Professional Perspective

Data Issues in Life Science Collaborations

Megan R. Baca and Georgina Jones Suzuki, Ropes & Gray

Bloomberg Law

Data Issues in Life Science Collaborations

Contributed by Megan R. Baca and Georgina Jones Suzuki, Ropes & Gray

The life sciences industry is entering a new era of collaboration deals with the onset of big data usage in recent years. Biotechnology, pharmaceutical, university, technology, insurance, private equity, and other players are increasingly collaborating on data-centric deals, as they seek to leverage insights from data to improve drug discovery, patient treatment, and clinical decision support software.

While data-centric life science collaborations are not fundamentally different from traditional life science collaborations, deal lawyers will need to address a number of legal considerations that are unique to them. This article provides a snapshot of the growing interest in data-centric life science collaborations and a checklist of key process and contract drafting points for the business development and legal professionals structuring these deals.

Market Landscape

The life sciences industry has seen an explosion in data-centered life science collaborations as more companies leverage artificial intelligence (AI), machine learning, bioinformatics, and other technologies to extract novel insights from big data.

Many trends are driving interest in data collaborations:

- Technology companies have increasingly entered the life sciences space as they partner with pharmaceutical companies to leverage data collected from wearables, mobile devices, and other technology to find commercial applications for their data.
- Analysis of vast quantities of genetic data has also skyrocketed as improvements in whole genome sequencing technology have made genetic analysis easier and more cost-effective.
- Many data collaborations grew out of the Covid-19 pandemic as institutions sought to quickly understand a novel disease and identify treatments.
- AI technologies progressively enable researchers to process data more quickly and accurately than humans. Report Linker currently estimates the AI in life sciences market to be valued at over $1 billion in 2020 and expects that market to grow to close to $5 billion by 2025.
- With growing interest in precision medicine to improve clinical decision-making, data will continue to be at the forefront of life science collaborations.

These and many other new innovations and applications will create opportunities for disruption in the market, and lawyers will need to be prepared for advising clients on such matters.

Drafting Considerations

Given the proliferation of data-centered life science collaborations, it is increasingly important that deal lawyers understand the legal issues unique to these types of arrangements.

Life science collaborations involving data do not involve profoundly different approaches than those in traditional life science collaborations. However, transactional lawyers should keep a number of process and contract drafting points in mind, particularly with respect to intellectual property (IP) and confidentiality.

Below is a checklist for business development and legal professionals involved in structuring life science collaborations involving big data assets.

Deal Structure

First, consider the collaboration structure. Parties should consult early with counsel to evaluate the advantages or disadvantages of any particular deal structure. While traditional life science collaborations through direct contractual licenses are common, joint venture structures are also used, particularly where parties wish to restrict uses of data to the joint venture entity due to privacy or regulatory concerns.

Data trusts are also becoming more prevalent. Under this structure, multiple parties can pool their data into a trust, which serves as a fiduciary for the data providers and oversees the proper use and sharing of the data. Data trusts can be used to overcome regulatory and ethical issues in sharing and processing of data, while fostering data interoperability between parties.

Definition of Data

A comprehensive definition for data should be included in the agreement. Unfortunately, there is no one size fits all approach to drafting a robust data definition. A data definition where a biotechnology company is providing genetic samples for data sequencing will look very different from a deal where an insurance company is contributing underwriting, health, and morbidity data to glean new insights.

It is helpful to build into a definition a broad description of data, such as "scientific, technical and other data" and then list out particular subcategories of data that may be relevant—a similar approach to standard definitions of the term "confidential information." Such subcategories may include, for example, computational validation data, genetic data, in vitro or in vivo data, biological or chemical data, pharmacological or toxicological data, manufacturing data, clinical test and quality control data, safety data, and so forth.

Parties might also consider cross-referencing a schedule in the definition, and then detail on such schedule particular datasets that would be within scope. Parties should consider that data is not pure intangible information—data can include the information embedded in a document, tissue sample, or medical image, for example. In addition, agreements should specify the format in which data will be provided to ensure that a receiving party can manipulate the data as needed for permitted uses.

Interplay Between Data and Intellectual Property

Traditional life science collaborations involve know-how, which is usually defined to include data. This begs the question—should data simply be treated as know-how under a collaboration?

The question ties into a broader philosophical debate as to whether data is a form of IP right. While many lawyers view data as a form of know-how, and while certain displays of data may be protectable under copyright, the U.S. legal framework does not formally recognize database rights like some other countries.

Many practitioners instead view data rights as more of a contractual right. For example, a data provider may not have any claims to ownership of data, since the underlying data subject ultimately controls the data.

Putting aside the underlying philosophical questions, data certainly should be treated like know-how in some regards. Parties should consider what data will be contributed to the collaboration, what type of licenses the data will be subject to, how the data will tangibly be transferred to the other party, and how improvements and derivatives to the data will be handled.

However, in deals involving significant data processing or data with independent economic value, many parties find it useful to separate out data from know-how. Separate treatment allows the parties to clearly define the treatment, use and ownership of data and its derivatives apart from general know-how.

For example, consider a deal where a platform company licenses technology, including know-how, to a licensee, and the licensee agrees to assign improvements it makes to the platform back to the platform company. In this scenario, the two parties may wish to agree to a very different framework regarding the treatment of data sets provided by either party to the other. Such distinction also makes it easier to draft compliance terms for regulated data, such as "protected health information" under the Health Insurance Portability and Accountability Act in the U.S.

Parties can consider creating a special term for specific data sets provided or generated under the agreement, such as "collaboration data," and carving out that term from general terms used to describe IP generated under the collaboration, such as "developed IP." Parties should then review the entire agreement carefully, using the correct terms in the appropriate provisions.

Data Improvements

In traditional IP contexts, an improvement typically captures any modifications, enhancements, or other improvements on a party's IP. However, with respect to data, it is important to make clear that an improvement does not include a mere rearrangement or alternative representation of data.

For something to constitute a true improvement, there typically needs to be a newly identified association or relationship between data sets. An improvement definition for data should be limited to these types of unique insights, rather than merely changing the window dressing of data. Parties should consider reflecting such a concept in the definition of improvement, or alternatively in any definition intended to capture improved data.

Training AI Algorithms

In traditional life science collaborations, an improvement made by one party to another party's IP is oftentimes retained by such other party. However, this model does not translate well with training of AI models. One party may use the other party's data to train its algorithms, but the AI platform cannot unlearn such data, since the data has been used to improve underlying algorithms.

Meanwhile, data owners are typically sensitive to AI platform owners retaining data indefinitely and potentially using it for the benefit of a competitor. Data owners generally want their data to only be used for the purpose of the collaboration.

One strategy that parties can use to address such conflicting interests is to prohibit incorporation of any data, or derivatives of such data, into an AI platform directly, such as permanently storing the data within the platform. However, the data may specifically be used as training sets to derive weights and parameters for machine learning algorithms, with the AI platform owner being free to use those new weights and parameters, without the data, both within and outside of the collaboration.

Alternatively, parties could agree that data may be retained and used for internal research directed to improvements to the AI platform, with the platform owner being free to use the improvement on the AI platform. However, it is critical that the agreement clearly state that the AI platform owner is free to use AI platform improvements, developed with or without such data, for any purpose.

Regeneration of Data Output

Parties should also be careful that confidential treatment and IP ownership of data output will not unduly restrict the owner of a technological platform.

Research results generated under a life sciences collaboration are often required to be treated confidentially by one or both parties, with one party owning such research results. However, in data collaborations, AI technology can be used to generate, for example, millions or billions of theoretical compounds, the vast majority of which do not meet screening criteria under the research plan.

Because the AI platform may regenerate such compounds outside of the collaboration, such as with future collaborators, the parties should ensure that data output that does not meet screening criteria is not subject to such confidentiality and ownership terms.

The owner of an AI platform should be able to regenerate such data output independently outside of the collaboration, so long as such party does not reuse the other party's input data to regenerate such output. The confidentiality and IP provisions of a collaboration should include appropriate carve-outs to permit such use.

Representations and Warranties

If a counterparty is supplying data or materials such as genetic samples, it is critical that robust representations and warranties be obtained from such counterparty around such data or materials.

For example, the counterparty should make representations that it has obtained and will obtain all necessary authorizations, consents and approvals of, and provided all necessary notices to, third parties such as data subjects, institutional review boards and regulatory authorities, that are required to be obtained or provided for the contemplated or actual use of such data under the agreement.

The counterparty should also make representations that its processing of data will be in accordance with applicable laws and regulations, as well as reasonable ethical, procedural, and quality control standards. The counterparty should also make representations that no third party will have a claim or interest in the other's platform technology by virtue of any use of such data, to the extent such data is sourced from a third party.

Parties should also agree to comply with specific privacy and security terms governing the use of such data, such as information security policies and procedures.

Confidentiality

Parties should carefully review confidentiality terms to see if standard terms apply to all types of data sets. It is critical that parties consider whether the data was generated by a contracting party, obtained from a private third-party source, derived from a data subject or acquired from the public domain. Confidentiality terms may need to be varied depending on the source of data.

For example, while the parties may be comfortable with confidentiality terms expiring within a set period following the end of a collaboration, the confidentiality of personal information should likely survive indefinitely.

Specialist Input

Data collaborations frequently involve thorny questions with respect to regulatory, health care, privacy, antitrust and other matters. In many cases, laws and regulations impose restrictions on the recipient's use of data, which can restrict information exchange and collaboration.

For example, transfer of personal information from the EU into the U.S. likely implicates the General Data Protection Regulation (GDPR) and other privacy regulations, which may require the parties to structure agreements in a way that takes into account controller or processor relationships.

In addition, where Clinical Laboratory Improvement Amendments (CLIA)-certified laboratories are implicated, contracts may need to specify that research-grade data must be validated prior to use for any clinical purposes.

A deal lawyer should always involve specialists from an early stage to navigate these issues, as specialist matters may shape the underlying structures of deals. When necessary, local counsel should advise on jurisdiction-specific matters.

Conclusion

Data collaborations in life sciences are here to stay and can present unique challenges for business development and legal professionals structuring these deals. Paying attention to the drafting considerations discussed above can help ensure that the intricacies of these deals are accounted for.