The Office of the National Coordinator for Health Information Technology

Legal and Ethical Architecture for PCOR Data

CHAPTER 2:

LEGAL AND ETHICAL SIGNIFICANCE OF DATA FOR PCOR

Submitted by:
The George Washington University
Milken Institute School of Public Health
Department of Health Policy and Management

TABLE OF CONTENTS

Chapter 2

Legal and Ethical Significance of Data for PCOR

INTRODUCTION

The most basic definition of "health information" is any information concerning the health of at least one person. When considering law and policy, however, the regulated information must be specifically defined. For example, the physical medical record, the content of the record, biological samples taken from a person, and data aggregated from many different people can all be considered "health information," but they may be treated differently under the law. Not all health information is subject to regulation, and information that *is* regulated may be subject to laws that overlap or directly contradict each other.

In order to understand the legal and ethical significance of the different data types used for PCOR and CER, a number of key questions must be asked. There is no single legal framework governing "health information"; rather, information may be subject to one or more statutes and/or regulations depending on the information's specific characteristics. For purposes of applying legal protections and restrictions, health information can be defined based on a variety of characteristics, such as its content, its source, and its form. These characteristics are not mutually exclusive so that multiple overlapping rights and obligations may apply to a particular record or piece of information.

The answers to these questions help determine the legally relevant characteristics of the data. Together, the questions and associated answers provide the foundation for this Architecture and are woven throughout the various components of the Architecture. In practice, the law is applied to specific facts or scenarios. However, there are common themes that provide a more generalized legal analysis and that may be extrapolated to support a broader legal and ethical framework. The themes identified and summarized below address the key elements of health information and provide an outline of the core concepts that support this Architecture. This structure should serve as a tool for consistent application of the Architecture across research data use scenarios for PCOR and CER.

This chapter explains fundamental concepts for organizing data according to categories and types so that legal requirements can be applied. The first section presents questions to ask in order to identify the key characteristics of health information used for PCOR, and the second section gives an overview of the most significant health information data types relevant to PCOR. The key characteristics include identifiability, content, subject, source, access, use/purpose, consent/authorization, security, and legal status. (These characteristics also appear in the Framework in Chapter 4). The data types include clinical data, administrative data, patient-generated health data (PGHD), patient reported outcomes (PROs), genetic information, biospecimens, surveillance data, and quality improvement data. Legal and ethical requirements will vary depending on the type of data sought or held by a researcher. Together, the key characteristics and data types are the fundamental features of data that must be identified to discover the legal and ethical requirements that may apply to the data in question and map the various decisions and actions that must be taken to ensure that the research is conducted in compliance with applicable federal and state laws and ethical requirements.

KEY CHARACTERISTICS OF DATA TYPES FOR PCOR

Identifiability

Identifiability refers to the ability to link information to particular individuals. Health data that contains identifiable information will likely fall within the scope of federal and/or state privacy laws that govern the use and disclosure of health information (e.g., HIPAA, Family Educational Rights and Privacy Act (FERPA), 42 C.F.R. Part 2). Because these laws do not contain a consensus definition of "identifiable," determining whether particular data are identifiable will depend upon the circumstances of its collection and storage as well as its content and what laws apply to its collection and use. Common elements that render data identifiable include names, postal addresses, and social security numbers. Data that have been de-identified are generally afforded less privacy protection and consequently may be easier to access or share.[1]

Considerations for Identifiability

- What information does the data contain? Does this information directly identify individuals?
- Has the data been de-identified? If so, how?
- Can the data be re-identified when combined with data from another source?
- Where was the data collected? How was the data collected?

Content

Content refers to the subject matter or substance of the data. This may include contact, demographic, medical, insurance, web behavior, and/or employment information. Data that contain identifying information will be subject to federal and/or state privacy laws. Data that include information regarding mental health, substance abuse, genetics, and/or HIV status might be subject to additional regulation, depending upon the source of the data and the purpose for collecting or using the data.

Considerations for Content

- Does the data contain health information?
- Does the data contain identifying information?
- Does the data include information on mental health, substance abuse, genetics, HIV, and/or other conditions granted special legal protection?

Subject

Subject refers to the person or thing that is the focus of the data. Human subjects' protection and/or privacy laws may affect the collection/use of any individual's data by limiting the data that can be collected. In particular, these laws may impose stricter requirements for data belonging to members of certain classes, such as minors, prisoners, individuals with limited mental capacity, and pregnant women.

---

[1] *See also* U.S. Department of Commerce National Institute of Standards and Technology (NIST). Appendix A, HIPAA Information De-Identification Reference at p. 3. In De-Identification of Personal Information [Internal Report 8053] (2015). *Available at*: http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf; NIST. De-Identifying Government Datasets: Second Draft [Special Publication 800-188] (2016). *Available at*: http://csrc.nist.gov/publications/drafts/800-188/sp800_188_draft2.pdf

Terms of service agreements may also affect data collection or use by limiting inclusion of certain types of information and/or information about certain classes of individuals.

Considerations for Subject

- Who or what is the focus of the data?
- Does the data contain identifying information about the subject and/or other individuals?
- Does the data pertain to a minor, a prisoner, a legally incompetent individual, or a member of another class that receives special consideration under the law?
- Does the user agreement or term of service agreement address subjects?

Source

Source refers to the person, entity, and/or setting in which the data originated or was collected. Persons that originate data may include: (1) patients (e.g., via a wearable device, patient-reported outcomes (PRO) survey); (2) providers (e.g., recording a measurement in a medical record, collecting biospecimens); (3) health plans (e.g., claims data); and (4) government agencies (e.g., Medicare payments, conducting public health surveillance tasks, collating encounter data into Medicaid data files). Settings in which data may originate include clinics, homes, laboratories, etc. Source may also refer to the persons or entity that shares data with another person or entity (e.g., a data repository, registry, or research network, or clearinghouse that provides data to researchers).

Considerations for Source

- Who generates or collects the data (e.g., a healthcare professional from an individual)?
- What is the setting in which data are collected?
- Is the data collection direct or indirect?
- Is the data generated or collected ancillary to another event (e.g., a clinic encounter), or does the data generation/collection occur as the primary event (e.g., an individual voluntarily submitting his or her data to a research network)?
- Is the data aggregated or combined with data from other sources, and if so, who aggregated/combined the data?

Access

Access (as used here) refers to the ability of a person or entity other than the individual subject(s) of the information to view, create, edit, or share data. Factors that impact a person's ability to access data (accompanied by a brief overview of the impact these factors may have on such ability) include:

- Content of the data (e.g., whether data contains identifying information subject to privacy laws, which will likely restrict access to the information, or includes information that is subject to special protections, such as substance abuse treatment information);
- Reason for accessing the data (i.e., access for certain purposes, such as for research, may require authorization/consent, whereas other purposes, such as treatment or administrative tasks, may not);
- Ownership interests (ownership issues are discussed in more detail below; ownership in general is relevant to access insofar as persons or entities may have the ability to limit access to data in which they have ownership interests); and
- Position or affiliation of the person(s) seeking to access data (privacy laws or data use agreements may limit data access to specified individuals).

Considerations for Access

- Does the data contain identifying information?
- What is the reason for accessing the data?
- What is the legal status of the data?
- What is the position/affiliation of the person seeking to access the data? Is there a legal relationship between the parties (e.g., employment)?
- Is there a contract governing access to the data? If so, what are the terms?
- Are there any reporting requirements associated with release and use of the data?

Use/Purpose

The intended use or purpose of the data collection will affect whether and how the data may be collected and used. Relevant uses/purposes for collecting/sharing data include patient care, research, claims processing, advertising/marketing, and personal uses.

Considerations for Use/Purpose

- What is the reason for collecting the data?
- What is the proposed use for the data?
- Who is collecting/using the data?
- Does the use/purpose involve sharing data with other individuals?
- What purpose was communicated to the subject(s) of the information?
- Who is requesting the data from the data source (e.g., patient, health plan, provider)?
- What disclosures and/or uses of the data are specified in the applicable notice(s) of privacy practices?

Consent/Authorization

Consent/authorization refers to the activities and documentation potentially required of researchers seeking permission to collect, use, or share data about an individual. Whether consent or authorization is necessary will depend upon the content of the data, the party collecting or sharing the data, the purposes for collecting or sharing the data, and relevant statutes and regulations. Consent/authorization procedures generally require notifying individuals of the intended uses and disclosures of their information and having individuals execute a document stating that they consent to or authorize the uses or disclosures of their information.

In general, the term “consent” is used to refer to informed consent to participate in research (a concept governed by the Common Rule). Authorization is used to refer to authorization given by an individual subject of information to an entity to disclose that information to a third party. Authorization is a term used in HIPAA, and here it is used to encompass all similar permissions (e.g., as they apply to Part 2, GINA, etc.).

Considerations for Consent/Authorization

- Is consent or authorization necessary prior to a researcher collecting, accessing, or releasing data?
- Is the data being collected/used for a study that is subject to Institutional Review Board (IRB) approval?
- Has an IRB waived the informed consent procedure applicable to participating in the research?

- Can persons withdraw their consent/authorization?
- Can persons opt-out of particular uses for their information (e.g., commercial use)?
- Did an agent (e.g., parent, person with medical power of attorney) give consent or authorization?
- Was consent/authorization a condition of receiving something else (such as medical treatment or payment)?
- Was consent/authorization combined with permission for something else (such as medical treatment)?

Security

Security refers to the means by which data is protected from unauthorized use or access. Security measures generally include technical, administrative, and physical safeguards. Technical safeguards include items such as encryption, firewalls, passwords, antivirus software, and SSL/TLS transmission. Physical safeguards include measures that limit an individual's access to facilities, workstations, and devices that house data or may be used to access data (e.g., policies that limit server room access to authorized personnel). Administrative safeguards include plans and policies for identifying security risks, preventing security breaches, monitoring security, remedying security breaches, and training employees on proper security procedures.

Considerations for Security

- What is the form or medium of the data?
- Where is the data held?
- Who can access the data?
- What technical, physical, and administrative safeguards are employed to secure the data?
- What are the researcher's data management obligations once the data has been obtained?

Legal Status

Legal Status refers to rights and responsibilities related to the data that may be triggered by ownership rights, agency principles, and/or contractual obligations. Legal status determines who may assert rights to that information. Individuals or entities with ownership interests may grant, restrict, or deny access to information. Contractual obligations, such as data use agreements, vendor contracts, or terms of service agreements, may apply. Principles of agency may give a researcher the rights and obligations of the healthcare organization that employs him or her. Finally, some state laws, such as consumer protection and patient privacy laws, may confer rights and responsibilities with respect to access to data or data held by researchers.

Considerations for Legal Status

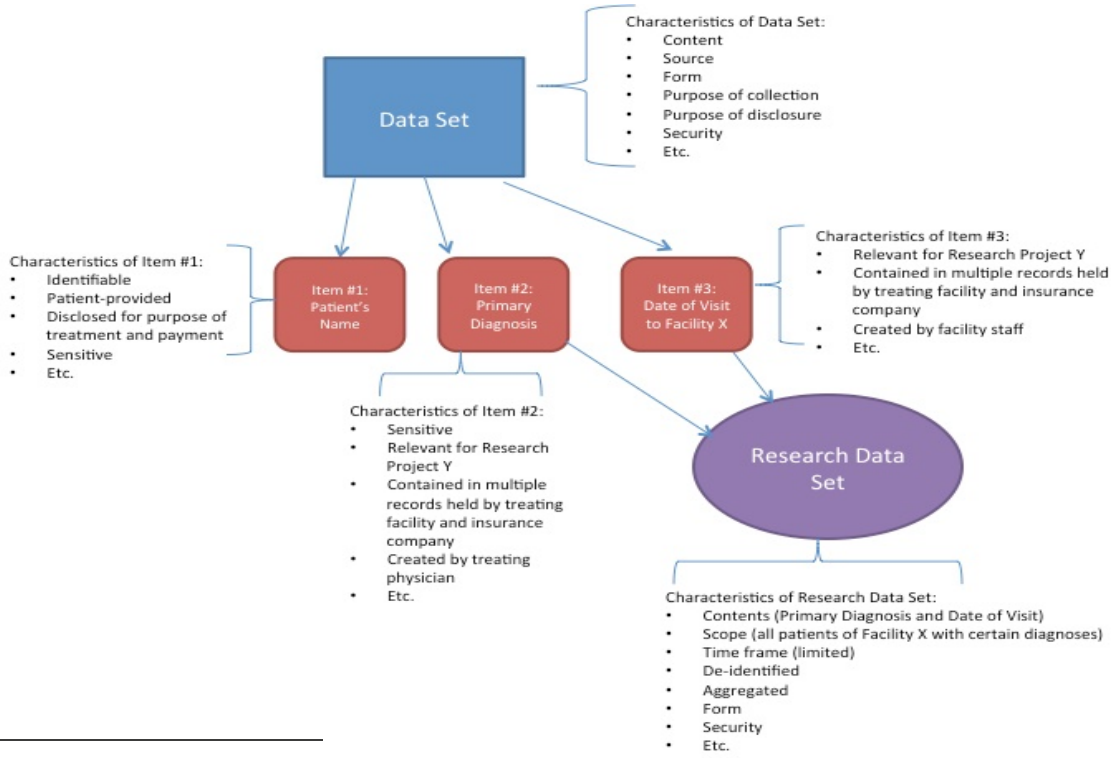
- Who owns the data?
- Are there contracts in place that govern how the parties to the contract or their agents may use the data?
- What is the position/affiliation of the person seeking access to the data? Is there a legal relationship between the parties (e.g., employment)?
- Does the state law grant ownership rights or rights to restrict access?

TYPES OF DATA RELEVANT TO PCOR

Given the scope and depth of PCOR, it necessarily involves the collection and use of a wide range of health data types often from multiple sources. When considering data "types" that are relevant to a legal and ethical analysis, note that any characteristic of a piece of data or set of data may be a common factor that will allow the information to be sorted by that factor as a particular type. The characteristic may be related to the data's content, source, form, or purpose, among other things.

Frequently, attempts to map data types will yield categories that only fit in the context of a particular transaction or reflect just one aspect of the data, such as "claims data" or "patient-generated data." These terms identify the source of the information but don't capture all of the characteristics necessary to map all data relevant to research to all of the legal requirements and ethical principles that may apply. This project maps the legally relevant characteristics of data in addition to defining common data types that reflect one or more common characteristics. The questions described above can be used to draw out the various characteristics of a given data set or individual data elements. As illustrated below, the characteristics may pertain to a collection of data, such as a research data set, or individual elements of the data, such as a particular patient's diagnosis. Individual components of a data set may have a characteristic, such as sensitivity or identifiability, which may not be shared by other components. If the data is collected into a different form, such as by aggregation, the new data set may have characteristics that are different from the characteristics of the source data. Even if the data is drawn from the same field in a database of multiple individuals' data, its characteristics may vary from one individual to the next. For example, the content of a field such as "primary diagnosis" may cause the information to be treated differently if the diagnosis is substance abuse-related.

Figure 1: Data Characteristics[2]



[2] Graphic created by GW Team (2016).

As noted above, a certain characteristic may be identified that pulls together many different pieces of data into a common data "type." This common characteristic is typically some significant aspect of the data, such as source or purpose. Defined below are the common data types that are identified in the health policy literature.

Clinical Data

Clinical data refers to data related to a patient's health, health status, and/or treatment that are collected orally or electronically via a patient-provider interaction in a clinical setting. Clinical data also may be collected as part of a clinical trial. A treating provider or research institution maintains clinical data (electronically or on paper), even though the data may initially be generated outside of the physical clinical setting (e.g., via a telehealth visit or a device such as a remote blood pressure monitor). Clinical data may be found in the following locations:

- Electronic Health Records
- Electronic Medical Records
- Paper-based Medical Records
- Biospecimens
- Clinical Trials
- FDA-Regulated Medical Devices and Technologies

Administrative Data

Administrative data refers to data collected and/or used primarily for administrative (e.g., record-keeping purposes, payment purposes). Administrative data typically includes patient demographic information, payment information (e.g., health plan information, claims), and other related information. Similar to clinical data, administrative data may be found in:

- Electronic Health Records
- Electronic Medical Records
- Paper-based Medical Records
- Practice Management Systems

Sources of administrative data include private and public payers (e.g., private health plans, Medicare, Medicaid) and providers.

Patient-Generated Health Data (PGHD)

PGHD is health-related data created, recorded, or gathered by or from patients (or patients' family members or other caregivers) in nonclinical settings.[3] PGHD may include a patient's health history (reported by the patient, family members, and/or caregivers), treatment history (including medications, biometric data, symptoms, and lifestyle choices), and/or other personal health-related information. This information is distinct from the clinical data discussed above (i.e., information created during or through provider encounters in clinical settings). While clinical data is generally protected by a variety of federal laws governing individually identifiable health information (e.g., HIPAA, Part 2—see discussion of these laws in Appendix A), PGHD is not subject to the same federal protections. PGHD is distinct from clinical

[3] HealthIT.gov. "Patient-Generated Health Data" (*last updated* April 26, 2017). *Available at:* https://www.healthit.gov/policy-researchers-implementers/patient-generated-health-data.

data in that: 1) patients, not providers, are primarily responsible for capturing this information; and 2) patients are solely in control of how and with whom to share this information. Examples of PGHD include data collected or generated by these devices: blood glucose monitoring or blood pressure readings using home health equipment (and which do not automatically transmit to an EHR) and exercise or diet tracking using a mobile app.[4] PGHD can be extremely helpful for both treatment and research purposes. For example, it can help providers track how patients are doing between medical visits, allow patients and providers to collect information on an ongoing basis rather than solely at clinical visits, and provide information useful to treat—as well as prevent—chronic diseases.

There is a rapidly evolving array of technologies designed to enable patients to collect their health information beyond the clinical setting and share that information with providers and researchers. Examples of these new technologies and applications include: Personal Health Records (PHRs), "wearables" such as Fitbit and Jawbone devices; and applications such as Apple Health, diabetes trackers, and the OneTouch Verio Sync Meter that uses Bluetooth to send data to a person's iPhone and then generates reports and data that can be shared with healthcare providers. In particular, Apple, Google, Nike, and Under Armour are investing in wearable technology that syncs with their health-tracking platforms. Devices used to test blood glucose levels, cholesterol, oxygen levels, etc., that can connect to networks have also been developed. Furthermore, the continued development of the Internet of Things[5] should lead to widespread use of fully networked devices. As such, the amount, variety, and quality of available data should increase in the future as more people use wearables, more devices join the Internet of Things, and technology continues to improve.

Patient Reported Outcomes (PROs)/Patient Reported Outcome Measures (PROMs)

According to the Food and Drug Administration (FDA), a PRO is "a report coming directly from the patient (i.e., study participant) about the status of a patient's health condition without amendment or interpretation of the patient's response by a clinician or anyone else."[6] PROs can be collected during clinic visits or at home through various web tools or mobile devices either for treatment or for clinical trial participation (e.g., PROMIS Assessment Center,[7] PatientViewpoint,[8] REDCap Research Electronic Data Capture[9]). PROs are distinguishable from PGHD (discussed above) because a researcher or clinician typically initiates collection using a scientifically validated survey or instrument, whereas collection of PGHD is generally patient-initiated and does not require a validated instrument or survey.

---

[4] HealthIT.gov. "Patient-Generated Health Data" (*last updated* April 26, 2017). *Available at:* https://www.healthit.gov/policy-researchers-implementers/patient-generated-health-data.

[5] Jacob Morgan. "A Simple Explanation of 'The Internet of Things'" Forbes.com (May 13, 2014). *Available at:* http://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#45f87ae56828.

[6] *See* U.S. Department of Health and Human Services Food And Drug Administration (FDA). Guidance For Industry, Patient-Reported Outcome Measures: Use In Medical Product Development To Support Labeling Claims, at p. 32 (2009). A*vailable at:* http://www.fda.gov/downloads/Drugs/.../Guidances/UCM193282.pdf.

[7] Assessment Center. "What is Assessment Center" (2013). A*vailable at:* https://www.assessmentcenter.net/.

[8] PatientViewpoint. "Welcome" [log-in credentials required]. A*vailable at:* https://www.patientviewpoint.org/Login.aspx (*last visited* March 16, 2016).

[9] Research electronic data capture (REDCap). "Homepage". *Available at:* http://project-redcap.org/ (*last visited* March 16, 2016).

Genetic Information

Genetic information is information about an individual's genetic makeup and the genetic makeup of an individual's family members, as well as information about the manifestation of a disease or disorder in an individual's family members (e.g., family medical history).[10] Family medical history is included in the definition of genetic information because it is often used to determine whether someone has an increased risk of getting a disease, disorder, or condition in the future. Genetic information also includes an individual's request for, or receipt of, genetic services or participation in clinical research that includes genetic services by the individual or a family member of the individual, and the genetic information of a fetus carried by an individual or by a pregnant woman who is a family member of the individual and the genetic information of any embryo legally held by the individual or family member using an assisted reproductive technology.[11]

Personal Genomics/Direct-to-Consumer Genetic Testing is an emerging market that allows individuals to submit a DNA sample in order to determine their disease risk/carrier status, find relatives, determine paternity, etc. If direct-to-consumer genetic tests can avoid FDA scrutiny and gain widespread consumer interest, then personal genomics is another market that could generate significant amounts of relevant PCOR/CER data.

Biospecimens

Biospecimens, which are a type of genetic information, include tissue, blood, urine, or other human-derived material. A biospecimen can comprise subcellular structures, cells, tissue (e.g., bone, muscle, connective tissue, and skin), organs (e.g., liver, bladder, heart, and kidney), blood, gametes (sperm and ova), embryos, fetal tissue, and waste (urine, feces, sweat, hair and nail clippings, shed epithelial cells, and placenta). Biospecimens may be collected in a clinical setting, patient home, or other site. For example, some biospecimens (e.g., urine, saliva, hair follicles, sperm, etc.) may be safely collected and stored by an individual within the privacy of their own home. Many research studies allow at-home collection because it is more efficient than clinic visits. Individuals also may voluntarily donate samples to biobanks/biorepositories.

Surveillance Data

Surveillance data refers to the health information collected to facilitate the planning, implementation, or evaluation of public health activities. Government agencies/organizations (e.g., CDC, World Health Organization (WHO)) are often the party responsible for collecting and disseminating surveillance data. Examples of surveillance data sources include the CDC's Behavioral Risk Factor Surveillance System (BFRSS),[12] which continually monitors chronic health conditions, use of preventive services, and risk behaviors, and the FDA's Sentinel Initiative, which monitors the safety of FDA-regulated products.[13]

---

[10] Adapted from definition of "genetic information" set forth in GINA Title I (2008).

[11] 29 U.S.C. § 1191b(d)(6) (2017).

[12] U.S. Department of Health and Human Services Centers for Disease Control and Prevention (CDC). "Behavioral Risk Factor Surveillance System" (*last updated* August 25, 2017). *Available at:* http://www.cdc.gov/brfss/.

[13] FDA. "FDA's Sentinel Initiative" (*last updated* December 14, 2016). *Available at:* http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm2007250.htm.

Quality Improvement Data

Quality data refers to information collected to assess the performance of healthcare providers and/or health plans and the results of these performance assessments to improve the quality of care delivery. Various government and non-government agencies/organizations (e.g., The Centers for Medicare & Medicaid Services, The National Committee for Quality Assurance, etc.) develop and maintain quality measures for use by private or public payers. Existing measures focus on a broad range of quality-related factors such as patient safety, patient-centered care, care coordination, and affordability.[14] Although there is no indication that quality improvement data is included in the scope of PCOR research, the increasing creation of quality improvement data could inform future research efforts.

[14] *See* National Quality Forum (NQF). National Priorities Partnership and the National Quality Strategy (2011). https://www.qualityforum.org/Setting_Priorities/NPP/Input_into_the_National_Quality_Strategy.aspx.