Law360 - Expert Analysis

9 Data Considerations for Investment Management Attorneys

By **Nathan Greene** March 11, 2019, 3:00 PM EDT

Law360 (March 11, 2019, 3:00 PM EDT) --
The volume and variety of data available today is growing at a staggering pace, and costs of data storage and processing continue to fall. These trends are at the root of the data revolution, and their impacts can be seen in many investment management domains.

Data powers algorithmic-based trading strategies. Data powers customer facing "robo-advice" and chatbots. Data powers compliance and risk software. For industry lawyers and compliance officers this is both "nothing new" and an accelerating source of questions.

To start with the basics, what is data?

Data 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. Other phrases include derived data, which refers to the idea that new data are created through manipulation of data, and metadata, the idea that data markers (e.g., when and how data was created) can be as important as original-source data.

Types of data 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; online search statistics; news or social media "sentiment" analysis; etc. — that the industry coined the umbrella term "alternative data" to capture the landscape.

Can data be protected?

Whether your company generates its own data or acquires data from someone else in a

commercial transaction of some sort, a headline question becomes how to protect it — with protection suggesting two related goals, namely (1) avoiding literal loss, theft or corruption and (2) establishing legal protections against infringement or misuse.

As to the former (avoiding loss), this is principally an information security matter, with safeguarding sensitive digital data being central to every firm's cybersecurity efforts. As to the latter (legal protections), this is principally a contractual matter. Detailed and thoughtful data protection and ownership terms need to be considered in connection with potentially every vendor contact, every customer contract, a firm's terms of employment and employee manuals, its website terms of use pages, and so on.

Of course this is a two-way street. At the same time as you are seeking to ring-fence your data contractually, other parties — such as your vendors — likely are nibbling at the perimeter, laying claim to data generated under their relationships with your firm. Protecting data thus means being sure that any third-party claim is consistent with your firm's view of the relationship.

There also may be intellectual property bases to protect data, notably as a trade secret. In general, however, data is not patentable or copyrightable. That said, systems for analyzing data, especially if grounded in technology, can be patentable, and how a database is arranged, organized and presented can be copyrightable.

Is data regulated?

There is no comprehensive legal and regulatory approach to data that applies across jurisdictions today. Instead, a worldwide patchwork of often conflicting laws and regulations apply. Here are a few of them:

- ***Privacy:*** Many jurisdictions seek to protect "personal data" or "privacy" associated with individuals, especially names, addresses, government identification numbers and the like. Populations deemed especially vulnerable, such as children or the elderly, often are given special data protections. Personal health and financial records, gender orientation information, political and religious affiliations, and other special categories of personal information can have heightened protection too.

- ***Governmental Data:*** There is a natural presumption that governmental data, especially in democratic societies, is intended to be "open" and accessible to the public. In fact, this is not guaranteed, and permitted uses of governmental data can be context-specific. For example, some public data sources may be presented with the disclaimer that they are intended for research and other noncommercial purposes. There also are a variety of instances when governmental data are explicitly nonpublic and restricted, e.g., in connection with governmental contracts or studies and approvals that have not yet been announced.

- ***National Security:*** National security-specific data are also higher risk, and the dividing line between national security and commercial considerations is increasingly blurred. 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.

- ***Website Data:*** "Web scraping" is the automated gathering of data from a third-party website. While the activity is widespread, a variety of legal claims may apply under U.S. 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 — 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 side.

Are data vendors regulated?

There have been a variety of proposals by the Federal Trade Commission to regulate data "resellers," but at present there is only very limited U.S. law specifically subjecting the sale of data to a comprehensive conduct or registration requirement.

That said, at least some data vendors have asked themselves whether they might be regulated as investment advisers. 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 U.S. Securities and Exchange Commission 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 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 (3) the information is not organized or presented in a manner that suggests the purchase, holding or sale of any securities. Given the profusion of data-based businesses today it is somewhat surprising that the last of these letters was issued in the 1990s.

Does data present material nonpublic information risk?

One of the threshold concerns for an investment manager purchasing data is that the data not carry the risk of tainting the manager with possession of material nonpublic information, or MNPI, and thus the possibility of being in breach of insider trading laws.

For example, consider a bank that is selling credit card transaction data. Might that data represent MNPI as to the bank by revealing the volume of credit card transactions the bank is handling? Or might that data represent MNPI as to a particular merchant by revealing sales information before it can be aggregated and publicly disseminated by the merchant?

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, 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 predict a retailer's overall sales figures and then trade in the securities of public companies when his program predicted a company's publicly reported sales results would surprise the market.

Among other things, the employee argued that the bank never saw more than 2 percent of a given retailer's credit card transactions, a basis to claim the data was nonmaterial. 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.

MNPI risk is one reason why some (but not all) firms prefer that the data they buy not be available to them on an exclusive basis. These firms also are attuned to the potential for broadly phrased concerns about data access as a fairness issue.

What is "data lineage" (and why should a compliance officer care)?

Data vendors tend to be unregulated businesses, but the data they sell might be regulated in a variety of ways — most prominently, it may be personally identifiable information ,or PII, protected under federal, state or non-U.S. privacy laws. "Data lineage" 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 through the course of its existence.

Understanding data lineage also can be a protection in mitigating insider trading risks. Under U.S. insider trading laws, one of the elements that a prosecutor often must prove is that the information was "misappropriated" in some way. For example, in the case of the bank employee who traded based on the bank's credit card data, prosecutors could readily say that he was not using the data as intended by his job description and therefore had misappropriated it within the meaning of insider trading law. Likewise, had the analyst sold that same data, it would have carried a serious flaw in its "lineage" (namely, it never belonged to the seller), exposing the purchaser to insider trading risks as well.

Should a firm diligence its data vendors?

In a word, yes. An investment manager purchasing data wants to be sure the vendor is alert to the same types of concerns that the manager has, that data lineage 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.

Is artificial intelligence relevant to the discussion? Is AI regulated?

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. Accordingly, the most prolific users of data

in support of investment programs likely have an AI component to their activities.

AI is not directly regulated but when used by a regulated business, such as an SEC-registered investment firm or a bank, regulators do have various expectations. Most importantly, regulators expect that AI, or any sophisticated software for that matter, has been tested and is reasonably well understood by its users, will continue to be tested and "fit for purpose" over time, and that its core operations and outcomes can be understood and explained to a firm's internal and external governance bodies (senior management, compliance and control functions, and regulators).

What are data ethics?

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 (e.g., legal and contractual requirements, technical capacity, social expectations, reputational considerations, etc.).

In Closing

The most important questions for a lawyer or compliance officer reviewing a proposed data initiative are: First, where will we get the data from, second, how are we going to use the data, and then, have we thoughtfully assessed related regulatory and contracting risks?

This primer helps troubleshoot issues based on the answers to those core questions. But there also are a host of "second order" questions not covered here, such as whether your contracts reflect state-of-the-art data terms in your favor (the answer is probably no), whether your compliance policies and procedures need updating (probably yes), and whether your existing disclosures to clients and investors need to change (maybe). In short, data is not solely a commercial issue; there is much for lawyers and compliance officers to do.

Nathan J. Greene is a partner at Shearman & Sterling LLP.