The Investment Lawyer
Covering Legal and Regulatory Issues of Asset Management
VOL. 26, NO. 6 • JUNE 2019

Big Data: A Legal and Compliance Guide for Investment Managers

By Nathan J. Greene

The volume and variety of data available to companies today is growing at a staggering pace. Costs of data storage and processing continue to fall. These trends are at the root of the data revolution, and their impact can be seen in many investment management domains. Data powers trading strategies. Data powers customer-facing "robo-advice" and chatbots. Data powers compliance and risk software.

Data as it is understood today exists largely as an intangible, literally a stream of digitized information. Yet it is also a high-value asset that can be used, bought and sold.

Data can be gathered from myriad sources and exists in both its "raw state" and in various states of organization or disorganization (or as data professionals prefer, various states of categorization or manipulation). Data can be presented with all links to the source intact or with various levels of de-identification, aggregation or anonymization. New data can be created through the manipulation of data or as metadata or data markers (for example, when and how data were created). This "derived" or "resultant" data can be as important as the original source data.

Types of data being consumed by investment professionals today are so broad—with investment strategies built on data tied to such diverse subjects as credit card spending; money transfer patterns; weather; traffic, port or other infrastructure activity; utility and cell phone usage; geolocation; online search statistics; news or social media "sentiment" analysis; and more—that the industry has coined the umbrella term "alternative data" to capture the landscape. Alternative data exists in such volume, and in such varied forms, that it is often accessible only by the application of sophisticated analytical techniques.[1]

With that background, this article is intended to equip legal and compliance professionals at investment management firms to think broadly about the implications of data to their businesses. Thinking through what data is, where it comes from, how it is used, how it is owned and controlled (or not), how it is protected and how it might be regulated (or not) should be part of today's basic legal and compliance mindset. Managing data will figure in a firm's contracts, its internal organizational decisions, and its dealings with counterparties, investors, service providers and regulators.

Data Protection

One of the first questions presented by data—whether self-generated or acquired from someone else—is how to protect it. A data protection strategy will be driven by a variety of related goals, namely (1) avoiding literal loss, theft or corruption; (2) establishing protections against third-party infringement;

and (3) compliance with specific legal requirements attaching to the data.

Avoiding Loss

Safeguarding sensitive digital data is central to every firm's cybersecurity efforts, and a raft of guidance is available discussing expectations of the US Securities and Exchange Commission (SEC) for investment manager cybersecurity policies. These include both survey findings from the SEC office that inspects investment managers and guidance that can be intuited from the allegations in the multiple instances when the SEC has sued firms for alleged cybersecurity deficiencies.[2] To date, SEC enforcement has focused on breaches involving exposure of client or customer personal information; there has been less interest in other, more general cybersecurity concerns for the investment management industry, such as protection of core operating systems or "crown jewels"-type information, ransomware deterrence and the like. That said, a worry for any organization that expands its data profile—even when the data involved may not be personal data—is the possibility that the organization may be a more attractive target for a hack, intentional sabotage of its data or data sources, or other attack.

Protection Against Infringement

Infringement here refers not to a specific term of art, but to a bundle of concepts akin to virtual trespassing or virtual appropriation. This builds on the reality that who generates, owns and controls data is not always clear. Thus, a thoughtful data strategy will focus in part on better delineating ownership and control.

In the first instance, this typically will be a question of contract. For a data-aware firm, detailed and thoughtful data protection and ownership terms need to be considered in connection with potentially every vendor contract, every customer contract, a firm's terms of employment and employee manuals, its website terms of use pages, and so on. And of course this is a two-way street. At the same time as the data-aware firm is ring-fencing its data contractually, its service providers and counterparties (likewise data-aware) are nibbling at the perimeter, laying claim to data generated as part of their relationships with the firm. Protecting a firm's data thus means being sure that any third-party claim is consistent with the firm's view of the relationship.

There also may be intellectual property bases to protect data, notably as a trade secret. While data generally cannot be patented or copyrighted, systems for analyzing data, especially if grounded in technology, might be patentable, and how a database is arranged, organized and presented might be protected by copyright.

Legal Requirements

As already suggested, the type of data most likely to carry a legal requirement to protect it is personal data associated with individuals, especially names, addresses, government identification numbers, and the like. At the leading edge of a comprehensive data protection regime is the European Union's General Data Protection Regulation (GDPR). An investment manager can find itself subject to the GDPR if it (1) is established, uses IT equipment or targets data subjects in the EU and (2) collects, organizes or holds information relating to identifiable EU residents.[3] There are specialized requirements that apply depending on the type of processing an investment manager undertakes, but this level of detail is beyond the scope of this article. Broadly speaking, investment managers would be required, under the GDPR, to use accurate personal information exclusively for a specified and legitimate purpose. The information must be protected and not stored for any longer than required for its stated purpose.[4] The investment manager also must have a legal basis for processing the data, for example, where they are required by law to do so or pursuant to consent by the data subjects.[5] Stricter rules attach to certain categories of data which are deemed by their nature sensitive. Such data include information relating to an individual's

criminal records, their race, political opinions, religion, etc.[6]

Only data relating to "identifiable natural persons in the EU" is in scope to the GDPR. This means, for example, that the regulation does not apply to personal data that have been effectively anonymized or pseudonomized to a level where the data no longer reveal individual EU residents. But whether anonymization or pseudomization has been effective depends on the facts and can be questioned.[7]

Non-EU investment managers also may find themselves indirectly affected by the operation of the GDPR, as the regulation prohibits the transfer of personal data to third parties (including members of the same group of companies) in cases where the data would lose the protection afforded to it by the GDPR. US investment managers (alongside investment managers from twelve other jurisdictions[8]) have more leeway as the European Commission has exempted them from the transfer prohibition by virtue of their adequate national safeguards.[9]

US data protection laws include state privacy laws, which, like the GDPR, are broad-brush and seek to protect a state's residents in any sphere of activity, and federal privacy regulations specific to financial services such as Regulation S-P and Regulation S-ID, both adopted by the SEC. The US state law drawing the most attention at present is the California Consumer Privacy Act (CCPA), which goes into effect in 2020 and will require new and heightened disclosures when a California resident's personal information is sold or resold. Because nearly every US business has California touch points, the law has the potential to *de facto* set new national standards. As in Europe, various US state and federal laws separately address special protections for populations deemed especially vulnerable, such as children or the elderly, or sensitive personal information tied to health and financial records, gender orientation information, and political and religious affiliations.

The risk associated with a breach of the data protection requirements under the GDPR and other data protection laws is not insignificant. From a purely financial perspective, the breach can lead to the relevant authorities imposing substantial fines—notably under the GDPR the ceiling for a fine is the higher of four percent of the investment manager's global revenue and €20 million. This fine would be in addition to any claim for damages by the affected data parties, potentially in the form of a class action. Reputational harm can be even more damaging, especially with the viral publicity these cases can receive.

Data Sourcing

While some large datasets are sourced by investment managers directly (see, for example, the web scraping discussion below), the more common practice is to buy data from third parties. As you would expect, a thriving marketplace exists in which various kinds of organizations collect and market data.

The core functions of a data marketplace are sometimes described as gathering, cataloging and—importantly—"curating" data to streamline accessibility for buyers. Some data vendors actively tout the fact that data need not leave their domains—meaning, for example, that an investment manager can access and manipulate data at a vendor's site without having to tackle the risk and expense of actual receipt and possession of data. Data vendors can range from the largest and most sophisticated global companies seeking to monetize data generated in their businesses to small, niche organizations to a host of intermediaries and middlemen.

Regulation

There have been a variety of proposals by the US Federal Trade Commission (FTC) to regulate data "resellers" (which would cover many data vendors), but at present there is only very limited US law specifically subjecting the sale of data to comprehensive conduct or registration requirements.[10] Two current initiatives should be noted. First, the CCPA (again, effective in 2020) will bring new disclosure requirements when a California resident's personal

information is sold or resold, which will inevitably affect data markets that include such data. Second, a variety of US Senate bills (proposed by both Democrats and Republicans) address data protection and could direct the FTC to perform a study of the data reseller industry and/or propose regulations.[11]

At least some data vendors have asked themselves whether they might be regulated as investment advisers. This is because investment firms tend to be among the most significant purchasers of data, and the data they buy often informs a firm's investment program—prompting at least the possibility that the data might be akin to a securities research report or similarly regulated content. In fact, there is a long history of data vendors approaching the SEC to ask exactly that question (am I an investment adviser?). That back-and-forth generated a series of SEC Staff interpretive letters over 30 years, which collectively stand for the principle that a data vendor is not an investment adviser so long as (1) the information provided is readily available in its raw state; (2) the categories of information presented are not highly selective; and (2) the information is not organized or presented in a manner that suggests the purchase, holding or sale of any security or securities. Given the profusion of data-based businesses today it is somewhat surprising that the last of these letters was issued in the 1990s.[12]

Vendor Diligence

Investment managers routinely assess their data sellers from a compliance and risk management perspective. An investment manager purchasing data from a vendor wants to be sure the vendor is attuned to the same types of concerns that the manager has, that data lineage (discussed below) can be properly confirmed, and that the vendor has some level of compliance infrastructure. Taken as a whole, the investment manager would like to be sure the vendor has an appropriate understanding and respect for both the various rules and contracts that might govern the vendor's rights in the data and the regulated context in which the investment manager is operating.[13]

An investment manager also may wish to understand how the vendor's business practices more generally might present business and reputational risks. For example, some (but not all) firms prefer to interact only with vendors who offer data on a non-exclusive basis—meaning the same data purchased by one firm can be purchased by others on more or less the same terms. This interest in non-exclusivity tends to be driven by both fairness considerations like those outlined at the end of this article and the risk of receiving material non-public information in an exclusive relationship.

Data Lineage

Understanding "data lineage" (or "data provenance") is critical to diligencing a dataset. The term refers to the concept that the purchaser or user of data should know enough about the chain of ownership to confirm the data was legitimately collected and appropriately managed and protected through the course of its existence. Understanding data lineage can be an important protection in mitigating the insider trading risks covered later in this article (because data that is properly obtained and transferred over its lifecycle generally cannot be said to have been, using the rubric of US insider trading law, "misappropriated"). In the ideal case, the investment manager trading on information derived from data will be able to confirm that the data was obtained legally and with third-party consents where applicable, that the further transfer of the data was likewise legal and consented to, and that disclosures associated with these consents were appropriate and at least contemplate use of the data for commercial or business purposes, including sale.

Web Scraping

"Web scraping," also called crawling or spidering, is the automated gathering of data from a third-party website. Scraped data is an increasingly important component of the investment research programs at many asset managers. Its applicability

to funds aside, scraping is critical to many business processes and is therefore in wide use. But permissibility of the practice—and associated legal risk—remains unclear. A variety of legal claims may apply under US law to unauthorized scraping, including breach of contract, copyright infringement, trespass and other torts, and state and federal laws specific to website access. Federal law, enforceable both criminally and civilly, specifically protects websites from unauthorized access, with that phrase potentially extending the law's protections to any website whose terms of use forbid or limit automated scraping of data from the site.[14] Given the legal overlay, investment managers that use scraped data often have compliance policies and procedures associated with web scraping.

Government Data

There is a natural presumption that governmental data, especially in democratic societies, is intended to be "open" and freely accessible to the public. In fact, this is not guaranteed and permitted uses of governmental data, even when it can be readily accessed, can be context specific. For example, some public data sources may be presented with the disclaimer that they are intended for research and other non-commercial purposes. There are also a variety of instances when governmental data might be explicitly non-public and restricted, for example, in connection with governmental contracts, studies and approvals that have not yet been announced.

The tension between those principles—"open government" versus access and use restrictions—is illustrated in various ways. As a modest example, there have been claims brought against the US federal courts for charging fees for court records; claimants argue the fees infringe on their right to access public information.[15] Still more fraught, the SEC and US Department of Justice have brought cases involving so-called "political intelligence" operations, generally referring to the collection of government information before it is widely disseminated. The first high-profile case resulted in a settlement with the SEC in which a political intelligence firm agreed to enhance its policies and procedures for handling sensitive government information.[16] In a later case, the Department of Justice prosecuted and convicted four individuals—a government insider, a political intelligence consultant, and two portfolio managers at a significant investment manager—for alleged insider trading involving information from a government health insurance rate-setting office regarding upcoming reimbursement rule changes.[17]

National Security

National security-specific data are also higher risk, and the dividing line between national security and commercial considerations is increasingly blurred. Outside the United States, national security data can be especially difficult to divide from commercial data. As an example that could give rise to concern in any country, imagine a data collection program gathering public information on critical infrastructure such as dams, power plants and the like; imagine further that the data are then transferred outside the host country. Such a program may be entirely innocent but still could be misconstrued and generate national security concerns and governmental investigation. National security issues are magnified in jurisdictions with significant state ownership of what otherwise would appear to be traditional commercial enterprises.

Insider Trading

A threshold concern for an investment manager purchasing data is that the data not carry the risk of tainting the manager with possession of material non-public information (MNPI) and thus the possibility of being in breach of insider trading laws. This concern is not unique to the United States, as, across the pond, investment managers operating within the UK or on a UK-regulated market[18] also run the risk of acquiring data constituting inside information (the UK equivalent of MNPI). A particularly user-friendly example of inside information

is a prospective large-scale transaction of XYZ Corp. kept under wraps: the information relates to XYZ Corp., it is specific, it has not been made public and, upon becoming public, it is likely to affect the price of the XYZ Corp's shares.[19] Unfortunately, big data does not slot itself as neatly within the definition of inside information or MNPI. Consider, for example, a bank's credit card transaction data. Might that data represent inside information or MNPI *as to the bank* by revealing the volume of credit card transactions the bank is handling? Or might that data represent inside information or MNPI *as to a particular retailer* by revealing sales information before it can be aggregated and publicly disseminated by the retailer? This uncertainty, while an interesting intellectual exercise, does in fact entail significant risk for investment managers, who may face both civil and criminal consequences for being in breach of insider trading laws.[20]

This latter risk is highlighted by an insider trading case brought by the SEC against a bank employee (a fraud detection analyst) who, in the ordinary course of business, had access to real-time information on credit card transactions processed by the bank. The employee allegedly developed a software program based on that data that permitted him to extrapolate a retailer's overall sales figures and then trade in the securities of that retailer when his program predicted the retailer's sales would vary from its publicly reported forecasts (for example, disappoint or positively surprise the market). Among other things, the employee argued that the bank saw only a very small percentage of a given retailer's credit card transactions, a basis to claim that the data he had was non-material. But the court rejected that defense and accepted the premise that the credit card transaction data, on those facts and when used in that manner, constituted MNPI.[21]

The typical issue faced by an investment manager is, of course, much more nuanced. For example, assume that fluctuations in a company's hiring activity might influence a trader's decision whether to buy or sell the company's securities. On that basis, a traditional word of mouth insider "tip" ("just heard that XYZ Corp. has pulled all of its recruiting searches, could mean they're not growing anymore") might be readily understood as carrying potential risk. But what about when that same fact—that XYZ Corp. has pulled all of its recruiting searches—instead can be divined from a mountain of job and recruiting search data housed by or visible on online jobs websites? Because there is no "tip," and the corresponding information might be obscured or aggregated within a larger data set covering many companies, it is understandably less likely to set the same alarm bells ringing. But sophisticated investment management consumers of data will ask questions intended to confirm the data was legitimately obtained without any violation of a duty of confidentiality or loyalty along the way.

Artificial Intelligence (AI)

Automated and AI-based applications are used throughout the industry. Marketing applications ingest social media and other source data to identify and profile customers. Chatbots interact with customers in service and marketing capacities. Quantitative programs trade in securities and derivatives markets, often at speeds and volumes far in excess of human trading. Other automated programs identify and research anomalies to support risk management, fraud detection, anti-money laundering profiling, and other control processes.

There are many different understandings of what constitutes artificial intelligence, all of which are beyond our scope here. That said, AI techniques are regularly deployed in analyzing large data sets and connecting data with trading software and the other commercial applications just described. The most prolific users of data in support of investment programs likely have an AI component to their activities.

Regulatory views on AI are still early stage and evolving, but the past two years has seen an upswing in pronouncements.

US Treasury Report

One of the broadest and most comprehensive discussions specific to AI in financial services is a 2018 report prepared by the US Treasury Department.[22] The report opens by observing that AI investment by financial services firms is accelerating and that AI innovations are driving efficiencies for firms and improved outcomes and choices for their customers. The report expresses concern, however, that "black box" systems are inconsistent with traditional regulatory expectations of transparency and auditability for industry activities. Treasury also suggests that AI presents a variety of two-edged sword risks: for example, trading will become ever faster and more efficient, but potentially at risk of new bouts of extreme volatility; or new tools might help root out rogue traders, money launderers, cyber criminals and other bad actors, but bad actors surely will challenge systems with their own sophisticated applications as well.

SEC/FCA Guidance

The SEC has not directly spoken as to how investment managers and other SEC-regulated firms should consider their use of AI.[23] But the agency has brought a number of enforcement actions involving failures by firms to properly vet and implement complex investment models (generally also alleging related failures to disclose weaknesses or limitations in the models). The overall impression from the cases is that the SEC expects that a firm (1) should carefully test technology before it is rolled out, (2) should continue to test over time, (3) should understand and be able to explain the technology's core operations and outcomes to the firm's internal and external governance bodies (senior management, compliance and control functions, and regulators), and (4) when relevant to customers or shareholders, should be transparent as to risks that might be presented by reliance on the technology.[24] This package of concepts is sometimes referred to as "model governance," referring to the governance and control frameworks that wrap around development and use of complex quantitative models.[25]

In the UK the Financial Conduct Authority pioneered a regulatory "sandbox" to enable firms to test technologically innovative products in a controlled environment.[26] The program is at the same time aimed at giving the regulator valuable insight on how these new technologies are being applied and should be regulated. In addition, following its survey on Technology and Cyber Resilience, the FCA published its findings identifying key areas for development. Mainly the FCA recommended that the firms should (1) develop effective third-party risk management; (2) endeavor to better appreciate the connection between cyber risks and other conduct issues; (3) aim for increased familiarity of their board members with information technology in order to foster the board's long-term ability to manage cyber risk; and (4) promote the development of in-house knowledge on cyber issues.[27]

US Federal Reserve Guidance

Another widely cited source of regulatory guidance on AI in financial services came in a speech by a member of the board of governors of the US Federal Reserve, who suggested "existing regulatory and supervisory guardrails"—and especially existing guidance on risk management when using complex models—provide a sufficient starting point. In other words, new US banking regulation specific to AI might be required in the future, but not yet.[28]

OECD Guidelines

The Organization for Economic Cooperation and Development, a transnational organization of which the United States is a member, is set to publish AI guidelines shortly. The goal is to establish international norms around such topics as transparency and accountability for AI, auditability and human control of AI, management of bias in AI, privacy and appropriate sourcing of data underlying AI, and more.[29]

Source Code

Investment management regulators have taken different tacks over time with respect to demanding access to sensitive source code when supervising businesses deploying AI or other sophisticated software applications. As an indication of how concerned some parties are that source code will be mishandled by the government (the highest order concern being that a company's intellectual property might be stolen by hackers or even bad actors inside the government), the US Congress debated the Protection of Source Code Act, which would have prohibited the SEC from accessing source code at SEC regulated businesses without obtaining a subpoena. The House passed the bill, but it appears to have died in the Senate at the end of 2018.

The same issues animated the US Commodity Futures Trading Commission, which grappled with the question in the course of developing its Regulation AT (referring to automated trading). The regulation would have given the CFTC access to quantitative trading software source code at CFTC regulated firms, but opponents argued, first, that due process protections require a subpoena before access and, in any event, that the CFTC is ill equipped to protect sensitive intellectual property from loss or theft.[30] The CFTC abandoned the initiative when Republican appointees became the majority on the CFTC following the election of President Trump.[31]

Robo-Advice

The rise of model-based approaches to delivering customized investment advice to a wider audience at lower cost is often termed "robo-advice." The gist of the service is that after the client completes a detailed online questionnaire an algorithm should be able to provide a reasonably tailored investment program to the client without the expense of human judgment and handholding. Various regulatory questions have been posited regarding robo-advice, with an emphasis on being sure (1) the client understands the limitations of the service, (2) the questionnaire used to interact with the client is appropriate, complete and thoughtfully designed to gather the right feedback, and (2) the algorithm is properly tested and maintained.[32]

There will also be next generation robo-advice models that draw on new sources of data and pose new questions. Consider a service that mines social media or online search activity for greater insights into the client's circumstances. In one version of the service, it could cross-check learning from a client's social media accounts against the questionnaire and highlight potential inconsistencies. In another version, the questionnaires themselves might be made "smart" and adapt seamlessly to the client, even asking different types of questions based on that social media learning (in the same way that different users of many online services can see quite different versions of the service tailored to the individual). In another version, the service would pitch additional products based on that learning (for example, suggesting college or health savings accounts, annuities or other offerings based on apparently relevant personal information).

An extension of the robo-advice model like the one just described presumably could deliver an even more efficient and tailored version of the service, but at some cost to the client's privacy expectations and no doubt with room for error.[33] As investment managers marry their services to increasingly diverse pools of personal data, these considerations—and questions of additional disclosures or safeguards (and, indeed, ethics as outlined below)—will come to the fore.

RegTech

How data and AI inform a firm's control functions, especially around regulatory compliance, often is referred to as RegTech. The idea is simply that technology, especially when it can analyze data and surface anomalies and correlations more efficiently than human eyes and intuition, must be part of today's compliance officer toolkit.

A significant driver for investment in RegTech is the perception of an arms race. Regulators trumpet

their success in developing quantitative and risk analytic processes that crunch industry data and guide their regulatory inspections, rulemaking and other initiatives. Meanwhile, compliance officers and industry executives are riveted by the possibility that their regulators might "know their data better than they do."

Accelerating implementation of RegTech then becomes necessary simply to keep up.[34] The reality, though, is that firms that implement RegTech solutions have not always found them to be well suited. Developing a solid checklist to evaluate RegTech tools is critical to success.

For example, a firm might ask: Do the tool's designers truly understand the regulatory issue they are solving for, who else has road-tested the tool, is the intellectual property underlying the tool in order, can the tool interface with legacy systems at the firm (or often just as important, the firm's service provider), what data sources does the tool draw on and can it cleanly ingest the firm's data (or, again, service provider data), does the tool create new data exposure or security risks, and what redundancy and business continuity protections are available? Contracting questions might go to the provider's level of product support and customization, licensing terms, openness to audits, and insurance and indemnification.

Data Ethics … and Fairness More Generally

It has been common over many years for firms that make heavy use of data to speak of their "data ethics." This is sometimes referred to as embodying the principle that the question for a firm is not whether it can (operationally or legally) put data to a particular use, but whether it *should* (whether doing so is "right"). Data ethics policies are intended to ensure that an organization has a governance framework to answer that question and, in doing so, considers a broad range of factors (for example, legal and contractual requirements, technical capacity, social expectations, reputational considerations, and the like).[35]

Illustrative of the can/should dichotomy is a speech by a former SEC official, who held AI out as a potent tool in developing actionable insights for the agency's examination and enforcement programs. But the official pointedly added "… algorithms can't then prepare a referral to enforcement. And algorithms certainly cannot bring an enforcement action. The likelihood of possible fraud or misconduct identified based on a machine learning prediction cannot—*and should not*—be the sole basis of an enforcement action" (emphasis added).[36] In other words, AI insights inform enforcement thinking, but when it comes to whether to invoke government authority in a way that implies or actually alleges wrongdoing (which is what a subpoena or enforcement action does), we simply should not give an algorithm the last word.

Another version of the can/should dichotomy is illustrated by a UK regulatory white paper on privacy, which put the issue this way: "… big data analytics is sometimes characterized as sinister or a threat to privacy or simply 'creepy' … because it involves repurposing data in unexpected ways, using complex algorithms, and drawing conclusions about individuals with unexpected and sometimes unwelcome effects."[37] Said differently, a correlation that is statistically significant and relevant may be sufficiently difficult to explain and justify that, perhaps, one should not act on it—to do so being, either actually or in perception, "creepy" or "wrong." Consider the science fiction movie *Minority Report*, in which police arrest suspects before they commit a crime based on the visions of psychics. Data can produce conclusions that may feel equally problematic to those targeted. For example, the US FTC found evidence that credit scores for certain groups of people were lowered on the basis of repayment histories of other people with similar preferences in retail stores.[38] In the EU, the GDPR does partially address this concern, by giving a data subject the right to object to its personal data being used for

profiling purposes.[39] This opt-out approach, however, does not really address the question of what a firm *should* do with the data it is legally allowed to process.

More broadly, there is a long-running, philosophical debate around how "fair" financial markets should be and what fairness means in this context. In the case of insider trading law, for example, "fairness" is premised on the idea that some types and sources of information must be in the public domain to be fair game for a trader. The cited UK regulatory white paper, on the other hand, links "fairness" to expectations, proposing that data should not be used for purposes outside the reasonable expectations of the data subjects.[40]

As a well-publicized example from 2013, the New York Attorney General investigated a news organization selling advanced access to economic survey data. Under pressure, the organization changed tack and set guidelines establishing more uniform access rules. Commenters at the time recognized the philosophical tension between maintaining "fairness" and lawful information advantage and segmentation, with a *New York Times* article highlighting the latter considerations. Said the paper: "The race to get information first has been a part of financial markets at least as far back as the carrier pigeons that delivered news of the Napoleonic Wars to London." Likewise, "news providers of all sorts give preferential access to articles to their own subscribers."[41] Where then should the "fairness" line be drawn?

That case also deserves special note because of its association with New York's Martin Act, which a succession of New York Attorneys General have used to great effect in bringing securities fraud actions.[42] The Martin Act generally prohibits "fraud" and "fraudulent practices" in connection with the offer, sale or purchase of securities, but it differs from common law fraud. Common law fraud typically is understood as involving misrepresentations or omissions and to require scienter (or intent) to defraud. By contrast, the Martin Act does not require scienter. Courts instead have referred to its scope as prohibiting "deceitful practices contrary to the plain rules of common honesty."[43] Whether principles of "deceit" and "common honesty" should be stretched to address simple unfairness in the markets is, of course, doubtful. But the broad language of the Martin Act and a history of activist Attorneys General give pause.

Data Governance

Related to the core concepts of data protection and data lineage is the broader idea of data governance. At its most basic, data governance is intended to ensure data quality within a firm. The program is dedicated then to the nuts and bolts of maintaining availability of and access to data, data consistency, data mobility, data integrity and data protection. When dealing with regulated data or a regulated organization holding and using data, as in the case of investment managers, the data governance program will connect to and may overlap with the firm's compliance program. Data governance also can address a firm's view on more philosophical questions, like those of ethics and fairness.

Service providers (administrators, transfer agents, custodians) tend to have a central role for investment management businesses. Accordingly, an investment manager's data governance program is likely to contemplate significant service provider connectivity and data transfers.

As with any organizational program, a data governance program requires an "owner," who is ultimately responsible for its implementation. In larger or data-centric organizations, there may be a chief data officer. For investment managers, responsibility most likely will sit with a chief technology officer, chief information security officer, chief operating officer or chief risk officer.

Conclusion

The data revolution is fundamentally reshaping how investment managers deliver their services. As it does, the range of regulatory, disclosure and

contracting considerations to which investment management lawyers and compliance officers must attend will continue to multiply.

Lawyers and compliance officers at data-aware firms will bring a wide-angle lens to their consideration of the legal, contractual and ethical issues associated with data. A good start is to ask a series of broad questions and then relate those answers back to the themes covered in this article. Who at the firm is using data, what kind of data, from where is it obtained, how is it being held and manipulated and for what purposes, how transparent is this, what do the firm's contracts say about data, and how do a firm's data practices connect to its broader control and governance principles.

Financial services regulators meanwhile have themselves become prodigious accumulators and consumers of data. Regulators also are taking an interest in their regulated firms' data practices. To date, however, they have found their regulatory constructs to be sufficiently flexible and principles-based, such that there is little in the way of new rules specific to using and dealing in data by regulated firms. Whether that holds over the longer term is an open question, especially if these data practices ultimately become more constrained by the states, other federal regulators (like the FTC), non-US regulators or others.

Mr. Greene is a Partner in the New York office of Shearman & Sterling LLP. The author thanks his colleagues for their contributions to the article. Oliver Linch, Chrisangelina Lo, Emma Maconick, Wilf Odgers, and Barney Reynolds provided thoughtful input to the discussion of privacy regulation, notably the European Union's GDPR and California's CCPA.

NOTES

1 For a discussion of the variety of data that can be used by investment firms*, see, e.g.*, "Hedge Funds See a Gold Rush in Data Mining," *Financial Times* (Aug. 28, 2017), available at *https://www.ft.com/content/d86ad460-8802-11e7-bf50-e1c239b45787*.

2 The SEC office responsible for inspecting investment managers is the Office of Compliance Inspections and Examinations (OCIE). OCIE published two cybersecurity "risk alerts" after conducting a series of targeted industry examinations on the topic. *See* OCIE's 2015 Cybersecurity Examination Initiative (Sept. 15, 2015), available at *https://www.sec.gov/ocie/announcement/ocie-2015-cybersecurity-examination-initiative.pdf*; and Observations from Cybersecurity Examinations (Aug. 7, 2017), available at *https://www.sec.gov/files/observations-from-cybersecurity-examinations.pdf*.

The most recent case brought by the SEC involved data breaches by a large broker-dealer. *See* "SEC Charges Firm with Deficient Cybersecurity Procedures," SEC Press Release (Sept. 26, 2018), available at *https://www.sec.gov/news/press-release/2018-213*.

3 Articles 3-4, Regulation (EU) 2016/679 (GDPR).

4 Article 5, GDPR.

5 Article 6, GDPR.

6 Articles 9-10, GDPR.

7 Disputes around the effectiveness of anonymization under the GDPR are already cropping up, as described in a report by the advocacy group Privacy International, available at *https://privacyinternational.org/advocacy-briefing/2434/why-weve-filed-complaints-against-companies-most-people-have-never-heard-and*. *See also* "Data Brokers: Regulators Try to Rein in the 'Privacy Deathstars,'" *Financial Times* (Jan. 10, 2019), available at *https://www.ft.com/content/f1590694-fe68-11e8-aebf-99e208d3e521*.

8 The European Commission has so far recognized the following jurisdictions, aside from the United States as providing adequate protection: Andorra, Argentina, Canada (limited to commercial organizations), Faroe Islands, Guernsey, Israel, Isle of Man, Japan, Jersey, New Zealand, Switzerland and Uruguay. Further information about the adequacy decisions of the European Commission is available at *https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/*

adequacy-protection-personal-data-non-eu-countries_en.

9 Articles 44-45, GDPR.

10 Vermont recently implemented a data broker registration law focused on resale of consumer data. Press release available at *https://ago.vermont.gov/blog/2018/12/13/attorney-generals-office-issues-guidance-on-data-broker-regulations/*. Past efforts by the FTC to regulate data sellers stalled after a series of studies and recommendations sponsored by the agency from 2012 and earlier. *See, e.g.*, "Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers" (Mar. 2012), available at *https://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers*.

11 One example is the American Data Dissemination Act, sponsored by Sen. Marco Rubio (R–FL) and introduced Jan. 16, 2019. Press release available at *https://www.rubio.senate.gov/public/index.cfm/2019/1/rubio-introduces-privacy-bill-to-protect-consumers-while-promoting-innovation*. There also remains the possibility of administrative agency or executive department rulemaking. The National Telecommunications and Information Administration within the US Commerce Department has had an open statement of regulatory intent in the Federal Register since September 2018. That notice refers to harmonization and advancement of privacy regulation and is available at *https://www.federalregister.gov/documents/2018/09/26/2018-20941/developing-the-administrations-approach-to-consumer-privacy*.

12 *See, e.g.*, *Missouri Innovation Center, Inc.*, SEC No-Action Letter (Oct. 17, 1995); *Media General Financial Services, Inc.*, SEC No-Action Letter (July 20, 1992); *Charles Street Securities, Inc.*, SEC No-Action Letter (Nov. 28, 1989); *Butcher & Singer*, SEC No-Action Letter (Jan. 2, 1987).

13 A working group at Alternative Investment Management Association has prepared a due diligence questionnaire (DDQ) tailored for use by investment managers with their data vendors. The DDQ is not publicly available but bears some resemblance to DDQs in wide use with so-called expert network providers.

14 The most widely cited federal law in this area is the Computer Fraud and Abuse Act (CFAA), which makes it unlawful to "intentionally access" a computer or website without authorization or in a manner that "exceeds authorized access." There are numerous US state law analogs.

15 For a description of the cases and the legal and policy arguments, *see* Op-Ed: "Public Records Belong to the Public," *N.Y. Times* (Feb. 7, 2019), available at *https://www.nytimes.com/2019/02/07/opinion/pacer-court-records.html*.

16 *In re Marwood Group Research, LLC*, SEC Release No. IA-4279 (Nov. 24, 2015), available at *https://www.sec.gov/litigation/admin/2015/34-76512.pdf*.

17 "'King of Political Intelligence' Sentenced to Prison for Insider Trading," *Bloomberg* (Sept. 13, 2018), available at *https://www.bloomberg.com/news/articles/2018-09-13/-king-of-political-intelligence-gets-one-year-in-insider-case*. (The defendants remain free pending appeals.)

18 § 62(1), Criminal Justice Act 1993.

19 § 56(1) Criminal Justice Act 1993; Article 7, Regulation (EU) 596/2014 (MAR).

20 § 61(1), Criminal Justice Act; §§ 122I, 123, 123A, Financial Services and Markets Act 2000.

21 *SEC v. Huang*, No. 16-2390 (3d Cir. 2017), available at *https://law.justia.com/cases/federal/appellate-courts/ca3/16-2390/16-2390-2017-04-10.html*.

22 "A Financial System that Creates Opportunities: Nonbank Financials, Fintech and Innovation, U.S. Department of the Treasury," Report to President Donald J. Trump (July 2018), available at *https://home.treasury.gov/sites/default/files/2018-08/A-Financial-System-that-Creates-Economic-Opportunities---Nonbank-Financials-Fintech-and-Innovation_0.pdf*. An earlier and also very detailed report is by the Financial Stability Board, a global regulatory group. *See* Press Release, "FSB considers financial stability implications of artificial intelligence and machine learning" (Nov. 2017), available at *http://www.fsb.org/2017/11/*

artificial-intelligence-and-machine-learning-in-financial-service/.

[23] Various SEC personnel have spoken regularly about the SEC's own use of data and AI. *See, e.g.*, "From the Data Rush to the Data Wars: A Data Revolution in Financial Markets," Speech by SEC Commissioner Kara M. Stein (Sept. 27, 2018), available at *https://www.sec.gov/news/speech/speech-stein-092718*; and "The Role of Big Data, Machine Learning, and AI in Assessing Risks: a Regulatory Perspective," Speech by Scott W. Bauguess, Acting Director and Acting Chief Economist, SEC Division of Economic and Risk Analysis (June 21, 2017), available at *https://www.sec.gov/news/speech/bauguess-big-data-ai*.

[24] For a discussion specific to SEC expectations of quantitative trading techniques, *see* "SEC Enforcements Against Quant Managers Show a Pattern," *Shearman & Sterling FinTech Blog* (Jan. 15, 2019), available at *https://fintech.shearman.com/sec-enforcements-against-quant-managers-show-a-pa*.

[25] The FINRA report on digital advice, *infra* n.32, includes a lengthy discussion of FINRA's expectations for model governance.

[26] Further details on the FCA's use of the regulatory sandbox is available at *https://www.fca.org.uk/firms/regulatory-sandbox*.

[27] "Wholesale Banks and Asset Management Cyber Multi-Firm Review Findings," FCA (Feb. 18, 2019), available at: *https://cdn.wide-area.com/acuris/files/hedge-fund-law-report/industrymaterialsii/Wholesale%20banks%20and%20asset%20management%20cyber%20multi-firm%20review%20findings.pdf*.

[28] "What Are We Learning about Artificial Intelligence in Financial Services?," Speech by Lael Brainard, US Federal Reserve Board of Governors (Nov. 13, 2018), citing SR Letter 11-7 and SR 13-19/CA 13-21, available at *https://www.federalreserve.gov/newsevents/speech/brainard20181113a.htm*.

[29] "OECD Moves Forward on Developing Guidelines for AI," Press Release, available at *http://www.oecd.org/going-digital/ai/oecd-moves-forward-on-developing-guidelines-for-artificial-intelligence.htm*.

[30] Statement of Dissent by Commissioner J. Christopher Giancarlo Regarding Supplemental Notice of Proposed Rulemaking on Regulation Automated Trading (Nov. 4, 2016), available at *https://www.cftc.gov/PressRoom/SpeechesTestimony/giancarlostatement110416*.

[31] "US Regulator Declares 'Dead' Move to Seize HFT Code," *Financial Times* (Oct. 4, 2017), available at *https://www.ft.com/content/068ce050-a922-11e7-93c5-648314d2c72c*.

[32] SEC Staff Issues Guidance Update and Investor Bulletin on Robo-Advisers (Feb. 23, 2017), available at *https://www.sec.gov/news/pressrelease/2017-52.html*; FINRA Report on Digital Investment Advice (Mar. 2016), available at *https://www.finra.org/sites/default/files/digital-investment-advice-report.pdf*.

[33] To illustrate the room for error, not every purchaser of diapers has a child, not everyone searching for wedding venues is getting married, and not everyone talking about retirement on social media is in fact ready for a change in their investment profile.

[34] The perceived data analytics arms race does not just involve the risk that regulators will know a firm's data better than the firm does. As increasing volumes of information about investment managers are required to be made public, and at increasing frequency, there is also risk that competitors, journalists, academic researchers and others are drawing conclusions about a firm based on its public data. Forward thinking about those constituencies is part of the race.

[35] As one prominent example in a related field, a condition of the sale of the AI firm DeepMind to Google reportedly involved the establishment of a specialized board to oversee the ethics implications of how DeepMind's AI would be used by Google. *See* "Whatever Happened to the DeepMind AI Ethics Board Google Promised?," *The Guardian* (Jan. 26, 2017), available at *https://www.theguardian.com/technology/2017/jan/26/google-deepmind-ai-ethics-board*.

[36] Bauguess speech, *supra* n.23.

[37] "Big Data, Machine Learning and Data Protection," UK Information Security Office (Sept. 4, 2017),

available at *https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf*.

38 "Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues," FTC Report (Jan. 2016), available at *https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report*.

39 Articles 4, 21(1), GDPR.

40 UK Information Security Office report, *supra*, n.37.

41 "Regulators Examining Sales of Early Financial Data," *N.Y. Times* (July 8, 2013), available at *https://dealbook.nytimes.com/2013/07/08/regulators-examining-sales-of-early-financial-data/*.

42 N.Y. Gen. Bus. Law §§ 352-c & 353.

43 *People v. Federal Radio Corp.*, 244 N.Y. 33, 38-39 (1926).