

TMT INSIGHTS

From the Debevoise Technology, Media & Telecommunications Practice

Derived Data: Contracting Considerations and Market Practices

“Big Data” is big business. In practically every industry, including finance, healthcare, technology and media, a wide range of data sets are collected and mined for valuable business intelligence and insight. Twitter, for example, recently reported that it derived over \$300,000,000 in annual revenue from data licensing, including its “firehose” of over 500 million daily tweets. Large data sources like the “firehose” can be used for many purposes, from identifying breaking news events to measuring consumer sentiment.

Agreements governing the processing and use of licensed data take several forms. The most common type of data agreement is a license by which a data supplier grants a customer the right to use specified data for the customer’s business purposes. In the financial industry, for example, some trading firms have obtained a license to satellite images of parking lots (and other similarly esoteric data sets) to help them predict a retailer’s holiday-season sales.

A second category of data agreements is, in essence, a type of services arrangement. In these agreements, a company provides its own data to a third party for analytics processing. A retailer, for example, might use a data analytics firm to identify correlations between its sales and advertising spend.

Data consortiums (and other multi-participant environments, such as trading platforms) represent a hybrid situation that implicates aspects of both the data licensing and data processing modalities. Consortiums collect data from multiple industry participants to create statistics and other analytics that are then licensed back to the participants and, in most cases, third parties. The terms and conditions surrounding the consortium’s data collection and aggregation activities share many characteristics with a data processing arrangement. By contrast, the consortium’s distribution of processed data to its members and third parties is, in many respects, a species of data licensing arrangement.

At the core of every data agreement are a number of key provisions that govern data ownership and use. Of particular interest, is the right to create and use “derived data,” also known as “processed data” or “resultant data,” *i.e.*, new data sets created from the original licensed data by aggregation, algorithmic manipulation or other processing.

In a standard data licensing context, the licensor’s principal concern with a licensee’s creation of derived data is that an ostensibly new data set may be similar enough to the original to serve as a substitute, and diminish the market, for the licensor’s products. Licensors may also take the view that they should be entitled to capture all commercial value embodied in the data as a product, as distinct from any insights or other commercial value that the licensee can obtain from analyzing the data. To address these concerns, licensors may contractually limit a licensee’s right to create and then use derived data, or even prohibit altogether the licensee’s creation of derived data.

One way that licensors can address the potential substitutability of derived data for their products is by defining “derived data” as data from which the licensor’s original data cannot be “reverse engineered.” Alternatively, or in addition, the definition may more explicitly require that the new data set be sufficiently processed that it cannot serve as a commercial substitute for the original. Data that does not meet these thresholds remains the property of the licensor.

From the licensee’s perspective, a principal focus in any data agreement should be that the definition and treatment of derived data sets is consistent with licensee’s intended use. It is important for the licensee to have counsel that understands the commercial rationale for the transaction, including whether the client will be creating derived data and, if so, how the client intends to distribute or use this data. Counsel should then review the agreement (often a form served up by the data supplier) to ensure that it does not include provisions inconsistent with that commercial rationale.

The issues raised by derived data are somewhat different in the data processing modality. From the vendor’s (*i.e.*, data processor’s) perspective, an important value driver may be the right to aggregate the data of all its customers to create derived data that can be mined for industry insights and trends. The customer’s (*i.e.*, the data supplier’s) principal concern is ensuring that its proprietary data (and the know-how it embodies) is protected from disclosure and to avoid breaching confidentiality or other legal obligations it may have toward third parties. This can be done by anonymizing the data before supplying it to the vendor or by requiring that the vendor itself anonymize the data. The customer should propose standards for the vendor’s use of customer data (*e.g.*, that a minimum number of other customers be included in the aggregation). It should also request an indemnity from the vendor, to protect it from liabilities resulting from vendor’s use of the customer data in the aggregated data set. As a commercial point, the customer may also negotiate some non-monetary consideration for contributing to the derived data set, such as free access to any resulting report that the data processor commercializes.

As noted, data consortiums represent a hybrid of the data processing and data licensing modalities. Similar to a data processing arrangement, the consortium documentation will often provide that each contributor retains ownership of its individual contribution of data. The consortium will own the aggregated set of contributed data and, in some cases, may own all contributed data in anonymized form. In its role as a licensor of data products, the consortium may have the same concerns with respect to its members creating and using derived data as any other licensor: it will seek to prevent its members, the licensees, from creating derived data that can replace the consortium's products or diminish their value.

Derived data rights can be some of the most hotly negotiated issues in a data agreement. Careful attention to the numerous complexities of these agreements is critical to making sure they achieve the parties' business objectives.

Contributors

Henry Lebowitz

Partner – New York
hlebowitz@debevoise.com
+1 212 909 6588

Jeffrey P. Cunard

Partner – Washington, D.C.
jpcunard@debevoise.com
+1 202 383 8043

Michael A. Diz

Partner – New York
madiz@debevoise.com
+1 212 909 6926

Jonathan E. Levitsky

Partner – New York
jelevitsky@debevoise.com
+1 212 909 6423

Jim Pastore

Partner – New York
jjpastore@debevoise.com
+1 212 909 6793

Michael Schaper

Partner – New York
mschaper@debevoise.com
+1 212 909 6737

About Our Practice

Debevoise's far-reaching TMT practice is a top choice for clients seeking trusted advisors for complex deals or high-stakes cases. Highly ranked in a variety of categories in the technology, media and telecommunications space, Debevoise has what *Chambers Global* describes as a TMT practice that is "full-service and creative, and has a global presence." For more information please visit our [Technology, Media and Telecommunications practice page](#).