ABA Cyberspace Law Committee, Cloud & Enterprise Technology Subcommittee

CLOUD COMPUTING AND CLOUD MARKETPLACE TOOLKIT PROJECT

Data License Checklist

Candace M. Jones

A wide variety of material, including data, can be sourced in marketplaces sponsored by cloud service providers (and myriad other online sources). This checklist can be used to facilitate review of data licenses in any commercial setting. In the context of a cloud provider marketplace, getting access to communicate with the data licensor for negotiation may be challenging. Indeed, information about the dataset may be limited, complicating assessment of legal risk and fitness of the data for their intended business purpose.

Where information is limited or concerns cannot be addressed through direct negotiation, the client organization should pay particular attention to the intended use of the licensed data and its internal controls for managing the licensed data to the permitted use. Rule of thumb: Data that will be used internally for a limited purpose in a manner in which the data can be contained and readily destroyed and replaced (if needed) generally creates less risk to the licensee/user than data that are to be used as an integral component of an essential internal operation or distributed to customers.

Lawyers should keep in mind a common misconception that readily accessible published data may be freely used—without charge or other limitations. Not so. Be prepared to encounter this misconception in your client organizations.

With that background, review data license agreements with these ten points in mind. The first anticipates information about the transaction developed outside the license agreement. The remaining points relate to license terms.

No. 1: Data Licensor

The practices of the data licensor in collecting or creating the licensed data are foundational to the integrity of the data and the rights that the licensor purports to convey to the licensee (data user). The data licensee will need to determine whether or not the data are fit for the licensee's intended purpose.[1] The licensee will also have to consider whether the manner of collection or creation is consistent with underlying expectations of privacy (for personal information) and confidentiality (for business and other proprietary information). Use of licensed data by the licensee could be impaired if its licensor fails to comply with applicable law or upstream license agreements and consents.[2]

The proliferation of data collection channels offers intriguing new material for study and analysis. Companies rushing to capitalize on those new channels may not be as diligent about obtaining or documenting the provenance of the data or consents obtained in their collection process. Synthetic data, which are generated (created), not observed, may raise considerations about quality/veracity or bias, for example, that the licensee will want to have clearly documented.

---

[1] *Fit for purpose* provides context for assessing "quality." Just like beauty is in the eye of the beholder, the quality or condition of a dataset tends to be in the eye of the user. For new data licenses, the project team may be well-advised to consider a trial license or similar arrangement offering those who need the data an opportunity to review and test them.

[2] Consider the decision of the Federal Trade Commission ("FTC") in *In re Everalbum, Inc.*, FTC Docket No. C-4743 (May 6, 2021). Everalbum used images of customers' faces for purposes beyond the scope of customer consents. Among other things, the FTC ordered Everalbum to destroy work product Everalbum created using those images.

In short, the client's business and legal representatives should have a baseline understanding of the data to be licensed, the data's source, and the purpose for which the data are to be used.

No. 2: Users

Who needs access to the licensed data and for what purpose? These client expectations need to inform the license definition of *permitted users*. The universe of permitted users could be as narrow as named individuals or as broad as anyone employed or otherwise working in the client enterprise. Pay attention to affiliates and service providers that may need access to data. For example, an organization that outsources its information technology ("IT") function should be covered to the extent that licensed data are stored or processed in systems supported by the outsourced IT service provider. Also, consider the need for control functions, like legal and audit, to have access to licensed data.

Practice Note: As a model for allowing roles for which limited access may be required, consider exceptions that the client organization typically seeks to include in confidentiality provisions it encounters. For potential users outside the organization, see the note below about "distribution."

No. 3: Uses

Permitted uses may also be broadly or narrowly defined. Consider both currently planned uses and foreseeable uses. The data licensor may also specifically prohibit certain uses or impose conditions on use. For example, as machine learning projects picked up, some data licenses evolved to expressly prohibit the use of licensed data as training data for machine learning applications.

General Conditions

If the licensor imposes conditions on use, can the licensee manage those conditions effectively? For example, if licensed data may only be stored in and accessed from the United States, is that restriction consistent with the intended use of the data and the infrastructure in which the data will be stored and processed? Do license terms purport to restrict material created by the licensee in which licensed data are used? Managing licensed data and the material in which they are incorporated is a continuing obligation; the licensee needs to monitor for changes in systems and uses that could run afoul of license conditions.

Publishing

Except for public domain data, most data licenses limit publication to preserve the market for the licensor's data product. While some licenses prohibit publishing altogether, many permit some publication of licensed data subject to conditions that limit the amount or format of the data published.

Limits on the amount of data tend to be stated imprecisely, such as an amount not so substantial in quantity or character that the published data would substitute for a direct license from the licensee. The license may also permit periodic, ad hoc publishing but specifically prohibit regular, systematic publishing. Similarly, the license may permit publishing in a static format (like a pdf file) but not in a dynamic format (like a spreadsheet or comma-separated values ("CSV") file from which data can be readily extracted for reuse).

Publishing may also be subject to attribution or nonattribution requirements and disclaimers. Some licensors insist on being identified as the source of licensed data. Some insist on no attribution for the source of data. Be wary of agreeing not to identify the source of licensed data. There are many reasons an organization may want or need to identify sources of data that it uses, including reporting to regulators and compliance with laws or commercial practices that promote transparency.

When the licensor perceives risk of claims from a downstream data user—for example, a user of financial information—the data licensor may insist that the information be published with a disclaimer of the accuracy or completeness of the data and any liability for use of the data in making financial decisions.

Distribution

Any regular or systematic distribution of licensed data is generally beyond the scope of typical data license agreements. In fact, regular or systematic distribution is usually prohibited expressly. If regular distribution of any quantity of licensed data is planned, the rights and obligations regarding distribution should be covered in the license or a separate distribution agreement.

No. 4: Derived Data/Output

Licensed data are often an input for a calculation, analysis, aggregation, or other manipulation (change) that transforms the licensed data for the licensee's intended use or product. The licensee-created data are referred to with labels such as *derived data* or *output*. Licenses sometimes reserve broad rights for the licensor, including the right to have or "own" derived data. Licensor claims to derived data will almost always be showstoppers.

On the other hand, it is reasonable for licensors to require that derived data not be a ready source for others to get to the underlying licensed data. For this reason, the definition of *derived data* often includes an element that the derived data cannot be reverse engineered to extract the underlying licensed data.

No. 5: Noninfringement

Has the licensor warranted that it has the right to license the data for the permitted uses free of claims by third parties? It should. This warranty addresses in a general way the privacy and confidentiality considerations noted above in item No. 1. If a licensor refuses a noninfringement warranty, pay extra attention to the indemnification provisions. The licensor should indemnify the licensee from third-party claims alleging infringement, misappropriation, or noncompliance with privacy obligations or other consents.

No. 6: Term

Consider the duration of the license from both a project budget perspective and an operations perspective. If the licensee intends to use data in a manner that will continue beyond the stated term, then opportunities to renew (with some price protection mechanism, if possible) should be considered.

Also pay attention to the notice period in renewal or nonrenewal and termination provisions. Relatively short notice periods for renewal or termination, e.g., thirty days, may be consistent with other commercial contracts. Is that a reasonable period for the licensee to wind up its use of the licensed data? The notice provisions related to term and termination should take account of the intended uses of the data.

No. 7: End-of-Term Obligations

When the license ends, what happens to the licensed data? Storage and processing need to be managed with end-of-term obligations in mind. Data licenses often require licensed data to be purged at the end of the term. The licensee should consider the extent to which it may need to archive licensed data for regulatory or audit purposes or to support or validate work done with the licensed data.

Practice Note: As a model for tempering purge requirements, consider exceptions the client organization typically seeks to include in confidentiality provisions that it encounters: "Upon termination, Recipient must return or destroy confidential information in its control except [list of exceptions such as legal requirement to retain, impracticality of purging from backup files, etc.]."

No. 8: Handling/Confidentiality

Licensees will be responsible for controlling data consistent with the parameters of the license, e.g., access only to authorized users for permitted uses. Beyond that, data licensors may impose specific

handling or security requirements, such as geographical restrictions on storage and processing. Any specific handling requirements should be vetted with the project team.

Review confidentiality provisions carefully. If the licensed data are included in the definition of *licensor confidential information*, any publishing and sharing permitted by the license should be carved out of confidentiality restrictions to avoid internal conflict in the license terms. Similarly, if the licensee is required to provide the licensor any confidential information, including, for example, excerpts of derived data/output or downstream products, confidentiality obligations should be reciprocal.

No. 9: Warranties

Warranties? More realistically, licensees will find broad disclaimers of any warranty, including the accuracy, completeness, and ongoing availability of the data. As noted above, licensors should be pressed to warrant that they have the right to license the data free of third-party claims.

Short of a warranty, licensors may undertake to provide information that can help licensees manage any errors or changes in the data. Those practices may be set out as contractual obligations in the license; more often, they are not. Assistance with data quality questions and issues is often left to customer service practice as time and resources permit.

Research about such notices and support practices can be valuable even if the licensor does not agree to undertake them as obligations under the license. For example, does the licensor notify licensees periodically about data errors that have been corrected in the dataset? Does the licensor provide support—that is, access to someone who can answer questions about the data and chase down potential errors or gaps? If the licensee notifies the licensor of a potential error or gap, does the licensor say it will undertake to investigate and respond or at least implement a standard procedure to monitor for errors and gaps? If the licensor has a break or change in a data series, does the licensor have an obligation to provide prior notice and explanatory material?

Practice Note: Similar kinds of remedial provisions have long been included in software licenses. The support and problem reporting/response provisions in software licenses may be useful models for thinking about support for licensed data.

No. 10: Indemnification

A common compromise makes (a) the licensor responsible for indemnifying the licensee against third-party claims for infringement of proprietary or personal (privacy) rights and (b) the licensee responsible for indemnifying the licensor against third-party claims that the licensee used the data in a manner that exceeds the license scope and for harm caused by the licensee's downstream products, services, or derived data.

As with all indemnification obligations, the wherewithal of the indemnitor factors heavily in the weight that can be placed on the provision. Evaluating risk embodied in the indemnification provision should be informed by the information-gathering exercise referenced in tip No. 1 above.