Towards Best Practices for Licensing Data

By

Gillian M. Fenton, Esq., CLP
Executive Director
LST Strategies LLC, Washington DC, USA
gillian@lifesciencetransactions.com

ABSTRACT

Data as an asset class, and data licensing, are of increasing importance as we enter the age of artificial intelligence (AI). This paper explores the challenges and unique issues presented to the licensing professional when approaching a license of human biomedical data. Covered topics include intellectual property paradigms potentially relevant to data; restrictions on data use, including restrictions arising from informed consent forms in clinical trials; ingestion of biomedical data by AI models; privacy concerns; and an array of potential license clauses that are unique to data licensing. As the field is evolving rapidly, there is as yet no settled consensus on what constitutes best practices for data licensing – this paper offers thoughts on the stepping stones along the path to best practices.

Introduction

Biomedical data and data sets are incredibly valuable assets to their owners and developers, but are not formally recognized in the United States (or in many other countries) as a form of intellectual property (IP). Hence, the contractual provisions of data license (or data use) agreements form the first line of protection and certainty for determining the rights of data owners (licensors) and data users (licensees). While principles of contract and intellectual property laws constitute one layer of consideration for business strategies regarding data license agreements, there is also a complex body of laws and regulations concerning the generation, storage, use and disposition of biomedical data that must also be considered. Understanding the legal environment of data licensing is becoming increasingly important with the advent of artificial intelligence (AI) and the exponentially increasing applications of AI in a wide range of fields - from basic biomedical research to diagnosis/prognosis to image interpretation to modeling biological phenomena to biomanufacturing to personalized medicine, with many other possibilities in development (numerous use cases are reviewed in Torkzadehmahani et al., 2022).[1] This paper will explore some of the issues arising when negotiating and drafting data license agreements in life science industries.

Some organizations have invested in developing guidelines for standardizing data sharing agreements so as to foster collaboration and innovation.[2] However, experts have noted that a single data sharing agreement template is unlikely to be agreed upon, given the wide diversity of potential data applications across multiple industry sectors. Nevertheless, development of best practices around certain structural aspects – which inform a number of agreement clauses – is beneficial and facilitates transactions (and therefore, collaboration and innovation). Organizations that have developed guidelines and/or templates include the Linux Foundation, Microsoft, Responsible AI Licenses, Open Data Commons, Creative Commons, Japan Patent Office, the United Kingdom government, and the Singapore Information Media Development Authority.[3] However, these resources may not be suitable for biomedical data. NCATS (National Center for Advancing Translational Sciences) hosted a Biomedical Data Translator project[4] seeking to standardize biologic and biomedical data sets so as to facilitate their use and re-use by the research community.[5] This initiative was conducted primarily by scientists, not licensing professionals, and favors so-called permissive licensing, which covers

[1] Privacy-Preserving Artificial Intelligence Techniques in Biomedicine - PMC (Torkzadehmahani *et al.*, 2022).

[2] intellectual-property-expert-preliminary-report-on-data-and-AI-model-licensing.pdf (The Global Partnership on AI, November 2022).

[3] *Ibid.*

[4] https://ncats.nih.gov/translator (National Center for Advancing Translational Sciences).

[5] An analysis and metric of reusable data licensing practices for biomedical resources | PLOS One (Carbon *et al.*, PLOS One 14(3), 2019).

only permissions for reuse, transformation and redistribution in return for only attribution to the data source.[6] Notably, permissive licensing does not require business negotiation or the advice of legal counsel, and it does not realize financial value from the data. Overall, the NCATS project illustrates a lack of licensing sophistication in the biomedical science community.

Data Licensing

Data licensing has been an enduring interest of the IP licensing community, cutting across the life science and high-tech sectors.[7] However, while many licensing professionals are familiar with customary frameworks for patent licensing in their own industries, practices in different industries can be unfamiliar. Furthermore, license agreements are used in a wide variety of contexts, from product development to settlement/freedom to operate. It is not surprising in light of these factors that we may struggle to adapt to the novel issues presented by data as an "IP-like" intangible asset class – and to the even more specialized issues arising from biomedical data. Data and data sets are unlikely to meet the criteria for patentability so principles of patent licensing are ill-suited for use in data licensing agreements. Instead, data are similar to unique biological material in that data may be protected as trade secrets, with corresponding types of protection appearing in the few data licensing agreements that are currently publicly available. Copyright may also be relevant, depending on the complexity and configuration of the data set. A blend of licensing clauses derived from trade secret and copyright laws may therefore be an appropriate solution for data licensing. Given these adjacencies, I will explore the use of trade secret and copyright principles as applied to data.

To protect valuable data as a trade secret, just as for other types of trade secrets, it would be prudent to use a combination of administrative, legal, and technological controls,[8] and to clarify to employees that the data are considered confidential and proprietary information. Governing principles are found in the Uniform Trade Secrets Act as adopted by various states;[9] the Economic Espionage Act;[10] and the Defend Trade Secrets Act.[11] Similarly, Europe provides sui generis protection for databases.[12] Some courts have extended tort law concepts originating in harms to tangible property to data (for example, trespass to chattels and conversion).[13] Website scraping may be asserted as trespass via use of a computer system without, or in excess of, authorization, resulting in damages. Conversion may be asserted via unauthorized taking of information on computers/websites. Such claims may invoke the Computer Fraud and Abuse Act[14] (CFAA), which covers unauthorized access of computers used in interstate commerce. However, not all courts have so extended tort law or the CFAA.

Copyright, in comparison to trade secret and tort law, will not protect individual data points but could be relevant to a data set as long as there is at least a modicum of creative expression in the data arrangement and it has been fixed in a tangible medium. Copyright may also be a useful fallback or 'gap filler' for situations where the data are not confidential or trade secret protections cannot be implemented. However, the applicability of copyright protections to the use of data for training AI models is currently an area of controversy and the law is likely to continue to evolve. At least 40 lawsuits have been filed on whether the use of data to train AI models is an infringement of copyright, or is within the fair use defense, with mixed outcomes to date. It is therefore timely that the U.S. Copyright Office has released a report on this topic.[15] The punchline on use of copyright law to protect biomedical data in the context of AI/algorithm training is that licensing practitioners should be wary and should keep abreast of new developments in the law. Hence, data licensing agreements that rely exclusively on copyright type protections are not recommended.

Of course, licenses covering human-derived biomedical data must also implement privacy law concepts. The approach used for preserving privacy in a human-derived data set will dictate the mechanics of how the data is used and accessed; this is reviewed in Torkzadehmahani et al., 2022.[16] Presently, there are three potential approaches to privacy preservation: cryptography (homomorphic encryption and secure multiparty computation), differential privacy, and federated learning. These techniques vary in terms of whether the data stay with the owner and are accessed remotely by the licensee, or the data are transferred to the licensee for use/analysis and then destroyed after use. Cryptography approaches "bring the data to computation" whereas federated learning approaches "bring computation to the data." Differential privacy is a means of introducing randomness, even errors,

[6] *Ibid.*

[7] Viewpoints: Thinking Outside The Lab: Healthcare Data, Licensing and AI - LES USA & Canada (Chitra Kalyanaraman, 2023).

[8] For example, training requirements, physical and electronic access barriers, and security procedures.

[9] Trade Secrets Act - Uniform Law Commission.

[10] 18 USC § 1839(3).

[11] 18 USC § 1836.

[12] EU Directive 96/9/EC.

[13] Data as IP and Data License Agreements (1).pdf (Glazer, Lebowitz and Greenberg, Practical Law Practice Note #4-532-4243 (2017)).

[14] 18 USC § 1030.

[15] Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-Publication Version (May 2025), summarized in 5 Takeaways from the Copyright Office's Report on Generative AI Training | Copyright Alliance (Blog post by Rachel Kim, May 29, 2025).

[16] *Supra,* fn. 1.

to the data for purposes of stymieing attempts to re-identify data subjects. Torkzadehmahani et al. also discusses the practicalities of "computing overhead," i.e., the burden incurred by performing AI analysis of the data. It appears that no universally applicable method has emerged representing best practices for preserving privacy while training AI on biomedical data sets; some use cases will be most amenable to one of the foregoing approaches, whereas others will require a different approach, or even a combination of approaches. It may even be the case that a synthetic data set should be generated on the basis of the real data set, thereby protecting the real data from being accessed in the course of performing AI investigations. This complexity means that the licensing professional should enquire about the use case, including how the analysis will be performed and if it will require a transfer of data. The data protection provisions of the license agreement should be customized to fit the circumstances.

Licensing Considerations

The licensing professional approaching a data license should consider what activities will be carried out by the licensee – and what activities should not be carried out – and construct the data licensing agreement to cover the corresponding topics, including obligations, duty of care, and risks – especially risks that are unique to data. In one hypothetical example, the parties enter into a licensing agreement in which the licensee will take over development of a clinical-stage biologic drug. The clinical data package needs to be transferred to the licensee – the parties should specify the details of data delivery, including formatting, encryption type, data file type(s), and level of security of the portal or other platform utilized for data transfer. The licensing team will need to consult with regulatory affairs and the IT teams of both parties. Critically, the party drafting this license agreement must consult the Informed Consent Forms (ICFs) under which clinical trial subjects participated in clinical studies that gathered data from them.[17] The ICF will dictate key restrictions on permitted use(s) of the data, and the time period in which such use(s) can be carried out. Thus, the license should specifically define the licensee's permitted purpose for use of the data,[18] and obligate the licensee to restrict its use to the specified purpose. Upon data transmission, the receiving party should confirm receipt and useability of the data; this would be a good milestone event for payment purposes. Data may be analyzed or processed (implying transformation) during the course of a permitted use, which may produce derivative data. The licensing professional should also consider whether and/or to what extent the data may be used to train an algorithm or model. Data processing (with or without AI) may include a number of specific types of activities, such as reproduction, distribution, communication (e.g., to the public), adaptation, modification or combination with other data/information.[19] Data processing in the context of an AI use case may differ from general data processing.[20] The data license agreement should specify which party owns the outputs of data processing, and whether ownership is subject to any ongoing obligations to the licensor (and to the clinical trial participants). Finally, the agreement should state a protocol for handling and disposal of the data once the agreed upon purpose has been achieved, or the time period specified for use has run out. Because our example data license is for the purpose of completing development and commercialization of a biologic drug, other milestone and payment terms can reflect the traditional approach of aligning financial consideration with progress on the regulatory path to drug licensure, followed by commercialization.

There are some publicly available data and/or AI license templates available on the internet that are more data-focused and general than the foregoing example, but most of these are not focused on the specific challenges of biomedical data. One notable exception is the National Institutes of Health (NIH) Data Use Certification Agreement or DUC.[21] The DUC is organized to track with procedural steps taken by the parties: after referencing the NIH Genomic Data Sharing Policy,[22] it describes an application process by which the prospective licensee submits a data access request including a description of the proposed research to be performed using the data. Interestingly, the data access request is submitted jointly by the licensee's principal investigator and IT director and sets out attestations of compliance with privacy and other laws, the licensee organization's IT environment for data security (including whether cloud computing will be used) and other factors corresponding to representations and warranties. The DUC requires the licensee to limit accessibility of the data to specific personnel, proscribes actions that entail risk of identifying data subjects, and endorses the use of Certificates of Confidentiality.[23] The term of a DUC is limited to one year, and the license is not assignable or transferable. Publication of scientific results derived from use of the data is encouraged; however, the DUC lacks robust terms for patent and other intellectual property protections, since the NIH goal is to preserve accessibility and availability

[17] For general guidance on research practices involving human subjects, see Federal Policy for the Protection of Human Subjects ('Common Rule | HHS.gov.

[18] Data Licensing: Best Practices for Licensing your Business Data | BRION RAFFOUL LLP - IP Lawyers (Blog post by Gray and Tyhurst, January 21, 2025).

[19] End User License Agreements | NASA Earthdata (August 28, 2025).

[20] *Supra*, fn. 18.

[21] Universal_DUC.pdf (National Institutes of Health (NIH), 2023).

[22] NOT-OD-14-124: NIH Genomic Data Sharing Policy (August 27, 2014).

[23] Certificates of Confidentiality (CoC) | Grants & Funding (August 10, 2024).

of scientific advances. The NIH DUC also lacks any financial terms. It is therefore up to individual private sector contracting parties negotiating data licenses to develop views as to the financial value of the data, of the use case under consideration, and of what value inflection points are most relevant for financial terms.

New Concepts and Definitions

Separately, negotiators and drafters should consider the following key concepts and definitions for incorporation as appropriate into biomedical data license agreements. The categories and list below are of course not exhaustive, but should serve as a useful starting point.

- Original Data is the specific data or data set to be provided by the licensor. In the case of data from human subjects, it may be necessary to define the original data as Source Data that is processed by the licensor or an independent expert retained by the licensor to deidentify the data subject/source. This can be accomplished using Certificates of Confidentiality and following the process defined therein. The resulting anonymized or pseudonymized data set could then be defined as Licensed Data or Input Data. This process is one way of transforming the Original Data into a synthetic data set.[24]
- Conversely, the data resulting from processing by the licensee using an AI or other computational model or algorithm could be defined as Results, Processed Data, or Output Data. A critical question is: can the Original Data be traced or recreated from the Output Data? The acuity and consequences of this risk should drive the choice of privacy protection methodologies discussed above.[25] One possible solution, which adds a further layer of privacy protection, is to limit the licensee's rights in Output Data to commercialization of downstream products from which the Original Data cannot be derived or extracted.[26] If treated as under copyright law, Derived Data would be a derivative work, subject to rights of the licensor.
- The AI or other algorithm deployed for data processing (analysis) is often called a Model. Initially, the Model is an Untrained Model completely separate from the Input Data, but following processing the Model may be transformed into a Trained Model which may or may not be separable from the Input Data and/or the Output Data/Results. Ownership and control of the Model versus the Trained Model therefore needs to be thought through carefully in light of the specific circumstances.[27]
- In the case of generative AI or Large Language Models (LLM), there is an intermediate stage that also should be considered when developing an ownership framework: the Prompt crafted by the licensee and used to query the Model.

Data Ownership (Custody) and Use

- The different types of Data should be specifically defined. The licensor should require the licensee to acknowledge in the license agreement that the Original or Input Data are owned by the licensor and are valuable to the licensor (thereby invoking trade secret law, assuming that the data are confidential). If the data set has been collected, compiled, or arranged by the licensor in a way that supports the application of copyright protection, especially if the data have been published, the licensee should acknowledge that circumstance.
- Next, the licensee should consider the current state and provenance of the Input Data set: does the licensor maintain the data in a secure manner, in compliance with cGXP requirements[28] as applicable? Perhaps the most critical element of provenance of human-derived biomedical data is the Informed Consent Form (ICF) under which the data were originally collected. The nature and scope of permissions granted by data subjects will constrain all downstream activities performed with or on the data. The same is true for human biological samples. Best practices may require attaching a generic copy or template of the ICF or of the portion relating specifically to permissions granted for use of the data and/or samples to the license agreement in future research, development, and/or commercialization.
- Ownership of data would be expected to follow the principles of personal property ownership, such that the developer or originator would have title and therefore the power to alienate title. This implies that the default approach would be that the licensor owns the Original Data / Input Data and the licensee owns the Results / Output Data, however in actual practice this is not straightforward. For example, there are heightened sensitivities in the case of data (and biological samples) derived from human subjects, who have increasingly recognized legal rights under a range of U.S. and foreign statutes

[24] EX-10.24 (Data License Agreement of February 28, 2017 between eRx Network, LLC and Change Healthcare, Inc., available via the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database maintained by the Securities and Exchange Commission) provides a good example of a HIPAA compliant procedure for de-identification of healthcare data.

[25] It is critical to note that, in licenses entered into before the age of AI, both parties may be exposed to unacceptable risks of loss of confidentiality and/or breach of privacy laws if the agreement is not sufficiently robust to address the currently understood risks of applying AI. If new AI use cases are desired, it would be prudent for the parties to review their legacy agreement carefully and consider an amendment updating its provisions to reflect current best practices.

[26] *Supra,* fn. 19.

[27] *Supra,* fn. 2.

[28] GxP Compliance (University of Miami Office of the Vice Provost for Research and Scolarship; GxP is a shorthand term collectively denoting Good Manufacturing Practices (GMP), Good Laboratory Practices (GLP) and Good Clinical Practices (GCP) with the leading lowercase "c" indicating "current"). While not an official source, this webpage provides a succinct resource for all GXPs.

and regulations.[29] Types of human data may include data arising from clinical trials, diagnostics or other assays, genetic or genomic data, performance metrics from wearable or medical devices, etc. It may be preferable to characterize the licensor's rights in human data in terms of custody rather than ownership, given the recent evolution of privacy laws, and the agreement may need to address the licensee's obligations if custody is transferred - the licensee should agree to comply with restrictions in the ICF as well as other restrictions imposed by the licensor and by applicable law. An example would be the obligation to remove data derived from a specific human subject if that subject revokes consent or otherwise requests deletion of their data. Wolf et al.[30] propose a novel solution to this conundrum: they propose instituting a decentralized electronic system modeled on nonfungible tokens (NFTs) or blockchain to preserve confidentiality of the human data/biomaterials donor yet permit the donor to continuously monitor use of their materials in subsequent research.

- Human data may include or implicate Personally Identifiable Information (PII) and/or Personal Health Information (PHI), such that a range of privacy laws, reviewed in Glaser et al.,[31] will attach: the Gramm-Leach-Bliley Act, HIPAA, Health Info Tech for Economic and Clinical Health Act (HITECH), Fair Credit Reporting Act, Children's Online Privacy Protection Act, relevant state laws (California, Connecticut, and an increasing number of others), GDPR[32] in Europe and its counterpart in the United Kingdom (UK),[33] and PIPEDA[34] in Canada. These laws may impose obligations and restrictions on data even if the license agreement is silent, so it behooves the licensing professional to research relevant laws and regulations and draft the license agreement in terms that are not rendered void or unenforceable by applicable law. For example, the UK Information Commissioner's Office (ICO) recently found that 23andMe failed to require strong passwords, to require unique usernames rather than email addresses, to require multifactor authentication (especially when downloading raw genetic data), to fingerprint devices used for logins, to notify customers when accounts were accessed from new/unrecognized devices, and to provide access to login histories. Further, the ICO determined that 23andMe should have conducted breach simulations or penetration tests and taken steps to address vulnerabilities.[35] All of the foregoing should be considered for inclusion in data licensing agreements.

Restrictions on Access To and Use of the Data

- Critically, the data license agreement should specify a standard of care with respect to stewardship of the data, addressing issues of integrity, security, privacy and confidentiality. For agreements between U.S. entities, reference to the Cybersecurity Framework[36] promulgated by the National Institute of Standards and Technology (NIST) and to a minimum acceptable tier of organizational compliance may be an appropriate solution. For cross-border agreements, the equivalent citation would be to ISO/IEC 27001[37] with specification of a minimum level of implementation. Importantly, both the NIST and ISO standards provide procedures to be followed in the event of a data breach, along with obligating the parties to report actual or suspected breaches to each other and to cooperate on identifying a root cause of the breach and remedying the same. This obligation should attach to all types of electronically transmitted, transmissible, and stored data – not just to biomedical data or human data, although storage and processing of human health data can dramatically increase risk exposure.[38]
- The licensor should secure the right to audit the licensee's systems for data stewardship, including physical access barriers, administrative, legal and electronic systems. If human biological samples are also being licensed or otherwise transferred, audit rights should extend to the systems utilized for the samples as well, and the data license agreement should obligate the licensee to institute and maintain a system, such as a logbook, for traceability of the samples. The logbook should include an inventory of the samples as received, along with their utilization, consumption and disposition. Storage and shipping systems should provide a historical record of any deviations from recommended conditions, such as temperature.
- For both privacy and trade secret reasons, the data license agreement should prohibit disassembly or decompiling (reverse engineering) of the data. Especially in the case of anonymized, pseudonymized,

[29] Disrupting the biospecimen "treasure trove": Practice, precedent, and future directions | Science, Wolf *et al.* (21 August 2025), 389 *Science* 784-786 (Issue 6762).

[30] *Ibid.*

[31] *Supra,* fn. 13 (this reference cites the U.S. statutes listed above).

[32] General Data Protection Regulation (GDPR) – Legal Text.

[33] The UK GDPR | ICO.

[34] *Supra,* fn. 16.

[35] 23andMe Fine Signals ICO's New GDPR Enforcement Focus - Law360 (Churchill and Gibbs, July 30, 2025). For a U.S. legal perspective on 23andMe and proposed best practices for protection of human genetic data, see Ram *et al.* (September 11, 2025) 389 (Iss. 6765) *Science* 1092-1094 (The precarious future of consumer genetic privacy | Science).

[36] Cybersecurity Framework | NIST.

[37] ISO/IEC 27001:2022 - Information security management systems.

[38] See Illumina To Pay $9.8M To Resolve Cybersecurity Qui Tam Case - Law360 (Konnath, July 30, 2025), which illustrates the need for parties in control of, or processing, data to have adequate cybersecurity protections.

or otherwise deidentified human data, all activities that could intentionally or inadvertently cause recompiling to discover the data subject's identity should be prohibited.

- The data license agreement should include a confidentiality clause. Terms for confidentiality should be crafted to fit the specific circumstances and risks, notably the use of AI.[39] This means that the license drafter should consider whether uploading data to an AI, such as a Large Language Model (LLM), constitutes a disclosure or entails a risk of disclosure. It may also be impossible to retrieve or destroy data that has been processed using an LLM, raising the need to caveat any otherwise applicable obligation to return or destroy the data upon termination or expiration of the license agreement. A conservative approach to this risk would be to prohibit the use of AI for confidential data (in which case the data license agreement should set out a clear definition of "AI"); if processing via AI is essential to the purpose of the data license agreement, the drafter should consider requiring the licensee, its vendors (if any), and its sublicensees (if any) to disclose to the licensor all AI tools that will be used. This provides an opportunity for the licensor to consent or reasonably withhold consent to the use of specific tools.

Reps & Warranties Specific to Data and its Uses

- While it is customary for licenses to contain a representation that the licensor has title to the licensed subject matter, or at least the ability to grant the rights and licenses in the agreement, it is critical for a licensor of data to have a clear and verifiable chain of title to (or custody of) the data from its origination up to the effective date of the data license.

- Other representations and/or warranties to consider include those of data quality, consistency, interoperability, accuracy, and freedom from bias. In each case, the licensing professional should devote considered thought to defining each term for the context of the license agreement and the data use case. For example, what does data 'quality' mean? No standard definition has been developed to date. It may mean reliability, relevance, representativeness, completeness, lack of harmful bias, being free of third-party IP rights, or something else. These considerations point up the need for 'data quality' to be defined in relation to the purpose of the agreement.[40] Another representation to consider, where the data will be used in the course of seeking U.S. Food and Drug Administration (FDA) approval of a regulated product, is that the data are compliant with relevant cGXP standards and other FDA requirements.[41]

- Conversely, the parties to a data license should carefully consider data-specific disclaimers. The broadest being that the data are provided 'as is' or with 'no warranty of any kind.' If this approach is unacceptable, the parties may consider whether to exclude specific representations and/or warranties, such as those of accuracy, completeness, authenticity, usefulness, timeliness, reliability, and appropriateness.[42] Other disclaimers that are more familiar to those in high tech industries but will perhaps be unfamiliar to life sciences professionals include statements that the data are free of bugs, errors, defects or omissions.[43]

Liabilities & Indemnities and their Limits

- Well-drafted data license agreements should also include provisions for indemnities that are specific to data and its use. A classical approach is to consider what types or classes of third parties are relevant, and which of the parties is best positioned to handle risks arising from each class. The licensor would usually be better placed to manage risks arising upstream of the license, including claims made by data subjects prior to the effective date or not arising from actions taken by the licensee. Other risks that are best shouldered by the licensor may include claims made by prior data controllers or contributors. Conversely, the licensee is likely to be better positioned to handle risks arising downstream of the licensing transaction, including claims arising from the licensee's actions and its collaborators and customers.

- Data use cases involving AI are particularly challenging: AI-unique risks include hallucination, inaccurate or made-up results, defamation, revelation of source data, and more. Yet the licensee may not fully appreciate how an AI such as an LLM processed the data to produce undesirable results. A recent article by Belmont provides a thoughtful analysis of the fundamental incompatibility of the law of indemnity (based on causation) with AI (based on probabilities).[44]

- Exclusions or limitations of liability should also be considered. The parties may agree to exclude liability for the supply of erroneous data, disruptions in

[39] 7 AI-Specific Confidentiality Clauses (*Contract Nerds* Blog post by Heller, August 12, 2025).

[40] *Supra*, fn. 2.

[41] 7 Considerations For Conducting Drug Clinical Trials Abroad - Law360 (Berman *et al.*, May 6, 2025).

[42] *Supra*, fn. 13.

[43] *Supra*, fn. 19.

[44] The AI Output Problem: Rethinking Indemnity in the Age of Generative AI (*Contract Nerds* Blog post by Belmont, July 15, 2025).

data transmission, low quality interpretative work, or destruction/loss or alteration of data that may potentially cause damage to a business or otherwise.[45] Another exclusion to consider is misuse or unlawful use of the data by the licensee. While not limited to data licenses, prudent licensors should negotiate for the licensee to maintain liability insurance coverage, with the licensor as a named insured.

Valuation and Financial Terms

At present, very little publicly available information (scholarly or in the form of disclosed agreements) exists to guide licensors and licensees as to the value of biomedical data or guidance for structuring the financial terms of data license agreements. It therefore behooves licensing practitioners to carefully consider the relevant market, benefits arising from use of the data by the licensee, and other factors relevant to valuation. Indeed, it may be necessary to conduct a valuation of the data as an asset and/or of the use case that is the subject of the license agreement. A discounted cash flow analysis based on data-enabled product development may be appropriate, and other potential approaches to valuation should be considered. As noted above, the paucity of publicly available data licenses may constrain the use of a market/comparables approach to valuation. As far as financial terms go, payments to the licensor should be keyed to data milestones reflecting success in the transaction, such as receipt of the data and confirmation of its accessibility/interrogability, completion of a study utilizing the data, regulatory filings for a product based on use of the data, and other financial terms customary for product development licenses (if applicable).

Summary & Conclusions

In our current knowledge and AI driven economy, data has emerged as a novel asset of significant value. Yet, data does not fit neatly into any of the classical categories of IP rights, necessitating creativity on the part of those negotiating and drafting data license (or use) agreements. As there is presently a dearth of publicly accessible agreements or guidance to use as precedent, this paper sets out some thoughts on the approach to data licensing that hopefully may lead the field to develop best practices for data licensing.

ACKNOWLEDGEMENTS

The author would like to thank Ian DiBernardo and James Malackowski for critical review of the manuscript.

[45] *Supra,* fn. 2.