Legal Considerations for Sharing of Health Data
For Non-Commercial, Machine Learning Research

A White Paper from the University of Chicago Booth School of Business’s Center for Applied Artificial Intelligence 

---

¹This White Paper has been prepared by attorneys from Ropes & Gray LLP in collaboration with the University of Chicago Booth School of Business’s Center for Applied Artificial Intelligence. This White Paper does not constitute legal advice and does not create an attorney-client relationship with Ropes & Gray LLP. If you have questions regarding the topics discussed in this White Paper, you should consult your own counsel.
**Table of Contents**

I. Introduction ............................................................................................................................. 1

II. HIPAA .................................................................................................................................. 1
   A. De-Identification of PHI Prior to Use or Disclosure .......................................................... 3
   B. Patient Authorization for Use or Disclosure of PHI ......................................................... 4
   C. IRB/Privacy Board Waiver of Authorization for Use or Disclosures of PHI .................. 5
   D. Limited Data Set Disclosure of PHI Under Data Use Agreement ..................................... 5
   E. Reviews of PHI Preparatory to Research ...................................................................... 6
   F. Research on Decedents’ Information .............................................................................. 6

III. Substance Use Disorder Patient Records Confidentiality ............................................. 6

IV. State Privacy Laws ............................................................................................................... 8

V. Common Rule and Informed Consent ................................................................................ 9
   A. The Act of Sharing Data Is Not Human Subjects Research ............................................. 10
   B. Exemption for Secondary Research within HIPAA Covered Entities ........................... 10
   C. Sharing Data That Have Been Collected in a Primary Research Study ......................... 11
      i. Working with ICFs Under Which Existing Research Data Sets Were Collected .......... 11
      ii. Developing ICFs that Facilitate Sharing for Secondary Research ............................... 13
      iii. Certificate of Confidentiality ..................................................................................... 16

VI. Content of Data Use Agreement ...................................................................................... 16
   A. Liability of Data Contributors and Recipients ............................................................... 17
   B. Intellectual Property Issues ............................................................................................ 19
      i. Existing Intellectual Property ..................................................................................... 19
      ii. New Intellectual Property ......................................................................................... 22
      iii. Termination .................................................................................................................. 28
   C. Data Security Provisions ................................................................................................. 30

VII. Process Considerations for Review of Data Use Agreements .................................... 30

VIII. Third-Party Intermediary Models of Data Sharing ....................................................... 31
   A. Data Aggregator Platform ............................................................................................... 31
   B. Data Sharing Network with Central Organizer ............................................................. 32

IX. Conclusion .......................................................................................................................... 33
I. Introduction

Non-commercial, machine learning ("ML") research aimed at improving clinical care requires health data that are located across the United States at both academic and non-academic medical institutions ("Medical Institutions"). The data are most valuable when they are shared and aggregated for research projects, both among Medical Institutions and other institutions engaged in ML research (collectively, "Research Institutions"). Typically, such existing health data have been gathered in one of two circumstances, each of which has distinct implications for research—(i) primary research studies,\(^2\) such as drug or medical device trials, and (ii) standard-of-care clinical treatment and associated payment activities.

In this white paper, we analyze the regulatory, contractual and operational requirements and challenges that emerge when Medical Institutions share health data with Research Institutions for the purpose of conducting such ML research. We have focused solely on U.S. laws and regulations, addressing federal law and discussing examples of certain state laws. Of course, when data are collected from or shared with institutions outside of the United States, additional laws, such as the European Union General Data Protection Regulation ("GDPR"), could apply.

In Section II through Section V of this white paper, we describe the regulatory requirements that apply to such data sharing arrangements, as well as the operational considerations that these requirements raise. In Section VI through Section VII, we describe the content of data use agreements and the process considerations for such agreements. These sections all focus on the legal and contractual frameworks that apply to data sharing among Research Institutions. In Section VIII, we consider how the legal, contractual, and operational considerations are affected in data sharing models that rely on third-party intermediaries.

Understanding these requirements and challenges, in turn, can lead to improvements in the collection, sharing and use of health data to promote research. These improvements can be brought about, for example, by promoting education about data sharing requirements across institutions and developing template consent documents and data sharing agreements.

II. HIPAA

The federal Health Insurance Portability and Accountability Act of 1996, as amended and with its implementing regulations ("HIPAA") applies to the use and disclosure of protected health information ("PHI") by "covered entities." PHI generally is individually identifiable health information, and covered entities are health plans, health care clearinghouses, and health care providers that transmit any health information in electronic form in connection with a transaction covered by HIPAA; primarily, these are transactions related to health insurance reimbursement.\(^3\) Medical Institutions almost certainly will be covered entities and thus subject to HIPAA’s restrictions when they seek to receive, use or share PHI for ML. However, HIPAA would not apply to non-covered entities that use or share PHI for ML, such as many independent Research Institutions that either perform no health care services or that provide only health care services not

---

\(^2\) When data from primary research are used in a subsequent, distinct research study, we refer to the second study as "secondary research."

\(^3\) 45 C.F.R. § 160.103.
billed to third party payors. For example, there are many Research Institutions that operate clinical laboratories that perform no billing of third party payors and thus are not subject to HIPAA.

Covered entities are permitted to use and disclose PHI for certain activities without the authorization of the patient. Such activities include treatment, payment and health care operations. HIPAA also permits a covered entity to engage a third party, known as a business associate, to perform these activities on its behalf. De-identification of PHI, data analysis, and quality improvement are all common health care operations activities for which covered entities engage business associates.4

Because research, unlike quality improvement, is not a “health care operation,” it does not give rise to a business associate relationship, even if a researcher is hired by a covered entity to perform a research project.5 However, because many research-adjacent activities such as quality improvement, creation of limited data sets (discussed further below), and de-identification are considered to be business associate activities, in certain cases a single individual or entity may be a business associate for certain purposes and also a researcher for other purposes. For example, a Medical Institution may engage a third party to structure certain data sets on its behalf as a business associate. If the third party wishes to use the structured data to perform research, however, an additional permission under HIPAA will be required to permit the research.

Quality improvement and research are similar, but distinct, activities, and covered entities sometimes struggle to discern in which category a particular activity falls. HIPAA, like the Common Rule (discussed below), defines “research” as “a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.”6 Thus, the touchstone for research is whether it gives rise to findings that can be extended across one or more populations. The United States Department of Health and Human Services (“DHHS”) has stated that:

> We understand knowledge to be generalizable when it can be applied to either a population inside or outside of the population served by the covered entity. Therefore, knowledge may be “generalizable” even if a research study uses only the protected health information held within a covered entity, and the results are generalizable only to the population served by the covered entity. For example, generalizable knowledge could be generated from a study conducted by [the Centers for Medicare and Medicaid Services (“CMS”)], using only Medicare data held by [CMS], even if the knowledge gained from the research study is applicable only to Medicare beneficiaries.7

---

4 45 C.F.R. § 164.501.
5 DHHS, HIPAA FAQ 239 (“Disclosures from a covered entity to a researcher for research purposes do not require a business associate contract, even in those instances where the covered entity has hired the researcher to perform research on the covered entity’s own behalf.”) (last accessed Aug. 12, 2020).
6 45 C.F.R. § 164.501 (emphasis added).
Federal regulations do not offer a definition of quality improvement activities, beyond noting that they include “outcomes evaluation and development of clinical guidelines,” and exclude activities in which “the obtaining of generalizable knowledge is . . . the primary purpose of any studies resulting from such activities” (i.e., research is expressly not a quality improvement activity). In practice, quality improvement activities are often taken to mean systematic, data-guided activities that are designed to bring about immediate (or nearly immediate) improvements in a program, process or system. Thus, whether a systematic, data-guided activity represents research or quality improvement may turn on whether the primary purpose is to generate generalizable knowledge that can be extended across one or more population(s), or instead is primarily intended to bring about immediate (or nearly immediate) improvements in a program, process or system.

Several pathways permit covered entities to use or share PHI for research. We discuss below the following six such pathways: (i) de-identification of PHI prior to use or disclosure, (ii) patient authorization for use or disclosures of PHI for research, (iii) IRB/privacy board waiver of authorization for disclosure of PHI, (iv) disclosure of PHI in the form of a “limited data set” subject to a data use agreement, (v) reviews of PHI preparatory to research, and (vi) use of decedents’ PHI. While these pathways have existed with relatively little modification since the Privacy Rule first took effect in 2003, they can still cause confusion for institutions engaged in the sharing of data for research purposes.

The way in which these requirements apply to business associates often is another source of confusion. As described above, a business associate agreement is not required for a third party entity to perform research. However, if a business associate agreement properly binds a third party because the third party is assisting a covered entity in payment or health care operations activities, that business associate can only perform research functions if it obtains permission from the covered entity to use or disclose PHI for research purposes and is additionally able to satisfy one of the below pathways that allow use or disclosure of PHI for research.

A. DE-IDENTIFICATION OF PHI PRIOR TO USE OR DISCLOSURE

PHI can be “de-identified,” and thus converted into data that are no longer subject to HIPAA, pursuant to two mechanisms set forth in the HIPAA Privacy Rule: (i) the “safe harbor” method, which requires the removal of 18 enumerated identifiers and (ii) the “expert determination” method, which permits a statistical expert to render an opinion that data have been de-identified such that the risk of re-identification is “very small.” HIPAA permits PHI to be considered de-identified even when certain “re-identification codes” are assigned to de-identified information, provided that the codes are not derived from PHI and the covered entity assigning the codes does not release the key to re-identify the codes to a third party. In many secondary research scenarios, researchers need access to data elements that must be shed from a data set during the de-identification process, meaning that de-identification is not a feasible solution.

---

8 45 C.F.R. § 164.501.
9 45 C.F.R. § 164.514(b).
10 45 C.F.R. § 164.514(c). This approach parallels the Common Rule approach, discussed below, in which the use of coded identifying information or specimens does not constitute human subjects research.
Fortunately, HIPAA provides covered entities with several well-established methods by which PHI, including some or all identifiers, may be used or disclosed for research purposes, and described next in this White Paper. A covered entity may engage a business associate to de-identify PHI on the covered entity’s behalf, and the covered entity may even engage an intended recipient for research purposes to be a business associate for the purpose of de-identification of PHI. However, because business associates may only use PHI as permitted or required by their business associate agreements, they must obtain contractual permission from a covered entity in order to de-identify data before the business associate can use such data for research.

B. PATIENT AUTHORIZATION FOR USE OR DISCLOSURE OF PHI

Covered entities that use or disclose PHI for research purposes generally obtain the authorization of the data subject for the use or disclosure of such PHI. The subject’s authorization is typically combined with an informed consent form in the research context, often with one signature block. The authorization can be drafted broadly to permit secondary research; however, under the authorization’s terms, it must be drafted with sufficient specificity such that it would be reasonable for the data subject to expect that his or her PHI would be used for the contemplated research project. We provide model language for such an authorization in Appendix A.

We note that in secondary research researchers have sometimes had difficulty obtaining PHI from a covered entity after presenting a valid HIPAA authorization to the provider. This is because the posture of a HIPAA authorization is that it permits, but does not require, the health care provider to disclose the requested information. Therefore, health care providers can decline to provide the requested information even when an individual executes an authorization that satisfies HIPAA requirements. However, effective April 5, 2021, the information blocking regulations promulgated under the 21st Century Cures Act will restrict a provider’s ability to refuse to disclose electronic PHI to a researcher when the covered entity is permitted by law, such as through an authorization, to do so. See 45 C.F.R. pt. 171. Specifically, the information blocking provisions will prohibit any practice that (i) except as required by law or covered by an information blocking exception, (ii) is likely to interfere with access, exchange, or use of electronic health information (“EHI,” which generally refers to HIPAA electronic PHI) and (iii) that is undertaken with the requisite intent. The requisite intent for health care providers is knowledge that the practice is unreasonable and is likely to interfere with access, exchange, or use of electronic health information, which would appear to be met if a health care provider declined to provide EHI to a researcher who presents the health care provider with a valid authorization.

11 DHHS, Health Services Research and the HIPAA Privacy Rule, NIH Publication No. 05-5308 (May 2005) (“A covered entity may hire the intended recipient of the de-identified data as a business associate for purposes of creating the deidentified data. That is, a covered entity may provide a business associate that is also the deidentified data recipient with PHI, including identifiers, so that the business associate can deidentify the data for the covered entity. However, the data recipient, as a business associate, must agree in its business associate agreement to return or destroy the identifiers once the de-identified data set has been created.”).
12 45 C.F.R. § 164.502(a)(3).
C. IRB/PRIVACY BOARD WAIVER OF AUTHORIZATION FOR USE OR DISCLOSURES OF PHI

When the research cannot practically be conducted if the researchers are required to obtain the subject’s authorization, the covered entity may use or disclose PHI for research if it obtains a waiver of authorization from a cognizant IRB or from a privacy board meeting HIPAA’s requirements. Such a waiver must be based on a finding that, in addition to the impracticability of obtaining the data subject’s authorization, (i) the use or disclosure involves no more than a minimal risk to individuals’ privacy and (ii) the research could not practically be conducted without access to and use of the PHI. A privacy board is similar to an IRB but has as its sole function the evaluation of requests for waivers of HIPAA authorization. Most Medical Institutions rely on their IRB to grant waivers of authorization and do not form a separate privacy board.

Waivers of authorization are often used in secondary research, as it often would not be feasible to obtain the data subjects’ individual authorizations. Typically, well-established IRBs maintain template forms that researchers can use to request a waiver of authorization. When IRBs grant waivers of authorization, this greatly streamlines and expedites the administrative process for researchers.

Notably, a covered entity can rely on a waiver of HIPAA authorization from an external IRB (including another Medical Institution’s IRB, another Medical Institution’s privacy board, or even an external commercial IRB) to disclose PHI for research purposes; put differently, the waiver need not be granted by the covered entity’s own IRB. This option of relying on an external IRB’s grant of a waiver of authorization can greatly facilitate sharing of PHI in instances in which multiple institutions rely on a single IRB for a research project involving multiple institutions or in which only the recipient institution is “engaged” in the research (as discussed further in Section V, infra) and thus only the recipient institution’s IRB reviews the research.

D. LIMITED DATA SET DISCLOSURE OF PHI UNDER DATA USE AGREEMENT

HIPAA also permits covered entities to disclose PHI for research purposes without authorization or a waiver of authorization if the data are considered a “limited data set.” In this case, however, HIPAA requires that the disclosing covered entity enter into a data use agreement containing a satisfactory assurance that the recipient will use or disclose the PHI only for the purposes set forth in the data use agreement, that the recipient will employ appropriate safeguards to protect the data, and that the recipient will not attempt to re-identify any of the data subjects. HIPAA defines a limited data set as a data set from which 16 enumerated direct identifiers have been removed. As with de-identification, as discussed above, a covered entity may engage a business associate to create a limited data set on the covered entity’s behalf, and the covered entity

---

14 45 C.F.R. §§ 164.508, 512.
15 45 C.F.R. § 164.514(e)(3)-(4).
16 45 C.F.R. § 164.514(e)(2).
may even engage the intended data recipient as the business associate to which the creation of the limited data set has been delegated.\textsuperscript{17}

Important for research uses, limited data sets may retain certain data elements that would have to be removed for data to be considered de-identified under HIPAA. In some cases, these data points may be useful to retain for research purposes—the city or town, state, and zip code elements of an address; elements of dates (including both dates of birth and dates of service); and certain other unique identifying numbers, characteristics or codes. It is important to recall, however, that HIPAA defines research purposes, for which a limited data set may be used, as the creation of “generalizable knowledge.” Knowledge is considered “generalizable” when it can be applied to a population. There is not specific guidance addressing whether ML algorithms meet this standard, so the risk of relying on a limited data set for data sharing for ML algorithms is that a regulator could attempt to take the position that ML was not a permissible “research” basis for sharing data under HIPAA but instead was some another use of data (e.g., product improvement not resulting in generalizable knowledge) that falls outside of the definition of research.

\textbf{E. REVIEWS OF PHI PREPARATORY TO RESEARCH}

HIPAA permits researchers to review PHI in a manner “preparatory to research,” such as to identify eligible subjects, without the authorization of the persons whose PHI is reviewed. In these cases, the researcher must provide certain representations to the covered entity at which he or she is reviewing information preparatory to research.\textsuperscript{18} These representations include that (i) use or disclosure is sought solely to review PHI as necessary to prepare a research protocol or for similar purposes preparatory to research, (ii) no PHI is to be removed from the covered entity by the researcher in the course of the review; and (iii) the PHI for which use or access is sought is necessary for the research purposes.

\textbf{F. RESEARCH ON DECEDETS’ INFORMATION}

HIPAA’s limitations on the use and disclosure of PHI continue for fifty (50) years following an individual’s death. Notably, however, HIPAA permits researchers to use PHI of deceased individuals, without the authorization of the decedent’s personal representative, when the covered entity obtains two representations from the researcher.\textsuperscript{19} These representations are (i) that the use or disclosure sought is solely for research on the PHI of decedents and (ii) the PHI for which use or disclosure is sought is necessary for the research purposes. The covered entity may also request, in its discretion, that a researcher provide documentation of the death of the individuals whose PHI is sought.

\textbf{III. Substance Use Disorder Patient Records Confidentiality}

Federal law provides heightened privacy requirements that apply to substance use disorder patient records held by “federally assisted” substance use treatment programs. Federally assisted

\textsuperscript{17} DHHS FAQ 249 (last accessed Aug. 13, 2020) (“In addition, the covered entity may hire the intended recipient of the limited data set as the business associate for this purpose [of creating a limited data set] in accordance with the business associate requirements.”).

\textsuperscript{18} 45 C.F.R. § 164.512(i)(1)(ii).

\textsuperscript{19} 45 C.F.R. § 164.512(i)(1)(iii).
programs are broadly defined to include those that participate in the Medicare program, receive federal financial assistance in any form, or enjoy tax-exempt status.\textsuperscript{20} These requirements, generally known as “Part 2” requirements for the location at which they have been codified within Title 42 of the Code of Federal Regulations, apply in addition to HIPAA when a federally assisted program is conducted by a HIPAA covered entity.

Part 2 often causes confusion for data sharing efforts and Medical Institutions sometimes exclude Part 2 records entirely from research projects involving data sharing. Part 2 is more restrictive than HIPAA in certain respects—for example, Part 2 contains prohibitions on re-disclosure of Part 2 information. Also, many entities struggle in their electronic health records to identify Part 2 records when they offer both Part 2 and non-Part 2 services. This can limit important research, especially given the recent focus on research related to the opioid crisis, for which much important data is subject to Part 2.

Helpful for the research community, the Part 2 regulations have, in the past five years, been the subject of multiple rulemakings that have reduced the burden on researchers by harmonizing some aspects of Part 2 with HIPAA’s requirements. Before these changes, the Part 2 regulations had existed largely unchanged for decades after they were issued in the 1970s. Most recently, the Substance Abuse and Mental Health Services Administration (“SAMHSA”) issued a final rule amending the Part 2 regulations in July 2020.\textsuperscript{21} The final rule has adopted some changes that will likely be helpful for the research community. First, the final rule has revised the Part 2 consent requirements to permit patients to give broad consent to a wide range of entities without naming a specific individual at the entity who will receive the information, as had previously been required.\textsuperscript{22} Second, the final rule has changed Part 2’s research requirements to align them more closely with those of the Common Rule (as defined below in Section V) and the HIPAA Privacy Rule and to streamline research disclosures in such instances. For example, the final rule permits a disclosure of Part 2 information for scientific research on the basis that the disclosing entity is a HIPAA covered entity and the disclosure is subject to HIPAA, even if the recipient is neither a HIPAA covered entity nor subject to the Common Rule. (Nevertheless, the disclosing entity must, as a HIPAA covered entity, have a separate allowable basis under HIPAA to justify the disclosure for research purposes.)

Also, the statute under which Part 2 is promulgated was recently amended by the Coronavirus Aid, Relief and Economic Security Act (“CARES Act”), which also seeks better to align Part 2 standards with those of HIPAA. The CARES Act changes the statute and directs DHHS to issue new regulations. These are expected to lead to further modifications of some of the amendments adopted by the July 2020 final rule.\textsuperscript{23} The amended statute will no longer require patient consent for each disclosure of Part 2 information, but will instead permit disclosure for treatment, payment and health care operations to the extent consistent with HIPAA.\textsuperscript{24} Of course, as noted above, HIPAA does not regard research as treatment, payment, or health care operations.

\textsuperscript{20} 42 C.F.R. § 2.12(b).
\textsuperscript{22} Id. at 42,987.
\textsuperscript{23} Id.
\textsuperscript{24} Senate Bill 3548 § 4221.
Thus, while the CARES Act likely will not affect greatly the use of Part 2 information for research purposes, it may also contribute to the lessening of confusion regarding Part 2’s application.

IV. State Privacy Laws

Some states have data privacy requirements that are more stringent than those found in HIPAA. These state laws often apply to specific types of health data such as genetic data, HIV/AIDS information, and mental hygiene information. For example, New York and many other states require authorization for disclosure of a genetic test result and most states require a specific form of authorization for the disclosure of an HIV test result. Given the variation in laws between states and the fact that such laws often lack the substantial guidance that is available for federal legislation such as HIPAA and Part 2, research projects involving data from multiple states can be impeded due to the patchwork of state laws.

Among state privacy laws, the California Consumer Privacy Act of 2018 (“CCPA”) has drawn substantial attention. CCPA is a broadly sweeping, GDPR-style law that applies to the extent that information, held by a for-profit institution, is not protected by HIPAA or certain other applicable privacy laws. Also, a law was enacted in September 2020 with immediate effect to amend CCPA to exempt: (i) personal information gathered for research, as defined in HIPAA, if the research is conducted in accordance with the HIPAA Privacy, Security, and Breach Notification rules, the Common Rule, the International Council for Harmonization Good Clinical Practice Guidelines, or the human subject protection requirements of the United States Food and Drug Administration, or (ii) information from which identifiers linking the information to individuals has been removed pursuant to the HIPAA de-identification standard. Importantly, the law requires that, effective January 1, 2021, certain statements must be included in any agreement for the sale or license of de-identified patient data, including that the de-identified information includes de-identified patient information, that re-identification is prohibited, and that the purchaser may not further disclose the de-identified information to any third party unless the third party is bound to same or similar restrictions. The new law also requires businesses selling or disclosing de-identified health information to notify consumers online of such practice, which is notable because HIPAA covered entities and others that de-identify information typically are not otherwise required to notify patients of disclosures of de-identified information. Additionally, a ballot initiative was adopted by California voters in November 2020 to amend CCPA. The measure does not extend CCPA to nonprofits, but makes various other changes, including creating a California Privacy Protection Agency.

Thus, CCPA is unlikely to apply, in most respects, to the nonprofit Research Institutions anticipated to engage in the ML research initiative. However, the novelty of CCPA, which took effect on January 1, 2020 and has been enforced only since July 1, 2020, along with the ballot

---

25 See, e.g., NEW YORK CIV. RIGHTS L. § 79-L; CAL. HEALTH & SAFETY CODE § 120980.
26 See CAL. CIV. CODE § 1798.100.
27 See Assembly Bill 713. Note, however, that the law also prohibits the re-identification of personal information that is derived from patient information unless such re-identification is carried out in compliance with a statutorily permitted purpose, including among others, research conducted in accordance with the Common Rule. Thus, re-identification of the data in research would bring the research back into the ambit of other applicable laws.
initiative and statute amending it, have caused confusion regarding the law’s applicability to research studies. If for-profit, non-HIPAA covered entities were to play a role in a ML research project going forward, CCPA could be implicated. In cases in which the CCPA applies, it places requirements on persons collecting personal information to provide notices to individuals from whom they collect the data. Further, CCPA affords data subjects various rights with respect to their data, including rights of access and deletion, among others.

V. Common Rule and Informed Consent

The federal Policy for the Protection of Human Subjects (generally known as and referred to herein as the “Common Rule”) regulates research involving human subjects conducted by, carried out with the support of, or otherwise subject to regulation by DHHS, which includes the National Institutes of Health, and the other federal departments and agencies that adhere to it, including the National Science Foundation. However, the Common Rule’s impact is even broader because, as a matter of policy, many Research Institutions require compliance with the Common Rule for all human subjects research in which their researchers are engaged, regardless of source of funding for their research projects.

Under the Common Rule, the use of identifiable data pertaining to a human subject for secondary research purposes is considered research involving a human subject for purposes of the Common Rule. One core requirement of the Common Rule is that human subjects research, unless it is considered exempt from the Common Rule, be conducted only upon approval by an Institutional Review Board (“IRB”) and upon research subjects’ consent or an IRB’s waiver of the consent requirement. Historically, institutions required human subjects research conducted at the institution to be reviewed by an IRB operated by the institution. However, the Common Rule permits an institution to rely on the review and approval of human subjects research by an “IRB that is not operated by the institution,” provided that the Research Institution and the organization operating the IRB “document the institution’s reliance of the IRB for oversight of the research and the responsibilities that each entity will undertake to ensure compliance with the requirements” of the Common Rule. When an institution relies on an external IRB, such external IRB is often called a “central IRB,” and the agreement between the institution and the central IRB is referred to as a “reliance agreement” or an “authorization agreement.” A central IRB can be the IRB of another Research Institution or an IRB operated by a commercial organization that exists for the purpose of providing IRB review services. Further, there now exist several consortia of Research Institutions, including SMART IRB, that have adopted standardized reliance agreements. Use of these standard reliance agreements can further streamline the contracting process required to implement external IRB review.

In this section, we describe the ways in which these requirements commonly affect research on already compiled sets of data. The Common Rule was the subject of major revisions that largely

29 45 C.F.R. § 46.102.
30 45 C.F.R. § 46.109. The requirements for an IRB to waive the consent requirement are very similar to those required for an IRB to waive the authorization requirement under HIPAA.
31 45 C.F.R. § 46.103(e).
32 https://smartirb.org/.
took effect in January 2019 and introduced new pathways to ease the sharing of data for secondary research purposes, the most important of which we discuss below.

A. THE ACT OF SHARING DATA IS NOT HUMAN SUBJECTS RESEARCH

The release by a Medical Institution of already-gathered data to another party for secondary research is not itself a research activity that triggers the need for IRB review and approval under the Common Rule. This is because regulatory authorities do not interpret the act of releasing data as rendering the data contributor “engaged” in the research.33 Ensuring that institutional officials and IRBs are familiar with this fact can speed data exchange and research by anticipating and addressing potential objections.

Notably, however, the recipient Research Institution may nevertheless need IRB review and approval for the secondary research uses of the personal data released to it. Further, Medical Institutions contributing data should ensure that release for secondary research would not be impermissible under the terms of the research informed consent form (“ICF”) or other applicable law, such as HIPAA and state privacy laws. Often, HIPAA covered entities releasing data rely upon a waiver of HIPAA authorization from the recipient institution’s IRB.

B. EXEMPTION FOR SECONDARY RESEARCH WITHIN HIPAA COVERED ENTITIES

The recipient’s use of identifiable data in secondary research would constitute human subjects research subject to the Common Rule. The Common Rule contains an exception that permits secondary research uses of data within a single HIPAA covered entity, which on its face appears to be of limited utility for data sharing. However, when both the disclosing and receiving entities are HIPAA covered entities, as some Research Institutions participating in the ML research that are also Medical Institutions are likely to be, the Common Rule exempts, in some, but not all, circumstances, such secondary research from the requirements of the Common Rule. Importantly, when available, this exception eliminates the need for IRB review of the study and the informed consent of the subject.

Subregulatory guidance issued by the DHHS Secretary’s Advisory Committee on Human Research Protections (“SACHRP”) indicates that the exception is intended to allow separate HIPAA covered entities (such as two Medical Institutions) to conduct collaborative research sharing identifiable personal information with one another.34 This implies that the both the disclosing and recipient institution would need to be engaged in the research for this exception to permit the use and disclosure of the information to be governed under HIPAA rather than the

34 See SACHRP, Recommendations on the Interpretation and Application of § __.104(d)(4) the “HIPAA Exemption,” https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-b-december-12-2017/index.html. It is worth noting that the Common Rule exemption does not disapply the HIPAA requirements for disclosure of PHI, and thus the disclosing HIPAA covered entities would need to obtain from an IRB or Privacy Board a waiver or alteration of authorization permitting the disclosure to one another, unless a HIPAA authorization disclosure form or a waiver or alteration of authorization have previously been obtained.
Identifiable personal data that are collected originally in primary research studies can be subject to additional strictures imposed by the ICF and any research agreement (e.g., a clinical trial agreement) between the institution conducting the research and the sponsor or funder of the research. The common law applied by courts regards as a contract the ICF between the research participant and the institution and its researcher carrying out the research. This can be problematic for secondary research because, if subjects provide consent to a specific, limited set of uses of their personal data, their ICFs may be read to preclude the use of the data for other, non-specified purposes. Because multiple ICFs, often with varying language, are used across even a single study’s sites over time, analyzing whether the consents permit sharing of the data for secondary research can be a time consuming, costly exercise.

i. Working with ICFs Under Which Existing Research Data Sets Were Collected

Below, we describe frequent categories of ICF language, and the extent to which each category may permit or impede data sharing of existing data sets gathered in primary research studies.

a. ICF Is Silent Regarding Secondary Uses of Data and Specimens

Commonly, an existing ICF does not address secondary uses of data and specimens. When forms are silent, secondary uses of data and specimens are likely permissible if the data are modified such that they are not individually identifiable. Because no explicit or implicit promises were made to the contrary, de-identifying the data (thereby taking them outside the requirements of the Common Rule and HIPAA) and using the data for secondary research purposes would likely not be construed by a court as violating the terms of the ICF. As a matter of interpretation, institutions could look to the entire ICF to determine whether they believe it fairly could be read, in totality, not to preclude secondary uses of identifiable information and, if so, seek a waiver of authorization from a cognizant privacy board or IRB (thereby also taking the data outside the requirements of the Common Rule and HIPAA).

b. ICF Explicitly Prohibits Secondary Uses of Data and Specimens

An ICF may explicitly prohibit secondary uses of data and specimens. This is common with older ICFs, many of which were not written to accommodate future secondary uses of data. For example, the ICF might provide that “data and specimens collected in this study will not be used for any purposes other than the study for which they were collected.”

Because the investigator has made a promise to the data subject that his or her data will not be used for other purposes, identifiable data should not be used for secondary research purposes. It is true that, under both the existing and revised Common Rule, one could de-identify the data and
specimens before performing secondary research, thereby removing the data and specimens from the jurisdiction of both the Common Rule and HIPAA. However, such de-identification would appear contrary to the specific promise that was made to the subject as a matter of general consent principles, under which the ICF represents a contract between the research participant and the site and researcher. De-identifying the specimens and data does not obviate or relieve the promise to which the institution bound itself in the ICF, which explicitly advised the subject that their data would not be used for a secondary purpose. Moreover, de-identification and re-use in this scenario arguably would violate the ethical principle of “respect for persons,” one of the three principles outlined in the *Belmont Report* that underlies the Common Rule. Accordingly, data provided under a consent that explicitly prohibits secondary uses of data and specimens should not be used for secondary research.

c. **ICF Selectively Permits Secondary Uses of Data and Specimens**

An ICF may selectively permit secondary uses of data and specimens for particular types of secondary research, stating, for example, “following completion of the present study, data and specimens collected in this study may be used for research related to cancer” (i.e., not ML applications generally). Similarly, some ICFs state that secondary research will not be undertaken for “commercial uses.” Of course, most biomedical research could arguably be considered to have commercial uses if its intent in any way includes the development of a marketable product, including an ML product. Data and specimens gathered under such a consent could be removed from the applicable Common Rule and HIPAA protections through de-identification; however, doing so does not vitiate the explicit promise earlier made to the subjects that the specimens and data would be used only for research on cancer or non-commercial uses, as the case may be.

Therefore, it is often thought prudent in such cases for the researcher to limit the secondary uses to those related to cancer or to non-commercial uses, as applicable, because it is not clear that a court, applying general principles of consent, would view a data subject as having understood and agreed that even an anonymized version of his or her data or specimen would be used for research unrelated to cancer or for commercial uses.

d. **ICF and Research Protocol Contradict One Another Regarding Secondary Uses of Data and Specimens**

In some cases, the ICF and the protocol may contain contradictory or inconsistent statements regarding secondary uses of data and specimens. For example, the ICF may state that, “data and specimens collected in this study will be used solely for research related to cancer,” while the protocol may state “data and specimens from this study may be used by the study sponsor or third parties for secondary research on any topic.”

As a prospective matter, the protocol and the ICF should always be reviewed to ensure consistency of terms between the documents. In the event of a conflict, when seeking to use data for secondary research, it is generally better to follow promises made in the ICF, as the ICF would

---

typically be viewed by a court as the instrument through which the institution bound itself contractually when it presented the document to the subjects and obtained their signature on it.

**ii. Developing ICFs that Facilitate Sharing for Secondary Research**

The above frequently encountered limitations on data sharing from existing research data sets also provide lessons for drafting ICFs for prospective primary research studies when the data gathered may have some utility for future secondary research. Following these techniques could be helpful when designing ML studies or when designing studies in which data are collected that are intended to be used in a secondary research project studying ML. We provide in Appendix A sample language that can be included in consent forms and HIPAA authorizations used in connection with research studies to permit wide sharing of data for research.

**a. Broad Consent in Primary Research ICFs**

The revisions to the Common Rule that took effect in January 2019 introduced a regulatory pathway for “broad consent” to be obtained for non-exempt storage, maintenance, and research uses of identifiable information and biospecimens, which we refer to in this memorandum as “regulatory broad consent.” The regulatory broad consent is less useful than it first appears, as it also introduces several restrictive requirements greatly limiting its utility for the research community. These are discussed in greater detail below.

Prior to the Common Rule revisions, the research community had relied on a theory of broad consent in a general sense for many years, by inserting into ICFs a broad permission for secondary research uses of data and, in some cases, biospecimens. These “non-regulatory broad consents” were, and continue to be, widely recognized as effective in facilitating secondary research. The adoption in the revised Common Rule of a very specific, regulatory form of “broad consent” impliedly recognizes the long-standing practice of including a non-regulatory broad consent in ICFs. Importantly, the revised Common Rule’s adoption of regulatory broad consent should not be construed to preclude secondary uses of data and biospecimens under non-regulatory broad consents. **Accordingly, recipients of federal funding that are subject to the Common Rule will in most cases wish to continue to rely on non-regulatory broad consent, rather than on the new regulatory broad consent contained in the revised Common Rule.** SACHRP has advised that it believes that “non-regulatory broad consent” forms “signed before and after the effective date of the Final Rule:

. . . will continue to be effective. After the effective date of the Final Rule, a consent for future research using identifiable data and/or biospecimens must be analyzed as before, to assure that the consent has sufficiently detailed the future research uses, even if that consent would not be sufficient to meet ‘broad consent’ as used in the Final Rule. . . .**

The enduring availability of non-regulatory broad consent is important due to the restrictiveness and limitations of the regulatory broad consent. The most notable such limitation is that if a researcher offers someone a regulatory broad consent form and the person refuses to

---

provide regulatory broad consent, the revised Common Rule provides that a cognizant IRB may *never* grant a waiver of the consent requirement for secondary research uses of that individual’s identifiable information, thus effectively preventing such information from ever being used for secondary research purposes. Researchers could, of course, de-identify the data and then use those data for secondary research, but the revised Common Rule’s requirement could potentially confound important secondary research that requires or would benefit from use of identifiable information. Because of this limitation on regulatory broad consent, Medical Institutions using the regulatory broad consent would be required to track subjects’ refusals to provide broad consent, so the institution could ensure that a waiver of consent would never be granted for use of that individual’s information. This tracking would prove a burdensome process, requiring extensive and seamless information technology capabilities, which would often only be available in large Medical Institutions that have a centralized electronic health record in which such tracking could reliably and accurately be done. Further, if a significant percentage of persons to whom regulatory broad consent is offered refuse to give it, then a dataset could be rendered unusable for secondary research simply because of selection bias.

The revised Common Rule does not delineate whether an individual’s silence, non-responsiveness, and/or express declination to give regulatory broad consent would constitute “refusal to consent” for purposes of determining whether an IRB can later waive consent for the storage, maintenance, or secondary research use of an individual’s identifiable private information or identifiable biospecimens. While express declination to give broad consent is plainly “refusal to consent” under the revised Common Rule, the meaning and legal import of a person’s silence or non-responsiveness to a request to provide broad consent is less clear. SACHRP has recommended that in the context of offering a regulatory broad consent, HHS interpret “refusal to consent” to include only a person’s express declination to give broad consent, as demonstrated by an individual’s unambiguous written or oral communication to that effect. SACHRP recommended that an ICF seeking regulatory broad consent expressly state that failure to respond (i) will not be treated as a refusal to consent, (ii) will not prevent researchers from seeking a waiver of consent or pursuing an exemption for secondary use of personal data, and (iii) will not act as affirmative regulatory broad consent. SACHRP also recommended, as a best practice, that for persons who are offered regulatory broad consent through a medium other than face-to-face contact (e.g., by internet or mail), researchers attempt to provide a second offer of regulatory broad consent to participants who have not responded within a reasonable timeframe.

In sum, the non-regulatory broad consent, expressed as consent to data sharing for secondary research, with research purposes and goals broadly stated, remains the preferable method of enabling broad secondary research uses of clinical and research data.

b. Required Statements

The revised Common Rule prescribes set statements that must be included in ICFs with respect to secondary research uses. To the extent that new studies will be conducted from which data will be shared for secondary research purposes, it is important to include the appropriate statement for data sharing within the applicable research informed consent form(s). This is true even when a Medical Institution relies on non-regulatory broad consent for research. Specifically,

---

37 45 C.F.R. § 46.116(f).
new ICFs for any research study that collects identifiable private information or identifiable biospecimens must include one of two statements:

(1) A statement that identifiers might be removed from the identifiable private information or identifiable biospecimens and that, after such removal, the information or biospecimens could be used for secondary research studies or distributed to another investigator for secondary research studies without additional informed consent from the subject or the legally authorized representative, if this might be a possibility; or

(2) A statement that the subject’s information or biospecimens collected as part of the research, even if identifiers are removed, will not be used or distributed for secondary research studies.38

Thus, Medical Institutions can promote the quantity of data available for secondary research by ensuring that institutions sharing data include the first statement above in their research ICFs to preserve the possibility of removing identifiers from identifiable data or biospecimens collected in research so that the data and/or biospecimens may be used for secondary research purposes. These changes to the Common Rule were motivated by the famous Henrietta Lacks case and are designed to improve public awareness that their data and biospecimens can be de-identified, thereby removed from the applicable regulatory regimes, and used for research without specific consent or IRB review and approval.

c. Withdrawal of Consent

It is also important to consider subjects’ potential withdrawal of consent when drafting an ICF that contains a non-regulatory broad consent to facilitate secondary research. U.S. law requires that, upon a subject’s withdrawal of consent, researchers must cease collecting additional data, but may continue to use subjects’ personal data collected up to the point of withdrawal to preserve the integrity of the research.39 The ICF should put subjects on notice of the effect of withdrawal on data collection and use.

d. Review of Definition of Identifiable Private Information

The revised Common Rule introduces a process by which HHS and the other Common Rule departments and agencies will evaluate, on a four-year cycle, the definition of “identifiable private information” within the Common Rule and whether there are certain technologies that when applied to biospecimens or information generate information that is inherently identifiable.40 The Common

38 45 C.F.R. § 46.116(b)(9).
39 OHRP, Withdrawal of Subjects from Research Guidance (Sept. 21, 2010), https://www.hhs.gov/ohrp/regulations-and-policy/guidance/guidance-on-withdrawal-of-subject/index.html (noting that FDA requires data collected up to withdrawal to be retained for the trial to be scientifically valid and that HIPAA’s right for individuals’ to revoke their authorizations contains an exception to the extent a covered entity has taken action in reliance on that authorization, which in the context of research permits the continued use and disclosure of PHI already obtained pursuant to an authorization, as needed to protect the integrity of the research).
40 45 C.F.R. § 46.102(e)(7). With the recent change of presidential administrations, it is possible that there may be new interest in evaluating the definition of “identifiable private information.”
Rule revisions state that whole genome sequencing is expected to be one of the first technologies to undergo this analysis.\textsuperscript{41} These changes will be important for conducting secondary research because they will likely make it more difficult, and in some cases impossible, to remove secondary research from the jurisdiction of the Common Rule by shedding identifiers from the personal data to be used.

iii. Certificate of Confidentiality

Data collected in a primary research study also may be subject to a certificate of confidentiality ("Certificate of Confidentiality"). Certificates of confidentiality have caused confusion in some data sharing efforts, in particular because while Certificates of Confidentiality have existed since the 1970s, the 21\textsuperscript{st} Century Cures Act, passed in 2016, introduced several changes affecting their use. Specifically, the Public Health Service Act (the "PHS Act"), as revised by the 21\textsuperscript{st} Century Cures Act, \textbf{requires} the Secretary of DHHS to issue certificates of confidentiality to persons engaged in certain types of federally funded research and empowers the Secretary, in his or her discretion, to issue certificates of confidentiality for other research studies, including studies without any federal funding. The type of research a Certificate of Confidentiality protects is research that collects \textbf{"identifiable, sensitive information,"} a term which is defined broadly as "information that is about an individual and that is gathered or used during the course of research and through which an individual is identified or for which there is at least a very small risk, as determined by current scientific practices or statistical methods, that some combination of the information, a request for the information, and other available data sources could be used to deduce the identity of an individual."\textsuperscript{42}

Certificates of Confidentiality require researchers to withhold the name or identifying characteristics of subjects from individuals not connected with the conduct of the research. Importantly, a Certificate of Confidentiality confers immunity on such researchers from being legally required to produce the identity of research subjects in any federal, state, or local civil, criminal, administrative, legislative or other proceedings. However, the PHS Act contains an important exception that Certificates of Confidentiality do not bar disclosures for secondary research to the extent such disclosures are otherwise consistent with other applicable law, such as HIPAA or the Common Rule. Confusion about this point can result in delayed or disrupted data sharing, particularly if various stakeholders do not understand that data sharing is not precluded thereby.

VI. Content of Data Use Agreement

The preceding sections of this White Paper concerned primarily the application of specific laws and regulations to data sharing activities. The primary concern with parties seeking to share data in each instance under these laws and regulations is first, determining the applicability of the various laws and regulations, and second, having determined which laws and regulations apply to a given data sharing scenario, parties must develop a pathway for the sharing of data in accordance with those legal standards. The topics addressed in the remainder of this memorandum – liability of data contributors, intellectual property, and security standards – are less driven by regulatory


\textsuperscript{42} 42 U.S.C. § 241.
mandates and instead in large part by the negotiating positions of the parties. It may be helpful to keep this distinction in mind when reviewing this Section VI and the following Section VII and to recall that disposition of many of the issues discussed in them will be influenced as much by business concerns as by legal concerns.

A. LIABILITY OF DATA CONTRIBUTORS AND RECEPIENTS

Both Research Institutions that contribute and those that receive research data face concerns about potential liabilities and reputational and institutional risks stemming from the misuse of such data. For example, data contributors may worry about the possible misuse of data contributed for research purposes by parties interested in re-identifying particular individuals from the dataset. Accordingly, data contributors may seek to obtain certain representations and warranties that contributed data will be used only for research purposes. If contributed data are disclosed to persons or entities not under color of law – for example, if they are disclosed with neither consent of the research subjects nor with a waiver of authorization – then data contributors (and even, in some cases, data recipients) could face potential liability under HIPAA and state laws for breach notifications with respect to identified personal data, or in some cases partially de-identified PHI in the form of a limited dataset. In the case of a breach of PHI, which would include an unlawful disclosure of PHI to a third party, HIPAA requires notification to affected individuals, and in certain circumstances DHHS.43 The likelihood that a breach would result in mandatory breach notification to data subjects is lessened when data minimization techniques are employed to reduce the number of identifiers included in a data set. This is because the fewer the number of identifiers, the less probable that PHI would be considered to have been “compromised,” and low probability of compromise based on a risk assessment evaluating certain factors is the test for whether in specific instances, breach reporting is required under HIPAA.

Data recipients could face reputational harm if they receive and use low quality or falsified data. Accordingly, data recipients should put in place quality assurance measures to ensure that potential data contributors are bona fide and that their data merit inclusion in the data sets to be used for secondary research. For example, an apparently low quality, or possibly fictitious, dataset recently led to the retraction of two papers on COVID-19 from prominent journals.44 The researcher generally is best poised to review the quality of the data, but Research Institutions can assist researchers by making available a data review board of experts. Data contributors are sometimes concerned that their data will be misused by data recipients, thus leading to possible claims against the data contributor.

To offset data quality risks and the potential of misuse of contributed data by recipients, parties may seek indemnification provisions in their data use agreements. Data contributors and recipients should review such terms in data use agreements carefully to confirm they offer an appropriate level of protection. Research Institutions that are part of state governments are sometimes unwilling or unable to provide significant, or even any, indemnification to other parties as a result of constitutional or statutory limitations. Positions on indemnification vary by state. For example, the University of Massachusetts takes the position that the Massachusetts Constitution

---

43 45 C.F.R. § 164.404(a).
precludes state institutions from offering any indemnification to contractual counterparties without a supermajority vote of the Massachusetts General Court. On the other hand, the University of Texas and the University of Utah recognize an ability, albeit limited under their states’ laws, to agree to indemnify parties, including in certain research-related agreements. The variability of these positions across states and institutions, and the need to understand whether asserted limits arise from bona fide restrictions of state law or are simply negotiating positions, can lead to delays in perfecting data sharing agreements.

Furthermore, data use agreements may include “limitation of liability” provisions that are designed to limit the parties’ potential damages under the agreement in the event a legal claim or dispute subsequently arises. These provisions may initially appear to be “boilerplate” terms that may not attract the attention of the data contributors or data recipient in the course of negotiating data use agreements. In many circumstances, these provisions may be mutually beneficial to the parties, as both data contributors and data recipients may be incentivized to limit their liability under the agreement. Data contributors may be concerned with liabilities arising from errors or faults in the data provided (or erroneous representations made by the data contributor regarding the data), while data recipients may be concerned with liabilities arising from their use of the data, whether in a permitted or prohibited manner. Parties may seek in these provisions to disclaim expressly any liability for incidental, special, indirect or consequential damages, although losses stemming from breaches of confidentiality are sometimes carved out of the exclusion (i.e., breaches of confidentiality could result in a right to indemnification for incidental, special, indirect or consequential damages). In addition, indemnification obligations may be carved out of the limitation of liability.

Nevertheless, several issues should be evaluated before accepting such limitation of liability provisions. First, data contributors and data recipients should review these liability limitation provisions carefully to confirm whether such disclaimers are unilateral (i.e., limit only the liability of only the data contributor but not the data recipient, or vice versa) or bilateral (i.e., limit the

45 University of Massachusetts at Amherst, Office of the General Counsel, Memorandum re Indemnification Authority of the University (Aug. 17, 2017), https://www.umassp.edu/sites/umassp.edu/files/content/resources/general-counsel/Memorandum-Indemnification_%2817AUG2017%29.pdf (expressing the opinion that the Constitution of the Commonwealth prohibits state institutions such as the University from contractually agreeing to, or providing for, the indemnification of any entity that is not an agency of the Commonwealth, because such promise of indemnification is a “pledge of credit” requiring legislative approval).

46 The University of Texas System, Explanation of Indemnification Limitations and Insurance, https://www.utsystem.edu/offices/general-counsel/explanation-indemnification-limitations-and-insurance (last accessed Jul. 28, 2020) (“Unless the Texas Legislature by resolution permits a suit against the State, the State is not authorized to pay claimants and likewise is unable to indemnify a third party for the expenses associated with defending against such a claimant since such an indemnity would be an unauthorized grant of public monies to the putative indemnitee.”); see also The University of Texas System, Indemnification Sample Clauses, https://www.utsystem.edu/offices/general-counsel/indemnification-sample-clauses#IND6.0; Governmental Immunity Act of Utah, UTAH CODE § 63-30d-101 et seq.; Research Agreement By and Between Surgi Vision, Inc. and the University of Utah, SEC EDGAR filing, https://www.sec.gov/Archives/edgar/data/1285550/000119312511353987/d269096dext1026.htm (“...nothing in this Agreement shall be construed as a waiver of any rights or defenses applicable to the Institution under the Act, including without limitation, the provisions of Section 63-30d-604 regarding limitation of judgments. Subject to the provisions of the Act, University agrees to indemnify, defend and hold harmless Sponsor, its directors, officers, agents and employees against any actions, suits, proceedings, liabilities and damages that may result from the negligent acts or omissions of University, its officers, agents or employees in connection with this Agreement up to the limits of the Utah Governmental Immunity Act.”).
liability of both parties). Second, as a practical matter, data contributors and data recipients that are academic or nonprofit institutions should confirm with their organization’s legal and/or compliance departments whether their organization has developed any policies as to whether they accept, or are permitted to accept as a matter of law, these provisions. Such organizations may have developed policies stating that they will not accept certain limitations of liability that apply to their counterparties, on the basis that a limitation of liability that reduces a counterparty’s potential liability to the organization would effectively mean that the organization would bear any risk that falls outside of such liability limitation. In such cases, state institutions often have developed standard contractual language that disclaims liability beyond the limit of what the applicable state statute permits. Finally, data contributors may propose that breaches of certain terms in the data use agreements, such as breaches of confidentiality, will not be subject to any limitation of liability, based on the principle that such a breach would be so fundamental to the purpose of the data sharing arrangement that the data recipient should, as an ethical and legal matter, be liable to the fullest extent permitted under applicable law.

**B. INTELLECTUAL PROPERTY ISSUES**

Intellectual property rights can often be a highly contentious and heavily negotiated subject of negotiations for a data sharing arrangement. Legal scholars often describe property ownership metaphorically as a “bundle of sticks”\(^47\) or “bundle of rights,”\(^48\) suggesting that ownership of any tangible or intangible property may consist of a bundle of specific entitlements that can be discretely partitioned and further subdivided amongst more than one person or entity.

Accordingly, intellectual property rights and ownership in a data sharing arrangement can be broken down and further examined by both (i) the particular subject matter of the intellectual property at issue and (ii) the particular rights at issue. Furthermore, data contributors and data recipients must consider the means in which such rights are conferred in the course of the data sharing arrangement, such as by transfer of ownership, grant of a license, contingent rights such as an option, right of negotiation or right of first refusal, or various other forms of contractual rights.

i. **Existing Intellectual Property**

Prior to entering into a data sharing arrangement, data contributors will own certain intellectual property rights in and to the data it intends to disclose to the data recipients (such intellectual property, “**Existing Intellectual Property**”). In data use agreements, contributors may seek to provide a significant amount of specificity in describing what rights in and to its Existing Intellectual Property it is conferring to the data recipients.

a. **Ownership of Existing Intellectual Property**

Data sharing arrangements typically do not contemplate that there will be a transfer of any ownership rights in and to the Existing Intellectual Property from the data contributor to the data recipients.

---


recipient. Instead, data use agreements may provide that notwithstanding anything in such agreement to the contrary, the data contributor will retain ownership of its Existing Intellectual Property, and nothing in the agreement will be interpreted to suggest that any such transfer of ownership has occurred.

b. **Licenses under Existing Intellectual Property**

Data contributors may confer limited rights under the Existing Intellectual Property to the data recipients by granting a license with respect to the Existing Intellectual Property to the data contributors. A license differs from a transfer of ownership in that a license only confers limited, contractual rights in and to the Existing Intellectual Property. A license should provide:

1. Whether the rights conferred to the data recipient are exclusive or non-exclusive (i.e., whether the data contributor may grant the same license under the Existing Intellectual Property to a third party);

2. The specific intellectual property rights that are the subject of such license;

3. The scope of such license (i.e., how is the data recipient allowed to use such Existing Intellectual Property);

4. Whether such license is irrevocable or perpetual (i.e., such that the license grant would survive the termination of the data sharing arrangement and cannot otherwise be revoked or rescinded by the data contributor); and

5. Whether the data contributor may sublicense or otherwise transfer its rights under the license to another person (e.g., a third party, a party affiliated with such data contributor, etc.).

For example, data use agreements may include a license grant from the data contributor to the data recipient under the Existing Intellectual Property that: (1) is non-exclusive; (2) clearly and unambiguously describes the scope of the data being shared to the data recipient; (3) succinctly describes the permitted uses of the data (e.g., for academic research purposes only); (4) is revocable and non-perpetual; and (5) is not sublicensable or otherwise transferable without the consent of the data contributor. In the case of (3), data use agreements may limit the use of such data to specific, enumerated activities described in a written research plan, rather than broadly granting data recipients the right to conduct “research,” as this term is vague and may suggest to the data recipients that certain ancillary activities relating to research may be permitted, such as disclosing the data in publications resulting from such research or conducting research activities intended to support the commercialization of such data. In addition, data use agreements may describe, for clarity purposes, what activities do not fall within the scope of the license granted to the data recipients, and therefore, absent any other rights obtained by the data recipient in and to such data, the data recipient would be in breach of the data sharing agreement if it were to conduct such activities.

Finally, to provide additional protection for the data contributor, it may propose a negative covenant that expressly prohibits the data recipient from using the Existing Intellectual Property outside the scope of the license allowed by the data use agreement. A breach of this covenant by
the data contributor would provide a basis for terminating the data use agreement, as well as an additional breach of contract claim, in each case, as a result of the breach of this covenant.

c. Use of Existing Intellectual Property With Third Parties

Data recipients may erroneously assume that they have the unencumbered right to share the data contributor’s data with third parties, such as other collaborating institutions or service providers. Unless, however, the relevant data use agreement expressly permits the data recipient to share the data, data recipients should not assume that they have obtained the right to do so. Rather, unless and until the data have been disclosed to the public, data use agreements will typically contemplate that the data are “confidential information” of the data contributor, and accordingly the data recipient agrees to keep such data strictly confidential and not to use the data for purposes other than those expressly permitted in the data use agreement.

If the data recipient cannot conduct its intended research activities without sharing the data with a third party, then this issue should be raised to the data contributor in the course of negotiating the data use agreement, and the parties should discuss their respective risk tolerance for that onward sharing. The data contributor may permit the data recipient to share the data with third parties under certain limited circumstances or to certain third parties that have been expressly identified, subject to customary limitations and covenants assumed by the data recipient. For example, the data recipient would be required to cause any third party to comply with all obligations under the data sharing agreement (as if the third party were the data recipient), the third party would be required to agree to contractual confidentiality and intellectual property terms that are identical to or no less restrictive than those terms set forth in the data sharing agreement, and the data recipient would assume all liability for the third party’s breach of these obligations and terms. In addition, data contributors may seek to become a third party beneficiary under any such agreement between the data recipient and such third party, which would afford the data contributor the right to enforce such agreement against the third party (and to seek legal redress against such third party in the event of a contractual breach), even though the data contributor is not a party to that agreement.

d. Encumbrances and Subsequent Rights

In the course of granting rights under the Existing Intellectual Property to a data recipient, the data contributor should be mindful of what existing rights it has already granted to other third parties, as well as whether any third parties already possess certain rights under the Existing Intellectual Property. Data contributors may have needed to collaborate in the past with third parties in order to generate the data, or data contributors may have received funding from third parties (including governmental authorities) to generate the data. In each case, the data contributors may have also granted rights under the Existing Intellectual Property as a result of cooperation or collaboration with such third parties.

A common issue encountered by data contributors is the Bayh-Dole Act, which provides that, among other things, in exchange for federal funding provided to certain organizations, including universities, such organizations must grant a non-exclusive, royalty-free license to the United States government under certain inventions arising from the use of such funding.\(^{49}\) While

\(^{49}\) 35 U.S.C. §§ 200-212.
the existence of such a license under the Bayh-Dole Act would not prohibit a data contributor from granting another non-exclusive license to a data recipient, it may present issues in the rarer circumstances where the data contributor is asked for, and seeks to grant, an exclusive license to the data recipient.

Similarly, data contributors should consider the potential economic value of the data and Existing Intellectual Property, and whether such economic value may be affected by entering into a data sharing arrangement. While data sharing arrangements between Research Institutions are often limited to using the data for academic, research, or other non-commercial purposes, the data may nonetheless have utility for commercial purposes, such as for the development of a commercial product. Accordingly, the data contributor may, at a later time, explore entering into transactions with commercial entities that include the license, sale, transfer or other disposition of such data (and related Existing Intellectual Property) to such entity in exchange for monetary or non-monetary consideration. Under such a transaction, the commercial entity may seek comprehensive, exclusive rights to such data and relating Existing Intellectual Property, and the economic value of the foregoing may be correlated with the extent to which such data have been kept strictly confidential by the data contributor, as well as whether other third parties (including any previous data recipients) have already secured rights to use and disclose the data. As a result, conferring non-commercial rights under data to a data recipient may affect the future potential commercial rights to, and value of, the data. Economic value may be further affected by new data and other intellectual property rights that are generated by the data recipient as a result of its use of the data, which is addressed further in Subsection ii below.

ii. New Intellectual Property

Through the course of using data provided by a data contributor, data recipients may make, conceive, discover, reduce to practice or otherwise generate new intellectual property rights ("New Intellectual Property"). Such New Intellectual Property may take the form of patentable inventions, know-how, and/or trade secrets relating to such know-how. In addition, parties may seek to secure exclusive or non-exclusive ancillary rights related to the foregoing, such as the right to disclose the foregoing to the public, or the right to enter into negotiations to acquire, license or obtain commercial rights to the foregoing. While some data use agreements may be completely silent as to any such New Intellectual Property, such rights may also be subject to extensive negotiation between data recipients and data contributors.

a. Opposing Interests in New Intellectual Property

In the course of negotiating a data use agreement, data contributors and data recipients may raise arguments as to (i) in the case of the data contributor, why the data contributor should obtain broad rights in and to the New Intellectual Property conceived by the data recipient, and (ii) in the

\[30\] Jurisdictions have differed on a universal definition for “know-how” (and whether know-how is distinguishable from a trade secret), but in the United States, know-how is typically defined as secret or proprietary techniques and information, including (without limitation) formulas, patterns, devices and compilations of information. Carlos M. Correa, Legal Nature and Contractual Conditions in Know-How Transactions, 11 Ga. J. Int'l & Comp. L. 449 (1981). Available at: https://digitalcommons.law.uga.edu/gjicl/vol11/iss3/3.
case of the data recipient, why the data contributor should receive limited rights (or in some cases, no rights at all) in and to such New Intellectual Property.

Data contributors may raise the arguments that (1) had the data contributor not shared its data with the data recipient, the data recipient would have not been able to generate the New Intellectual Property, (2) as a matter of economic fairness, if the data contributor is granting rights under its data to the data recipient (and in many cases, at little or no cost to the data recipient), then the data recipient should be obligated to grant reciprocal rights under its New Intellectual Property back to the data contributor, as the data recipient would otherwise be acting as a “free rider,” contrary to the spirit of academic collaboration, (3) if the data contributor were to receive no rights under the New Intellectual Property from the data recipient (and the data presumably required significant time, effort and costs to generate), then such an arrangement would create a disincentive for the data contributor to share (or in some cases, even generate) additional data in the future, and (4) the data recipient should not have the right to restrict the data contributor’s freedom to operate with respect to using the data and/or the Existing Intellectual Property (as a result of the data recipient inventing New Intellectual Property that is necessary or useful to exploit such data or Existing Intellectual Property).

Meanwhile, data recipients may raise arguments that (1) because research activities will often involve reviewing and analyzing a number of data sets from different organizations, a data contributor cannot prove with reasonable certainty that the data recipient would have not been able to generate its New Intellectual Property without the use of that particular data contributor’s data, (2) because the data recipient may be accessing the data for academic, non-commercial purposes, arguments as to economic fairness are inappropriate or irrelevant, as the data recipient does not stand to gain any commercial value from the use of the data,\(^5\) and (3) there would also be a “chilling effect” on the data recipient’s use of the data if it knows that any New Intellectual Property it generates will be subject to contractual encumbrances owed to the data contributor.

\[b. \quad \textbf{Scope of New Intellectual Property}\]

Data contributors and data recipients must first determine an appropriate definition of “New Intellectual Property” itself, before even discussing what rights in and to such intellectual property will be granted to each party. On one hand, a broader definition of New Intellectual Property may consist of any and all intellectual property rights conceived, developed or otherwise generated by the data recipient in the course of using the data contributor’s data or otherwise in conducting the research activities relating to the use of such data. Generally, a broader definition of New Intellectual Property would favor the data contributor, if such definition is used to define the scope of any license or other rights the data contributor would receive from the data recipient with respect to such New Intellectual Property.

Conversely, a narrower definition of New Intellectual Property may consist of intellectual property rights conceived or generated by the data recipient that specifically relate to the data, specifically arise from the use of the data, and/or are necessary for the data contributor to use or practice the data (or the Existing Intellectual Property relating thereto). Accordingly, if a research project consisted of analyzing and reviewing a number of data sets from a number of sources, not

---

\(^5\) Commercial rights relating to the data and New Intellectual Property are addressed further in Section d below.
all intellectual property rights conceived or developed by the data recipient in the course of conducting such research project may be considered New Intellectual Property (as certain of such intellectual property rights would not “specifically relate” to the data). In such a scenario, however, it may be difficult for parties, in practice, to draw a fine line between what data “specifically arose” from the use of certain data sets versus others, or what intellectual property rights “specifically relate” to certain data sets versus relating to multiple data sets. We provide in Section 1 of Appendix B sample language that can be included in data use agreements to define “New Intellectual Property.”

c. Rights to use New Intellectual Property

Once a definition of New Intellectual Property has been agreed to by the data contributor and data recipient, the parties will need to discuss what rights the data recipient will grant to the data contributor under New Intellectual Property. As in the case of Existing Intellectual Property, data recipients may also grant a limited license (rather than a transfer of ownership), or a “grant back license,” under its New Intellectual Property to the data contributor. Similar issues arise in determining the scope of such license, e.g., (1) whether such license will be exclusive or non-exclusive, (2) the scope of the rights to use such New Intellectual Property, (3) whether such license is irrevocable or perpetual, and (4) whether such license is expressly sublicensable or transferable.

The scope of the grant back license may determine the outcome of the other issues described above. For example, the data recipient may agree to grant to the data contributor a license to use the New Intellectual Property for any purpose. If data recipients, however, are concerned with the broad nature of this license, it may instead propose that the data contributor may only use the New Intellectual Property for certain limited purposes (e.g., non-commercial or research purposes). As with many of these issues, its outcome may be determined by the relative leverage or bargaining power between the two parties.

Once the scope of such license has been agreed upon, parties must also determine whether such license will be exclusive or non-exclusive, and whether such license will be revocable or perpetual. If the grant-back license is not irrevocable or perpetual, then the license can be limited to a certain time period or terminate simultaneously with the termination of the applicable data sharing agreement. Finally, parties may also determine whether such license can be sublicensed by the data contributor, or whether such sublicensing rights may be subject to certain limitations. Each of the foregoing positions will be dependent upon the context of the relationship and the data at hand, including the proprietary nature of the data and relating Existing Intellectual Property, as well as their potential commercial value.

As a caveat to the above, parties may conflate a license from a data recipient under New Intellectual Property with an obligation for the data recipient to disclose New Intellectual Property to the data contributors. In actuality, the two concepts are distinct from one another, as a licensor can grant a license under certain intellectual property rights without having the affirmative obligation to disclose tangible forms of such intellectual property to the licensee. As a result, data contributors and data recipients should discuss whether any such form of disclosure will be appropriate given the relationship at hand and, if so, how rapidly the disclosure must be made and how fulsome it must be and in what form. We provide in Section 2 of Appendix B sample license
language in which a data recipient would grant rights to the data provider under its New Intellectual Property.

d. Contingent Rights and Commercial Rights

If the data contributor is willing to accept certain limited rights to the New Intellectual Property, whether by limiting the definition of New Intellectual Property or limiting the scope of its grant-back license from the data recipient, then the data contributor may also seek certain additional ancillary rights in and to the New Intellectual Property.

While the data sets may have initially been generated for academic, non-commercial purposes and will be subsequently used by the data recipients for similar, non-commercial purposes, the data (and both the Existing Intellectual Property and New Intellectual Property relating thereto) may nonetheless possess potential commercial value. According, researchers and technology transfer offices at each such organization may later seek to convey such rights to a for-profit entity, which in some cases may be founded by or affiliated with the researchers themselves. As a result, data contributors may seek to capture these commercial rights relating to New Intellectual Property, as the value of the data and Existing Intellectual Property may be significantly diminished if New Intellectual Property is to be commercialized. Some potential forms of contingent rights relating to this issue include:

(1) An exclusive option to obtain an exclusive commercial license under such New Intellectual Property;

(2) An exclusive option to enter negotiations for such an exclusive commercial license;

(3) A right of first negotiation to negotiate such an exclusive commercial license; or

(4) A right of first refusal to obtain such an exclusive commercial license.

Option (1): Exclusive option to obtain a license

Option (1) above provides the data contributor an exclusive, automatic right to receive a commercial license under the New Intellectual Property at its election. The terms of such commercial license will have already been pre-negotiated and pre-determined by the Parties, and until the data contributor exercises the option (or if the option earlier expires), the data recipient would be prohibited from granting the same commercial license to another third party, or any other rights to a third party that would conflict with the option granted to the data contributor. Exclusive options may be time-bound in nature and would therefore will expire within a specified period of time (as agreed to by the data recipient and data contributor in the data use agreement). Of the contingent rights listed above ((1)-(4)), option (1) would provide certainty to the data contributor, as

52 In the context of a license grant, the terms “commercial” and “non-commercial” are often undefined in the applicable data use agreement or license agreement. Accordingly, it is possible that ambiguity may arise in determining what is a “commercial” use of data versus what is “non-commercial” in nature. License agreements with for-profit entities often define “Commercialization” as activities directed to the marketing, advertising, promotion, distribution or sale of a product. Some licensors, however, may seek to define “commercialization” as any activities with a for-profit entity, which would accordingly prohibit the licensee from sharing the data or relating intellectual property rights with any for-profit entity.
it will know from the outset of the data use agreement that it can obtain rights to the New Intellectual Property at its election. But from a practical perspective, option (1) would also require the parties to fully-negotiate a commercial license before any New Intellectual Property is ever conceived, and the economic value of such New Intellectual Property may be difficult to ascertain before it is ever generated.

Option (2): Exclusive option to enter into negotiations for a license

Option (2) differs from option (1) in that the data contributor would not receive an automatic right, at its election to receive a commercial license from the data recipient. Instead, the data contributor would only receive a right to negotiate with the data recipient for such a license. Accordingly, option (2) would also contemplate that the parties would exclusively negotiate with one another for a limited period of time after the data contributor exercises its option, and after such period expires, the data recipient would be free to enter into negotiations with other third parties with respect to such a commercial license. Option (2) allows the parties to determine the economic value of the New Intellectual Property at a later time (which option (1) does not afford to the parties). Option (2), however, would not ensure that the parties will come to an agreement on such a commercial license. Accordingly, data contributors may also request a most favored nations clause in addition to option (2), which is described further at the end of this Subsection (d). We provide in Section 3 of Appendix B sample language in which a data recipient would grant a data contributor an option to negotiate such a license under its New Intellectual Property.

Option (3): Right of first negotiation

Option (3) provides similar rights to the data contributor as option (2), except with respect to the timing of the negotiations between the data recipient and data contributor. Rather than the data contributor having the right to enter into exclusive negotiations with the data recipient at any time during a certain period (as is the case in option (2)), the data recipient would notify the data contributor once it desires to enter into negotiations for a commercial license with a third party (or has received interest from a third party for such a commercial license), at which point the data contributor would have the right to enter into negotiations with the data recipient before the data recipient is permitted to negotiate with such third party. At such point, the data contributor may elect to enter into such exclusive negotiations, typically for a limited negotiation period, or it may decline to enter into such negotiations, such that the data recipient may freely enter into negotiations with a third party. As compared to option (2), option (3) may provide more flexibility to the data recipient in negotiating commercial licenses for its New Intellectual Property. Like option (2), however, option (3) would not ensure that the parties will later come to an agreement on such a commercial license. Accordingly, data contributors may also request a most favored nations clause in addition to option (3), which is described further at the end of this Subsection (d).

Option (4): Right of first refusal

Finally, a right of first refusal is similar structurally to a right of first negotiation, except the sequencing of negotiations between the data contributor and another third party are reversed. Under a right of first refusal, the data recipient is permitted to negotiate the terms of a commercial license with a third party, but prior to entering into such a license, the data recipient must first offer such identical terms to the data contributor, and the data contributor will have the right to match such
terms (and enter into the license) or refuse such terms (and permit the data recipient to enter into the license with the third party). Option (4) may create practical difficulties for data recipients, as a third party may not be incentivized to enter into negotiations with the data recipient if it knows that any terms it proposes must be brought to the data contributor before the third party can enter into such a license.

Each party’s willingness to entertain the foregoing contingent rights may be dependent upon the type of entity or institution that is entering into the data use agreement (e.g., an academic or nonprofit institution may be less accustomed to a right of first negotiation, as compared to a commercial entity). Accordingly, the data recipient’s legal department or technology transfer office should be notified before any of these rights are granted so that they may consider which rights may be most appropriate given the specific factual and legal circumstances of the proposed data sharing and use.

**Additional Considerations: Most Favored Nations Clause**

In connection with options (2) and (3) above, parties may also propose including a most favored nation clause following the expiration of the applicable option or right of first negotiation. Under a typical most favored nation clause, after the expiration of the applicable option or right of first negotiation, the data recipient would be prohibited from granting a commercial license to a third party under terms that are less favorable (to the data recipient) than those previously offered by the data contributor for a certain period of time. This mechanism is often instituted to prevent the data recipient from unreasonably refusing to enter into the commercial license with the data contributor, simply because it would prefer to grant such commercial license to a third party instead. For example, consider a scenario in which the data contributor and data recipient enter into negotiations for a commercial license, but the parties are unable to enter into such an agreement. If the data contributor last offered the data recipient economic terms that included an upfront license payment of $10,000, then a most favored nation clause would prohibit the data recipient from subsequently granting the same commercial license to a third party for $9,000.

e. **Publication Rights**

Another form of ancillary rights relating to New Intellectual Property is each party’s rights to publish such data and New Intellectual Property. Both the right to make such publications and the timing thereof can be critical to each party, as any trade secret contained in Existing Intellectual Property or New Intellectual Property will cease to be a trade secret once it has been publicly disclosed,\(^{53}\) and any invention within such intellectual property will cease to become patentable if it has been disclosed to the public.\(^{54}\) If both the data contributor and data recipients are academic or nonprofit Research Institutions, then each party may seek rights to publish the data or intellectual property relating thereto. Therefore, coordination and cooperation will often be required between the parties with respect to publications.

---

53 See, *e.g.*, the Uniform Trade Secrets Act, which has been adopted in some form by 48 state legislatures, as well as the District of Columbia, Puerto Rico and the United States Virgin Islands.

54 35 USC § 102.
Assuming that the data contributor is willing to permit the data recipient to publish the results of the data recipient’s research, data contributors will often seek certain customary review and comment rights of such publication prior to the date on which it is published. While some data contributors may seek a more fulsome right to review and comment on such publications in all respects, a more tailored approach would be to limit such review to (1) the scientific accuracy of the publication, (2) whether any unregistered intellectual property rights of the data contributor are included in such publication, and (3) whether any confidential information of the data contributor is included in such publication. Data contributors may also request the right to remove any such unregistered intellectual property rights or confidential information, or the right to delay the publishing of such publication for a limited period of time in order for the data contributor to seek patent protection (or other relevant protections) with respect to such intellectual property. Moreover, the data recipient may seek to include proscribed time periods for this review process, as well as a deadline for the data contributor to provide its comments, so as not unreasonably to delay publication.55

Other publication issues may include authorship or attribution rights for the data contributor. Some scholars have proposed models of “data authorship” to acknowledge data contributors, but these proposals are normative, and do not necessarily reflect current publication standards of intellectual effort required for determining authorship.56 Thus, the parties may stipulate that they will apply authorship and attribution principles set forth by a neutral, accredited resource, such as the International Committee of Medical Journal Editors’ Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals.57 We provide in Section 4 of Appendix B sample language that can be included in data use agreements regarding potential publication arrangements between data recipients and data contributors.

iii. Termination

Data use agreements should provide that each of the parties involved will have the right to terminate the data use agreement under certain enumerated circumstances. In negotiating the termination provisions, data contributors and data recipients should evaluate both the causes of termination and the effects of termination that result accordingly.

a. Causes of Termination

While specific termination rights may be negotiated between parties on a case-by-case basis, termination rights can fundamentally be bifurcated between termination rights for cause versus

55 With respect to restrictions on publication, an additional consideration is that 15 CFR § 734.8 (often referred to as the “Fundamental Research Exclusion”) permits individuals to perform certain basic research activities without triggering application of the Export Administration Regulations provided that any “technology” or “software” that arises during such research is intended to be published. Restrictions on publication can eliminate application of the “fundamental research exception” unless they are limited to prepublication review to ensure that the publication would not compromise patent rights and prepublication review by a sponsor of research to insure that publication would not inadvertently divulge proprietary information that the sponsor has furnished to the researchers.

56 Barbara E. Bierer et al., Data Authorship as an Incentive to Data Sharing, 376 NEW ENG. J. MED. 1,684, 1,685 (Apr. 27, 2017).

termination rights for convenience. The former provides that a party may terminate the data use agreement as a result of the material breach of the agreement (or in some cases, the breach of a particular provision of the agreement), which termination right will often be subject to an advanced written notice period and an opportunity for the breaching party to cure such breach. The latter provides that a party may terminate the data use agreement at its discretion, subject to a prior written notice period.

Data use agreements may provide both the data contributor and data recipient the right to terminate the agreement for convenience, subject to an advanced written notice period (e.g., 30 or 60 days). In addition to a termination for convenience right, data use agreements may also provide that either party may terminate the agreement as a result of the material breach of the agreement by the other party.

Data recipients may raise concerns that a reciprocal termination for convenience right would favor the data contributor on balance. Assuming that all rights and obligations of each party under the data use agreement will terminate upon termination (which we discuss further in the “Effects of Termination” section below), the data recipient’s research activities could be halted at any time by a data contributor’s exercise of its termination right for convenience. Meanwhile, upon termination the data contributor may lose the right to use the data recipient’s New Intellectual Property (if the data recipient has granted a license to the data contributor under such New Intellectual Property), unless such license is expressly perpetual (such that its grant-back license would survive the termination of the data use agreement). Accordingly, while parties may initially propose a reciprocal termination for convenience right in a data use agreement, data recipients should consider how material the data contributor’s data are to its research as a whole, and whether such research would be impractical or impossible to complete if it were to lose the right to use such data.

b. Effects of Termination

Either party’s willingness to accept flexible termination rights may be contingent upon the effects of termination that are set forth in the agreement. Many data use agreements remain silent on the effects of termination, which would accordingly mean that all rights and obligations of each party under the agreement, including the licenses granted to each party under the Existing Intellectual Property and New Intellectual Property, would terminate concurrently with the termination of the agreement. This approach, however, may create practical concerns for either party.

For example, if a party is in material breach of a data sharing agreement, it may be presented with the difficult decision to either (1) terminate the agreement but also lose its rights under the agreement, or (2) forego exercising its remedy to terminate in order to retain its rights under the agreement. On the other hand, a termination structure that allows the terminating party to retain all of its rights under the agreement may create adverse incentives for a party to seek opportunities to terminate the agreement, in order to obtain an economic “windfall” following termination.

As a hybridized approach to the above, some data use agreements may contemplate that the licenses granted to the parties will not fully terminate or fully survive the termination agreement, but will instead survive in a modified or abridged manner. For example, certain rights of the data recipient may survive the termination of the data use agreement solely with respect to those research
activities that have already been completed as of the effective date of termination—therefore, the data recipient would be prohibited from using the data sets to conduct new research activities, but could continue to exercise some of its rights under the agreement with respect to research activities that have already been completed (e.g., rights to publish the results of such already-completed activities).

C. DATA SECURITY PROVISIONS

In recent years, there has been increase in the prevalence of data use agreements’ containing data security standards, predominantly in agreements related to limited data sets received from the federal government. Note that when limited data sets are shared pursuant to a data use agreement, HIPAA itself requires only a general provision in the data use agreement that the limited data set recipient will “[u]se appropriate safeguards to prevent use or disclosure of the information other than as provided for by the data use agreement.”\(^{58}\) The limited data set provisions of HIPAA do not require the data use recipient, for example, to comply with the HIPAA Security Rule.

The Centers for Medicare and Medicaid Services (“CMS”) has in recent years required a level of protection consistent with the National Institute of Standards and Technology (“NIST”) moderate impact controls when providing limited data sets to external parties. NIST’s low, moderate, and high impact categories originated from its Federal Information Processing Standards publication 199 for federal information system security.\(^{59}\) More recently, private organizations have begun pushing their data recipients to accept similar standards. Various factors appear to motivate private organizations to press for certain standards. Some organizations have become more sophisticated about NIST and ISO standards, and others appear already to be under certain obligations to maintain the “moderate” standard.

Compliance with the NIST moderate impact controls can prove challenging for many organizations, and each institution must analyze the requested standards in light of its own data security infrastructure and sophistication. This is a highly technical exercise that is typically conducted not by attorneys, but by the information security staff of a given Research Institution. Some organizations rely on internal staff to crosswalk the standards themselves, whereas others hire a security consultant to conduct this exercise for them, sometimes the same security consultant who assists with a HIPAA covered entity’s “security risk assessment,” which is a required element of a HIPAA compliance program.

VII. Process Considerations for Review of Data Use Agreements

Data agreements can involve high transaction costs. Some of these arise endogenously from the data use agreements around which negotiations center, and other costs emerge exogenously from the review processes in place at Research Institutions. Typically, both sources of delay and cost can be reduced by the development and widespread circulation of robust, well-drafted templates among the likely pool of Research Institutions that will participate in research as data contributors or recipients. Widely circulating such templates would reduce the need to re-negotiate

---

\(^{58}\) 45 C.F.R. § 164.514(e)(4)(ii)(C)(2).

each agreement so extensively when each new data sharing opportunity arises or as each new data sharing request is received.

Research Institutions’ autonomy results in their following different approaches to handling the review of data use agreements. These agreements are handled variously across institutions by (i) offices of technology transfer, (ii) the office of general counsel, or (iii) a separate sponsored projects office. In addition, when PHI is involved, sometimes the Research Institution’s HIPAA Privacy Officer must review the agreement. Data sharing typically is expedited when institutions develop clear internal lines of responsibility for data use agreements so as to avoid the situation in which the agreement, and responsibility for it, is shuffled among multiple offices. Research Institutions that frequently collaborate can establish work groups and encourage one another to delineate clear processes for the review of data use agreements as a manner of improving the speed of agreement review. In addition, it can be helpful for members of the office of technology transfer, which focuses on intellectual property issues but also often conducts “first line” review of data use agreements, to be well-versed in HIPAA and human subjects research regulations. Such offices may lack familiarity with these key regulations, and thus when questions arise regarding HIPAA or human subjects research regulations, the technology transfer office must “escalate” the agreement to the office of general counsel and/or the privacy office, which can lead to significant delays in finalizing an agreement.

VIII. Third-Party Intermediary Models of Data Sharing

There are two primary models by which third-party intermediaries facilitate data sharing—as (i) a data aggregator or platform that collects, structures and holds the data or (ii) an organizer of a data transfer network. An example of the former model is the Vivli Platform,61 and an example of the latter model is the National Patient-Centered Clinical Research Network (“PCORnet”).62 In short, the involvement of a third-party intermediary does not greatly change the legal analysis outlined above, although such intermediaries must identify how they fit within the relevant legal framework. We address both scenarios below.

A. DATA AGGREGATOR PLATFORM

If the platform is created as a separate legal entity to house the data, we would not expect that it would be a HIPAA covered entity. Because research, and sharing data for research, are not considered to be within the treatment, payment, or health care operations purposes for which PHI may be used under HIPAA absent patient authorization, covered entities that contribute PHI to a platform require a separate legal basis to permit the data sharing, just as they would if sharing the PHI directly with a researcher at another institution or using the PHI internally for research. When authorization is relied upon as the basis for sharing the PHI, the language should be ensured to be broad enough to permit the platform to hold the PHI before sharing it with researchers. To the extent that the data to be shared include Part 2 records, to which more restrictive requirements apply, it would be necessary to identify to patients the platform as an additional entity that might

---

60 Sponsored projects offices are sometimes themselves subdivided between industry-funded and government-funded research.
62 See Data-Driven, PCORnet (last accessed Aug. 11, 2020).
receive the records. Also, if the platform were formed as a for-profit entity, its holding of the data could implicate the CCPA, which could require affording certain rights to data subjects, including limitations of sale of data and complying with access and deletion requests.\(^63\) Whether IRB approval and other Common Rule requirements for research would not apply to the data aggregator’s activities would turn on whether the platform is itself “engaged” in research or whether it is merely performing a service for those that are “engaged” in the research. From a HIPAA standpoint, if the data aggregator’s function is merely to store data for research purposes, it may function as a business associate, whereas the parties that actually conduct research using the stored data would not be considered business associates.

While the use of a third party platform or intermediary may initially appear to be a convenient resource for a data contributor or data recipient, the addition of a third party adds additional complexities from a contractual perspective. In most cases, the platform will enter into a data contribution agreement with the data contributor and a data use agreement with the data recipient. Meanwhile, the data contributor and data recipient will not enter into a direct agreement with one another, so neither party has contractual privity with respect to the other. Because the data recipient and data contributor lack contractual privity with one another, the platform will be responsible for causing the data recipient and data contributor to comply with their contractual commitments regarding the data sets, and each such party may typically seek recourse (with respect to any contractual claims) only against the platform if the other party fails to comply with such terms. This issue can be mitigated by the inclusion of a third party beneficiary clause in either agreement, which would provide that a third party to an agreement (e.g., a data contributor in the case of the data use agreement between the data recipient and the platform) can enforce the terms of such agreement against one of the parties to the agreement.

The inclusion of a platform aggregator may add complexities to the negotiation of a data use agreement. Because the data contributor will not be negotiating directly against the data recipient, the platform will need to ensure that both the data contributor agreement and data use agreement include terms that are acceptable to the data contributor, the data recipient and the platform itself, even though not all three of these entities will be parties to each agreement. In addition, if the data use agreement is entered into after the data contributor agreement or vice versa, then the platform may not be able to agree to terms that conflict with such already-executed agreement, which may delay, or in some cases even halt, negotiations between the platform and either the data recipient or data contributor.

**B. DATA SHARING NETWORK WITH CENTRAL ORGANIZER**

Even when a central entity does not serve as a data platform, sometimes central entities have held data sharing initiatives to facilitate and streamline their sharing efforts. In such models, the data often are transmitted from the data contributor institution to the data recipient institution, in which the same considerations discussed in the immediately preceding subsection with respect to Data Aggregator Platforms would apply. However, if the central entity provides only support to the data contributors and recipients, and does not access, receive or transmit the data, the analysis outlined above in the body of the memorandum—in which the data contributors and data recipients directly exchange data and must ensure compliance with applicable laws and regulations—remains

\(^63\) 11 CAL. CODE OF REGS. §§ 999.306, 999.313.
unchanged. Such networks often operate through the entry of data use agreements to permit the sharing of limited data sets between the different organizations that participate in the network.

IX. Conclusion

Sharing of data for ML secondary research raises a number of legal, operational and contractual challenges. Because of these challenges, Research Institutions that contribute and receive data for such research purposes confront, both within their own institutions and in counterparties, many points of confusion related to many disparate issues. However, legal avenues exist to permit such sharing, and better awareness of these avenues, and appropriate operational design around them, can facilitate data sharing initiatives.
APPENDIX A

SAMPLE CONSENT AND AUTHORIZATION PROVISIONS
FOR SECONDARY RESEARCH

This Appendix demonstrates how a primary study research consent form can be revised to incorporate language favorable to sharing primary study data for secondary research studies. Terms should be conformed to those used in the primary study consent form. We have [bracketed] those terms that typically differ in consent forms. Blue underlined text indicates text that should be inserted. Red strikethrough font indicates text that should be deleted.

Consent Form

[Study Site and Sponsor] may access and use your information and biospecimens to:

• Fulfil legal and regulatory obligations to ensure the study is conducted according to good clinical practice
• Make required disclosures to institutional review boards or government or regulatory authorities;
• Seek approval from government or regulatory authorities to market [study drug, device, medicine];
• Guarantee the integrity of the study;
• Ensure high standards of quality and safety of [Sponsor’s] products to advance public health and scientific research in the public interest;
• Publish the results of this and other studies; and
• Improve the quality, design and safety of this study and other research studies, including developing diagnostic products and tools.

Data and biospecimens collected in this study will not be used for any purposes other than the study for which they were collected. OR Following completion of the present study, data and biospecimens collected in this study may be used for research related to cancer. OR Data and biospecimens collected in this study will be used solely for research related to cancer.

Your information may be used by:

• [Sponsor] and its representatives (including its affiliated companies);
• People and/or organizations providing services to or collaborating with [Sponsor];
• Any organization that obtains all or part of [Sponsor’s] business or the rights to the product under study;
• Other researchers;
• The IRB overseeing this study; and
• Government or regulatory authorities, including the United States Food and Drug Administration and authorities located in other countries.

[Sponsor] and other companies, institutions or researchers may use your [information or biological samples], from which all identifiers that could readily identify you have been removed, in the future to support and advance other scientific research projects, including research related to the development of machine learning. At this time, we do not know the specific details of these research projects; however, your information and any biological samples could be used in combination with data from other sources, not related to you or this study.

Reasonable safeguards will be used to protect your information and/or biological samples used in any future research and may include: (a) limiting access to information and/or biological samples to individuals bound by duties of confidentiality; (b) removing additional identifiers from the data; and (c) obtaining approval of ethical review boards, when required by applicable law.

HIPAA Authorization

By signing below, I authorize my personal health information to be used and disclosed as described [above/in the informed consent form], including for future research studies for which the specific aims are not currently known. I understand I have a right to receive a copy of this [HIPAA Authorization].
APPENDIX B

SAMPLE DATA USE AGREEMENT PROVISIONS REGARDING RIGHTS TO NEW INTELLECTUAL PROPERTY AND PUBLICATIONS

The sample data use agreement terms in this Appendix are provided merely as a basic starting point and should be tailored to meet the specific requirements of the contracting parties. These terms should not be construed as legal advice for any particular facts or circumstances, nor should it be construed to suggest that any particular terms contained herein are “customary” or consistent with “industry standard” approaches given the fact-dependent nature of most agreements.

1. Definition of New Intellectual Property

As between the Parties, Recipient will own all rights, title and interests in and to data, discoveries, developments, inventions (whether patentable or not), improvements, methods of use or delivery, processes, know-how, or trade secrets that are generated, conceived, reduced to practice or otherwise made by or on behalf of Recipient as a result of the conduct of the Project or as a result of the use of the Data (“New Intellectual Property”).

2. License to New Intellectual Property

Recipient agrees to grant and hereby does grant to Provider a perpetual, non-exclusive, fully-paid up, royalty-free, sublicensable, irrevocable, worldwide license under any New Intellectual Property to use for any purpose.

3. Exclusive Option to New Intellectual Property

Recipient agrees to grant and hereby does grant to Provider an exclusive option to negotiate in good faith the terms of an [exclusive], royalty-bearing, worldwide license under any New Intellectual Property to use [for any purpose] (the “Option”). Such exclusive option may be exercised by Provider by providing written notice thereof to Recipient at [any time prior to the end of the Agreement Term] (the “Option Exercise Period”). If Provider so exercises such option in accordance with this Section, then Recipient agrees to negotiate exclusively in good faith with Provider for up to [__] days (or such mutually agreeable longer period) following the date of such exercise notice (the “Negotiation Period”) regarding commercially reasonable terms for such a license. In the event Provider does not exercise its option within the Option Exercise Period, or the Parties are unable to agree to terms regarding such a license during the Negotiation Period, then Provider’s Option shall expire and the terms of this Section shall be of no further force or effect.]


a. Recipient may publish study results, subject to review and comment rights from Provider:

Recipient is encouraged to make publicly available the results of the Project. Before Recipient submits a paper or abstract for publication or otherwise intends to publicly
disclose information about the results of the Project, the Recipient shall submit to Provider a copy of such publication at least [thirty (30)] days prior to its intended presentation or submission for publication. Provider shall have the right to make comments on such publication to review the scientific accuracy of such publication, to review such publication for patentable subject matter, or to request deletion of the confidential information of Provider (including, without limitation, anonymized patient-level data, research specifications or clinical trial/study protocols, reports, specifications, computer programs or models and related documentation, know-how, trade secrets, or business or research plans). Recipient will consider in good faith any such comments provided by Provider, and Recipient shall not include any such confidential information of Provider for which the Provider has requested deletion. In addition, upon written request from Provider within such [thirty (30)]-day period, Recipient shall not submit, publish or present such publication until the Provider is given up to [sixty (60)] days from the date of such written request to seek appropriate patent protection for any proprietary information of the Provider in such publication or presentation that Provider reasonably believes may be patentable.

b. **Recipient cannot publish study results without Provider’s prior written consent:**

The Parties acknowledge that scientific publications must be strictly monitored to prevent any adverse effect from premature publication of the Data or the results of the Project. Accordingly, Recipient shall not publish, present, or otherwise disclose, and shall cause its Authorized Persons not to disclose, any material related to the Data or the Project without the prior written consent of Provider.

c. **Parties determine publication policies at a later time:**

The Parties shall develop policies and procedures ("Publication Policies") for any publication with respect to the results of the Project, which policies and procedures shall be consistent with the Parties’ respective policies and procedures for publication and disclosure of the results of data of this nature. Notwithstanding the foregoing, Recipient shall provide Provider the opportunity to review each of the Recipient’s proposed abstracts, manuscripts or presentations (including information to be presented verbally) that relate to the Project, at least [thirty (30)] days prior to its intended presentation or submission for publication, and Recipient agrees, upon written request from Provider given within such [thirty (30)]-day period, not to submit such abstract or manuscript for publication or to make such presentation until the Provider is given up to [sixty (60)] days from the date of such written request to seek appropriate patent protection for any proprietary information of the Provider in such publication or presentation that it reasonably believes may be patentable. Provider also shall have the right to require that any such proprietary information be deleted prior to such publication or presentation.