a type of fund that is exceedingly sophisticated in its use of data: **the quantitative hedge fund**. This is your best potential customer. You can think of them as alpha hunters. To deliver on the bold promises they make to their investors, quantitative hedge funds will invest heavily in data.

That is not to say that other buy-side investors won't purchase data but the quants are a good place to start. However, this is not an industry where you can cold call an asset manager and suggest selling them FTP access to your data. These professionals are notoriously difficult to reach. More importantly, there is a lot more that goes into selling an investor a data asset.

If you want Wall Street to take notice of your data, here are few key considerations:

- **Data Productization** — you need to understand how Wall Street evaluates, transforms and uses data
- **Infrastructure and delivery** — you need to have the infrastructure to reliably deliver your data at the same time every day, 365 days a year, in a format consumable by your client
- **Distribution** — you need an "in" with this crowd to get them to look at your data in the first place

The rest of this document covers each of these considerations in detail.

BUILDING A DATA PRODUCT

Data is always a means, never an end. In the specific case of Wall Street customers, the end is making better trading and investment decisions. Your data is valuable insofar as it can help Wall Street clients — traders, investors, portfolio managers — make better decisions about stocks, bonds, currencies, commodities, economic variables and more.

But most raw data assets are very far from offering direct insight into market decisions. To offer that value, they must be “productized”: converted from their original raw form, into a new form that has clear and current value for a Wall Street audience. This productization process has several stages:

- 1 Evaluating data quality:** The first stage in productization is evaluating data quality. Data must be clean, consistent, accurate, timely, well documented and well structured; without these basic attributes, it is impossible for an investor to use a given dataset.
- 2 Detecting predictive power:** Assuming your data meets basic quality criteria, the next step is to look for a “signal” — a pattern in your raw data that predicts the behavior (price) of some financial market variable (asset).

- 3 Establishing commercial value:** Ideally, the signal discovered in the previous step can be translated into a data-driven trading or investment strategy that produces positive returns. The larger these potential returns, the more valuable the data.
- 4 Packaging the data for consumption:** Analysts and investors typically do not want to work with raw data; so it must be converted into a format that is preferred by them: often an index or a subset of the original data.

We explore each of these in greater detail below:

1. DATA QUALITY

In the absence of high-quality data, even the best analysis fails. Hence a basic requirement for salable data is that it fulfills certain quality criteria. Consider the following criteria:

Accuracy: The data must be as accurate as possible. Gaps, spikes, errors and outliers in the data all make it less trustworthy and hence less valuable.

Consistency: The data must be methodologically consistent and uniform both across space (i.e., across companies, geographies, products or whatever subject is covered) and across time (i.e., history). A dataset whose underlying calculation methodology changes from month to month, or from company to company, is hard to work with.

Documentation: All fields and indicators need to be precisely defined. Clients do not know your data as well as you do; educate them. Precise definitions help testers correlate your data against the right “target” variables.

Ticker Mappings: Many datasets contain information about public companies. Ideally, this information should be labeled with the tickers of these companies, to allow easy analysis across multiple datasets. The ticker is a unique identifier (e.g., AAPL) and is much easier to map across datasets than a company name (e.g., Apple, Apple Corp., Apple Inc., Apple Computer). Note that CUSIP, ISIN and OpenFigi are acceptable in place of tickers.

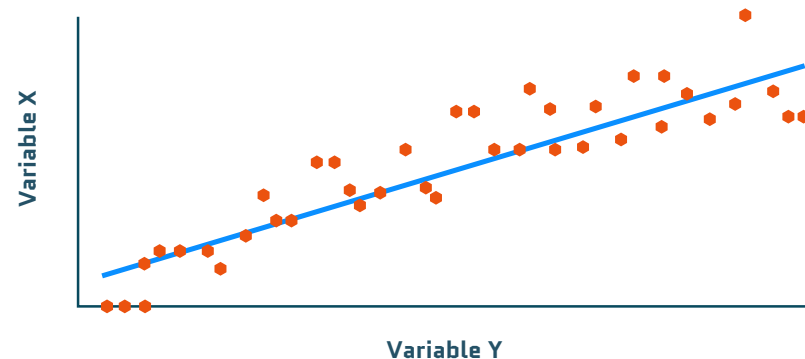
Historical Publication Timestamps: When testing the historical performance of a dataset, investors need to know exactly when each data point became available, to avoid look-ahead bias. For an investor to use weekly sales data, for example, she needs to know when each week’s number becomes available — which may be before, during or after the week in question. This is one of the most common reasons that prospective data partners fall short.

2. PREDICTIVE POWER

Assuming an acceptable level of data quality, the next question is to find out if the data has “predictive power”. Intuitively, we want to know if the data holds some pattern or signal, which can be used to predict the subsequent movements of an asset price or other similar financial indicator. This prediction must then hold up to a variety of statistical tests.

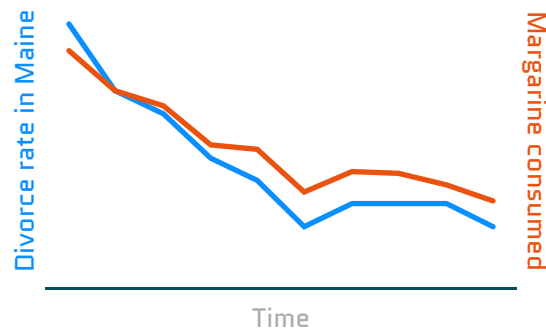
Some of the factors that go into evaluating predictive power are described below:

Correlation: The basic indicator of predictive power is correlation. If a dataset exhibits a relationship with a stock price, economic indicator or some other market variable, then it has the potential to be valuable. The correlation must be statistically significant, which will be measured through rigorous analysis and testing.



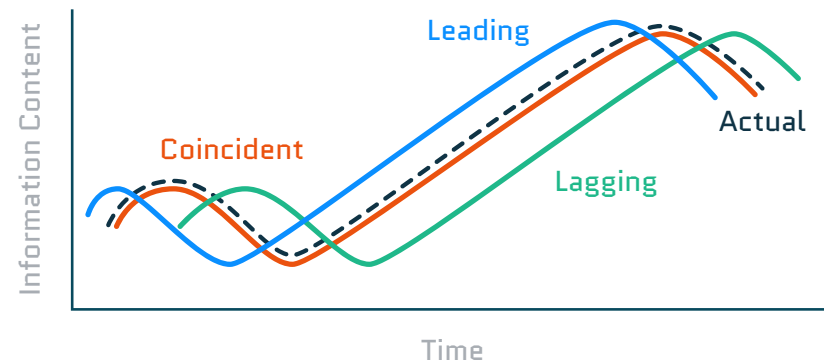
An example of a strong positive correlation between two variables, x and y. This graph shows that they move in tandem — if one rises, so does the other. Your data, for example, might be correlated with a stock’s price movement.

But beware of false correlations. For example, new car insurance policies are correlated with new car sales; this is an obvious, intuitive correlation, which justifies using the former to predict the latter. Rainfall in Mongolia might be correlated with IBM sales; however, without an intuitive explanation as to why, this is likely to be discarded as a statistical coincidence.



The divorce rate in Maine is highly correlated with the per capita consumption of margarine but it makes no intuitive sense. Source: <http://tylervigen.com/spurious-correlations>

Leading Indicator: Alternative data is valuable to Wall Street only insofar as it predicts the behavior of some market variable. Contemporaneous correlations are insufficient; the signal must be available sufficiently ahead of market action so that one can trade and profit from it. Unfortunately, many of the correlations we discover are either coincident or lagging; hence, they do not provide insight for investors.



Leading indicators predict the behavior of some market variable. Coincident or contemporaneous indicators occur at the same time of the market variable's movement. Lagging indicators occur afterward.

Length of History: All else being equal, a dataset with a longer history is preferable to one without. A long data history allows users to statistically test the predictive powers of the data. Less than two years of historical data is usually insufficient unless your signal is extremely powerful. Four years is generally considered acceptable; eight years is the gold standard. However, keep in mind that investors are trying to figure out how to adapt their models to shorter histories, so they can take advantage of more of the alternative data that is available.

Completeness of Coverage: The best datasets cover as much as possible of the relevant “universe” — whether that’s an asset class, a product or some other unit. A dataset with information about stocks should cover all stocks in a defined market. A dataset with information about products should cover as many products as possible. If you have a dataset that predicts the performance of just one retailer, you will only have a handful of customers for it.

No Revisions or Backfilling: Data that is revised after the original publication date is worthless for backtesting. When simulating behavior at a given time, only the data available at that time can be used. Revisions and backfilling negate this purpose. Take consumer transaction datasets, for example. When new consumers are added to the dataset's survey base, their past transactions are included. This changes the historical data and makes rigorous analysis a lot harder. If this data is to be added, it must be properly version controlled.

No Bias: Different types of biases can exist in a dataset. Here are a few examples to avoid:

- **Survivorship bias:** Historical data that includes only those companies that still exist today, omitting shuttered businesses
- **Look-ahead bias:** Running a test or simulation that uses data that was not available during the period being analyzed
- **Selection bias:** Choosing a subset of your space (list of assets, list of indicators, time period) not at random; the subset might not be representative of the whole

Statistical Significance: Length of history, strength of correlation, width of coverage and intuitive economic priors are all tools to evaluate statistical significance. The predictions made by alternative data are inevitably noisy, but the more you can prove significance for a given dataset, the more trustworthy and valuable that dataset becomes.

3. COMMERCIAL VALUE

Let's say you've checked every box up until this point. You'll now want to ask yourself: Does my dataset have commercial value? You could have a perfectly clean, reliable, predictive dataset — but if investors cannot make money off it, then you cannot sell it. Here are some factors that determine the commercial value of a dataset:

Data Edge: Your customers will look to gain an edge from your data. It needs to be either faster or more accurate than what they are using today or it must offer a unique insight previously unavailable to them. This seems simple and obvious but one thing that often gets overlooked is just how accurate Wall Street's estimates actually are. These estimates come from analysts who have been honing their craft for decades. It's not sufficient to make an accurate prediction; you have to do it faster and better than Wall Street can.

Monetization Strategy: Customers should be able to convert the edge offered by your data into trading or investment profits, via a clear and straightforward monetization strategy. The more direct the connection between your data and a profitable trading strategy, the more valuable your data.

Deep Market: The best monetization opportunities are found in large, liquid markets. Data that predicts the behavior of small or illiquid securities is inherently less valuable. For example, your data may be predictive of penny stock movements. But very few professionals will be interested in trading penny stocks, which

means your data may not find a hedge fund audience.

Uniqueness and Replicability: The more unique your data, the more valuable it can be. Are there others who can replicate either the data you have or the signal you are likely to produce? Are other versions of your data available? Are there proxies that achieve the same purpose? If the answer to all these questions is no, then you likely have a very valuable data asset.

Exclusive Access: If everyone in the market has access to a given dataset, it won't have much value. For this reason, it's important to restrict access to a few select customers. Exclusive distribution partnerships help create a sense of value for clients.

Table Stakes Potential: New and alternative datasets typically begin life as exclusive assets available only to a select few, at a very high price. Gradually, knowledge of the data diffuses through the market and its value diminishes. However, a few datasets survive and become "table stakes": they are no longer exclusive but become required purchasing by everyone in the market. At that point, participants who don't have these datasets are at a disadvantage. The most valuable datasets are those that have the potential to become "table stakes".



4. PACKAGING DATA FOR CONSUMPTION

Most Wall Street analysts and investors typically do not want to work with “raw” data; they prefer well-defined, compact datasets that they can work with using their existing toolkit. For this reason, it is important to package raw data assets into consumable data products for Wall Street. Some of the factors that go into this are described below:

Index or Estimate Creation: Raw data assets, such as satellite image files, shipping manifests and consumer transaction receipts, while rich in information, are not easy to use for most Wall Street traders and portfolio managers. So these datasets must be reduced to more tractable forms: indexes, estimates or other easily manageable numbers. The key is to achieve this reduction without stripping the underlying data of relevant information content.

Marketing Collateral: While hedge funds are among the most sophisticated investors in the market, the world of alternative data is still new to them. Each alternative dataset presents a brand new analytical, economic and trading-strategy challenge. It is critical to educate these customers on the usage and scope of this data, via onboarding documents, white papers, demo spreadsheets, example scripts and other marketing collateral. Empirically, we have found that rich and sophisticated database-level marketing collateral dramatically increases the value of a given dataset.

Compliance Issues: Wall Street investors are justifiably paranoid about issues of compliance: specifically, violating inside information

(MNPI) and privacy (PII) regulations. An important part of the data productization process is to ensure that no such issues arise; the data product should be legally and logistically clean, with clear IP ownership and no question of MNPI/PII violations, auditable and access-controlled. Ensuring compliance while being careful not to over-scrub the data such that it loses signaling power is a hard balance to get right.

Productizing a dataset is thus a challenging endeavor. Information quality, predictive power, commercial value and consumable packaging are all necessary to create a dataset that customers will pay for. And even then, the work is not done; the next step is to build an infrastructure that can deliver this data efficiently and reliably.



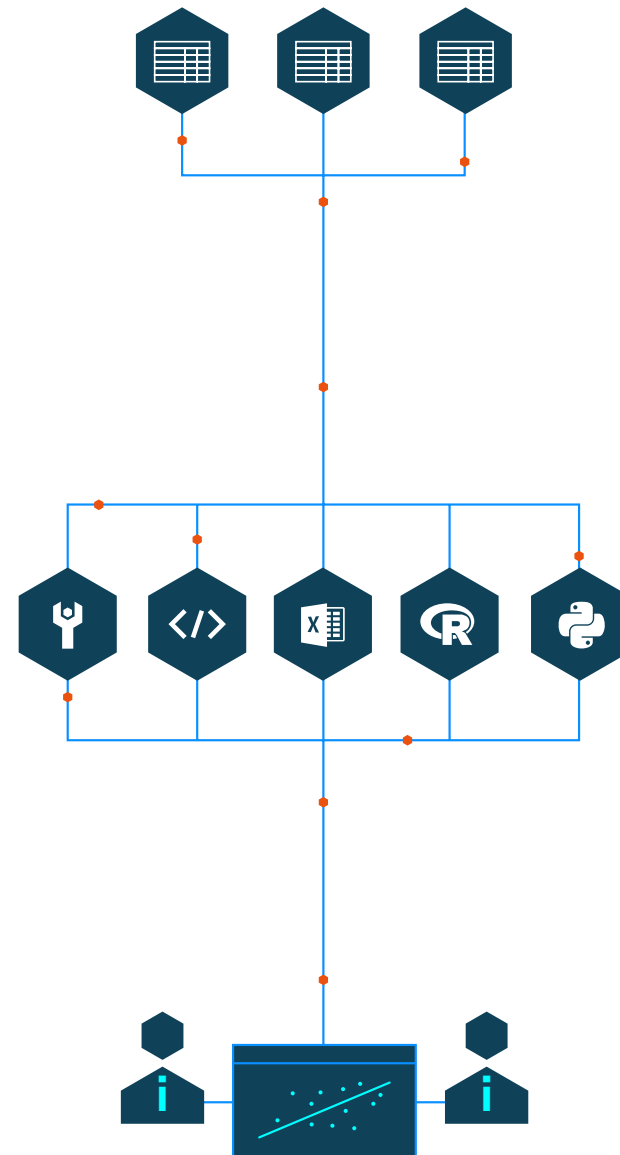
DELIVERING YOUR DATA

Delivering your data isn't as simple as transferring files via FTP. Quantitative analysts and professional investors use specific tools in their data analysis activities, which means they require a data delivery suite to match. The favored analyst tools are R, Excel and Python, while the preferred method of getting data into these applications is via RESTful API.

APIs should be capable of delivering a variety of different data points that are pertinent to an investor's trading strategy. This could be anything from sales estimates to GPS coordinates. Moreover, a robust API will need to allow for a large number of daily calls and have no service disruption at any point.

If you want to monetize a data asset, you will need to commit to reliably delivering the data at every single interval specified (hourly, daily, weekly, etc.) without fail. Investors will take large positions based on your data and will need complete faith that the data will always be where it needs to be. Errors in delivery can potentially cost investors thousands, sometimes millions of dollars in trading opportunities; thus, they need a trustworthy and reliable data source.

APIs also should be properly documented. Likewise, database pages should feature consistent formatting that detail the column headers, update frequency, ticker mappings (if needed), history and access. Crucially, this infrastructure and its associated instructions should constantly be updated to account for ever-changing company information, technologies and investment tools.



MARKETING AND SELLING YOUR DATA

The proliferation of hedge funds over the past three decades has created a group of active managers who make their wealth partly by keeping their circles closed. Even after you have successfully productized your data assets, you will still be challenged to penetrate the coterie that is Wall Street — your key demographic.

Here are a few strategies to consider:

STRATEGIC PARTNERSHIPS

A large part of monetizing data — at least within the context of hedge funds — is to get your foot in the door with the people who matter. Like all networking, one of the best ways to get considered is to have a trusted relationship with someone who is already connected with the hedge funds. If you are serious about monetizing data, strategic partnerships are key.

MARKETING CHANNELS

While the digital channels you are already using will come into play as you market your data, one way to get on Wall Street's radar is through traditional marketing. We find that our customers respond well to placements in top tier media outlets, including **The Wall Street Journal**, **The Financial Times**, **The New York Times** and **The Economist**. Whether print or digital, appearing in such publications will get you noticed and offer a dose of credibility to your brand.

The right events may also help. This audience attends conferences and seminars to stay abreast of new developments in the data world. Some of the bigger firms may even employ “data hunters” — people who go to industry-specific events around the world in search of new datasets. The trick is to pick the right ones: Avoid those that are focused on capital introduction or those that serve a dual purpose of education and accreditation. Neither of these tends to attract the right people.

DATA COLLATERAL

Don't underestimate the need for good collateral. While your customer base is sophisticated and quantitatively-inclined, there is no substitute for clear documentation that explains how to:

- Understand your data (structure, columns and attributes)
- Access your data
- Write code using your data
- Infer predictive signals from your data

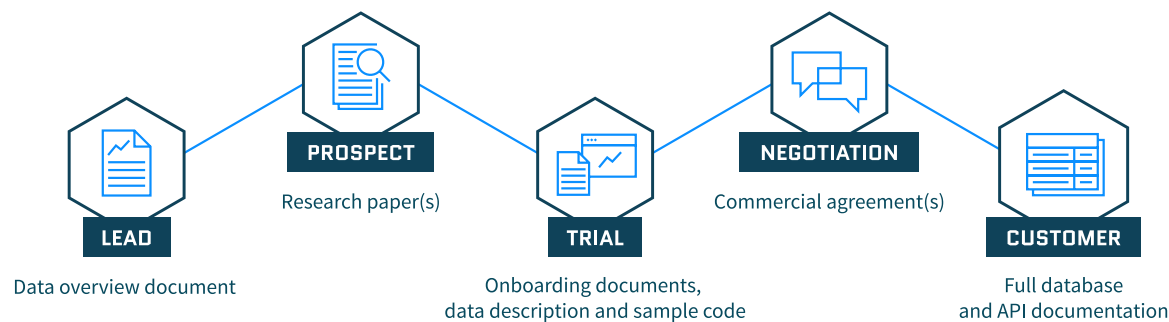
These documents, housed in an easily accessible document management system, will be required reading for your customers and will be necessary to jump-start any trial agreements you have in place. Nurturing your hedge fund customers is as important as nurturing any other audience to whom you might market — if not more so. They are busy professionals with competing interests. Even with a game-changing dataset before them, Wall Street clients will still have a hard time prioritizing the trial process.

THE SALES PROCESS

Even if you are confident that your data is significantly predictive of some aspect of the economy, don't expect a short sales cycle. Hedge funds have resources and infrastructure deployed methodically around existing trading strategies and they will want to understand how your data fits into their ecosystem. This does not happen overnight.

Lengthy trials are the norm here — you should plan for 90 days and be flexible if your prospect comes back to you multiple times with questions or concerns about your data. You will need to help your prospects climb a steep learning curve; remember that this is a brave new world and investors are seeing the type of data you have for the first time. You'll want a dedicated resource who is a subject matter expertise on your industry and technically capable with your data to shepherd a prospect through the trial; a salesperson alone is not likely to be enough. The collateral you produce should be distributed at key points throughout the sales process.

Finally, know your audience. Try to meet with them face to face — physical visits are invaluable. While challenging, the more specifically you can match your data to your audience, the better. If your data predicts macro-economic signals, for example, you'll want to focus on funds that take a macro approach to trading.

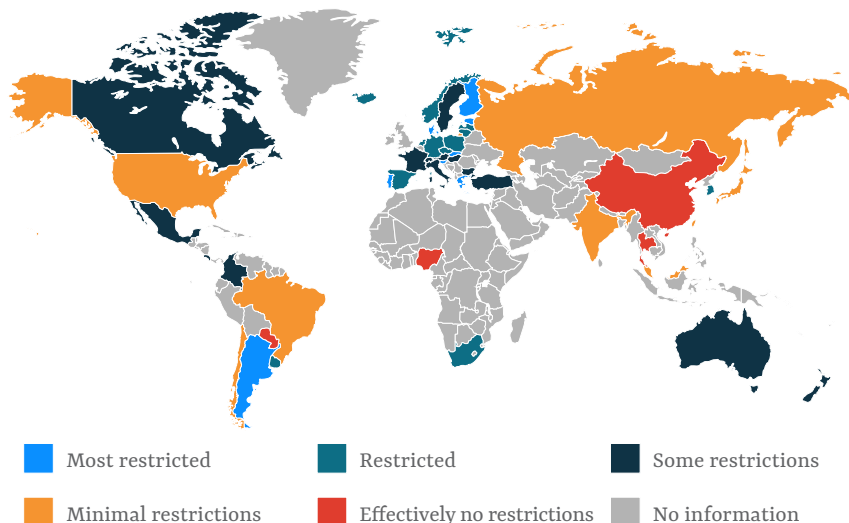


Match your collateral distribution to the sales funnel

PROTECTING YOURSELF AND YOUR DATA

Without fail, all potential data partners wonder about privacy. The media has made much of datasets used by Wall Street that are not “anonymized”, but the reality is that Wall Street does not care one bit about an individual’s information. They only care about data that moves markets; that is data in the broad aggregate.

The key concept you will need to understand here is personally identifiable information (PII). The data you deliver to your Wall Street customers cannot contain any personally identifiable information at all. You must scrub this information out of your dataset and **furthermore make it impossible to reconstruct.**



Privacy and data protection by country. Source: Forrester

Some countries, such as those in Europe, have adopted strict data protection laws that provide clear directives on how companies can process and disseminate data. However, other countries, including China, have complex legal and regulatory systems that must be deciphered and navigated carefully, as they do not always come together to form a coherent guide for data partners. With the international complexities of data privacy legislation and the constantly evolving landscape of big data, compliance must be a chief concern for any company considering data monetization.

Information privacy laws in the US have been legislated on a sectoral basis. For example, the Health Insurance Portability and Accountability Act (HIPAA) governs the national standards for electronic healthcare data transactions. More recently, in April 2017, the Trump administration signed a congressional resolution to overturn the internet privacy protections created by the FCC during the Obama administration, allowing broadband internet service providers (ISPs) to “track and sell a customer’s online information with greater ease”.²

The European Union rolled out sweeping harmonized data protection controls in 2012 through the General Data Protection Regulation (GDPR). This regulation will become binding upon all member states by May 2018; however, given the UK’s departure from the EU, it is not yet known whether it will diverge from the GDPR.³

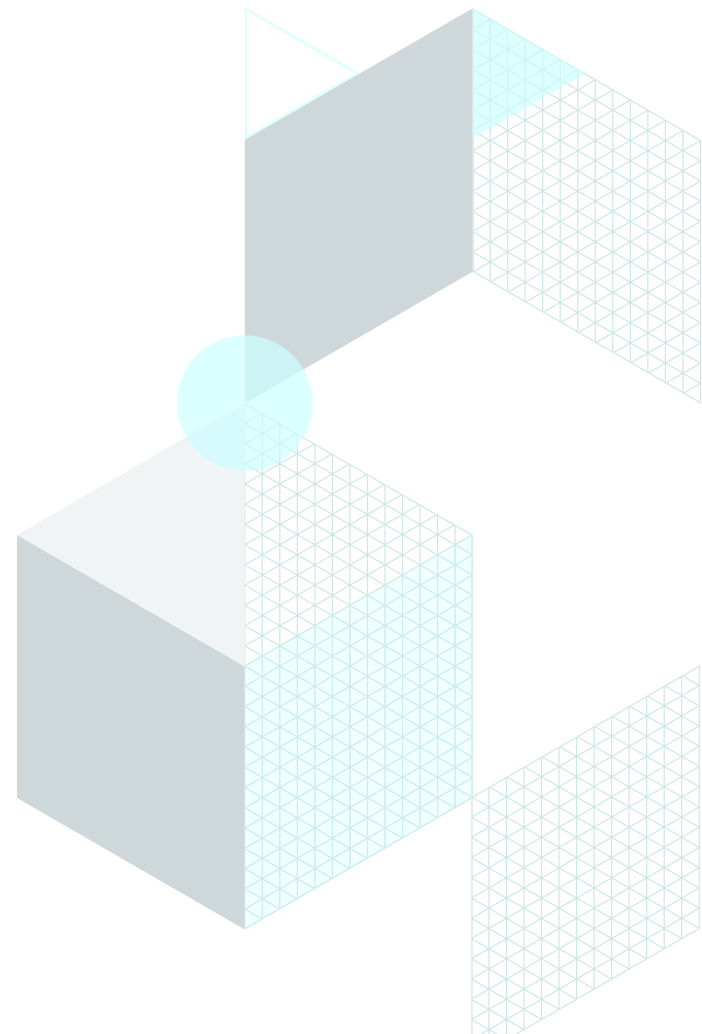
² https://www.nytimes.com/2017/04/03/technology/trump-repeal-online-privacy-protections.html?_r=0

³ http://www.eversheds-sutherland.com/global/en/what/articles/index.page?ArticleID=en/Data-Protection/brexit_impact_on_gdpr_170616

These are just some examples that illustrate the importance of understanding the legislation, especially if you have a global product. The feed you deliver should be properly stripped of any personally identifiable information at the consumer level. This is any data that can identify a specific individual, be it their name, phone number, address or even gender.

One promising notion to keep in mind, however, is that Wall Street is not interested in PII in any way. They are much more interested in the aggregate trends than the fact that Lucy Morgan, 27, of Louisville, Kentucky just bought six family-sized bars of Hershey's chocolate using her MasterCard at Costco. What Wall Street wants to know is how many people are buying Hershey bars across the US this quarter compared to last quarter, how many new credit cards has MasterCard issued in the last month or how many daily transactions Costco has handled compared to Walmart. They care in the aggregate, not at the individual level. They also have their own compliance policies and regulations to contend with and you can be certain some of these will be written into the contracts they sign with you.

In addition to protecting the privacy of your dataset, you may want to protect your firm's privacy as well. There are precedents in place to sell data anonymously and, in our opinion, this is wholly acceptable and should be an easily negotiable aspect of your contract with an asset manager.



PRICING YOUR DATA

How do you know how much to charge? The industry is nascent and pricing is not widely understood; in many cases it is as much art as it is science. However, there are a few guidelines that are useful to keep in mind:

DATA ATTRIBUTES

The data attributes outlined previously — quality, predictive power, uniqueness — are directly related to price. The better a dataset scores on these evaluative attributes, the higher a price it can command in the market.

ASSETS UNDER MANAGEMENT (AUM)

The single strongest indicator of how much a client will pay for a dataset is how big the client is. This is not just because larger clients can afford to pay more (though that is certainly the case); it is also because larger clients can derive more profit from the same data.

Assume a given alternative dataset is expected to generate 1% in “excess” returns for a client. That dataset would be worth more to a hedge fund managing \$5 billion (where 1% = \$50 million) than it would be to a fund managing just \$100 million (where 1% = \$1 million).

Note that we are talking here about “excess” returns: These are the returns that accrue from using the alternative data, above

and beyond the returns that the manager would have made without that data.

RETURN ON DATA INVESTMENT

Given the risks and uncertainty involved in all investing, hedge funds will expect a 10x to 20x return on a data investment. So if a fund expects to make \$1 to \$2 million in excess trading profits by using your data, they should be willing to pay you \$100,000 for that data.

Of course, the same dataset can be sold multiple times, to multiple funds, at no extra cost to you, the data partner. (The cost of duplicating a data asset is zero). However, this imperative must be balanced against the need to keep your data closely held and hence expensive.



TIERING

One way to maximize income while remaining narrow is to build a variety of data products from the original raw data asset. You can slice it and dice it differently, depending on the profile of the firms you are targeting: fundamental versus quantitative, small versus large, hedge fund versus investment bank, fast versus slow access and so on. This allows you to sell essentially the same data product multiple times without diminishing its alpha.

EXCLUSIVITY

In some cases a potential client may ask you for exclusivity — i.e., to sell your data to them and no one else. This is a fascinating quirk of the hedge fund world — the fewer the people who have access to a dataset, the more an individual firm can profit from it. Consider offers of exclusivity if the bid is near your expected total revenue from selling the dataset to multiple customers.



CASE STUDIES

Every company today is a data company. Business processes everywhere are becoming digitized. Firms like Walmart and Target know exactly what you search for and what you end up buying. Other companies are intimately involved in payroll, transactions, delivery and every other stage of the commercial pipeline. Every single action that these firms take is recorded and stored for analysis. Because of this, nearly all the companies in the world, regardless of size, have their finger on the pulse of some little corner of the economy.

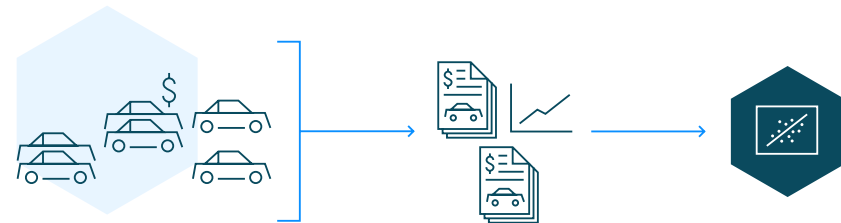
As such, valuable alternative datasets come from all kinds of industries. Below you will find several case studies of alternative data from different industries being used to predict an aspect of the market today.

AUTO INSURANCE

Auto insurance providers can give investors a daily count of how many new car insurance policies they sell, sliced by manufacturer. This data can be used to estimate total daily new car sales per manufacturer. Anyone trading auto equities would be interested in this kind of data.

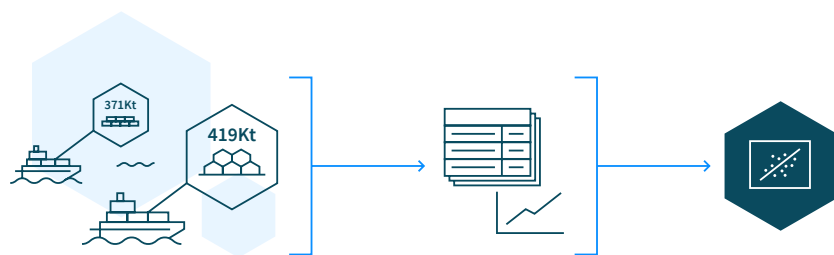
The volume of new car sales, which is almost perfectly correlated with the number of insurance policies issued for new cars, is a key driver of performance for auto manufacturers. Armed with this data, investors can take a position in a specific auto stock

ahead of consensus estimates. It's this information advantage that allows investors to generate the elusive alpha of which we have spoken.



COMMODITIES

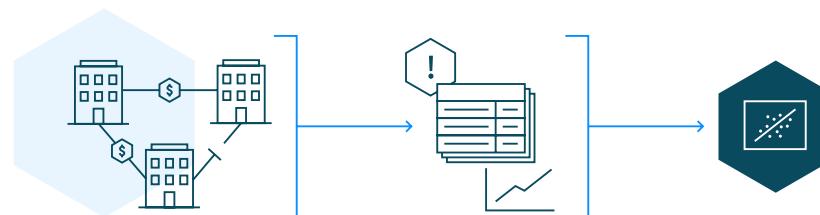
Did you know that China is the world's leading producer of crude iron ore? And yet it is also the biggest importer of usable ore from the world's two leading producers Australia and Brazil.⁴ Knowing this will highlight the importance of shipping and logistics in transporting a commodity from one country to another. Investors studying these assets have historically relied upon monthly reports to inform production volume and demand but innovations in data science have led to datasets that can infer commodity transportation from automatic identification system (AIS) data combined with port and berth mappings.



BUSINESS HEALTH METRICS

A business service and commercial data provider collects detailed transaction data from thousands of American companies in the United States. As such, they maintain the largest database of commercial activity in the world, focusing on inter-company payment patterns, specifically, payment amounts, timing, delays and delinquencies.

Using this data, investors can learn if a company is paying all its debts on time and how their payment patterns are evolving. Consistent late payments can be evidence of strength and the ability to squeeze suppliers. But sudden increases in late payments may be evidence of weakness, namely cash-flow problems. This data can be used to gauge debtor distress — a leading indicator of stock underperformance that can help inform a simple trading strategy.



⁴ https://minerals.usgs.gov/minerals/pubs/commodity/iron_ore/global_iron_ore_data.pdf

NEXT STEPS

So how do you take advantage of this \$7 billion opportunity? Throughout this document, you've learned what it takes to monetize your data assets for a Wall Street audience. You know that your data must be clean, error-free, predictive and unique. You must have a delivery infrastructure in place to produce reliable feeds 365 days a year in formats investors are accustomed to. You will need to invest in marketing and selling the data specifically to a financial audience. You must scrub out personally identifiable information. And you must consider the pricing economics behind your assets.

You are now equipped with the basics to determine whether you want to move forward with this potential revenue stream. You may choose to move forward by leveraging your internal resources. Or you can consider partnering with Quandl.

Quandl is a fast and effective pathway to transforming your data assets into recurring revenue.

We already have 200 partners on our platform, thanks to:

- A direct channel to top-tier financial services organizations
- The technical infrastructure to deliver your data at the high standards Wall Street demands
- The domain expertise to properly package your data for maximum appeal on Wall Street

- A nimble team capable of bringing products to market fast
- A brand associated with excellence and innovation in data and data delivery

We offer all aspects of the productization, infrastructure, delivery, marketing and sales discussed in this white paper.

PRODUCTIZATION

Quandl's data science team will evaluate your data and work with you to ensure that it can be converted into a compact asset that can be packaged and sold to a Wall Street audience. While it is up to our clients to determine exactly how your data can be worked into their trading strategies, we will prepare, document and package your data into an asset they can work with almost immediately.

INFRASTRUCTURE AND DELIVERY

Our platform makes data available by API, through custom libraries for R, Python and Excel, and via a digital ecosystem that the financial community already uses. No matter where the data originates, once it is on Quandl, we make sure that it's clean, uniform, well documented and easy to use. Our uptime rates exceed 99.9% and we have a team of engineers and data curators in place to maintain your data feeds, alert you when necessary and keep our connection pipes up and running 24/7/365.

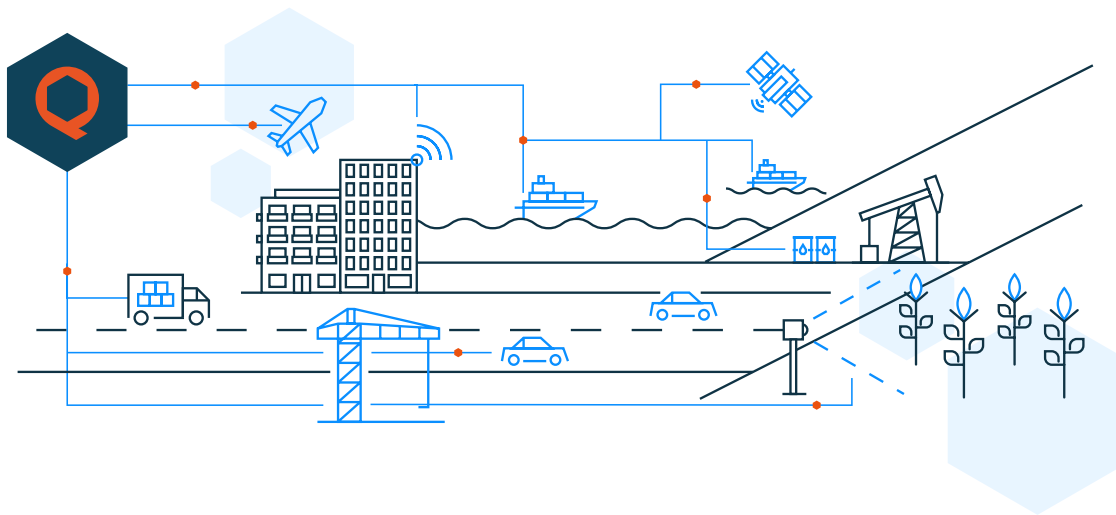
MARKETING AND SALES

Quandl has built a channel into Wall Street methodically and thoughtfully over five years. We started with founders who came from Wall Street themselves and understand the landscape as insiders. We grew organically from there.

Today, the buy-side relies on us to provide them with alpha-generating datasets and will often recommend that data providers work through us to get to them.

We have received inbound inquiries from several hundred data providers, large and small. Few of these sell data as their primary source of revenue. But they're all sitting on valuable, untapped data assets and they're leveraging Quandl to monetize their excess capacity. Note that we select our partners based primarily on the quality and uniqueness of the data they bring to the table.

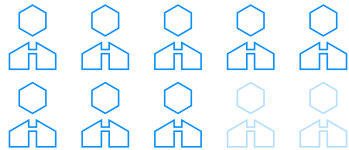
If you think that your business has been collecting alpha-generating data, please talk to us. We have a simple, confidential data evaluation protocol that lets us examine your data and come back to you with a roadmap to full deployment and monetization.



ABOUT QUANDL

Designed for professionals, Quandl is the only platform for financial and alternative data. We have more than 200,000 financial analysts and hundreds of data partners transacting on our platform. Our partners include companies from industries as diverse as insurance, machinery, freight and biotech.

TODAY OUR USERS INCLUDE:



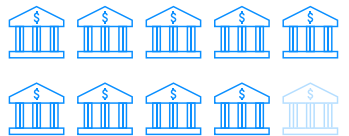
8/10

Largest hedge funds in the world



7/10

Largest asset managers in the world



9/10

Largest investment banks in the world

We bring undiscovered data from non-traditional publishers to investors seeking unique, predictive insights. We leverage exclusive relationships to deliver these alpha-generating datasets to our customers.

Traditional financial data is already a 27-billion-dollar industry. We expect non-traditional, unorthodox, long-tail data to grow to a comparable size. In the next decade, a larger share of returns in the finance industry will be driven by new and untapped data sources. And therein lies the opportunity for both Quandl and our partners.

OUR FOUNDERS



Tammer Kamel,
CEO & Co-founder

Tammer founded Quandl in 2011 and continues to serve as its CEO. He focuses on client and partners relationships. Prior to founding Quandl, Tammer spent 15 years as a technologist, risk manager and portfolio manager at Citibank and Simplex Asset Management.



Abraham Thomas,
CDO & Co-founder

Abraham leads Quandl's data science team, which identifies, analyzes and transforms raw data into salable data assets. Before Quandl, Abraham was a senior portfolio manager and head of US bond trading at Simplex Asset Management, a multi-billion dollar hedge fund based in Tokyo.

⁵ <http://finance.yahoo.com/news/global-spend-financial-market-data-142000678.html>

Contact Us

Chris Harding, COO
chris.harding@quandl.com

 @quandl

 www.quandl.com

Quandl