RESEARCHFOCUS

MITIGATING LEGAL RISKS ASSOCIATED WITH ALTERNATIVE DATA

January 2018

Integrity Research Associates, LLC is an advisory firm founded in 2003 to analyze the global investment research industry. We serve a select number of hedge funds and asset managers as experts on external research sources used to enhance investment returns. We offer deep domain insights on research firms, mitigate research compliance risks, and monitor key research industry developments.

Integrity Research is the only firm which tracks the entire investment research industry. We follow over 4,000 research and data firms in the U.S., Europe and Asia, assisting clients in generating more value for their spending on external research. We pay particular attention to obscure, hard-to-find research sources and on innovative new research developments.

Our core services are recommending research and data firms which add value, managing the risks associated with external research through due diligence aids, and monitoring developments in the industry through our news and commentary. We also provide strategic consulting on research-related topics.

Integrity ResearchFocus is a registered trademark of Integrity Research Associates, LLC.

For more information, please contact:
Michael Mayhew – Michael. Mayhew@integrity-research.com – 646-786-6859

© Copyright 2018 by Integrity Research Associates, LLC.
Federal copyright law, Title 17 of the U.S. Code, makes it **illegal to reproduce this report** by any means and for any purpose, unless you have our written permission. Copyright infringement carries a statutory fine of up to **$100,000 per violation**. We offer a reward of $2,000 for information that leads to the successful prosecution of copyright violators.

Contents

Introduction 3
Growth of Alternative Data 4
Broader Investment Management Use 6
Risks 8
Insider Trading & Alternative Data 9
SEC vs. Huang 10
US vs European Insider Trading Law 11
Personally Identifiable Information (PII) 12
PII in Europe 13
Web Scraping 13
LinkedIn vs. hiQ Labs 15
Other Risks 16
Growing Media and Regulatory Scrutiny 17
Closer Examination by the Media 17
Increasing attention to PII 18
Growing Regulatory Attention 19
The Importance of Industry Standards 20

Introduction

As use of alternative data grows and the number of alternative data sources multiply, associated compliance risks will also increase. There is growing regulatory interest in alternative data, with some experts expecting prosecutors to bring cases. Media scrutiny is also increasing, with growing concerns about privacy violations, copyright infringements and whether alternative data creates an unlevel playing field for retail investors.

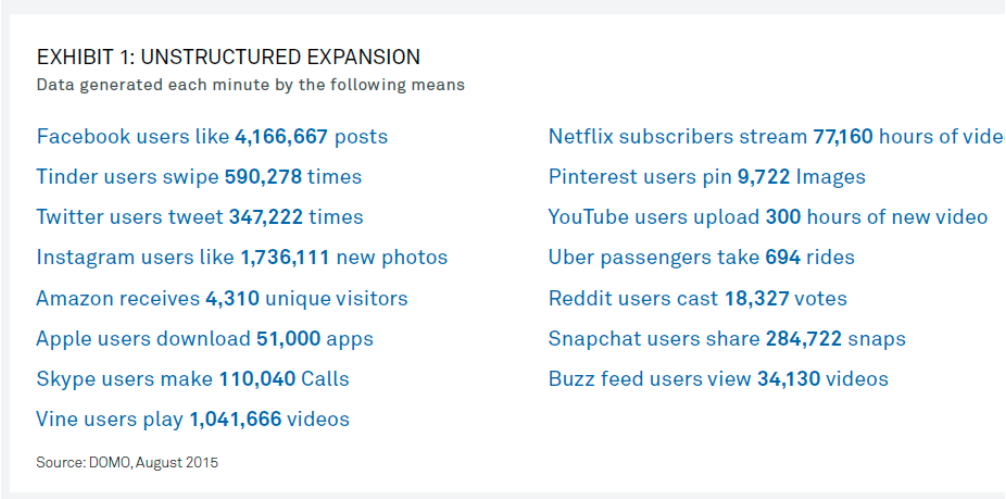
We believe that alternative data is at a place analogous to that of expert networks a decade ago. Like alternative data, expert networks gained increasing adoption by asset managers, particularly hedge funds, and there was a dramatic increase in the number and variety of expert networks. Poor practices at a minority of expert networks fed a series of insider trading investigations which severely impacted the industry, causing many expert networks to exit the business. Nevertheless, by developing consistent compliance practices across the industry, the more compliant expert networks survived, and are now once again on a strong trajectory of growth.

In this paper, we call for the alternative data industry, including buy-side consumers, sell-side users, vendors and those that originate alternative data, to take proactive steps to inoculate the industry from adverse media and regulatory scrutiny. By developing consistent and broadly adopted compliance standards, alternative data participants can reduce legal risks and compliance costs, not only for their individual entities but for the industry overall.

Growth of Alternative Data

We define 'alternative data' as the intersection of big data[1] and investment research. When applied to investment research, alternative data is the collection, cleansing, packaging, modeling and distribution of large unstructured and structured data sources to generate predictive insights and improved investment returns.

Alternative data is not a new concept. Quantitative asset managers have been utilizing differentiated data sources for decades. With the increasing proliferation of big data there are now more opportunities to mine investment insights. At the same time, fundamental asset managers are increasingly adopting quantitative techniques.



Source: Blackrock, "The Evolution of Active Investing"

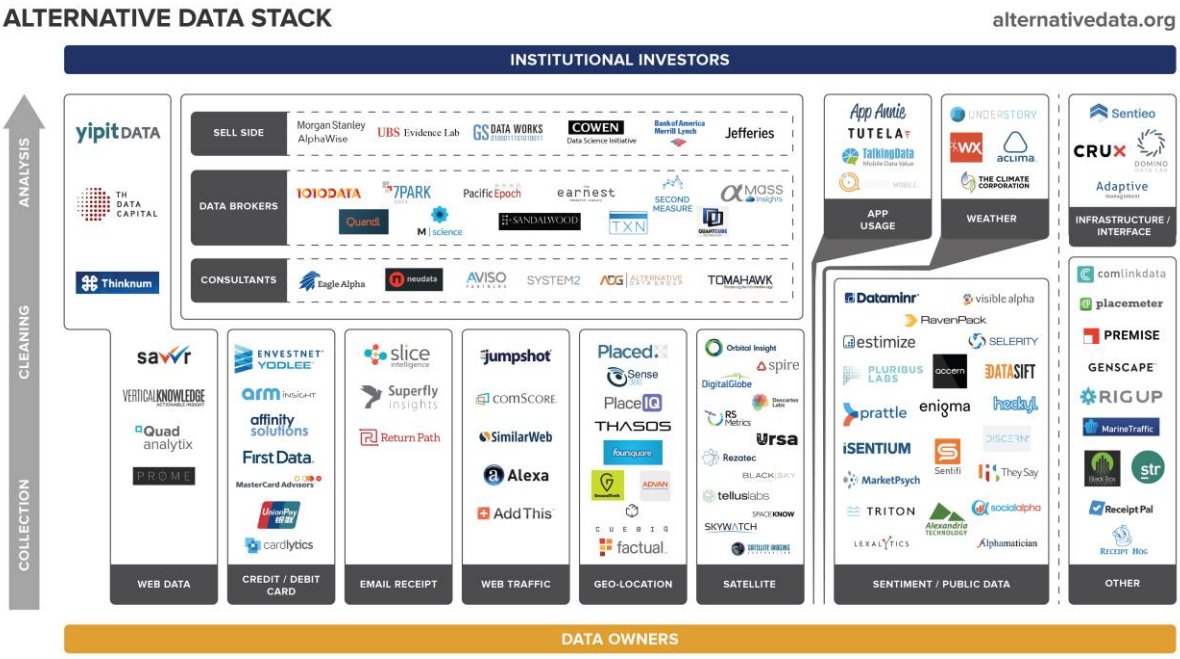
IBM estimated that 90% of the data available today has been created over the last two years.[2] According to Two Sigma, the world now creates an Exabyte of data every hour, which equates to about a billion gigabytes, much of which is unstructured. The Internet of Things (IoT) -- driven by the embedding of networked sensors into home appliances, collection of data through sensors embedded in smart phones, and reduction of cost in satellite technologies – will accelerate the production of data further.

With the increasing volume of data comes increasing diversity of sources and applications. Data derived from social media provides use cases for news, market sentiment and data mining. As

[1] Defined by Gartner Technology as "the information assets characterized by such a high volume, velocity and variety to require scientific technology and analytical methods for its transformation into value."
[2] Source: IBM, "Bringing Big Data to the enterprise".

millennials and the succeeding Generation Z continue to come into their own, the quality of social media data will continue to improve as it becomes even more central than it is today.

More broadly, the internet also generates data of investment value. 'Web scraping' can track pricing trends for anything being sold digitally, from groceries to automobiles. Analysts can evaluate new product introductions and product life cycles by scraping product reviews on consumer websites. Evaluating retail companies and other companies with strong digital business lines increasingly demands the use of alternative data. As more economic activity becomes digital, digital evaluation methods become mandatory.



Source: AlternativeData.org

Sensor data -- whether derived from satellites, smart phones, the IoT or other sources – is one of the fastest growing and increasingly valuable sources of alternative data. Satellite images and geolocation data from mobile phones are supplanting traditional 'channel checking' methods of mystery shoppers.

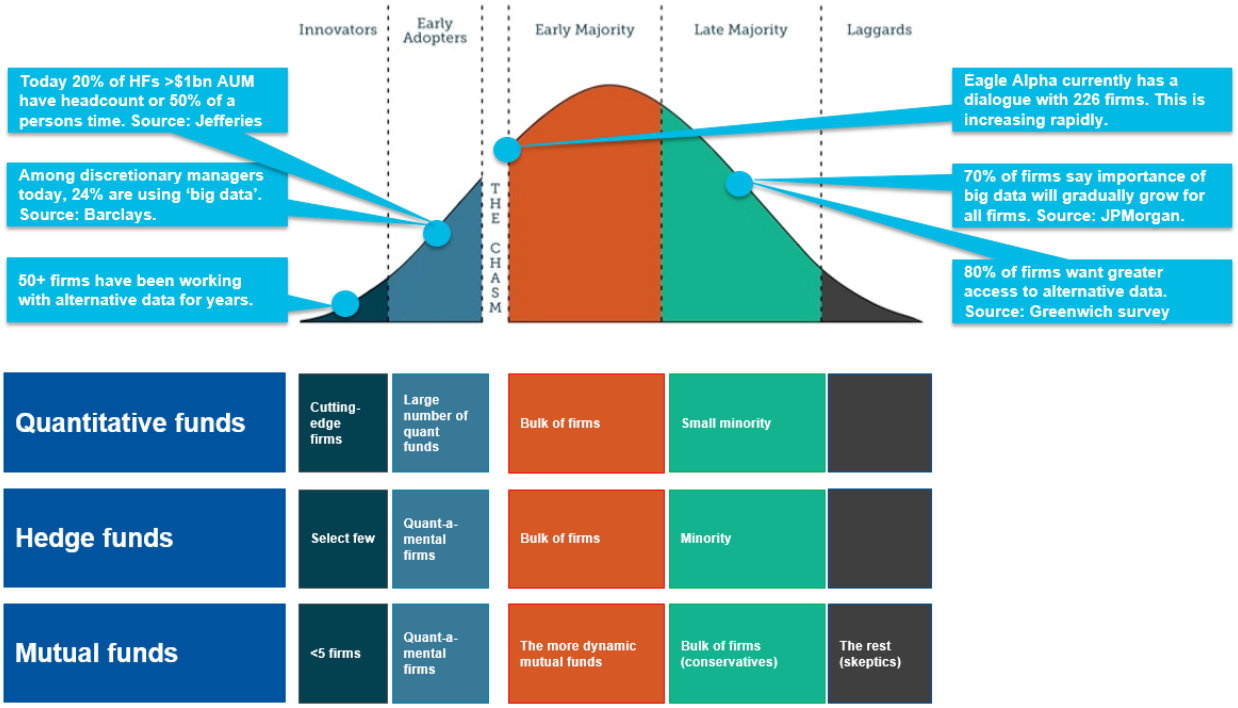
The global market for big data, related technology, and analytics is estimated at $130 billion and is expected to grow to over $200 billion by 2020.[3] The financial industry, estimated to have ~15% spending share, is one of the drivers of growth. JP Morgan estimates the investment management industry's spend on big data is in the $2-3 billion 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).[4]

[3] IDC, *Worldwide Semiannual Big Data and Analytics Spending Guide*.

[4] JP Morgan Quantitative and Derivatives Strategy, "Big Data and AI Strategies: Machine Learning and Alternative Data Approach to Investing", May 2017.

Broader Investment Management Use

Until recently, alternative data usage has been concentrated with quantitative investment managers such as Renaissance Technologies, Winton Group, D.E. Shaw, Two Sigma or WorldQuant.[5] A minority of discretionary hedge funds such as Point72 or Citadel have invested in integrating alternative data into their investment process.



Source: Eagle Alpha, "Alternative Data: Applications & Case Studies Version 2"

Increasingly, fundamental asset managers are utilizing alternative data. Paul Tudor Jones cut some of their fundamental traders in switching to a quant strategy where traders and computer are expected to work in tandem: "No man is better than a machine, and no machine is better than a man with a machine." (Paul Tudor Jones to his investment team, August 2016).

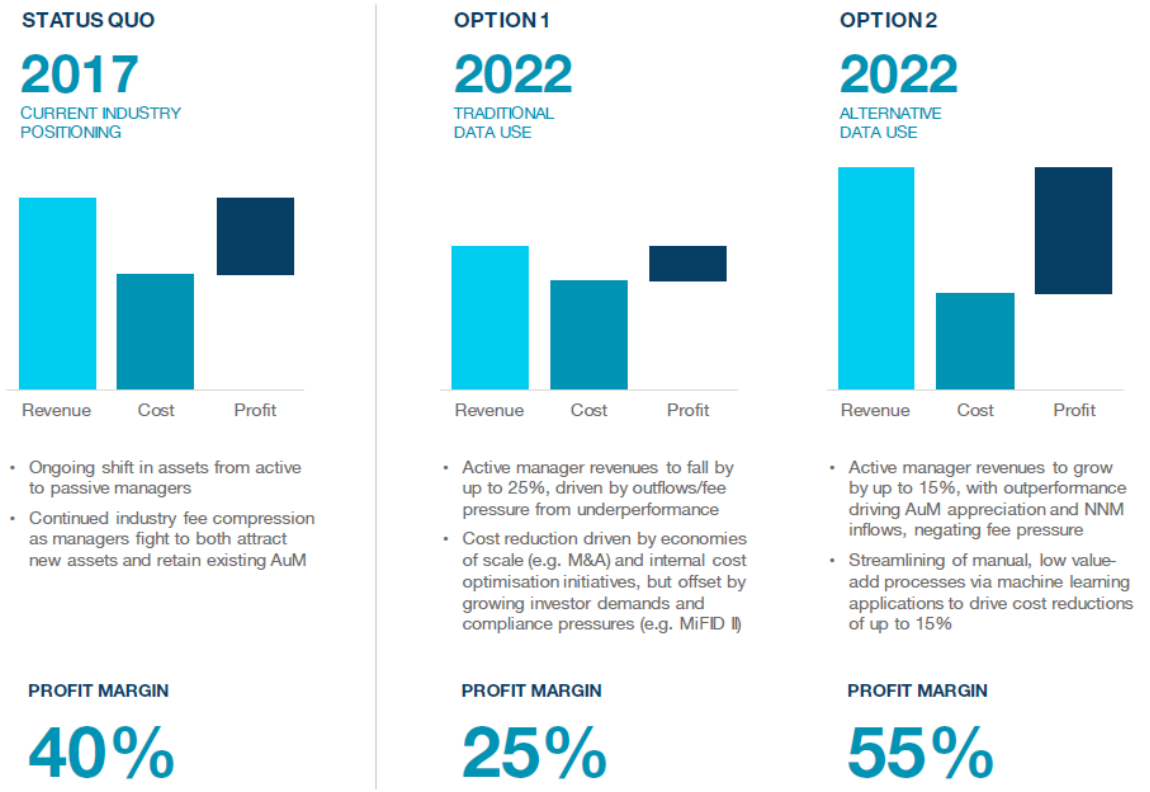
Jefferies, in a June 2017 white paper, stated that 20% of hedge funds with over $1 billion in AUM have a person dedicated to alternative data or a person is spending 50% of their time on alternative data.[6] Eagle Alpha estimated there are 150-200 firms in the early adoption phase and expect the bulk of hedge funds will utilize alternative data, along with some long-only managers as usage matures.[7]

[5] AQR was a notable exception, but according to the Financial Times has begun experimenting with alternative data. https://www.ft.com/content/3a8f69f2-df34-11e7-a8a4-0a1e63a52f9c

[6] Jefferies, "Quantifying Intuition: Mapping the Data Science Landscape in the Hedge Fund Industry".

[7] Eagle Alpha, "Alternative Data: Applications & Case Studies Version 2", September 8, 2017 https://s3-eu-west-1.amazonaws.com/ea-pdf-items/Alternative_Data_Report_Version_2.pdf

FIGURE 20: P&L IMPLICATIONS FOR ACTIVE MANAGERS



Consulting firm Quinlan & Associates predicts that asset manager profit margins will decline from 40% currently to 25% by 2022, caused by fund outflows from active managers and increasing fee pressures. However, large active asset managers which embrace alternative data might attain improved margins of 55%, assuming a 15% increase in revenues from improved asset flows caused by performance-enhancing alpha extracted from alternative data and a 15% decrease in costs through automation.[8]

In 2015, Blackrock warned that asset managers who do not adopt alternative data will be left behind: "We believe that in order to generate sustained alpha, investors should embrace acquiring, analysing and understanding the fast growing universe of data. Those who are unable to do so run the risk of falling behind in a rapidly changing investment landscape."[9]

[8] Quinlan & Associates, "ALTERNATIVE ALPHA: Unlocking Hidden Value in the Everyday", September 2017 http://www.quinlanandassociates.com/insights-alternative-alpha/

[9] Blackrock, "THE EVOLUTION OF ACTIVE INVESTING: Finding Big Alpha in Big Data" September 2015.

Risks

With proliferating data and broader use by investment managers come greater media, regulatory and legal risks. Systematic use of alternative data is relatively new, so it has not attracted critical scrutiny from media and regulators, but that is changing.

There are assorted risks associated with alternative data, which can be organized into a few major categories: exclusivity, insider trading, privacy violations and copyright infringements. In addition, there are other elements of risk such as poor data quality, difficulties in validating data, increased model risk, and the potential need for additional disclosures.

Alternative data that is available on an exclusive basis potentially heightens insider trading risk by increasing the materiality of the data and highlighting the non-public nature of the data. Irrespective of domicile, insider trading rules focus on information that is both material and non-public.

Key Legal Risks Associated With Various Types of Alternative Data

Alternative Data Type	Available on Exclusive Basis	Insider Trading	Privacy Law	Copyright Law
Web Data	Unlikely	Lower Risk	Lower Risk	Higher Risk
Web Traffic	Unlikely	Lower Risk	Lower Risk	Higher Risk
Sentiment Data	Unlikely	Lower Risk	Lower Risk	Lower Risk
Credit Card Transactions	Possible	Higher Risk	Higher Risk	Lower Risk
Email Receipts	Possible	Higher Risk	Higher Risk	Lower Risk
Geolocation	Possible	Higher Risk	Higher Risk	Lower Risk
Satellite	Possible	Higher Risk	Lower Risk	Lower Risk

Source: Integrity Research Associates

Consumer protection concerns relate to personally identifiable information which should be carefully quarantined and scrubbed by users of consumer-sourced information. Violations of copyright law and related issues surround the harvesting of publicly available data typically available through the internet.

Not all legal risks apply equally to all types of alternative data. Insider trading risks pertain most clearly to data which is non-public, which can be heightened by exclusivity agreements. Ultimately, the issue comes down to the consents obtained to collect the data and to distribute it to third parties.

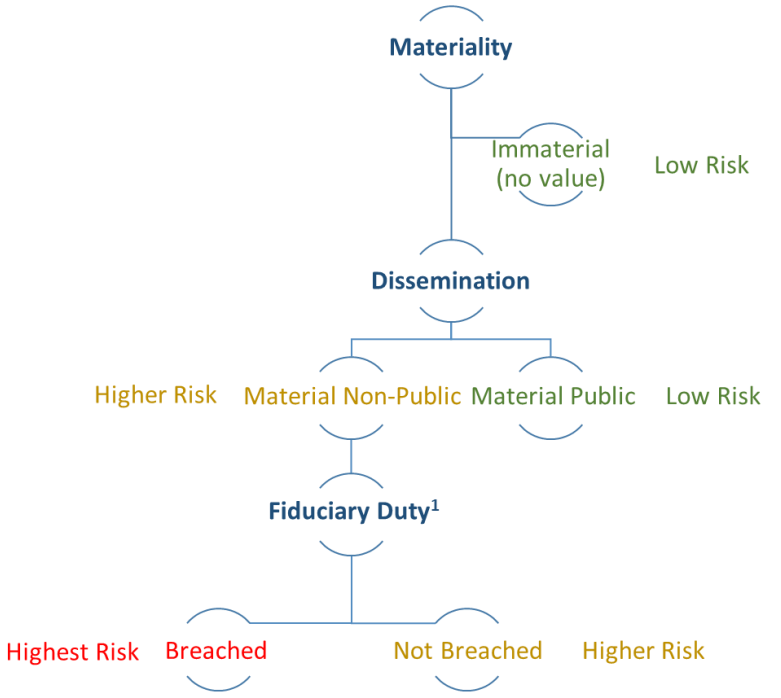
Risks associated with consumer protection laws apply to consumer data, but not to sensor data or certain types of web data. Copyright infringements and other injunctions against harvesting apply most directly to web data.

Insider Trading & Alternative Data

One of the reasons hedge funds have been attracted to alternative data was the perception that it has a low risk of being deemed inside information under insider trading regulation. Alpha-generating information is by definition material. Insider trading prosecutions are all ex-post, allowing prosecutors to focus only on successful trades, and the more successful the trade the more attractive it is to a prosecutor. However, if the data sources are public, such as social media or the web, it is difficult to classify the data as material non-public information (MNPI).

However, as more types of non-public alternative data such as credit card transactions or geolocation data have become prevalent, information is more easily classified as 'material non-public.' The issue is compounded when hedge funds pay large fees to keep the data exclusive.

Escalating Tests of Material Non-Public Information



[1]The test of fiduciary duty mainly applies to US insider trading regulation and may afford protection in non-US domiciles.

Source: Integrity Research Associates

The *Financial Times* recently related a story about Point72's chief market intelligence officer, Matthew Granade, who bragged (the FT's term) during a presentation at the London School of Economics that Point72 can get a real-time idea of how companies are doing long before their financial results are released.[10] When asked how this data could help Point72 if everyone had

[10] Financial Times, "Hedge Funds See Gold Rush in Data Mining" https://www.ft.com/content/d86ad460-8802-11e7-bf50-e1c239b45787

access to the same information, Granade reportedly responded: "The great thing about this area is you can arrange deals where you are the only ones who get it."

When alternative data is potentially both material and non-public, protection against US insider trading violations rests on whether the data was obtained in violation of a fiduciary duty. This may require consent to collect the information as well as consent that it be distributed to a third party. There have already be violations of the former.[11] With the growing proliferation of data, the likelihood of inadequate consents increases, thereby raising the risk of violations of fiduciary duty.

In some cases, asset managers have reportedly established thresholds of materiality governing their use of alternative data. Data that has too high a correlation with non-public information such as same store sales might be prohibited to use.

SEC vs. Huang

The US Securities and Exchange Commission has already successfully prosecuted an insider trading case involving alternative data.[12] The case involved two data analysts who obtained material non-public information by analyzing credit card transactions. Because they obtained access without the consent of the owner of the data, in this case Capital One Financial Corporation, they were prosecuted for insider trading and forced to pay over $18 million in disgorgement and penalties.

The SEC accused two data analysts employed by Capital One of making searches on Capital One's proprietary database of credit card transactions relating to least 170 publicly traded companies from November 2013 to January 2015. The defendants, Bonan Huang and Nan Huang (not related), used put or call options to trade retail stocks immediately before earnings releases.

In one example cited in the SEC complaint, after analyzing transactions involving outdoor goods retailer Cabela's Inc. Nan Huang purchased in February 2014 put option contracts on Cabela's stock for $51,890. Cabela's stock was trading around $70 at the time. A day later, Cabela's announced a 10 percent decrease in sales and its stock fell 8 percent to $64.26. Huang sold his options for $107,950 making a 108 percent return on his one-day investment, according to the complaint.[13] The SEC said the pair made an overall profit of $2.8 million on a $147,300 investment, a return of about 1,800 percent.

[11] FTC vs. Goldenshores Technologies, LLC, https://www.ftc.gov/news-events/press-releases/2013/12/android-flashlight-app-developer-settles-ftc-charges-it-deceived

[12] Securities and Exchange Commission v. Bonan Huang, et al., Civil Action No. 2:15-cv-00269 (E.D. Pa.), https://www.sec.gov/litigation/complaints/2015/comp23216.pdf

[13] Other examples: In April 2014, the pair made $142,000 after Coach Inc. announced lower than expected sales. In June 2014, they made more than $377,000 in one day after trading a positive surprise in Chipotle Mexican Grill Inc.'s earnings.

In January 2016, the SEC won a jury verdict[14] against one of the analysts, Nan Huang (the other analyst settled with the SEC agreeing to pay approximately $4.7 million in disgorgement and penalties to settle the SEC's allegations). Nan Huang lost an appeal to the U.S. Court of Appeals for the Third Circuit in April 2017, which required him to pay $13.5 million in penalties and disgorgement.

Nan Huang argued that the data obtained was not material because Capital One's transaction data represented an average of 2.4% of the companies' revenues and that he relied on a mosaic of other information in making his trading determination. The SEC presented evidence that the data had high correlations to revenues and argued that the information was material.

The Third Circuit court ruled that "Huang was able to make revenue projections approximately one month before the companies publicly announced their actual quarterly revenues. This insider information gave Huang early and nonpublic insight as to whether the companies were likely to under- or over-perform expert predictions and thus whether their stocks were likely to increase or decrease after the quarterly announcements."[15]

US vs European Insider Trading Law

The US and Europe differ considerably on insider trading rules. Whereas US regulators have tended to be more proactive in prosecuting insider trading violations, European insider trading rules are broader. The protections provided by US law potentially permitting material non-public information provided the information has not been misappropriated and no duty has been breached do not apply in Europe.

Under the EU's 2003 Market Abuse Directive and the market abuse regulation (MAR) that came into force in July 2016, European rules ban the use of any price sensitive or material non-public information, without any qualifying conditions for breach of fiduciary duty.[16]

Greenlight Capital was fined nearly £7.3 million ($11.6 million at the time) in 2012 by the UK's Financial Services Authority when it breached UK rules while believing it was in line with US standards.[17] Greenlight's portfolio manager, David Einhorn, was invited to a "wall crossing" conference call by Punch Taverns Plc. Einhorn refused to sign a confidentiality agreement associated with the call, but was permitted by the investment bank arranging the call (Merrill Lynch) to attend the call anyway. Einhorn subsequently sold Greenlight's holdings of Punch Taverns and was fined by the FSA, which argued that although no single statement of inside information made during the conference call, "reasonable investors are expected to interpret

[14] SEC Litigation Release No. 23445 https://www.sec.gov/litigation/litreleases/2015/lr23445.htm
[15] SEC v. Huang , 3d Cir., No. 16-2390, 4/10/17, https://law.justia.com/cases/federal/appellate-courts/ca3/16-2390/16-2390-2017-04-10.html
[16] For a comparison of European and US insider trading regulations, see Marco Ventoruzzo, "Comparing Insider Trading in the United States and in the European Union: History and Recent Developments" May, 2014 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2442049
[17] The Financial Services Authority, *Decision Notice to David Einhorn*, Jan. 12, 2012, http://www.fsa.gov.uk/static/pubs/decisions/dn-einhorn-greenlight.pdf

comments made to them in an appropriate manner, which may sometimes mean understanding more than the precise words spoken, or interpreting certain comments in light of the context."

Compliance officers at U.S. asset managers were shocked by the decision, in large part because Einhorn's refusal to sign the non-disclosure agreement would most likely have exempted him from prosecution in the US because he had no fiduciary obligation to the issuer.

Because of the differences in insider trading regulation, lawyers say it is harder to buy alternative data in the UK and Europe while some say exclusive data sets are effectively banned.

Personally Identifiable Information (PII)

Another major legal risk associated with alternative data is violation of consumer protection regulation. There are a few concerns for asset managers relating to PII depending on how PII is obtained, whether PII-level information is included in the data obtained by asset managers, and how PII data is anonymized.

US privacy laws are extensive, ranging from constitutional safeguards to tort law. Rules governing online data issued by the FTC in November 2016 are perhaps the most relevant guidelines for electronic data.[18] Importantly, the rules cover customer approval requirements for the use and disclosure of customer personal information.

If the consumer level data has been obtained illegally, there are few protections for end users such as asset managers. One concern centers on consents obtained from consumers. Popular weather app AccuWeather was caught sending geolocation data to a third-party data monetization firm, even when users had switched off location sharing.[19] AccuWeather and the data collection firm involved, Reveal Mobile, subsequently changed the application to conform to user preferences before any legal or regulatory actions were initiated, but the enforcement arm of the US Federal Trade Commission has brought actions in similar situations. In 2013, the developers of one of the most popular flashlight applications for Android mobile phones settled with the FTC for violating user location preferences.[20]

Consequently, many asset managers seek to receive data which excludes PII or to quarantine and purge PII, if received. For those who receive PII level information, there is the additional risk of the data being used inappropriately by rogue users, either internal or external.

In some cases, even PII data which has been anonymized can still be used to identify individuals if compared to other data and/or if users have access to specific transactions made by consumers. According to a paper published in 2013, researchers at MIT and the Université Catholique de Louvain, in Belgium, found in a 15-month study that, armed with an anonymized data set of

---

[18] FTC. "Protecting the Privacy of Customers of Broadband and Other Telecommunications Services", WC Docket No. 16-106, November 2016 https://apps.fcc.gov/edocs_public/attachmatch/FCC-16-148A1_Rcd.pdf

[19] ZDNET, "AccuWeather caught sending user location data, even when location sharing is off" http://www.zdnet.com/article/accuweather-caught-sending-geo-location-data-even-when-denied-access/

[20] FTC v. GOLDENSHORES TECHNOLOGIES, LLC https://www.ftc.gov/news-events/press-releases/2013/12/android-flashlight-app-developer-settles-ftc-charges-it-deceived

mobile phone users in a small European country, they needed just four interactions between users and network towers to uniquely identify 95% of people in the anonymized data set.[21] With just two data points, they could identify about 50 percent of users.

PII in Europe

The European Union has drafted a major update to its privacy laws -- General Data Protection Regulation (GDPR) -- which is due to come into force May 2018. The laws are expected to create a greater focus on transparency and organizations being accountable for what they do with personal data.

Jaap Tempelman, a lawyer at Clifford Chance, said funds needed to be "particularly cautious" about complying with European privacy regulation when using data culled from European users. He said: "The E-Privacy Directive says location data derived from telecoms can only be processed for very specific purposes related to providing and billing the telecoms service. If you want to use it for any other type of service you need specific consent from users."[22]

Giles Pratt, data privacy and cyber security partner at law firm Freshfields Bruckhaus Deringer, said hedge funds "either have to use the data in a way that is truly anonymized, or they will need to treat the dataset as personal data and so comply with all the rules and requirements of GDPR".[23]

Web Scraping

Web crawling is becoming broadly used as a source of alternative data and an investment research tool by both the sell side and by asset managers including hedge funds. However, firms utilizing web crawling or harvesting, need to manage the associated compliance risks due to an ever growing body of regulatory deliberation associated with the practice.

The courts find that operating a web-crawler is a legal and expected activity on the internet (Field v. Google[24]), however, financial firms need to minimize the potential headline risk as well as the potential legal costs associated with web crawling.

Prominent sites pay close attention to usage patterns and some are notorious for using legal means to stamp out non-human access to their site. Craigslist.com, for instance, has successfully sued several organizations for harvesting data from their site. Yet fewer than 50 web crawling cases have gone to US courts; a minuscule amount considering how widespread the practice has become.

---

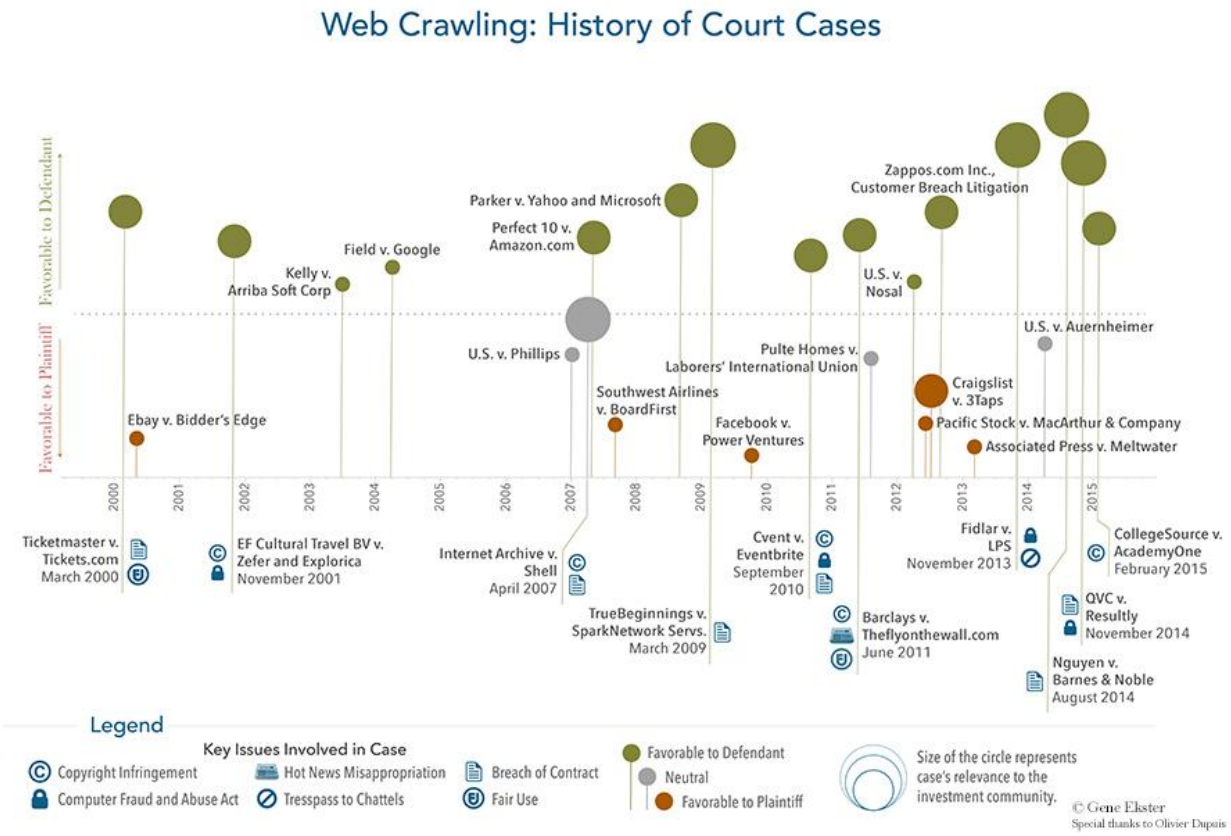
[21] Wired, "Anonymized Phone Location Data Not So Anonymous, Researchers Find" https://www.wired.com/2013/03/anonymous-phone-location-data/

[22] Financial Times, "Hedge Funds See Gold Rush in Data Mining" https://www.ft.com/content/d86ad460-8802-11e7-bf50-e1c239b45787

[23] Ibid.

[24] https://en.wikisource.org/wiki/Field_v._Google,_Inc.

Within the alternative data context, a relevant case is The Internet Archive v. Suzanne Shell.[25] The Internet Archive crawled Shell's site, which provided advice to alleged criminals with no stated commercial or competitive intent, its goal being to archive the information. Shell sued the Archive alleging breach of contract (TOU), civil theft and RICO violations. The issue was ultimately settled out of court, all the charges were dismissed except for breach of contract. Another relevant case is Fidlar v. LPS, where LPS, a real estate analytics company crawled Fidlar's lands records for analytic purposes.[26] Ultimately, the case was judged in favor of LPS.



The value of legal claims against web crawlers is low where the crawler does not crash or otherwise harm the website, and the crawled data is not used in competition with the website operator. Claims that are filed tend to be driven by crawlers which somehow damage the business of the data owners, whether directly or due to opportunity cost.

In many cases, courts have found that a crawler is not bound by the terms of use on a website, especially in case of a browser wrap. However, click-through agreements represent higher levels of legal risk. Ignoring cease and desist requests is similarly high risk.

[25] Internet Archive v. Suzanne Shell; 505 F.Supp.2d 755, Civ. No. 06-cv-01726-LTB-CBS (D. Colo., Feb. 13, 2007) http://www.internetlibrary.com/cases/lib_case456.cfm

[26] Fidlar Techs v. LPS Real Estate Data Solutions, Inc., No. 15-1830 (7th Cir. Jan. 21, 2016) http://caselaw.findlaw.com/us-7th-circuit/1723721.html

LinkedIn vs. hiQ Labs

In August, 2017, a US judge granted a preliminary injunction against LinkedIn in what promises to be a landmark case for web crawling.[27] The initial round was positive for firms which use web crawling, but the case is far from over.

Judge Edward Chen of the US Northern California District Court granted a temporary restraining order that forbids LinkedIn from blocking bots operated by hiQ Labs from harvesting data from public-facing user profiles. San Francisco-based hiQ sells human resources solutions designed to map employees' skill sets and predict attrition. It relies on publicly available information, primarily data from LinkedIn.

HiQ sued after receiving a cease and desist letter from LinkedIn, which blocked hiQ's bots and threatened to bring legal action. LinkedIn noted that its User Agreement prohibits various methods of data collection from its website, and stated that hiQ was in violation of those provisions. LinkedIn stated that any further access to LinkedIn"s data would violate state and federal law, including California Penal Code § 502(c), the federal Computer Fraud and Abuse Act ("CFAA"), 18 U.S.C. § 1030, state common law of trespass, and the Digital Millennium Copyright Act. HiQ accused LinkedIn of violating anti-trust laws.

The injunction sided with hiQ, finding that the block directly threatens hiQ's business and may even be anti-competitive. "The court concludes that based on the record presented, the balance of hardships tips sharply in hiQ's favor," Judge Chen wrote. The injunction dismissed LinkedIn's allegations that hiQ violated the anti-hacking provisions of the Computer Fraud and Abuse Act because no authorization is required to access the public profile pages on LinkedIn.

In September 2017, LinkedIn filed with the 9th Circuit Court of Appeals[28] seeking to vacate Chen's order. LinkedIn's argument centered primarily on whether web scraping violates the privacy of its users, which promised to be a more potent weapon in its battle to limit web scraping of its site.

In its appeal, LinkedIn argued that it has the right to protect the privacy of its users, and that hiQ lacks a valid antitrust claim. Reinforcing its privacy argument, LinkedIn received an amicus brief[29] from advocacy group Electronic Privacy Information Center (EPIC), a respected Washington-based non-profit dedicated to digital privacy.

EPIC's 27-page brief argued that the injunction against LinkedIn ignored the importance of privacy, which should have been the paramount consideration: "The lower court's injunction ignores a company's attempt to limit the collection and use of personal data it has obtained. This is at odds with the core purpose of modern privacy law."

---

[27] HIQ LABS, INC. v. LINKEDIN CORPORATION, Case No.17-cv-03301-EMC
https://regmedia.co.uk/2017/08/14/hiqlinkedintro.pdf

[28] HIQ LABS, INC. v. LINKEDIN CORPORATION, Case: 17-16783

[29] https://epic.org/amicus/cfaa/linkedin/hiQ-v-LinkedIn-EPIC-Amicus-Brief.pdf

EPIC characterized data scraping information pertaining to individuals as inherently violating the privacy of those individuals: "Companies such as hiQ Labs that scrape profile data undermine the privacy preferences of LinkedIn users. The tracking of user profile edits negates the user's 'Do Not Broadcast' choices. The lower court failed to recognize that a mandatory injunction prohibiting LinkedIn from protecting user profile information would directly harm users' interests."

HiQ's arguments, and an amicus brief filed by the Electronic Frontier Foundation, DuckDuckGo and the Internet Archive, have centered on the Criminal Fraud and Abuse Act, claiming it should not be used to stop users from accessing public content. Arguments in the case are scheduled to be heard March 2018.

Other Risks

Although much of the compliance attention is focused on the risks of insider trading, privacy regulation and copyright law, there are other potential risks associated with alternative data.

Data Quality Risk

One risk is associated with poor quality data. Because most alternative data is unstructured, the data needs to be processed in some form to become usable. Techniques depend on the nature of the underlying data. Textual data on the web tends to be filtered by natural language processing. Geolocation data often needs to be mapped by geo-fencing coordinates. The quality and efficacy of these techniques may vary significantly.

Quality concerns are magnified when the alternative data is unique or scarce, making it extremely difficult to validate its accuracy. According to Deloitte, "Since each alternative data set may be unique or scarce, investment teams may have difficulty finding a way to verify the accuracy of a data set. In some cases, detailed review of the procedures used to gather and manipulate the data may be an IM firm's best strategy to mitigate this risk."[30]

Immateriality

Alternative data may be scarce or hard to obtain but that does not necessarily mean it will contribute positively to the investment process. Similarly, data might be relevant for a certain asset class or industry sector, but otherwise useless. In addition, data that is of value for investors that have relatively short investment horizons may be less valuable for investors with longer investment horizons.

Model Risk

After the SEC levied a $240 million fine in 2011 against AXA Rosenberg for concealing model error, asset managers instituted controls to ensure investment models are performing within the

[30] Deloitte, "Alternative data for investment decisions: Today's innovation could be tomorrow's requirement" https://www2.deloitte.com/us/en/pages/financial-services/articles/alternative-data-for-investment-management-decisions.html

guidelines of their related investment policy statements and in accordance with client disclosures.[31] As alternative data inputs affect portfolio construction, the impacts on model efficiency need to be monitored.

According to Deloitte, alternative data is highest risk during the model input phase: "There is risk of complication at each point of the model revision—input, implementation, and output—with alternative data impacting the model perhaps most at the input phase." In addition, alternative data may be incorporated in the model incorrectly, creating irregular or inconsistent trading signals under certain conditions.

Disclosures

Some commentators believe that use of alternative data may require additional disclosures. Deloitte argued that "Alternative data is different enough from traditional data that asset managers should consider modifying investor disclosures about investment policy and processes."[32]

Growing Media and Regulatory Scrutiny

The proliferation of alternative data, greater usage by asset managers – particularly hedge funds – and the increasing number of risks associated with alternative data are combining to increase media and regulatory scrutiny of alternative data.

Closer Examination by the Media

Media attention on alternative data is increasing. Within the last eighteen months, major publications have written articles highlighting the growth of alternative data. Sources include

[31] SEC Administrative Proceeding File No. 3-14224 https://www.sec.gov/news/press/2011/2011-37.htm
[32] Deloitte, "Alternative data for investment decisions: Today's innovation could be tomorrow's requirement" https://www2.deloitte.com/us/en/pages/financial-services/articles/alternative-data-for-investment-management-decisions.html

The Wall Street Journal[33], Financial Times[34], Forbes[35], Newsweek[36], Institutional Investor[37], The Economist[38], The Australian Financial Review[39], among others.

Much of the coverage has been benign, but not all. A 2000-word article on alternative data published in August 2017 by the Financial Times was subtitled "Investors seek an advantage from tradable intelligence but legality of alternative data is in question."[40] In a September article, reporters for the Financial News proactively sought comments – largely negative – from consumer protection advocates and regulators on the growing hedge fund use of geolocation data.[41]

Increasing attention to PII

Consumer privacy advocates are increasingly raising concerns about hedge fund use of mobile phone users' location data, thanks to the media.

Frederike Kaltheuner, a policy officer at Privacy International, suggested that hedge funds using location data to inform their trading strategies had "unprecedented population-level insight about individuals' lives, communities and entire nations and markets" and questioned whether the practice was ethical.[42] Ed Johnson-Williams, a representative of Open Rights Group, a charity that protects citizens' digital rights, said: "While some people may be happy for their data to be used for public-interest research, they may not want their data to be sold on to contribute to a hedge fund's investment decisions."[43]

Griff Ferris, a researcher at Big Brother Watch, a civil liberties group, argued that because it is possible for data to be de-anonymized by combining it with publicly available records, even anonymized geolocation data is risky: "The collection of such personal and sensitive datasets

---

[33] The Wall Street Journal, "Wall Street's Insatiable Lust: Data, Data, Data" https://www.wsj.com/articles/wall-streets-insatiable-lust-data-data-data-1473719535

[34] Financial Times, "Hedge Funds See Gold Rush in Data Mining" https://www.ft.com/content/d86ad460-8802-11e7-bf50-e1c239b45787

[35] Forbes, "How To Predict Netflix Subscriber Growth With Alternative Data" https://www.forbes.com/sites/alapshah/2017/07/20/how-to-predict-netflix-subscriber-growth-with-alternative-data/#60b219216d73

[36] Newsweek, "Neudata Releases Version 2 of its 'Scout' Alternative Data Platform" http://www.newsweek.com/neudata-releases-version-2-its-scout-alternative-data-platform-656923

[37] Institutional Investor, "The Next Big Source of Alpha: Your Phone?" https://www.institutionalinvestor.com/article/b1505pw79hfj54/the-next-big-source-of-alpha-your-phone

[38] The Economist, "Why Investors Want Data" https://www.economist.com/blogs/economist-explains/2016/08/economist-explains-18

[39] The Australian Financial Review, "Hedge funds step up hunt for alternative data to detect emerging trends" http://www.afr.com/business/energy/electricity/hedge-funds-step-up-hunt-for-alternative-data-to-detect-emerging-trends-20170814-gxw8mp

[40] Financial Times, "Hedge Funds See Gold Rush in Data Mining" https://www.ft.com/content/d86ad460-8802-11e7-bf50-e1c239b45787

[41] Financial News, "Campaigners Sound Alarm Over Hedge Funds' Data Use" https://www.fnlondon.com/articles/regulators-campaigners-sound-alarm-over-hedge-funds-data-use-20170904

[42] Ibid.

[43] Ibid.

therefore poses a potentially very serious privacy risk, and both individuals and companies should be aware of this."[44]

Privacy issues surrounding social media and cell phone usage are potentially explosive, and a convenient match is the growing use of such information by secretive, highly paid hedge funds. Another potential theme for enterprising reporters is how the high cost of alternative data creates an uneven playing field for individual investors. The alternative data industry should be prepared for negative media attention attempting to cast alternative data as unethical or illegal.

Growing Regulatory Attention

Where media goes, regulators follow. Although the SEC has successfully enforced an insider trading prosecution involving alternative data – discussed above – regulatory attention to alternative data has been light so far.

The Consumer Financial Protection Bureau (CFPB), a relatively new US regulatory authority founded six years ago, is investigating the use of alternative data in the generation of credit scores. The CFPB issued a request for comment on the use of alternative data and modeling techniques in the credit process in February 2017 and closed the comment period in May, 2017.[45] The CFPB is issuing no-action letters to users of alternative data to generate credit scores.[46]

The UK Financial Conduct Authority, which regulates Britain's security markets, told reporters from Financial News that although alternative data is not an area the regulator is looking at specifically, the regulator does monitor data financial firms collect and how they use it.[47] If alternative data were deemed insider information, the FCA indicated it would pursue enforcement action.

The UK Information Commissioner's Office, an independent authority which regulates information rights, warned that location data is "a valuable source of information but its collection and use may be considered very intrusive," citing anonymized and aggregated data as "key safeguards."[48]

Some commentators expect cases to be brought against users of alternative data. Jonathan Streeter, a former federal prosecutor who led the insider trading case against Raj Rajaratnam, told the *Financial Times* that he expected prosecutors to bring cases involving alternative data: "I

[44] Ibid.
[45] Consumer Financial Protection Bureau, "Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process" https://www.consumerfinance.gov/policy-compliance/notice-opportunities-comment/archive-closed/request-information-regarding-use-alternative-data-and-modeling-techniques-credit-process/
[46] http://files.consumerfinance.gov/f/documents/201709_cfpb_upstart-no-action-letter.pdf
[47] Financial News, "Campaigners Sound Alarm Over Hedge Funds' Data Use" https://www.fnlondon.com/articles/regulators-campaigners-sound-alarm-over-hedge-funds-data-use-20170904
[48] Ibid.

don't know of a case that's been brought, but everyone anticipates that there will be one soon and [prosecutors] would like to bring one."[49]

Peter Greene, a partner at NY law firm Lowenstein Sandler said, "We have not seen the SEC take a hard look at it but ultimately regulators are going to want to understand this area and how data is procured."[50]

The *Financial Times* quoted an anonymous fund manager who cited rising legal risks surrounding alternative data: "It's like the Wild West and ultimately it will come under the purview of the regulators. I have yet to see a clean legal opinion on all this. And we've looked for one."[51]

New York attorney-general, Eric Schneiderman, who prosecutes financial fraud under the broad scope of the Martin Act, is cited by legal experts as a potential source of alternative data litigation. In 2013 the NY Attorney General's office brought action against Thomson Reuters for providing premium subscribers early access to consumer sentiment data.[52]

If risks associated with alternative data are increasingly low hanging fruit for journalists, a similar dynamic works for ambitious regulators. State level attorney generals, particularly those in New York and Massachusetts, have historically taken the lead in prosecuting cases with media potential. Once one regulator takes up the cudgel, others feel compelled to join the fray to protect their regulatory ambits.

The Importance of Industry Standards

In this paper, we argue that the expansion of alternative data combined with growing usage by asset managers is heightening the associated risks. New types and uses of alternative data have introduced new varieties of risks. More players consuming the data create more opportunities for inappropriate use.

In the early stages of alternative data usage, quantitative funds focused on large data sets covering broad sets of investments and asset classes. Quant funds demanded broad coverage and high frequency. Low latency news and sentiment data derived social media were based on public data, mitigating insider trading concerns, and contained small risk of privacy or copyright violations. The increasing use of transaction data, web scraping, and geolocation data have introduced additional risks to the alternative data landscape.

---

[49] Financial Times, "Hedge Funds See Gold Rush in Data Mining" https://www.ft.com/content/d86ad460-8802-11e7-bf50-e1c239b45787

[50] HFMCompliance, "How Should CCOs Deal With Alternative Data Risks", November 2017 https://hfm.global/hfmcompliance/analysis/how-should-ccos-deal-with-alternative-data-risks/

[51] Financial Times, "Hedge Funds See Gold Rush in Data Mining" https://www.ft.com/content/d86ad460-8802-11e7-bf50-e1c239b45787

[52] Integrity Research "Early Release of Sentiment Data Not Unlawful, But…" http://www.integrity-research.com/early-release-of-sentiment-data-not-unlawful-but%e2%80%a6/

At the same time, the profile of alternative data users has evolved from quant funds to long/short equity funds to long-only investors. The increasing focus on specific sectors and specific issuers has also heightened risks. Forty percent of hedge fund CCOs surveyed by HFMCompliance said they have turned down research over compliance concerns while a further 40% said that, although they have not turned down requests, they have asked for clarifications.[53] 20% said they have always been comfortable with the research they have received from providers.

If usage of alternative data has inconsistent oversight, vendors with poor compliance practices will also flourish. In this sense, the current situation with alternative data can be compared to the use of expert networks a decade ago. In 2009, when Integrity Research surveyed the expert network industry, there were over forty expert networks, with varying levels of compliance capabilities.[54] As subsequent enforcement actions indicated, at least one expert network was cutting compliance corners, and the resulting insider trading investigations created significant stress for the entire industry.

Nevertheless, leading expert networks had developed a set of compliance standards which we documented in our reports at the time. By having a consistent set of best practices which were adhered to by the larger, more responsible firms, the expert network industry was able to avoid direct regulatory oversight. Although there are fewer expert networks today than there were in 2009, the industry is setting new revenue records.

We therefore believe that it will be increasingly important for the alternative data industry, including buy-side consumers, sell-side users, vendors and those that originate alternative data, to take proactive steps to inoculate the industry from adverse media and regulatory scrutiny. By developing consistent and broadly adopted compliance standards, alternative data participants can reduce legal risks and compliance costs, not only for their individual entities but for the industry overall.

[53] HFMCompliance, "How Should CCOs Deal With Alternative Data Risks", November 2017 https://hfm.global/hfmcompliance/analysis/how-should-ccos-deal-with-alternative-data-risks/

[54] Integrity Research Associates, ResearchFocus: Expert Networks 2009 http://www.integrity-research.com/expert-networks-outlook-exclusive-report/