# Considerations for Data Sharing Between Health Departments and Health Systems Dataset Holders

October 2017

## Background

Health systems data sources have become increasingly important for public health programs in recent years, both because of insurance coverage expansion under the Affordable Care Act (ACA) and because of incentive programs and federal investments that help providers and programs build their data and informatics capacity. Increasingly, there are opportunities for public health programs to leverage health systems data – including Medicaid claims, All Payer Claims Databases, and Electronic Health Records (EHRs)/Health Information Exchanges (HIEs) – to augment public health surveillance and ultimately outcomes for a number of health issues, including hepatitis C (HCV).

To support health department programs to use health systems data, NASTAD has partnered with the University of Massachusetts Medical School to create a series of technical resources. These resources are intended to help health department hepatitis programs assess opportunities for using health systems data to augment surveillance and assess HCV prevention and treatment access and utilization. All of NASTAD's health systems data resources can be found on the Health Systems Integration Informatics page. For questions or more information about this work, please contact Amy Killelea or Alyssa Kitlas.

## Considerations for Data Sharing

The rapid advancement of health information technologies has led to the increasing availability of large health systems datasets housing public and private insurance claims data and electronic health records (EHRs). These datasets are in the custody of a variety of private and public entities, and can generally be categorized as either encounter-based (originating from health care providers, such as an EHR database in the possession of a hospital chain) or claim-based (originating from health care payers, such as a claims database in the possession of the state Medicaid agency). By analyzing, manipulating or combining these datasets, health departments can improve surveillance efforts, perform quality improvement activities, respond to public health emergencies and disease outbreaks, and increase access to care.

Accessing these datasets is often difficult, however. They generally contain personally-identifiable health information (PHI) and are subject to a variety of statutes, regulations and internal policies, the most significant of which is the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The HIPAA rules apply to all U.S. health insurers and to all U.S. health care providers that transmit data electronically to health insurers for payment purposes, and impose strict rules about when and how such a HIPAA-covered entity is allowed to release PHI to third parties. A health department investigator who understands the legal landscape governing the request of datasets, particularly the impact of HIPAA, will be much more likely to succeed in such a request.

This resource takes a step-by-step approach to help health department investigators to understand, at a high level, the various practical, technological and legal issues that arise from requesting the release of a dataset.

## Step One: Planning for the Data Request:

In order to knowledgably approach a dataset holder with a request for release of or access to a dataset, and to subsequently negotiate an appropriate data sharing agreement, the investigator first must make several determinations.

The Purpose of the Project: What is the ultimate purpose of the project? Many statutory and regulatory schemes governing the protection of data, including HIPAA, condition the disclosure of PHI on the purpose of that disclosure.

The most common purposes for a dataset release are:

1. to provide treatment to a patient
2. to obtain payment for health care services provided to a patient
3. for a HIPAA-covered entity's health care operations, such as financial auditing or quality improvement
4. for research, defined as a systematic investigation designed to develop or contribute to generalizable knowledge
5. for public health purposes including reducing costs, improving outcomes, and broadening access

This resource assumes that health departments most often request datasets for research purposes and for public health activities purposes, and, less often, for the purpose of providing treatment to patients.

Need for Identifiers: Does the investigator need a dataset that contains identifiers? As described more completely below, the HIPAA statutory scheme creates three levels of datasets:

- Fully de-identified: A fully de-identified dataset under HIPAA is one that has been stripped of all data elements that can identify an individual. These elements include obvious identifiers such as names, addresses, full face photographs, and social security numbers but also less obvious identifiers such as geographic subdivisions smaller than a state (with the exception of the first 3 digits of certain zip codes), and all elements of dates associated with an individual, including birth dates and dates of service, except for year – with the caveat that birth years that indicate

an age over 90 must be aggregated into a single category. Note: Race, ethnicity, national origin, gender, gender identity, and sexual orientation are not considered to be identifiers under HIPAA so may be included in a fully identified data set.

The release of a completely de-identified datasets raises no privacy or security concerns, and such datasets are often publicly available. However, the lack of identifiers within de-identified datasets makes it difficult or impossible to link them with other datasets, and the lack of geographical and chronological information makes these datasets unusable for many types of projects.

- Partially de-identified "Limited Dataset" (LDS): An LDS under HIPAA is partially de-identified. It has been stripped of all patient identifiers other than:
  - elements of dates (years, months, days, hours and minutes)
  - geographic subdivisions down to cities and five-digit zip codes
  - ages in years, months or days or hours

See below for a further discussion of the obligations of a HIPAA-covered entity when releasing an LDS.

- Fully-identified dataset: A fully identified dataset under HIPAA contains data elements that either alone or in combination with other data can identify an individual. These datasets are considered to contain PHI.

| Type of Dataset | Legal protections | Potential health department uses |
|---|---|---|
| **De-Identified Dataset** | Few legal restrictions; often publicly available | Reports on aggregate trends by payer for HCV prevalence or prevention and treatment utilization (purpose = research) |
| **Limited Dataset** | HIPAA; state-specific protections may apply | Reports by payer for HCV prevalence or prevention and treatment utilization, broken down by some demographic variables (purpose = public health) |
| **Fully identifiable Dataset** | HIPAA; other state-specific protections may apply | Data matching with HCV surveillance; provider outreach and/or patient-specific linkage to care (purpose = patient treatment) |

Use of the Dataset: Will the dataset be used on its own, with results coming from analyzing and manipulating the dataset? Or will the investigator seek to link two or more datasets to achieve results? Combining two or more datasets is a powerful analytic tool, but involves determining a common element upon which to match the datasets.

It also involves coordinating two or more requests for dataset disclosures and possibly negotiating two or more data sharing agreements.

Identifying the Dataset Holder or Holders: Which entity is in possession of the desired dataset? If the dataset holder is a health care provider or payer, it is most likely covered by HIPAA and the investigator needs to frame the data request accordingly. If the entity is not a HIPAA-covered entity – for example, a state agency which holds an All-Payer Claims Database (APCD) but is not itself a health care provider or payer – the investigator must determine which other statutory scheme applies to the data request, if any.

Intellectual Property: Who will own the intellectual property that results from the project? In some cases, the health department will be allowed legal ownership of the dataset after release, as well as all information derived from use of that dataset. In other cases, the health department might not obtain such independence; the data holder might condition the release of the dataset on its ability to continue to own the released dataset and control the result of the project. This issue is often one of the most vigorously negotiated in the applicable data sharing agreement.

Data Transmission: How will the investigator propose that the dataset be transmitted from the data holder? Particularly for large datasets, the most efficient and reliable method of transmission is often electronic, with the caveat that such transmission must be encrypted, (i.e., through a secure web browser session or via secure File Transfer Protocol (FTP)). It is also feasible for the encrypted dataset to be copied onto physical media such as a hard drive, flash drive or laptop drive and then transported physically to the investigator by mail or courier, with the package being carefully tracked during the process. Of course, the dataset holder will be interested in a transmission method that is least burdensome on its operations.

Data Storage: How will the dataset be stored once it is in the investigator's possession? Datasets containing PHI must be stored securely, preferably in an encrypted state, and all datasets need to be backed up regularly to guard against data loss or corruption. If the data holder continues to assert ownership over the dataset after release, it may seek to impose particular storage specifications in the data sharing agreement.

## Step Two: Understanding Privacy and Confidentiality Requirements

### The HIPAA Disclosure Rules

As described above, a HIPAA-covered entity is always allowed to disclose a dataset if the dataset has been stripped of all identifiers – a fully de-identified dataset. Such dataset no longer contains PHI and is no longer subject to the HIPAA rules. If a health department is able to make use of a de-identified dataset, it is often the easiest dataset to obtain.

If an investigator determines that using a fully de-identified dataset is not feasible, the investigator must understand the

laws that govern the disclosure of the identified dataset.

Since most available health datasets originate from a HIPAA-covered entity – either a health care provider (encounter-based data) or health care payer (claims-based data) – the HIPAA disclosure rules are the most important to understand.

The HIPAA disclosure rules governing PHI may be summarized as follows:

1. Datasets Containing PHI May Be Disclosed with Patient Consent: A HIPAA-covered entity is always allowed to disclose a dataset containing PHI for any purpose if it has obtained patient consent to do so. Unfortunately for health departments, obtaining such patient consent is often burdensome for the HIPAA-covered entity, and, in the case of large datasets, generally not feasible.
2. Datasets Containing PHI May Be Disclosed Without Patient Consent If the Disclosure is Required by Law: A HIPAA-covered entity may disclose a dataset containing PHI to a third party if the disclosure is required by law. This resource assumes that no law requires a HIPAA-covered entity to disclose PHI to the health department which seeks it, and that therefore this rule is inapplicable.
3. Datasets Containing PHI May Be Disclosed Without Patient Consent for Routine Business Purposes: A HIPAA-covered entity is allowed to disclose a dataset containing PHI without patient consent for these routine business purposes: (a) to provide treatment to a patient; (b) to obtain payment for services rendered; (c) to perform its own health care operations[1] and, in some situations, (d) to assist with the health care operations of another HIPAA-covered entity. This resource assumes that while health departments are (in most cases) not themselves HIPAA-covered entities (because they do not submit electronic bills to health insurers), they sometimes do perform activities that provide treatment to a patient.
4. "Limited Datasets" May be Disclosed Without Patient Consent for Research and Public Health Purposes: A HIPAA-covered entity is allowed to disclose a partially-de-identified dataset – an LDS – without patient consent for public health or research purposes, but only if the receiving entity signs an "LDS Data Use Agreement" with specific provisions in it. See below for the mandated contents of the LDS Data Use Agreement.
5. Fully-Identified Datasets Containing PHI May Be Disclosed Without Patient Consent for Research Purposes Under Certain Circumstances: A HIPAA-covered entity is allowed to disclose a fully identified dataset without patient consent for research purposes in these circumstances:

[1] HIPAA defines the term "health care operations" as the administrative, financial, legal, and quality improvement activities necessary to run its own business and to support the core functions of treatment and payment.

| When a HIPAA-Covered Entity Is Allowed to Disclose a Fully Identified Dataset for Research Purposes | | |
|---|---|---|
| **For an In-Custody Review Preparatory to Research** | **To Perform Research on Decedents** | **Pursuant to an IRB Waiver of Consent** |
| **Disclosure is allowed for the purpose of preparing a research protocol but only if the PHI does not leave the possession of the HIPAA-covered entity** | Disclosure is allowed if all patients in the dataset are decedents | Disclosure is allowed if an institutional review board (IRB) has authorized a waiver of the consent requirement |

A health department investigator should also be aware of an additional obligation imposed on the HIPAA-covered entity when disclosing PHI without consent for research purposes: it must track these disclosures so that it can inform an individual who requests an "accounting of disclosures" that such disclosure was made. The administrative burden of tracking released information for an "accounting of disclosures" can make a HIPAA-covered entity reluctant to release identifiable datasets for research purposes.

Note that there is no provision to allow a HIPAA-covered entity to release a fully identifiable dataset to a third party for public health purposes. A health department which seeks a dataset for public health purposes will need to request a de-identified dataset or an LDS, or will need to obtain a dataset from an entity which is not a HIPAA-covered entity.

## Other Privacy and Confidentiality Laws Applicable to the Requested Disclosure

HIPAA is not the only federal law that applies to personally-identifiable health information. For example, the Federal "Confidentiality of Substance Use Disorder Patient Records" regulations at 42 CFR Part 2 strictly limit how a holder of such records may disclose them to third parties. These regulations allow the disclosure of such data without patient consent only in certain medical emergencies, for limited audit and evaluation purposes, and for certain research purposes. The research disclosure rules are similar to the HIPAA rules in that they require IRB approval, but they are not exactly the same. A detailed discussion of these rules is beyond the scope of this resource.

Many states have enacted laws granting additional protection to substance use disorder records, and records related to other highly sensitive conditions such as HIV status, sexually transmitted diseases and/or mental health. Health department

investigators should assess the applicable state laws for data disclosure.

In addition, mandatory reporting laws and other administrative processes have resulted in state and federal agencies holding datasets containing PHI which are generally not subject to HIPAA or other laws mentioned above – a state agency holding an APCD is a perfect example of this, as is a state agency in possession of a database resulting from a law requiring hospitals to submit information about emergency department utilizations. These datasets are often easier to obtain than a dataset subject to the HIPAA rules, particularly in a de-identified state.

## Step Three: Understanding Data Sharing Agreements

Frequently – but not always – a dataset holder which releases information to a third party will require that third party to execute a legal agreement. The purpose, contents and title of that legal agreement will vary according to the particular circumstances. This resource will divide these legal agreements into two categories: the two types of agreements that are mandated by HIPAA and the other types that are not.

### Data Sharing Agreements Mandated By HIPAA

HIPAA requires a covered entity to enter into a legally mandated type of agreement in two situations: (1) A "Business Associate Agreement" when releasing PHI to a subcontractor or agent that it has hired to perform a function on its behalf and (2) An "LDS Data Use Agreement", when releasing an LDS to a third party for research or public health purpose.

### The Business Associate Agreement:

Whenever a HIPAA-covered entity hires a subcontractor to perform a function on its behalf, and the subcontractor requires PHI to perform that function, the subcontractor becomes a business associate of the HIPAA-covered entity, and the business associate must enter into a "Business Associate Agreement." HIPAA mandates the minimum requirements for such agreement, which include, for example, the following provisions:

- a provision establishing the permitted and required uses and disclosures of PHI
- a provision providing that the business associate will not use or further disclose the PHI other than as permitted or required by the contract or as required by law
- a provision requiring the business associate to report to the covered entity any use or disclosure of the PHI not provided for by its contract, including incidents that constitute breaches of PHI
- a provision requiring the business associate to make available to the federal government its internal practices, books, and records relating to the use and disclosure of the PHI
- a provision requiring the business associate to return or destroy the PHI at the termination of the contract
- a provision requiring the business associate to ensure that any subcontractors that will have access to PHI agree to the same restrictions and conditions that apply to the business associate

It is unlikely that a health department would ever be asked to sign a Business Associate Agreement when receiving a dataset from a HIPAA-covered entity, since the health department, when using a released dataset, would not be performing a service for the HIPAA-covered entity.

The LDS Data Use Agreement: A health department is much more likely to be asked to sign an LDS Data Use Agreement than a Business Associate Agreement. As described above, a health department who requests an LDS for public health or research purposes will be required to sign LDS Data Use Agreement as a condition of receiving the LDS. An LDS Data Use Agreement must contain the following provisions:

- a provision setting forth how the receiving entity may use and disclose the LDS
- a provision identifying the individuals who may use or receive the LDS
- a provision prohibiting the receiving entity from using or disclosing the LDS other than as allowed in the agreement or by applicable law
- a provision requiring the receiving entity to establish appropriate safeguards to prevent a non-permitted use or disclosure
- a provision requiring the receiving entity to report unauthorized uses or disclosures to the HIPAA-covered entity
- a provision requiring the receiving entity to require its subcontractors and other agents who receive a copy of the LDS to agree to the same terms and conditions the receiving entity has agreed to
- a provision prohibiting the receiving entity from attempting to identify the information in the LDS or contact the individuals

An LDS Data Use Agreement is not limited to these particular provisions, however, and in practice, these types of agreements often include additional provisions governing such topics as including intellectual property ownership, specifics of data transmissions, termination provisions and others. See below for a more thorough discussion of these types of provisions.

## Data Sharing Agreements NOT Mandated By HIPAA

Other than what is legally mandated by HIPAA (and perhaps other statutory schemes beyond the scope of this resource), a legal agreement is not required when a data holder releases a dataset to a third party. However, many data holders do, as a matter of policy, require the data recipient to sign a legal agreement before taking possession of a dataset, especially if the data holder wishes to retain rights to the dataset after it has left the holder's possession. Various data holders title these agreements in various ways: they might call it a Data Use Agreement, a Data Release Agreement, a Data Sharing Agreement, a Data Exchange Agreement, a Trading Partner Agreement, a

Memorandum of Understanding[2], or sometimes, just plain Agreement. In reality, it is never consequential how these non-mandated agreements are titled, [3] it is the contents of those agreements that are important to the parties. In a non-mandated agreement, the contents are entirely subject to negotiations between the two parties to the agreement, and such contents will depend on the particular needs and desires of the two parties. In general, however, provisions are subject to some general principles, described below. Whether a health department is negotiating additional provisions in a HIPAA-mandated agreement, or all included provisions in non-mandated data sharing agreement, the following are considerations for provisions to include:

## Intellectual Property

The intellectual property aspects of the health department's project are subject to negotiation no matter which type of agreement is at issue. It is critical for all parties involved in the data sharing to reach a clear understanding about the intellectual property rights that attach both to the original dataset and to the work product that is derived from the health department's use of the dataset. Specifically, the data sharing agreement should state:

- Whether the disclosing entity will continue to own the data after it is released, and, if not, whether it will have any rights to the data at all
- Which entity will own the work product that is derived from the receiving entity's use of the data, and what may be done with that derivative data
- Whether the data recipient will be acting independently or as an agent of the disclosing entity when using the released dataset

No matter how the intellectual property issues are resolved, it is almost always useful for the data sharing agreement to specify how the dataset will be transmitted to the receiving entity:

- What technology will be used for the transmission of the dataset
- When will the transmission be considered complete
- Which entity is responsible for the data during transmission
- What happens if the transmission fails

## Restrictions on Data Use

If the disclosing entity retains rights to the dataset after release or to the work product that is derived from the receiving entity's use of the dataset, the disclosing entity may wish to use the data sharing agreement to impose restrictions on the receiving entity's use of the data, such as:

- How the receiving entity may use and disclose the dataset while performing its investigation
- What security controls the receiving entity must apply to the data, including where the data will be stored, whether encryption is

[2] In some states, a data sharing agreement between two different state agencies within the same state is called a "Memorandum of Understanding." This term is in recognition of the fact that the agreement is not technically a legally enforceable contract but rather a writing that sets forth an understanding between two different departments within the same legal entity.

[3] For convenience, this resource will refer to any non-HIPAA-mandated agreement as a "data sharing agreement."

required, whether the data may be accessed from portable devices, whether different access roles are required, etc.
- How the receiving entity will educate its workforce about the privacy and security controls that apply to the data
- Whether the disclosing entity will be given the right to inspect the premises where the data are stored, and or to inspect books and records that are related to the receiving entity's use of the data
- Whether the receiving entity will be allowed to subcontract any of its allowable uses and disclosures to a third party and if so, how and under what circumstances

### Breach and Termination

Finally, most data sharing agreements contain standard contract language governing breach and termination of the agreement:

- Whether the receiving entity will be responsible for the consequences of any improper use, inadvertent disclosure, or privacy breach related to the data, and the extent of such responsibility
- Whether, if there is an alleged breach, the entity will be given an opportunity to fix the alleged breach
- A description of the circumstances under which either party may terminate the agreement and whether and when a notice period is required before termination
- Whether the receiving entity will be required to return or destroy the data when the permitted uses are concluded
- Whether the protections that apply to the data survive the termination of the agreement

Health departments who are seeking the release of a dataset should carefully consider all the above issues and develop a strategy about how these issues will be handled in any data sharing agreement that might be required by the data holder.

## Conclusion

There are increasing numbers of valuable health datasets available to health departments seeking to build their data and informatics capacity. These datasets generally originate either from providers and are encounter-based or from payers and are claim-based. Most of these datasets remain in the possession of providers and payers, both of which are HIPAA-covered entities. Sometimes though through various administrative processes or mandatory reporting laws, these datasets are in the possession of state or federal agencies which are subject not to HIPAA but to other statutory schemes.

Before approaching a data holder to ask for the release of a dataset, a health department needs to plan carefully, identifying how it plans to use the dataset, determining what level of identifiers it needs in the dataset, and defining how the dataset will be transmitted and stored.
It also needs to understand the laws, particularly HIPAA, governing how the data holder will be permitted to release the dataset.

Finally, health departments need to understand when they might be required to sign a legally mandated Business Associate Agreement or LDS Data Use Agreement, and when they might be required to sign a non-mandated data sharing agreement. In any of these situations, the health department must understand the various provisions that could be presented in these agreements, including intellectual property rights, limitations on use and disclosure, and post-termination responsibilities.

## Acknowledgments

This document was developed by NASTAD. NASTAD is a leading non-partisan non-profit association that represents public health officials who administer HIV and hepatitis programs in the U.S. and around the world. NASTAD collaborated with the University of Massachusetts Medical School (authors: Deborah Drexler and Abigail Averbach).

NASTAD would also like to thank the individuals at the state and city/county health departments who generously gave their time and content expertise for the development of this resource.

This publication was supported by a grant from Gilead Sciences.

Murray C. Penner, Executive Director
Shanell McGoy, Tennessee, Chair
October 2017