Practical Guidance

Data License Agreements

A Practical Guidance® Practice Note by Chad A. Rutkowski, BakerHostetler LLP

Chad A. Rutkowski
BakerHostetler LLP

This practice note outlines key provisions in data license agreements, including the license grant, sublicensing, derived and usage data, data delivery, confidentiality and security, data audits and controls, disclaimers, representations and warranties, indemnification, and the treatment of data post-termination. This note also discusses related issues such as the protectability of data and data compilations under copyright and trade secret law (which supports the notion of a "license" to data) and the lifecycle of data and data markets, which should inform the provisions of any data license agreement.

Certain topics are beyond the scope of this practice note, including:

- **Industry-specific considerations**. While this note is industry agnostic, be aware that data license deal terms and structure can differ significantly among industries. For instance, different considerations would apply to healthcare data, consumer purchasing data, sales data, industrial data, real estate data, etc.
- **Privacy laws**. This note does not address the impact of state or federal privacy laws. Note, however, that compliance with such laws can be a major driver of the value of data sets and impact provisions such as representations and warranties and indemnification.
- **An expansive definition of data.** This note presumes that the term "data" refers to information and facts. However, to a data scientist, "data" can consist of images, videos, and other protectable content (e.g., using photographs to train machine learning data models).
- **Other types of data agreements.** Data agreements are sometimes styled as "data usage agreements" or "data sharing agreements." Such agreements have many of the same features as data license agreements but tend to lack one key ingredient—a license grant. While it is certainly possible to draft a commercially reasonable, efficient, and effective agreement without a license grant, such agreements arguably do less to establish a company's data resources as an asset. This note thus presumes that the parties are approaching the data exchange as a protectable asset that is being licensed to the other party.

For a sample data license agreement and a related checklist, see Data License Agreement (Pro-Licensor) (Short Form) and Data License Agreement Checklist. For overviews of copyright and trade secret law, see Copyright Fundamentals and Trade Secret Fundamentals. For resources on intellectual property licensing, see Intellectual Property Licensing Resource Kit.

Initial Considerations – Data Lifecycle and Data Markets

To draft data license agreements with effective, impactful, and reasonable scope provisions, it is important to understand:

- How data is used by data scientists and analysts (i.e., the data lifecycle)
- The kinds of markets in which operationalized data is deployed

Each is discussed below.

Data Lifecycle

The lifecycle of data generally follows a predictable pattern regardless of industry, as set forth below:

- **Collection.** In this phase, data is collected as an adjunct to business operations or from purposeful acquisition. Such data might include:
  - Enterprise data (i.e., operational data and data generated by machinery and computer assets)
  - Customer data (including purchasing habits and payment history)
  - Record data (such as electronic medical records and financial records)
  - Event stream and real-time data collection (such as data collected from internet of things connected devices and machinery)
  - Data sets acquired from brokers, data trusts, web scraping, or other means
  - Usage data (e.g., user browsing history, searches, and information regarding user interactions with websites or applications, such as through cookies)
- **Storage.** In this phase, collected data is housed through cloud infrastructure and data warehouses for later use.
- **Formatting.** In this phase, raw data is cleaned and structured to enable further use via:
  - Schemas
  - Data models
  - Normalization techniques
- **Analysis.** Data is analyzed by refining workflows, applying algorithms and other techniques to the data, and/or training machine learning models to detect patterns and refine probability analysis. There are three basic kinds of analytics:
  - Descriptive analytics (describes the current state of the subject depicted by the data)
  - Predictive analytics (anticipates future states of the data subject)
  - Prescriptive analytics (recommends actions based on patterns revealed in the data)
- **Reporting.** This stage involves the creation of dashboards, data visualizations, and reports that enable further action.
- **Maintenance.** Ongoing maintenance may be necessary to ensure the data sets do not create unfair bias, that the accuracy of analytics is constantly audited and improved, and that the data is adequately protected.

Often in the lifecycle of data, it is necessary to combine sources of data to enable richer insights, or to modify the original data set to enable formatting and analysis. Variously known as "derived data," "aggregated data," or "resultant data," the ability to create these derivations is often the very *raison d'etre* of a data license agreement and informs the scope of the license. See "License Grant" under Key Provisions in Data License Agreements below.

Data Markets

The lifecycle of data generates entire ecosystems of economic actors—including data brokers, analysts, software providers, and data storage infrastructure vendors—whose activities combine to create different markets. These markets operate at varying levels of sophistication, as described below:

- **Mature markets.** These markets have robust business models around sourcing and licensing data, with provable return on investment (ROI) and successful legal enforcement. Examples include:
  - Social media data
  - Syndicated data in point of sales in retail and consumer packaged goods
  - AdTech data
  - Real estate data
  - Financial markets data
- **Developing markets.** These are emerging markets that share some of the hallmarks of the robust data markets, but without the consolidation or provable ROI and successful enforcement. While long term business planning may have begun, such markets are not yet fully operationalized. Examples include:
  - Healthcare data
  - Financial services data
  - Industrial data
- **Nascent markets.** These markets treat data as an afterthought and/or seek to obtain a near-term business advantage without a longer-term monetization or business goals in mind. Examples might include:
  - Trucking logistics
  - Agriculture
  - Public utilities
  - HVAC and plumbing equipment

Protectability of Data and Data Compilations

Generally, "[r]aw facts may be copied at will," and "information is not property unless some law makes it so." See Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 350 (1991); Silvaco Data Sys. v. Intel Corp., 184 Cal. App. 4th 210, 239 (2010), *as modified on denial of reh'g*, 2010 Cal. App. LEXIS 771 (May 27, 2010), *and disapproved on other grounds* by Kwikset Corp. v. Superior Ct., 51 Cal. 4th 310 (2011). However, when facts and information are collected, stored, and organized, the resulting compilations may be subject to legal protections and, in turn, licensed to others.

There are at least three avenues by which a licensor might claim ownership of data and/or data compilations:

- Copyright law
- Trade secret law
- Contractual covenants

Each is discussed below. For general overviews of copyright and trade secret law, see Copyright Fundamentals and Trade Secret Fundamentals.

(This note does not address the applicability of patent law. Strictly speaking, patents may protect methods and processes surrounding managing and analyzing data, but not the underlying data sets themselves. Thus, a patent license can only give the licensee the right to make, use, sell, or import an article of manufacture or a process/ method covered by the patent, not any underlying data. Note, however, that a patent license may be an important adjunct to a data license and may impact pricing and other deal terms. For more information on patent licensing, see Intellectual Property Licensing Resource Kit — Patent Licenses.)

Applicability of Copyright Law to Data Compilations

Copyright protects "original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device." 17 U.S.C. § 102. Compilations of information are entitled to copyright protection if the information was organized in a creative, original fashion. See Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 348 (1991) (holding that factual compilations may be sufficiently original if choices as to selection and arrangement are made independently by the compiler and entail a minimal degree of creativity).

Copyright conveys to the owner certain exclusive rights, including the rights of:

- Reproduction
- Display
- Distribution
- Adaptation (i.e., the right to create derivative works)

See 17 U.S.C. § 106 and Exclusive Rights of Copyright Owners. These rights may be licensed, including in the data context. For more on copyright licensing generally, see Intellectual Property Licensing Resource Kit — Copyright Licenses.

Copyright Registration of Databases

While registration is not necessary to obtain copyright protection or to license rights, there are various advantages to doing so (e.g., a registration is required to sue for copyright infringement and allows plaintiffs to seek statutory damages and attorney's fees if the work at issue was registered within three months of first publication or prior to the infringement). Thus, it is generally advisable to register copyrighted works. For more on registration, see Registration of Copyrights.

To register an electronic database with the U.S. Copyright Office, the contents of the database must be:

- Arranged in a systematic manner
- Accessed solely through an integrated information retrieval program or system that:
    - o Uses a query function to access the content
    - o Yields a subset of the content or organizes the content based on the parameters specified in each query

See U.S. Copyright Office, Compendium of U.S. Copyright Office Practices § 727.1 (3d ed. 2021). An application to register a database typically covers the selection, coordination, and/or arrangement of data, information, or files, but does not cover the data, information, or files themselves unless they are specifically claimed in the application.

The Copyright Office recognizes three discrete forms of authorship in compilations of data, any one of which can support a copyright registration:

- Selection authorship (i.e., choosing the material or data that is included in the database)
- Coordination authorship (i.e., classifying, categorizing, ordering, or grouping the material or data)
- Arrangement authorship (i.e., determining the placement or arrangement of the material or data within the database as a whole)

See U.S. Copyright Office, Compendium of U.S. Copyright Office Practices § 727.2 (3d ed. 2021). For these forms of authorship to warrant protection, they must be original to the author, meaning they must have been independently created and possess some minimal degree of creativity. Feist, 499 U.S. at 345. The threshold level of creativity is very low—it simply requires that the work "possess some creative spark, 'no matter how crude, humble or obvious' it might be." Id.

Case Law Involving Data Compilations

There are numerous cases supporting authorship in data compilations after *Feist*, including several that concern electronic databases. These include both circuit and district court decisions, as set forth below:

- **Circuit decisions.** At the appellate level, the Ninth, Eleventh, and Seventh Circuits have considered the issue. See Experian Info. Sols., Inc. v. Nationwide Mktg. Servs. Inc., 893 F.3d 1176 (9th Cir. 2018) (database of consumer names and addresses protectable by virtue of selection authorship in "culling data from multiple sources and selecting the appropriate pairing of addresses with names before entering them in the database"; however, defendant's copying of only 80% of database insufficient to infringe plaintiff's thin copyright); BUC Int'l Corp. v. Int'l Yacht Council Ltd., 489 F.3d 1129 (11th Cir. 2007) (section headings of electronic database protectable element of electronic database of yacht listings); Assessment Techs. of Wis., LLC v. WIREdata, Inc., 350 F.3d 640 (7th Cir. 2003) (grouping of 456 fields of publicly available real estate value information into 34 tables protectable as a compilation but copying of database to extract information—as opposed to copying of database format—was deemed a fair use).
- **District court decisions.** For relevant district court decisions, see, e.g., Metro. Regional Info. Sys., Inc. v. Am. Home Realty Network, Inc., 888 F. Supp. 2d 691 (D. Md. 2012) (copyrightable authorship in electronic database of real estate property listings uploaded by independent brokers where the plaintiff controlled the quality and accuracy of content); Snap-on Bus. Sols., Inc. v. O'Neil & Assocs., Inc., 708 F. Supp. 2d 669 (N.D. Ohio 2010) (vendor's data tree, link structure, and associational relationships in database management software sufficiently original to deny summary judgment to the defendant customer who allegedly infringed by extracting their own data from vendor's database); Madison River Mgmt. Co. v. Bus. Mgmt. Software Corp., 387 F. Supp. 2d 521 (M.D.N.C. 2005) (finding that defendant's electronic database, which imposed on the raw data at issue a new structure and metadata enhancements, was a copyrightable compilation and declining to find fair use as a matter of law where there were questions as to whether the plaintiff copied the database to take advantage of its structure); Corsearch, Inc. v. Thomson & Thomson, 792 F. Supp. 305 (S.D.N.Y. 1992) (selection, coordination, and arrangement of state trademark data in database protectable).

Courts have also found that copyright protection may subsist in the "containers" of data (i.e., software and certain non-literal elements of software) that port, manage, and/or organize the data, including:

- Database schemas (see DSMC, Inc. v. Convera Corp., 479 F. Supp. 2d 68 (D.D.C. 2007) and Digital Drilling Data Sys., L.L.C. v. Petrolink Servs., Inc., 965 F.3d 365 (5th Cir. 2020))
- SQL data structures (see Positive Software Sols., Inc. v. New Century Mortg. Corp., 259 F. Supp. 2d 531, 536 (N.D. Tex. 2003), *vacated in part on other grounds*, 2003 U.S. Dist. LEXIS 7175 (N.D. Tex. Apr. 29, 2003) (describing SQL data structures as "the organization of data into tables, the selection of column elements for the tables, [and] the names, data types, and sizes of column elements"))
- Selection and coordination of data fields (see Harbor Software, Inc. v. Applied Sys., Inc., 925 F. Supp. 1042 (S.D.N.Y. 1996))
- Link structure and navigational information for an online tools database (see Snap-on Bus. Sols., Inc. v. O'Neil & Assocs., Inc., 708 F. Supp. 2d 669, 685 (N.D. Ohio 2010))
- Data warehouse design and data models (see eScholar, LLC v. Otis Educ. Sys., Inc., 2005 U.S. Dist. LEXIS 40727 (S.D.N.Y. Nov. 3, 2005))

For additional guidance on the copyrightability of computer software, including databases, see Nimmer on Copyright § 2 A. 10.

Applicability of Trade Secret Law to Data and Data Compilations

Trade secret law is a patchwork of federal and state statutory and common law, and its specific application can vary accordingly. But there is general consensus around the definition of what constitutes a trade secret.

Under the federal Defend Trade Secrets Act (DTSA), for instance, trade secret is broadly defined as "all forms and types of financial, business, scientific, technical, economic, or engineering information . . . whether tangible or intangible, and whether or how stored, compiled, or memorialized physically, electronically, graphically, photographically, or in writing," if:

- The owner has taken reasonable measures to keep such information secret –and–
- The information derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable through proper means by, another person who can obtain economic value from the disclosure or use of the information

18 U.S.C. § 1839(3). The DTSA also lists various examples of potential trade secrets in the definition, including compilations. Id. For more information on the DTSA, see Defend Trade Secrets Act (DTSA) Fundamentals.

Definitions under the Uniform Trade Secrets Act (UTSA) and Restatement of Torts are similar, and generally support the notion that "compilations of information receive trade secret protection, even where the component parts contained therein are publicly available," so long as the compilation derives independent economic value from not being generally known or readily ascertainable. See, e.g., Ihs Global Ltd. v. Trade Data Monitor, LLC, 2021 U.S. Dist. LEXIS 101952, at *27 (D.S.C. May 4, 2021) (noting that both the DTSA and SCUTSA, South Carolina's version of the UTSA, have similar definitions for what constitutes a trade secret). Such compilations can include databases. See, e.g., Compulife Software, Inc. v. Newman, 959 F.3d 1288 (11th Cir. 2020); Duggan v. Am. Family Mut. Ins. Co., 2010 U.S. Dist. LEXIS 30691 (E.D. Wis. Mar. 30, 2010).

While trade secret law does not involve the same licensable "bundle of rights" as copyright law, trade secrets can be licensed without losing protection. See Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 486 (1974); Nova Chems., Inc. v. Sekisui Plastics Co., 579 F.3d 319, 327 (3d Cir. 2009). However, licensing trade secrets is often a fraught exercise, as revealing a trade secret to a licensee risks loss of the secret due to:

- Inadequate contractual protections
- Disclosure (inadvertent or otherwise) by the licensee to third parties
- Reverse engineering

Thus, when drafting a trade secret license, it is crucial to specify how the trade secret should be handled and used by the licensee and any other relevant persons (such as employees, contractors, or affiliates of the licensee).

Contractual Covenants regarding Data and Data Compilations

Data license agreements often do not specify what category of intellectual property, or IP, is being licensed. Many licenses simply recite that the licensee acknowledges the data is "proprietary" or otherwise "owned" by the licensor and then proceed to dictate the license terms. This is essentially the creation by fiat of a form of pseudo-intellectual property—the data resides on a server owned and controlled by the licensor, and the licensee cannot get access to it without acceding to the terms dictated by the licensor.

Even if the existence of genuine IP rights is questionable, such contracts can be effective and powerful tools in securing to the licensor many of the hallmarks of traditional IP licenses, such as:

- Use restrictions (e.g., restrictions on fields of use or against competitive uses)
- Royalty payments
- Audit rights
- Imposition of seat licenses or other restrictions on the number of users (and concomitant tiered pricing)
- Reverse engineering restrictions

Case law suggests that courts will enforce such contract terms, even where the IP that formed the basis of the license is itself expired or unenforceable. See, e.g., SAS Inst., Inc. v. World Programming Ltd., 874 F.3d 370 (4th Cir. 2017); Nova Chems., Inc. v. Sekisui Plastics Co., 579 F.3d 319, 328 (3d Cir. 2009) ("Trade secret licenses may endure even where the trade secret itself is destroyed by general disclosure.").

Regardless, the licensor might want to bolster the strength of these covenants by requiring the licensee to provide certain written acknowledgments, such as some or all of the following:

- The data is owned by the licensor
- The licensor expended significant resources to create and/or maintain the data (thus making the data valuable)
- The data and/or its compilation is original to the licensor

Key Provisions in Data License Agreements

Key provisions in data license agreements concern the following:

- License grant
- Sublicensing
- Derived and usage data
- Delivery of data
- Confidentiality and security
- Data controls and audits
- Disclaimers
- Representations and warranties
- Indemnification
- Treatment of data post-termination

Best practices for negotiating and drafting these provisions are discussed below.

License Grant

Too often data licenses allow the licensee to use the data (presumably under the theory that the data constitutes a trade secret) but with little guidance on scope, leaving open the question of what type of use is permissible. When drafting the license grant clause, remember that "rights" are being "licensed," such that the scope of the license should reflect both:

- The theory of intellectual property under which the data is being licensed (see Protectability of Data and Data Compilations above)
- The phase in the lifecycle of data in which the data is being deployed (see Initial Considerations – Data Lifecycle and Data Markets above)

Consider questions such as the following: Is the data collected in a database or other data set to which copyright might apply? Can it be reproduced? Distributed? Displayed to third parties? And perhaps most importantly, can derivatives be made of it through aggregation with other data sets, or by anonymizing, disambiguating, or cleaning the data such that the data is removed or otherwise modified, or by combining the data in other ways?

Once these questions are answered, the license grant can be tailored and more granularly restricted by treating the contemplated data lifecycle use in the same way that field of use restrictions are treated in traditional copyright licenses. "Field of use" can be defined either:

- In the license grant itself (if the contemplated use is simple and straightforward) –or–
- Elsewhere in the agreement

For instance, a license grant clause that resembles a traditional copyright license might read:

> Licensor hereby grants to Licensee a non-exclusive, non-transferable, non-sublicensable, revocable license to reproduce, display, distribute, make derivative works of, and use the Licensed Data during the Term and in the Territory but only for Permitted Uses.

(In this example, Permitted Uses is a defined term that equates to field of use restrictions.)

This short, straightforward sentence addresses a number of important deal points that are common to IP licenses. It also sets the table for establishing the rights, obligations, and limitations that can be more fully addressed elsewhere in the agreement, including:

- **Reproduction rights.** Will the licensee install an instance of the licensed data or data set on its own servers? If not, and the licensee will be accessing data through the licensor's portal, using a data feed, or accessing the data through an application programming interface (API), consider changing "reproduce" in the above license grant clause to "access."
- **Display rights.** Will the licensee be permitted to make dashboards, data visualizations, reports, or otherwise reveal the data to others? Note that the class of individuals that can access these displays can be addressed elsewhere in the agreement in a fashion that protects the confidentiality of the data. See "Confidentiality and Security" below.
- **Distribution rights.** Will the licensee be permitted to make the data, or distillations of the data by way of the "Reporting" layer of the lifecycle, available to third parties? As with the display right, the distribution right can be granted in the license grant clause but limited by other clauses in the agreement. Consider, for instance, whether any copyright or confidentiality/proprietary notices apply to the data, such that the licensee should preserve such notices for any downstream distribution.

- **Right to prepare derivative works.** Can the licensee combine the data with other data or otherwise aggregate it, anonymize it, add or remove data, or otherwise modify the data set?
- **Use.** Does the data have trade secret value as well as copyright value, such that "use" would be prohibited if not otherwise granted?
- **Licensed data.** Defining the data to be licensed can be a nuanced exercise. This may be as simple as identifying specific databases. Alternatively, the definition may have certain requirements or restrictions embedded within it, such as access to "anonymized data," or specifying that only a certain class of data is being licensed.
- **Permitted uses.** Permitted uses can be itemized in the license grant clause or defined elsewhere in the agreement. Defining the term may be preferable, as a definition can provide more detail on the scope of the rights being granted and can specify whether the data can be used for purposes such as:
    - Internal uses only
    - Providing analysis and recommendations to others
    - Creating specific aggregates or other derived data
    - Summarizing, reporting, or displaying to others
    - Any other uses appropriate under the circumstances
- **Exclusivity.** The grant clause should address whether the license is being granted on an exclusive or nonexclusive basis. If exclusive, the agreement should address whether that exclusivity is to a certain right (e.g., reproduction, derivative work, etc.); geographic area; or within a certain industry or other field of use.
- **Territory.** Given the decentralized nature of most data uses, territory is generally presumed to be worldwide. However, if the data at issue is personally identifiable information (PII), export-controlled information, or otherwise susceptible to governmental regulations, territorial restrictions can be important to comply with (or avoid) such regulations.
- **Term.** The term of the data license will often depend on the nature of the data being licensed and the scope of the permitted uses. Data that will be downloaded and reproduced by the licensee, aggregated, or used to enhance products or operations often needs to be licensed on a perpetual basis, as such data is hard to remove and decouple once it has been integrated into other data sets or used to enhance operations. At the other end of the spectrum, data that will be accessed by a data feed or API, or whose usefulness is temporally limited, is normally tied to the term of the agreement itself.
- **Revocability.** Licensees will often want the license to be irrevocable, which simply means that the license cannot be terminated early for reasons other than a material breach (such as delays in payment). The ability to revoke the license can create great disruption for the licensee, especially if the data is being aggregated or acted upon in day-to-day operations. Revocability is thus a potent lever for the licensor but is a term that most often is negotiated in favor of the licensee.
- **Transferability.** Data licenses are often nontransferable, especially if the licensor has trade secret interests in the data at issue.
- **Sublicensing.** The ability to sublicense may be important to the licensee if it lacks the sophistication to act upon the data and needs to allow access to analysts and third-party software vendors to obtain the benefit of the license. For more on this issue, see "Sublicensing" below.

If the data being supplied is unstructured and/or made available in a feed or other real-time stream, "use" may be the only appropriate licensed right under a trade secret theory. However, use is a broad concept. See, e.g., Compulife Software Inc. v. Newman, 959 F.3d 1288, 1313 (11th Cir. 2020) ("[A]ny exploitation of [a] trade secret that is likely to result in injury to the trade secret owner or enrichment to the defendant is a 'use.'") (quoting Penalty Kick Mgmt. v. Coca Cola Co., 318 F.3d 1284, 1292 (11th Cir. 2003) and Restatement (Third) of Unfair Competition § 40 cmt. c (1995)). Thus, you should still consider the same rights questions that apply to copyright licenses, such as:

- Can "derived data" be created?
- Can the data be shared with third parties in certain circumstances?
- How will the trade secret nature of the information be protected if the data can be reproduced, distributed, incorporated into other data sets, or displayed at the presentation layer?

One potential solution to licensing data while preserving trade secrets is dividing access to tiers of data depending on the status of the user. That is, depending on the nature of the data and the nature of the user, certain users might have access to all raw data, others might have access to only data aggregated in a fashion to remove confidential or personal information, and others might have access only to dashboards or reports that reveal specified slices of data.

The license grant can also specify that only a certain number of users can access or benefit from the use of the data, akin to "seat" licenses in software licensing. Such provisions do not necessarily enhance IP protections in the data, but rather, allow the licensor to charge more for increased usage. And, significantly, usage beyond the licensed number of seats may be considered not just a breach of the license agreement, but may also constitute copyright infringement and/or trade secret misappropriation.

For more on software licensing, see Software License Agreements: Drafting and Negotiating the Agreement. For more on copyright infringement, see Copyright Fundamentals — Enforcement; see also Copyright Enforcement Resource Kit. For more on trade secret misappropriation, see Trade Secret Fundamentals — Misappropriation.

Sublicensing

As noted above, the license grant may include a right to sublicense. Grant of this right is fairly common, as the skill sets necessary to act upon licensed data often require analysts, consultants, software providers, and other third-party vendors. But the more hands that touch the data, the more mischief that might result.

Thus, a licensor should ensure that the licensee retains ultimate liability for any breach committed by a sublicensee. The licensor may also want to require each sublicensee to agree in writing to be bound to the terms of the license, especially as regards:

- Preserving confidentiality
- Adhering to security requirements
- Using the data only for permitted uses

For more on the first two terms, see "Confidentiality and Security" below.

Derived and Usage Data

Derived data and usage data are often important drivers of a data license deal. Derived data may refer to a new data set created when the licensed data is added to, deleted from, or combined with other data. It might also refer to entirely new information or data that is generated from analyzing and using the licensed data. Usage data tends to be more passive in nature, generally referring to new data that is generated from a licensee's usage of systems and services provided by the licensor. Careful definitions are essential here—especially with respect to ownership of derived and usage data—and depend on the nature of the business deal at hand.

If the licensee is a data analytics and/or software analytics vendor, the licensee will often seek the right to use the licensed data to improve or enhance the vendor's services or products. Such use is not quite "derived" or "usage" data, but rather the techniques that are incident in handling the data. Careful drafting is again essential here. Key considerations include:

- Will the vendor use the licensor's data to enhance algorithms or train data models?
- Will copies of the licensor's data need to be maintained by the vendor indefinitely?
- Will the analytics provided by the vendor be provided to the vendor's other clients, including the licensor's competitors?

In some instances, cross-licenses or license-back provisions may be appropriate, which can both enable the licensee to get the benefit of its bargain but be limited in such a way as to not do competitive harm to the licensor.

Delivery of Data

The parties should consider how the licensee will get access to the data. Will there be an installation of the licensor's database? Access to an unstructured data warehouse that resides on a server or set of servers? Access through an API maintained by the licensor? The parties should ensure that the data is delivered in a fashion that will be useful to the licensee, while still maintaining the licensor's ability to control dissemination of its valuable data. Another key consideration is compatibility between the parties' various technologies.

Confidentiality and Security

Confidentiality

The confidentiality clause in a data license agreement can serve two important purposes. First, the clause creates the duties that are necessary for trade secret protection, and hopefully has the practical effect of actually preventing disclosure of confidential data. Second, however, is the fact that the licensee agrees to treat the data as confidential, whether it rises to the level of a trade secret or not. This important covenant is part of what creates the "IP by fiat" scenario described earlier in this practice note. See Protectability of Data and Data Compilations above. For an overview of trade secret law, see Trade Secret Fundamentals.

Confidentiality clauses in data licensing agreements tend to adopt many of the typical boilerplate terms of non-disclosure agreements (NDAs). However, standard NDAs can often be nonspecific on the level of care the receiving party must exercise in keeping the other party's information confidential, often merely requiring "such care as the receiving party treats its own confidential information" or words to that effect. Such language may not be sufficient in the data licensing context (depending on the type of data and the licensee), in which case, the licensor might want to require more specificity in how the data is handled. For instance, the licensor could require adherence to either:

- Specific industry protocols
- Best practices established by the licensor

Note also that standard NDAs generally terminate the data recipient's obligations after a certain time period. Such termination may or may not be appropriate in the data licensing context. On one hand, if trade secret information is being exchanged, then the obligation might need to be perpetual or for an extended period of time; on the other hand, the relevancy of the data at issue might have a short shelf life.

For more information on confidentiality agreements, see Confidentiality Agreements. For additional resources, see Confidentiality and Non-disclosure Agreements Resource Kit.

Security

Data security provisions often bolster and further define the licensee's confidentiality obligations. These clauses might specify the kinds of encryption, firewalls, passwords, protection software, and other technological protections that the licensee must use to safeguard the licensor's data. Note that encryption requirements can vary depending on the data's transmission status ("in transit" versus "at rest").

Data security clauses may also cover:

- Physical protection systems (such as locked doors, security guards, and cameras)
- Employee background checks
- Outside audits of data security procedures
- Internal audits and testing
- Penetration testing
- Repair of inadequate systems

When deciding which data security provisions to include in a data license agreement, the parties might wish to reference and incorporate relevant protocols such as:

- ISO/IEC 27002:2005 (code of practice for information security management jointly developed by the International Organization for Standardization and the International Electrotechnical Commission)
- Federal Information Processing Standard (FIPS) 140–2 (U.S. government standard that defines minimum security requirements for cryptographic modules in information technology products)
- NIST SP 800–44 (guidelines on securing public web servers issued by the National Institute of Standards and Technology)
- Microsoft's Windows Security Baselines
- OWASP's Web Security Testing Guide (comprehensive guide to testing the security of web applications and web services issued by the Open Web Application Security Project)
- CIS Controls and Benchmarks (various standards and security configuration guidelines issued by the Center for Internet Security)

The parties should also consider how data breaches will be handled, including the responsibility and timing for providing notice of data breaches, cooperation with each other and law enforcement in addressing such breaches, and procurement of insurance to cover related costs. For more information on these and related considerations, see Privacy and Data Security Considerations When Negotiating or Reviewing a Transaction or Agreement.

Data Controls and Audits

Data Controls

Practical controls on data usage will be driven in large part by the nature of the data being licensed and the regulations that may or may not apply. The more the data impacts PII, healthcare data, or financial data, the more attention will need to be paid to imposing restrictions, data handling requirements, and reporting obligations on the licensee.

Depending on the data at issue, compliance might be necessary with laws such as:

- Health Insurance Portability and Accountability Act (HIPAA)
- Health Information Technology for Economic and Clinical Health Act (the HITECH Act)
- Fair Credit Reporting Act (FCRA)
- Gramm-Leach-Bliley Act (GLBA)
- Children's Online Privacy Protection Act (COPPA)
- California Consumer Privacy Act (CCPA)

- Virginia Consumer Data Protection Act (CDPA)
- General Data Protection Regulation (GDPR)
- Export control laws

For more information on these laws, see HIPAA Resource Kit, Fair Credit Reporting Act (FCRA) and State Mini-FCRAs: Step-by-Step Guidance for Compliance, Gramm-Leach-Bliley Act (GLBA) Privacy Requirements, Children's Online Privacy Protection Act (COPPA) Compliance, California Consumer Privacy Act (CCPA) Resource Kit, General Data Protection Regulation (GDPR) Overview Resource Kit, and Export Compliance Programs.

Additionally, the combination of differing data security regulations, the reputations of certain countries in respecting privacy interests and/or seizing personal data, and the existence of export laws may require robust provisions regarding restrictions on cross-border transfers. These may include:

- Requiring specific identification of servers and their locations where data is stored
- Prohibitions against use of certain cloud storage platforms
- Specific indemnities, insurance coverages, and liability caps for violation of cross-border prohibitions
- Specific audit rights to ensure compliance

For more information on other countries' data privacy and security laws, see Data Protection and Privacy in International Jurisdictions.

Licensors may also seek to prohibit licensees from reverse engineering, duplicating, or otherwise reproducing the licensed data.

Audits

To ensure compliance with data control provisions, licensors may require licensees to grant access to their systems and records for auditing purposes. Note, however, that licensees may resist such access, as it can put the licensee's own proprietary information and technology at risk. Such audits can cause considerable disruption, so limitations on the frequency and timing of the audits, and the degree of bilateral cooperation, can require considerable negotiation. Moreover, such audits can be expensive. Costs are generally borne by the licensor unless a certain threshold of noncompliance (e.g., 10%) is discovered.

Disclaimers

Questions abound as to whether data qualifies as a "good" subject to the Uniform Commercial Code (UCC). See generally Chadwick L. Williams, Not So Good: The Classification of "Smart Goods" Under UCC Article 2, 34 Ga. St. U. L. Rev. 453 (2018) (available here). Regardless, most data licenses contain classic UCC warranty disclaimers, including disclaimers against express or implied warranties of merchantability, fitness for a particular purpose, and noninfringement. Disclaimers as to title (i.e., noninfringement) are less common, however, as licensees typically want to know the licensor has the rights being licensed and will want to include an express warranty to this effect. See "Representations and Warranties" below.

Disclaimers in data license agreements often include language more suited to the data context than merchantability. For instance, a disclaimer might state that the data is provided "as is" or "as available," or that the data or access to it will be "uninterrupted" or "error free." Licensors might also disclaim the following:

- Use of the data does not infringe the privacy or intellectual property rights of third parties.
- Use of the data will yield a particular result or insight.
- The data is complete, timely, secure, or meets a certain quality threshold.
- The data is suitable for particular analytics, forecasting, or simulating uses.

For more information on UCC warranties and disclaimers, see Uniform Commercial Code Article 2 Express Warranties and Uniform Commercial Code Article 2 Implied Warranties.

Representations and Warranties

While the rights to the data being licensed are obviously the heart of any data licensing deal, the extent to which a licensor is willing to represent and warrant the ability to grant those rights can be tied to:

- The type of data being licensed
- How the data is collected
- The industry in which the data is being licensed and used
- Payment and/or any other consideration demanded by the licensor
- The licensor's role in the data lifecycle (see Initial Considerations – Data Lifecycle and Data Markets above)

If a licensor is providing access to data in a more passive fashion (e.g., because it happens to have access to data as a byproduct of its other operations), it is more likely that the licensor will take a caveat emptor approach to licensing. If, however, a licensor holds itself out as a data broker, or otherwise promotes or markets the data it provides, then the licensee might reasonably expect more robust representations and warranties. The licensee should also think through why it's licensing the data from the licensor and craft licensor warranties to fit those reasons, if possible.

For instance, a licensor might represent and warrant that:

- It has the rights to grant access to the data for the contemplated use
- There are no third-party claims or rights that would prevent or interfere with the license grant
- The data:
  - Is accurate
  - Contains no malicious code or virus
  - Meets certain quality or suitability standards

It is also possible to hedge the representations and warranties. For instance, a licensor might warrant that it "has made reasonable efforts" to ensure it has the rights to grant access to the data for the contemplated use, and/or that the licensor "is not aware" of any third-party claims or rights that would prevent or interfere with the license grant.

A licensee will typically represent and warrant that the data will not be used in any fashion that violates the rights of consumers or third parties, particularly privacy rights and intellectual property rights, and will not be used in a fashion that violates any laws or regulations.

For more information on representations and warranties generally, see Commercial Agreement Representations, Warranties, Covenants, Rights, and Conditions and Representations and Warranties Drafting.

Indemnification

Indemnification is a risk-shifting mechanism that requires one party to a contract to cover any damages or settlement amounts, and to defend and absorb litigation costs, for a set of defined risks related to the contractual relationship between the parties. How the risks get shifted depends on a mix of equity and bargaining power. Typically, the party best able to control the risk indemnifies the other party, but the party with the greater bargaining power can modify that calculus.

For instance, if a licensor has sufficient bargaining power, it might require the licensee to indemnify it for any third-party claims arising out of errors in the data, even if such errors originated in the data provided by the licensor. The licensor could argue that "but for" the licensing relationship, it may not have been exposed to such claims. Conversely, the licensor might indemnify the licensee against third-party intellectual property claims if the data is alleged to infringe copyrights or trade secrets, as the licensee may not have been exposed to such claims but for its use of the data.

Note that indemnity triggers can be either:

- Broadly written (e.g., indemnification for any claims "arising out of the Licensee's use of the data")
- Narrow (e.g., tied to a breach of a specific representation or warranty)

For more information on indemnification provisions generally, see Indemnification Provisions in Commercial Contracts.

Treatment of Data Post-termination

One of the most hotly negotiated terms of a data license is what happens to the data when the license ends.

For some kinds of data—particularly real-time streamed data, such as financial market data—termination of access to the data will be sufficient, as the data by its nature becomes stale and unusable quickly. For other kinds of data, its use may have become bound up in derivative works, rendering return of the data either impossible or extremely disruptive. In such instances, it may be acceptable for the licensee to retain deidentified, anonymized, or aggregated data, but return the original data set. In other instances, the licensee may need to return, cease using, or destroy all instances of the data (including any derived, resultant, and usage data). For additional considerations regarding such data, see "Derived and Usage Data" above.

Chad A. Rutkowski, Partner, BakerHostetler LLP

Chad Rutkowski's practice focuses on the intersection of copyright and technology. As co-lead of BakerHostetler's Digital Transformation and Data Economy team, he assists clients in the midst of the "digital transformation," helping them identify and capture the intellectual property (IP) in their valuable data, algorithms, and enterprise software. Clients rely on him to help identify their IP, to build internal processes for its management, navigate open source and other open innovation strategies, and to create licensing programs that enable full value realization. He further protects that value in disputes and litigation, whether through enforcing his clients' IP rights or fending off attacks by competitors.

Chad is the principal author of the firm's Beyond Source Code web tool, which digests all court decisions addressing the scope of copyright protection in the non-literal elements of source code. Chad writes and presents regularly on the robust protections afforded by copyright law to software, data, artificial intelligence, and related innovations. He has helped his clients protect innovations in healthcare technology, water management, smart cities, population health management, and online consumer lending.

Chad's interest in how IP law has adapted to digital transformation began in his former role as a founder and business manager of a digital media publisher. Drawing from his professional and legal experiences, he co-leads the firm's Technology & IP Transactions and Outsourcing practice team and contributes regularly to the firm's IP blogs. Additionally, he is an active member of the Copyright Society of the USA and the ABA IP Law Section. Chad is AV Preeminent-rated by Martindale-Hubbell and has been voted by his peers as a Pennsylvania Super Lawyer and Rising Star.

LexisNexis.com/Practical-Guidance