The HEDGE FUND LAW REPORT
The definitive source of actionable intelligence on hedge fund law and regulation
Lowenstein Sandler
www.hflawreport.com Vol. 11, No. 2 January 11, 2018

TECHNOLOGY

A Fund Manager's Roadmap to Big Data: Its Acquisition and Proper Use (Part One of Three)

By Shaw Horton, The Hedge Fund Law Report

The exponential growth of IT systems has given researchers, governments, corporations and fund managers the ability to identify correlations and patterns from a combination of previously unlinked data sets with incredible speed. "Big data" often refers to the use of predictive analytics, which extract value from these data sets. Raw data can be collected from a variety of sources, including user interactions on the internet, satellite images, consumer transactions and industry trends.

Although only a small minority of fund managers comprehensively capture value from this data, spending on big data continues to increase with fundamental-driven investors seeking to enter the environment. Building an internal infrastructure to acquire and process raw data is a time-consuming and expensive undertaking. As a result, most fund managers look to third-party data vendors in an effort to not only generate alpha, but to respond to new regulatory requirements; reduce costs; and assist with other operational and managerial functions.

This article, the first in a three-part series, explores the big-data landscape and how fund managers can acquire and use big data. The second article will analyze issues and best practices surrounding the acquisition of material nonpublic information; web-scraping; and the quality and testability of data. The third article will discuss risks associated with data privacy, the acquisition of data from third parties and the use of drones, as well as ways fund managers can mitigate those risks.

For more on big data, see "Tips and Warnings for Navigating the Big Data Minefield" (Jul. 13, 2017).

What is Big Data?

The use of automated systems to store, retrieve, study, distribute and manipulate data has experienced, and continues to experience, tremendous growth. The rapid advancement in technology that has opened the door to big data or alternative data can be explained in part by Moore's law – the observation (or, in many ways, the target or self-fulfilling prophecy) that the number of transistors on a microchip will double every two years.

IBM defines big data as "a term applied to data sets whose size or type is beyond the ability of traditional relational databases to capture, manage, and process the data with low-latency. And it has one or more of the following characteristics – high volume, high velocity, or high variety." In other words, automated systems are now able to quickly identify correlations and patterns from a combination of different and once-unlinked datasets.

Humans can interpret these patterns to develop meaningful conclusions, which can have profound impacts on crime (e.g., the Los Angeles Police Department uses PredPol to predict and stop crime), healthcare (e.g., researchers can analyze DNA to predict disease patterns) and investing. "In many ways, alternative data is what we used to think of as research," noted Evan Schnidman, founder and CEO of Prattle. "Alternative data involves supplementing traditional research with unbiased data."

The raw data comes from a variety of sources, including the internet, social media, the Internet of Things, geospatial imagery, consumer transactions and user-generated location data. Emmett Kilduff, founder and CEO of Eagle Alpha, indicated that the most popular categories of alternative data among hedge funds are consumer transactions, geospatial imagery and sentiment.

The Internet, Public Statements and Data Exhaust

Broadly, the largest data sets are found on the internet. For example, Google processes more than 60,000 search queries every second and can use this data to analyze the individual or collective preferences of users. The Clearnet – i.e., the parts of the internet that are indexed by standard search engines – also provides vast troves of data, including statements individuals post on forums; job postings; web traffic; public court records; and public statements by governments and companies.

Machine learning can synthesize public statements through natural language processing (NLP). With NLP, computers read and analyze text to draw conclusions about the tone of a news article or company report, as well as the subjects on which companies focus over long time horizons. Goldman Sachs has

stated on its site that it uses this approach to assemble less-obvious relationships from clusters of companies that "appear together in news articles, regulatory filings or research reports."

Similarly, Prattle uses NLP to analyze public statements from central banks and companies to derive quantitative scores on, for example, the hawkishness or dovishness of certain central bank statements, as well as the bullishness and bearishness of corporate remarks. "Because we've built a unique lexicon for each individual company, we can pick up on nuanced linguistic differences between how companies communicate," Schnidman remarked. "We can analyze how each word, phrase, sentence or paragraph relates to each other word, phrase, sentence or paragraph, and tie the patterns in language to specific price movements."

Users also generate "data exhaust," or data created as a byproduct of their online actions (e.g., sites visited or links clicked). Single-sign-on frameworks allow companies to develop even more detailed profiles of users, particularly when this data is combined with cookies. Companies have also been able to generate and aggregate user data by adding engagement tools to websites (e.g., by allowing users to share content).

Social Media

Social media networks provide forums for users to interact generally (such as through Facebook, Twitter or Reddit) or on more specific topics (such as through Instagram, Flickr, LinkedIn or Meetup). By their very nature, these sites invite users to generate more data, thereby increasing the value of the sites.

Paradoxically, giving users greater control over their information-sharing practices has led to a decrease in privacy. Companies do not typically put these activities through rigorous internal ethical review processes given that the data are rarely made public and the impact on individual users is seen as small.

The social media sites themselves or other third parties can use this data for commercial marketing, research or other purposes. The data can be used to track, among other things, brand popularity and customer satisfaction. Global analytics company company CRISIL, in its May 2017 report on big data in asset management, noted that "Dataminr, which applies analytics on Twitter data, had revealed preliminary reports of Volkswagen's emissions scandal three days before the market reacted."

See "Best Practices for Investment Advisers Using Social Media to Mitigate Advertising Rule Violations and Other Risks" (Mar. 23, 2017)

The Internet of Things

The "Internet of Things," which refers to the interconnection of objects that are not easily identifiable or accessible via conventional search engines, is also a large source of data. These objects are used to, among other things, manage inventory, monitor electricity usage and make life easier at home (e.g., through smart home objects like printers, thermostats and lights).

Some search engines are making this data more accessible. For example, Thingful can locate the geographical position of objects and devices. This makes it easier for companies to synthesize personal data on the ways in which users interact with their connected devices. Shodan has gathered data on millions of devices, including servers and routers. The Washington Post reported[1] that sophisticated Shodan users can access "industrial control computers [and] the systems that automate such things as water plants and power grids."

Geospatial Imagery

Satellite imagery has also emerged as an enormous source of data for governments, construction companies and financial services firms. For example, SpaceKnow uses 2.2 billion satellite observations over 500,000 square kilometers to track 6,000 industrial facilities across China in order to generate an index of manufacturing activity. The index provides a counterpart to the official PMI numbers, which many investors view with a degree of suspicion.

"Synthesizing geospatial data into a country-wide index is a challenging task because not only do measurements have to be made on an ultra large scale, but the individual locations being monitored for change must be hand annotated to provide accurate insights," said Hugh Norton-Smith, vice president of financial services at SpaceKnow. Thus, while the data can provide general insights into a particular country's economy, it can also be used to analyze things like commodity production (e.g., through the tracking of crop density).

SpaceKnow has developed relationships with the largest satellite providers (e.g., DigitalGlobe, Airbus and Planet Labs) who offer different types of sensing platforms. While one entity may provide high resolution imagery, another may provide images from different spectral bands or provide more frequent overpass of locations.

The HEDGE FUND LAW REPORT
The definitive source of actionable intelligence on hedge fund law and regulation
Lowenstein Sandler
www.hflawreport.com Vol. 11, No. 11 March 15, 2018

"Geospatial imagery has been available in the public domain for some time now, so we have a high degree of comfort sharing insights derived from this with our clients," noted Norton-Smith. "As an alternative data source, geospatial has been around longer and is better understood than many other forms of alternative data, such as credit-card transaction reports or location tracking. We are one of a number of organizations that have been selling this information to hedge funds and commodity traders for some years now."

Consumer Transactions

Another major source of data is consumer and small-business transactions. Websites and apps like Mint, for example, aggregate data on individual users' credit-card and debit-card transactions, bank accounts, investment accounts and loan accounts. This allows aggregators, as well as companies that acquire the data, to analyze consumer behavior for specific industries and retailers.

The privacy statement of Intuit's Mint, for instance, states that the company may use user data for "research, including publishing or sharing combined information from many users. . . ." Even if a company promises not to sell aggregate data, it can change its privacy agreement at any time.

Location Data

Location data, particularly in the context of mobile devices, is another source that can provide powerful insights into consumer behavior. Google, for example, collects data on users' locations, ranging from IP addresses, GPS and other sensors (e.g., accelerometers and gyroscopes). Mapping data services like HERE or Waze also generate data, particularly in connection with the use of automobiles. Foursquare collects data on users' browsing histories, purchases and location "check-ins," which is then made available to third parties via its enterprise platform.

Other Sources

Other data sources include:

- healthcare or biometric information (e.g., data pertaining to the face, eyes, fingerprints and other body parts);
- scientific research;
- commerce information (e.g., the use of VHF radio transmissions to track the movements of hundreds of thousands of ships across the world);
- industry trends (e.g., the aggregation of historical research from industry analysts to identify patterns in specific sectors and products); and
- e-government (e.g., biometric passports, online services, voting systems, citizen participation tools and recordings of government sessions).

Acquiring and Using Big Data

In a 2015 study conducted by The Economist Intelligence Unit (EIU) and sponsored by Northern Trust, only 13% of asset managers indicated that they captured value from data "entirely." Although the use of alternative data has become more common, few best practices exist. As the survey notes, "asset managers are forced to expend resources on cleaning and preparing data that is often incompatible or even incorrect," and many organizations simply lack the analytical expertise.

According to consulting firm Opimas[2], spending on alternative data will exceed $7 billion by 2020. This rapid expansion is evident in the number of vendors that provide data sets to hedge funds and other asset managers. Norton-Smith explained that while SpaceKnow has traditionally provided data to quantitative hedge funds, they are working with an increasing number of traditional, fundamental-driven investors who are looking to overlay their fundamental investment processes with alternative data insights. He noted that this is driven both by the need for new alpha sources and also because of allocator expectations.

See "Ernst & Young 2017 Survey Examines Hedge Fund Strategic Priorities; Hedge Fund Offerings and Investor Allocations; Evolution of Front Offices; and Industry Risks" (Dec. 14, 2017).

Fund managers can collect and synthesize alternative data sets (often through the use of machine learning or cognitive computing) internally or through third-party data vendors. Subscription costs for many of the services offered by data vendors is often lower than building the requisite infrastructure internally. As a result, "the overwhelming majority of managers are acquiring data through vendors," said Peter Greene, partner at Lowenstein Sandler.

"The only method pursued in-house with any regularity is web scraping." Managers who collect data internally may still use third-party data to validate their own conclusions, explained Proskauer partner Jeffrey Neuburger.

In either case, however, methods for collecting the data include the following:

- purchasing or licensing it from the primary source (e.g., directly from a social media website through access to application-programming interfaces). In most instances, these data have a large user base. Some companies, however, are offering data to a limited set of users at significantly higher prices;
- web crawling (i.e., the process of systematically browsing the internet, typically for the purpose of indexing) and web scraping (i.e., the process of fetching, or downloading, and extracting information from a webpage); and
- acquiring raw data directly, for example through access to hardcopy archives that have been converted into easy-to-use digital formats like XML; through the use of beacons or sensors, such as in medical devices, road cameras or farm machinery; or through images captured from satellites or drones.

Eagle Alpha provides data in three forms: raw, processed and curated. Kilduff noted that to serve the entire global asset management industry, it is important that data providers address all three. "Many firms don't have the bandwidth to work with raw data, so they are happy to get processed or curated data."

In a recent publication on alternative data in the asset management industry, Deloitte noted that firms are primarily adopting alternative data to acquire information advantage; "any edge, even a narrow timing advantage, may yield a more effective trading signal, algorithm, or investment model," Deloitte stated. Greenwich Associates surveyed 23 hedge funds located in the U.S. and Europe in late 2016, asking respondents to identify key areas in which alternative data plays a role in the investment process: 61% said that they use data as a predictor for future market or sector movements; 48% for idea generation; 44% to research specific names; and 39% to find market mispricing and arbitrage opportunities. In other words, hedge funds can use alternative data to develop quantitative strategies; conduct fundamental analyses; analyze high-level trends and investment decision-making; identify systemic risks; make predictions; or measure liquidity or the impact of high-frequency trading.

There are, however, many advantages to big data that go beyond investing. In EIU's study, managers cited responding to new regulatory requirements, improving customer satisfaction, meeting new business goals, reducing costs, expanding efficiency, improving controls and tracking demographic trends of customers as reasons for investing in new data sources. In addition, Citi Business Advisory Services conducted a study on big data in investment management and found that compliance teams are using big data for eDiscovery or to help standardize compliant uses of social media. Furthermore, marketing teams are "looking to examine investor and distribution information to better target capital-raising efforts."

[1] "Cyber search engine Shodan exposes industrial control systems to new risks" (Jun. 3, 2012) is available here: https://www.washingtonpost.com/investigations/cyber-search-engine-exposes-vulnerabilities/2012/06/03/gJQAIK9KCV_story.html?utm_term=.cdfb2c719df7.

[2] Optimas' report, "Alternative Data – The New Frontier in Asset Management" (Mar. 31, 2017), may be read here: http://www.opimas.com/research/217/detail/

The HEDGE FUND LAW REPORT
The definitive source of actionable intelligence on hedge fund law and regulation
Lowenstein Sandler
www.hflawreport.com Vol. 11, No. 3 January 18, 2018

TECHNOLOGY

A Fund Manager's Roadmap to Big Data: Its Acquisition and Proper Use (Part Two of Three)

By Shaw Horton, The Hedge Fund Law Report

As fund managers increasingly turn to sophisticated data streams to boost investment returns and produce greater operational efficiencies, it is critical that they understand the legal and practical risks posed by the use of big data. Issues surrounding material nonpublic information (MNPI) pose the greatest threat to firms. Managers must understand not only the misappropriation framework under the Securities Exchange Act of 1934 (Exchange Act), but also how the New York State Attorney General (NYSAG) and regulators in the E.U. pursue insider trading claims. Additionally, whether engaging internally in web scraping or purchasing scraped data from third parties, managers must be conscious of contractual, intellectual property (IP) and tort claims that a site owner may allege against a fund manager. Finally, many of the largest challenges posed by the use of big data are practical or ethical in nature.

This second article in our three-part series on big data analyzes issues and best practices surrounding the acquisition of MNPI; web scraping; and the quality and testability of data. The first article explored the big-data landscape and how fund managers can acquire and use big data in their businesses. The third article will discuss risks associated with data privacy, the acquisition of data from third parties and the use of drones, as well as recommended methods for mitigating those risks.

For more on big data, see "Best Practices for Private Fund Advisers to Manage the Risks of Big Data and Web Scraping" (Jun. 15, 2017).

x

"The most important legal concerns relate to violations of the securities laws," stated Proskauer partner Jeffrey Neuburger. "Commercial issues can be worked out quietly, but issues with the SEC or the DOJ are in another league. While the SEC and DOJ haven't brought any cases yet in the alternative data space, they've been active in breach of duty cases and hacking cases, so I expect them to begin to focus on it more."

"The big brand-killer for any manager is a regulatory investigation that is alleging a breach of a securities law in pursuing trading strategies," continued Derek Steingarten, partner at K&L Gates. "The fund will be long out of business before a determination has been made as to whether the actions were appropriate or not."

Although there are no specific laws, regulations or SEC guidance that address the gathering and use of big data, trading on MNPI in this context may lead to insider trading liability under the antifraud provisions of Section 10(b) and Rule 10b-5 under the Exchange Act. The government must prove that a defendant:

1. purchased or sold securities;
2. misappropriated information "from a person or entity to whom he or she owed a fiduciary duty or other relationship of trust and confidence";
3. knowingly possessed MNPI; and
4. acted with scienter.

Under Rule 10b5-2, a duty of trust and confidence exists whenever:

- a person agrees to maintain information in confidence;
- there is a history, pattern or practice of sharing confidences "such that the recipient of the information knows or reasonably should know that the person communicating the [MNPI] expects that recipient will maintain its confidentiality"; or
- information is obtained from one's spouse, parent, child or sibling.

The HEDGE FUND LAW REPORT
The definitive source of actionable intelligence on hedge fund law and regulation
Lowenstein Sandler
www.hflawreport.com Vol. 11, No. 3 January 18, 2018

According to Deloitte, "the definition of material is . . . subject to interpretation with some firms relying on statistical testing to determine whether information is material or not." Moreover, "if an alternative data set is thought to be too predictive of normally protected information such as quarterly revenue, then some firms are steering clear of the data" altogether. See "How Can Hedge Fund Managers Distinguish Between Market Color and Inside Information" (Nov. 19, 2009).

Although many of the risks associated with the use of alternative data are specific to web scraping, managers must be particularly conscious about MNPI because it "applies across the use of all alternative data sets," stated Benjamin Kozinn, partner at Lowenstein Sandler. "MNPI is an existential risk to a firm and by far the most significant issue we deal with."

For more on insider trading, see "General Insider Trading Policies and Procedures May Be Insufficient for Hedge Fund Managers to Avert SEC Enforcement Action" (Nov. 3, 2016).

Misappropriation

In Carpenter v. United States, a columnist at the Wall Street Journal (WSJ) shared confidential information with stockbrokers prior to the publication of a column, knowingly in contravention of the WSJ's rules. The stockbrokers "bought and sold stocks based on the column's probable impact on the market and shared their profits" with the columnist.

The U.S. Supreme Court held that the WSJ had a property right in "keeping confidential and making exclusive use, prior to its publication, of the schedule and contents" of the column. Therefore, the columnist's activities constituted a scheme to defraud the WSJ because he violated his "fiduciary obligation to protect his employer's confidential information by exploiting that information for his personal benefit, all while pretending to perform his duty to safeguarding it."

For more on Carpenter, see "Lessons for Hedge Fund Managers From the Government's Failed Prosecution of Alleged Insider Trading Under Wire and Securities Fraud Laws" (Jul. 21, 2016).

Similarly, in SEC v. Huang, a data analyst at Capital One "downloaded and analyzed confidential information regarding purchases made with Capital One credit cards at over 200 consumer retail companies and used that information to conduct more than 2000 trades in the securities of those retail companies." The U.S. Court of Appeals for the Third Circuit found the defendant guilty of insider trading, holding that the information was material because it altered the total mix of information in the eyes of a reasonable investor and that the defendant misappropriated the information by violating Capital One's confidentiality policies.

For more on the misappropriation theory, see "Court to Rule on Novel Issue of Insider Trading Law in Case Against Leon Cooperman and Omega Advisors" (Mar. 30, 2017).

Courts have also found that misappropriation liability can arise from deceptive conduct such as hacking. The U.S. Court of Appeals for the Second Circuit (Second Circuit) held that "misrepresenting one's identity in order to gain access to information that is otherwise off limits, and then stealing that information is plainly 'deceptive' within the ordinary meaning of the word." This concept was illustrated in SEC v. Hong, where the defendants allegedly hacked into the networks of two law firms (through the installation of malware) to steal confidential information. The Second Circuit opined, however, that gaining unauthorized access through the mere exploitation of a weakness in an electronic code may not constitute deception.

Thus, it remains to be seen whether courts will regard the breach of a website's terms of use (TOU) or other contractual arrangements through the use of deception is sufficient for insider trading liability. Nevertheless, given that this would not be an enormous leap, in those situations "all that managers would have left is that the information obtained is public," stated Lowenstein Sandler partner Peter Greene. "The safest way to draw the line is to assume that information that can be obtained only through password access is non-public."

"People have been selling traditional data for decades and have been conscious of legal concerns, like MNPI," remarked Emmett Kilduff, founder and CEO of Eagle Alpha. "For better or worse, the law has not evolved. Therefore, while the analysis remains the same, it must be applied to a new context," continued Steingarten.

While fund managers must understand, for example, what actions constitute deception in the changing technological landscape, they should continue to employ the same techniques they currently use to prevent issues related to MNPI. These include, among other things, adopting a policy regarding insider trading; recordkeeping; implementing employee training programs; monitoring employees' personal securities trading; maintaining information barriers; and enforcing issues internally.

See "K&L Gates Partners Identify Eight Actions That Hedge Fund Managers Can Take to Avoid Insider Trading Violations (Part Two of Three)" (Nov. 20, 2014).

Duty to Others, Martin Act and E.U. Market Abuse Regulation

In both Carpenter and Huang, the courts left open whether owners of information can trade on their own confidential information. Owners must nevertheless be careful, as they may owe a duty to others. A credit card company may owe a duty to its customers, for example, because it has agreed to only use transaction information for certain purposes in its privacy policy.

Moreover, under the Martin Act, owners may be liable absent a breach of duty. The NYSAG may prosecute "fraud or misrepresentation in the public offer, sale and purchase of securities and commodities" when the government proves that the defendant engaged in a "misrepresentation or omission of a material fact or other conduct which deceives or misleads the public, or even tends to deceive or mislead the public." Unlike with securities law violations, the NYSAG need not show scienter or damages under the Martin Act.

Eric Schneiderman, the current NYSAG, has used the act to pursue "Insider Trading 2.0." For example, his office forced "Thomson Reuters to stop giving a select group of clients access to . . . market-moving information in the University of Michigan's Survey of Consumers two seconds before the rest of their subscribers saw it."

For more on the Martin Act, see "Newly Appointed Chief of New York's Investor Protection Bureau Describes Its Enforcement of the Martin Act and How Managers Can Avoid Prosecution" (Oct. 20, 2016).

In addition, owners must be cognizant of non-U.S. insider trading regimes. Notably, the E.U. Market Abuse Regulation (MAR) applies to companies with financial instruments admitted to trading in the E.U. MAR applies to transactions that take place in a third country and prohibits "persons in possession of inside information from using that information to deal or attempt to deal in financial instruments or to recommend or induce another person to transact on the basis of inside information." Unlike in the U.S., there is no requirement for a fiduciary duty or other relationship of trust and confidence.

See also "Ten Practical Consequences for Hedge Fund Managers of the FCA's Thematic Review of Asset Managers and the E.U. Market Abuse Regulation" (Mar. 19, 2015); and "E.U. Market Abuse Scenarios Hedge Fund Managers Must Consider" (Dec. 17, 2015).

Issues Related to Web Scraping

Breach of Contract

Web crawling and web scraping may violate a website's TOU, end-user license agreement or application programming interface. Website owners typically set forth these agreements in "clickwrap" or "browsewrap" form. Clickwrap agreements require users to affirmatively check a box acknowledging assent before they are allowed to proceed. Browsewrap agreements are typically posted as a hyperlink on the bottom of the website and require no user action.

When determining enforceability, courts look to whether the user had reasonable notice of and manifested assent

The HEDGE FUND LAW REPORT
The definitive source of actionable intelligence on hedge fund law and regulation
Lowenstein Sandler
Vol. 11, No. 3 January 18, 2018

to the agreement, as with any other contract. Courts generally do not enforce browsewrap agreements because mere use of a website does not signify assent. Enforceability, therefore, turns on whether the user had actual or constructive knowledge of the TOU.

In Nguyen v. Barnes & Noble, the U.S. Court of Appeals for the Ninth Circuit (Ninth Circuit) held that:

> where a website makes its terms of use available via a conspicuous hyperlink on every page of the website but otherwise provides no notice to users nor prompts them to take any affirmative action to demonstrate assent, even close proximity of the hyperlink to relevant buttons users must click on – without more – is insufficient to give constructive notice.

The U.S. District Court for the Western District of Virginia Roanoke Division, however, held that a user may have constructive knowledge of a website's TOU when: (1) she creates a "fictitious profile and email account" to carry out a prohibited activity; or (2) her own website contains a similar browsewrap agreement to that of the aggrieved party.

Clickwrap agreements, on the other hand, are routinely held as enforceable, even when the user does not read the agreement. Hybrid clickwrap-browsewrap agreements (i.e., agreements where the TOU are only visible via a hyperlink but where a user must indicate that he or she has read and agreed to the TOU) have also consistently been held as enforceable.

While it is impractical to review every website's TOU, managers should be aware that there are risks to engaging in web scraping without a comprehensive review. Nevertheless, "such claims will likely be pursued only where the manager attempts to compete with the website operator, infringes on the copyright of the material on the website, or impairs the performance of the site," Greene asserted, adding, "Given that these are unlikely, breaching a website's TOU is a business risk that a lot of managers may be willing to take."

Copyright Infringement

To receive copyright protection, a work must be original and fixed in a tangible medium. Although facts (such as the list price of a vehicle) are not copyrightable, creative facts (such as the Kelley Blue Book valuation of a vehicle) are copyrightable as long as the process is not entirely mechanical. Even a compilation of facts receives copyright protection where the author uses discretion in arrangement and inclusion.

Copyright owners have the exclusive right to copy or reproduce their works; prepare derivative works; distribute or sell their works; and display their works publicly. Thus, automated data collection via web scraping may infringe upon a site owner's copyright if it leads to the reproduction of copyrighted content (including underlying code or user-generated content). Even a small taking of the original work, if qualitatively significant, may be sufficient to trigger claims of copyright infringement.

For information on how hedge funds can protect their own IP, see "How Hedge Funds Can Protect Their Brands and IP: Pepper Hamilton Attorneys Discuss Trademarks and Copyrights (Part One of Two)" (Feb. 23, 2017).

To avoid liability in the course of web scraping, managers should refrain from collecting or manipulating data that contains creative elements. If a manager must copy copyright content, however, it may find safe haven in the fair use doctrine. In evaluating whether an infringement is fair use, courts look at:

1. the purpose and character of the use;
2. the nature of the copyrighted work;
3. the amount of the original work used; and
4. the effect the use will have upon the potential market for or value of the copyrighted work.

In Kelly v. Arriba Soft Corp, the defendant search engine displayed low-resolution thumbnail copies of photographs. The Ninth Circuit held the reproduction was fair use: "[t]he court deemed the use transformative because the thumbnails served an entirely different

function than the original images." Moreover, the defendant's use of the artistic works was "unrelated to any aesthetic purpose," and the thumbnails did not harm the market for or value of the photographs. In Associated Press v. Meltwater U.S. Holdings, Inc., however, the U.S. District Court for the Southern District of New York held that the defendant, in indexing Associated Press articles and providing excerpts to users, could not avail itself of the fair use doctrine. The automatic capture and republication of text without anything more (e.g., commentary or insight) was not transformative. Further, the defendant's use of the content to generate analytics did not render the initial indexing and excerpting as transformative.

Separately, in Ticketmaster Corp. v. Tickets.com, Inc., Tickets.com used a web scraper to extract information from internal Ticketmaster web pages. In doing so, the program temporarily loaded the information into random access memory. The defendant extracted the factual information, which was displayed on an internal web page, and discarded the rest. The U.S. District Court for the Central District of California (CD California) held that the momentary copying of copyrightable material constituted fair use because it was done with "the limited purpose of extracting unprotected public facts."

Thus, when copying copyrighted content, managers should seek to do so only momentarily and ensure both that the effect on the market value of the copyrighted material is small and that the amount and substantiality of the material used is limited.

Common Law Trespass

Trespass to chattels involves an intentional act that interferes with the tangible movable personal property of another. Courts have been reluctant to recognize this cause of action in the context of web scraping.

The CD California, in Tickets.com argued that there are "flaws inherent in applying doctrines based in real and tangible property to cyberspace." Nevertheless, the court recognized that the claim may have merit where there is "some tangible interference with the use and operation of the computer." It is conceivable that this could occur in situations where excessive automated data collection crashes a website, causes users to experience delays or limits a site's operational capacity.

Practical and Ethical Concerns

Some of the largest challenges related to the use of alternative data include identifying the correct data sources, recruiting the right talent, setting up an infrastructure and managing issues with budget allocations. The absence of industry-wide standards also means that "[e]xecutives are forced to allocate considerable time and labor to processing and scrubbing purchased data, dealing with incompatible formats and separating useful from non-useful data," states a 2016 report by the Economic Intelligence Unit and sponsored by Northern Trust.

In her book Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, Cathy O'Neil, a mathematician and former quantitative analyst at D.E. Shaw, warned that "the models being used today are opaque, unregulated, and uncontestable, even when they're wrong." Indeed, fund managers must be cautious that the models upon which they rely are substantiated by inputs or signals from the real world.

For example, in deriving scores for different communications, Evan Schnidman, founder and CEO of Prattle, emphasized that Prattle's theoretical framework can be tested by real-world evidence: individual stocks move based on what is said in earnings calls. Without this feedback loop, it becomes unclear whether correlations will hold. "There is a lot of data out there that has been under-analyzed or incorrectly analyzed due to spurious correlations," he explained.

These issues can be exacerbated when too many factors or data streams are used in developing models. For example, there may be times when a trend indicated by several small data sets disappears or is reversed once those data sets are combined into one large data set (a phenomenon known in probability and statistics as "Simpson's Paradox"). This can be problematic because, in such cases, managers must decide whether to use the aggregate or partitioned data.

It is important, therefore, that managers recruit teams that are heavily trained in statistical analysis. Although managers may have difficulty retaining or recruiting data scientists given a potential war for talent, this risk can be mitigated, in part, by relying on external talent pools.

For raw or processed data sets, hedge funds are insistent on backtesting, or data providers have zero chance of getting sales, noted Kilduff. Quantitative funds in particular are concerned about acquiring data sets that have several years of history and that are matched with hundreds of tickers.

In addition to testing models from the outset, managers may be able to reduce risk by regularly validating the models as the underlying regulatory and market environments change; consistently normalizing data from each source; and integrating internal teams and systems. "There has to be a partnership internally between traders, technology staff and compliance personnel. Everyone needs to understand the source of data and the risks associated with it," said Steingarten, adding, "There's no cookie-cutter solution for a firm." Creating policies and procedures is "a very fund-specific task based on the particular risk appetite of a manager," continued Neuburger.

The Federal Trade Commission also recommends that companies consider:

- the representativeness of their data sets (i.e., whether certain populations are underrepresented or overrepresented);
- whether data models account for biases and prejudices;
- the accuracy of predictions (i.e., which correlations are meaningful); and
- whether reliance on big data raises ethical or fairness concerns (i.e., even if managers are technically compliant with the law, whether they should nevertheless refrain from using certain data).

The HEDGE FUND LAW REPORT
The definitive source of actionable intelligence on hedge fund law and regulation
Lowenstein Sandler
www.hflawreport.com Volume 11, Number 4 January 25, 2018

TECHNOLOGY

A Fund Manager's Roadmap to Big Data: Privacy Concerns, Third Parties and Drones (Part Three of Three)

By Shaw Horton, The Hedge Fund Law Report

A fund manager's use of new technologies and processes to streamline its business and generate improved performance comes with significant risk, which is pronounced when using big data, as few best practices currently exist within the industry. One of the most significant concerns about big data involves the acquisition or use of personally identifiable information (PII). Although PII enjoys broad protection under U.S. law, the U.S. has adopted a sector-by-sector approach when dealing with data privacy. Many state laws impose even more stringent restrictions on the use of personal data, and the E.U. General Data Protection Regulation (GDPR), which will go into effect in 2018, provides a comprehensive and onerous framework for data tied to E.U. citizens. Managers must also understand how to deal with third-party data vendors, including how to conduct due diligence on and negotiate contractual provisions with those service providers. Finally, as growing numbers of drones are used to capture images, managers must recognize and comply with a web of federal regulations, as well as state laws, surrounding this use.

This third article in our three-part series discusses the risks associated with data privacy, the acquisition of data from third parties and the use of drones, as well as recommended methods for mitigating those risks. The first article explored the big-data landscape, along with how fund managers can acquire and use big data in their businesses. The second article analyzed issues and best practices surrounding the acquisition of material nonpublic information (MNPI); web scraping; and the quality and testability of data.

For more on the adoption by fund managers of new technology, see our three-part series on blockchain: "Basics of the Technology and How the Financial Sector Is Currently Employing It" (Jun. 1, 2017); "Potential Uses by Private Funds and Service Providers" (Jun. 8, 2017); and "Potential Impediments to Its Eventual Adoption" (Jun. 15, 2017).

Data Privacy

Personally Identifiable Information

PII enjoys wide protection under U.S. privacy and data security laws, while public personal information (i.e., information of a non-confidential, non-intimate nature) can readily be used. "PII is the second most important issue we think about," commented Benjamin Kozinn, partner at Lowenstein Sandler.

See "SEC Enforcement Action Illustrates Focus on Investment Adviser Obligation to Secure Client Information" (Jun. 23, 2016); and "Navigating the Intersection of ERISA Fiduciary Duties and Cybersecurity Data Breach Protections" (Jun. 29, 2017).

Definitions of PII vary by law and regulation. Broadly, however, the National Institute of Standards and Technology (NIST), in a guide to federal agencies, defined PII as "any information about an individual . . . that can be used to distinguish or trace an individual's identity . . . and any other information that is linked or linkable to an individual." Examples of PII include:

- names or aliases;
- personal identification numbers (e.g., Social Security numbers or credit card numbers);
- address information;
- asset information (e.g., internet protocol addresses);
- telephone numbers;
- personal characteristics (e.g., photographic images or fingerprints);

- information identifying personally owned property; and information linked to any of the above, including date of birth, race, religion, medical information or education information.

When dealing with PII, managers should ensure that data has been de-identified or anonymized. NIST defines de-identified information as "records that have had enough [PII] removed or obscured . . . such that the remaining information does not identify an individual and there is no reasonable basis to believe that the information can be used to identify an individual."

De-identified information can be re-identified through the use of a code or algorithm. NIST defines anonymized information as "previously identifiable information that has been de-identified and for which a code or other association for re-identification no longer exists."

Although fund managers should seek to obtain anonymized data, NIST noted that:

> [d]e-identified information can be assigned a PII confidentiality impact level of low, as long as . . . [t]he re-identification algorithm, code, or pseudonym is maintained in a separate system, with appropriate controls in place to prevent unauthorized access to the re-identification information [and t]he data elements are not linkable, via . . . reasonably available external records, in order to re-identify the data.

Nevertheless, experts at the Delft University of Technology have argued that "it is almost always possible to reconstruct links with individuals by using sophisticated statistical methods and by combining multiple databases that contain personal information." Cryptographic techniques such as homomorphic encryption, however, which "provides for the ability to compute on data while the data is encrypted," could have revolutionary impacts for preserving privacy.

"Ultimately, when obtaining data from a source, the first conversation needs to revolve around whether the data is personal in any way," noted K&L Gates partner Derek Steingarten. "Any breaches, for example, are not just a business issue but become a regulatory enforcement issue."

"Funds should not collect PII," stated Jeffrey Neuburger, partner at Proskauer, adding that he has not yet seen a situation where a fund needs PII. "Managers should be spot-checking data, and if there is any inkling that PII exists, they should deal with it appropriately."

"If a manager inadvertently receives PII, it can return the data; delete the data and include a deletion log for regulators; or quarantine the data in its information technology infrastructure. Regardless, managers must catalogue whichever solution they adopt in their compliance policies and procedures," advised Kozinn. "The risk of an enforcement action through the inadvertent receipt of PII is low when a manager can show the regulator that it followed its policies and procedures," added Lowenstein Sandler partner Peter Greene.

Sector-by-Sector Approach

The U.S. employs a sector-by-sector approach to protect data that are most sensitive and at risk. Some of the most important federal data privacy laws include:

- the Children's Online Privacy Protection Rule in connection with the online collection of information from children under 13 years of age;
- the Gramm-Leach-Bliley Act (GBLA) with respect to the collection, use and disclosure of nonpublic personal financial information. In a February 2017 comment letter to the Consumer Financial Protection Bureau (CFPB), the American Bankers Association recommended that the CFPB (1) ensure that consumer data be subject to the GBLA regardless of whether it is held by a bank or third party; (2) require third parties to provide "clear, detailed disclosures about how data will be used"; and (3) give consumers control over what information is shared;
- the Controlling the Assault of Non-Solicited Pornography and Marketing Act and the Telephone Consumer Protection Act, which govern the collection and use of e-mail addresses and telephone numbers, respectively;
- the Electronic Communications Privacy Act (ECPA),

which regulates the interception of electronic communications. For example, NebuAd settled a class action lawsuit for allegedly copying, transmitting, collecting, storing, using and altering private data from users via deep packet inspection to provide targeted ads. Notably, liability under the ECPA extends to the intentional use or disclosure of illegally intercepted communications where one knows the information used or disclosed came from an intercepted communication and has sufficient facts that such interception was prohibited;

- the Stored Communications Act, which prohibits certain unauthorized access to stored communications and records;
- the Computer Fraud and Abuse Act (CFAA), which regulates computer tampering and is discussed in further detail below; and
- the Health Insurance Portability and Accountability Act (HIPAA) in connection with the collection and use of protected health information (PHI) and electronic transmissions of medical data. Compliance with HIPAA, however, is only required for "covered entities" (i.e., health care providers, health plans and healthcare clearinghouses), "business associates" (i.e., persons or entities that perform certain functions, such as data aggregation, with respect to PHI and on behalf of a covered entity) and business associate subcontractors.

The Computer Fraud and Abuse Act

The CFAA prohibits intentional unauthorized access to a "protected computer" to, among other things, obtain information, cause damage or commit fraud. Unauthorized access is established when a user (1) obtains access without authorization; or (2) obtains access with authorization but uses that access improperly. The CFAA is routinely asserted by site owners as the basis for relief against data collectors.
In Facebook v. Power Ventures, the U.S. Court of Appeals for the Ninth Circuit (Ninth Circuit) held that a violation of a website's TOU, without more, is insufficient for liability under the CFAA. Instead, a defendant is liable only when she engages in "technological gamesmanship" (e.g., by concealing her identity through the manipulation of the User-Agent string in an HTTP request or by ignoring a website's robots exclusion standard (robots.txt)) or enlists a third party to access the site after permission has been explicitly revoked, such as through the issuance of a cease-and-desist letter or imposition of IP blocks.

The U.S. District Court for the Northern District of California, in hiQ Labs, Inc. v. LinkedIn Corp, analyzed the issue in connection with public data. HiQ collected information from public LinkedIn profiles via web scraping. LinkedIn sent hiQ a cease-and-desist letter, alleging violations of LinkedIn's user agreement, which prohibited scraping. LinkedIn also implemented technical measures to prevent hiQ from accessing the site.

The court distinguished the situation from Facebook, arguing that, in that instance, the defendant gained access to a portion of a website that was protected by a password authentication system. Consequently, the court held that "hiQ's circumvention of LinkedIn's measures to prevent use of bots and implementation of IP address blocks does not violate the CFAA because hiQ accessed only publicly viewable data not protected by any authentication gateway." The case has been appealed to the Ninth Circuit, and it remains to be seen whether hiQ will succeed.

These cases illustrate that in the context of web scraping, the CFAA may have little weight where the information obtained is public. Nevertheless, managers should continue to take precautions when scraping public information, such as complying with robots.txt standards. While compliance is voluntary, it is customary in the industry.

See "Protecting Hedge Funds' Trade Secrets: What a Difference a Year Makes" (Apr. 19, 2012).

The Federal Trade Commission Act

When a particular category of data is not covered by a specific law, the Federal Trade Commission Act (FTCA) governs. Section 5 of the FTCA is generally applicable to most companies acting in commerce and prohibits unfair or deceptive acts or practices.

An act or practice is unfair if it is likely to "cause

substantial consumer injury, the injury is not reasonably avoidable by consumers, and the injury is not outweighed by benefits to consumers or competition," the FTC stated in its January 2016 report entitled "Big Data: A Tool for Inclusion of Exclusion?" For example, a company is engaged in an unfair act or practice when it fails to reasonably secure consumer data or sells the data to someone that it knows or has reason to know will use the data for fraudulent purposes.

An act or practice is deceptive under Section 5 "if it involves a material misstatement or omission that is likely to mislead a consumer acting reasonably under the circumstances." Thus, if a company violates a material promise, such as to refrain from sharing data with third parties, it will likely be engaged in a deceptive act or practice.

Section 5 also prohibits companies from using previously collected personal data without consent in a manner that is materially different and less protective than what was initially disclosed.

State Laws and the Common Law

Managers must also be conscious of state privacy laws. Given that federal privacy laws do not preempt all state laws, managers may need to comply with both for specific kinds of data. For example, the California Online Privacy Protection Act applies to any person or company whose website, online service or mobile application collects PII from California consumers. A privacy law in Massachusetts (201 CMR 17.00) requires businesses to "develop, implement, and maintain a comprehensive information security program that is written in one or more readily accessible parts and contains administrative, technical, and physical safeguards."

Additionally, managers may be exposed to liability under common law. Individuals may sue to protect against intrusions of solitude; public disclosure of private facts; publication of facts that place the individual in a false light; and appropriation of name or likeness.

E.U. General Data Protection Regulation

The E.U. GDPR, which will go into effect on May 25, 2018, imposes on owners of data what some have interpreted as particularly onerous requirements. "The GDPR will have a profound effect on any data held that is tied to an E.U. citizen," noted Evan Schnidman, founder and CEO of Prattle. The GDPR applies even where a business is not based in the E.U.

Personal data is broadly defined under the GDPR as "any information relating to an identified or identifiable natural person." An identifiable natural person is "one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."

Under the GDPR, data must be processed fairly, lawfully and transparently; consent to the use of data must be "freely given, specific, informed and unambiguous," and in the case of sensitive personal data, "explicit"; data may only be processed for the purpose for which it was obtained; individuals have a right to be forgotten and a right to know if their data has been hacked; and organizations must appoint a data protection officer to oversee compliance. The GDPR also imposes enormous penalties on non-compliance.

Although the GDPR imposes many new requirements that do not exist in the U.S., Neuburger argued that for fund managers, "the GDPR is more relevant when talking about safeguarding information related to investors or employees." While he has not seen many managers roll out GDPR compliance systems globally, he stated that "in many cases it is easier to do so on an enterprise-wide level." Nevertheless, Greene cautioned that managers should "absolutely have E.U. counsel advise them when purchasing data where the original user is an E.U. resident."

See "Key Considerations for Fund Managers When Selecting and Negotiating With a Cloud Service Provider"

(Sep. 21, 2017); and "How Recent Data Breaches Have Affected the Cyber Insurance Market for Fund Managers" (Aug. 3, 2017).

APEC and OECD Privacy Principles

The Asia Pacific Economic Cooperation Privacy Framework and the Organisation for Economic Co-operation and Development Privacy Principles serve as guides for how fund managers can structure their data privacy policies. Specifically, where applicable, fund managers should:

- limit the collection of data. For example, data collected should be relevant to a specific purpose for which it is to be used;
- ensure data is accurate, complete and up-to-date to the extent necessary for the purposes for which it is being used;
- refrain from sharing or repurposing the data without consent or unless required by law;
- allow individuals to confirm whether the manager has any data relating to them; challenge data relating to them; and have the right to have that data amended or erased;
- clearly communicate developments, practices and policies with respect to personal data; and
- implement reasonably security safeguards to protect the data.

Acquiring Data From Third Parties

Negotiating Agreements; Due Diligence

Managers should make clear that they do not want certain kinds of information from vendors. They should "negotiate contracts with vendors to include representations that data sets, for example, do not include PII," said Steingarten. "There's certainly more flexibility in those negotiations, in contrast to other contracts, as many vendors are sensitive to the legal issues."

"A lot of these vendors are willing to take on higher levels of risk than the more established funds," cautioned Neuburger. "Some vendors will make representations readily, notwithstanding the fact that they may not even be true. Therefore, it is important that funds engage in appropriate levels of due diligence, especially in the case of new vendors without established levels of credibility."

For instance, managers should ensure that vendors are recertified at least annually, as well as conduct initial and ongoing diligence in examining whether data sets have been inadvertently transferred with PII. If there is any concern that vendors are insufficiently scrubbing data of PII, managers should create dedicated teams that clean the data before it is involved in the investment process.

See "Best Practices for Due Diligence by Hedge Fund Managers on Research Providers" (Mar. 14, 2013); and "Hedge Fund-Specific Issues in Portfolio Management Software Agreements and Other Vendor Agreements" (Aug. 4, 2011).

In addition, managers must take great care when negotiating contracts with third-party vendors given the importance of MNPI concerns, said Greene. "Managers must ensure that the data provenance is pure along the entire chain – i.e., from the original user to the purchaser of the data." To do this, "managers must negotiate a very robust set of representations that the seller has the right to sell the data for use in the financial services industry and for compensation and that, by doing so, no duty or obligation in the chain has been breached."

Beyond that, Greene continued, managers should ask to see underlying agreements where the vendor came into possession of the data. "What representations did the seller get from its source of data? How does the seller know that the source had the ability to transmit the data to the seller?" By extension, managers must only use the data as provided for in the contract with the data vendor.

Managers should also ask vendors about the sources of data or the technology used in collecting it. "Managers should tailor the representations for the particular type of data they are receiving," argued Neuburger. Thus, in the case of web scraping, a manager should ask whether the vendor considered a website's TOU. Further, in the case of drones (discussed in greater detail below), the manager should ask the vendor whether it is acting in

compliance with state and federal laws and whether it is abiding by the National Telecommunications and Information Administration's (NTIA) best practices.

Managers must also consider the backflow of information when negotiating contracts with vendors. If a manager acquires raw data from a vendor and manipulates it into a proprietary data set, the vendor may still have the right to use the enhancements and sell that enhanced data to other clients.

Finally, managers should negotiate representations that the vendor has not been sued and that the data are accurate. Managers may also seek to include indemnification or liability allocation provisions, and they should never rely on a vendor's legal analysis.

Exclusivity

Third-party data providers may be viewed as agents of a manager, which may give rise to vicarious liability. This risk is heightened in instances where managers use exclusive data sets.

Many data providers are hesitant to engage in exclusivity arrangements with funds, however. "Because this is such a grey area from a legal and compliance perspective, we don't want to be involved in these arrangements," said Emmett Kilduff, founder and CEO of Eagle Alpha. "Some funds are also unwilling to purchase exclusive data sets not only for legal reasons, but for potentially negative press and public relations reasons." Neuburger agreed, stating that while exclusive data sets are rare, custom data sets are not unusual.

Issues Related to Drones

Part 107 of the Federal Aviation Regulations requires, among other things, an operator to maintain visual contact with his or her drone or unmanned aircraft (UAS); restricts the time, altitude, speed and airspace of flight; requires an operator to obtain a drone license; and requires an operator to perform a preflight visual and operational check to ensure safety.

In addition, most states have enacted laws or adopted resolutions addressing UAS. For example, Oregon HB 3047 prohibits the operation of UAS over the "boundaries of privately owned premises in a manner so as to intentionally, knowingly or recklessly harass or annoy the owner or occupant" thereof. Likewise, Delaware HB 195 prohibits the use of UAS over events with more than 1,500 people in attendance, over critical infrastructure (including petroleum refineries; power plants; commercial port and harbor facilities; and government buildings) or over any incident where first responders are actively engaged.

Although the Federal Aviation Administration does not regulate how UAS gather data on people or property, it encourages operators to comply with the NTIA voluntary best practices. These include informing other of the use of a UAS; showing care when operating a UAS or collecting and storing covered data (i.e., "information collected by a UAS that identifies a particular person"); limiting the use and sharing of covered data; securing covered data; and monitoring and complying with evolving federal, state and local UAS laws.

The HEDGE FUND LAW REPORT
The definitive source of actionable intelligence on hedge fund law and regulation
www.hflawreport.com Volume 11, Number 4 January 25, 2018

Peter D. Greene
Partner
T: +1 646.414.6908
pgreene@lowenstein.com

Benjamin Kozinn
Partner
T: +1 212.419.5870
bkozinn@lowenstein.com