CODE

Balancing Privacy with Health Data Access

Roundtable Report

September 2019

Published by The Center for Open Data Enterprise, www.opendataenterprise.org (2019)

Table of Contents

Foreword by HHS Chief Data Officer Mona Siddiqui

You don't have to read the news too closely to realize that some of the most newsworthy elements every day relate to data privacy. In July 2019 alone, headlines described Federal Trade Commission (FTC) actions against Facebook for privacy violations, a record $230 million fine against British Airways for a website failure that compromised half a million customers' data, and a $123 million fine against Marriott for a 2018 data breach that hit 339 million customers. Some months before, Google also faced scrutiny and fines under the General Data Protection Regulation (GDPR) from France's regulator, with a $57 million penalty levied in January for "lack of transparency" and valid consent controls for users, among other issues.

At the same time, we also read stories that show the value of using data appropriately - and how connecting information leads to insights that can save lives:

- A study examined information from two large commercial insurance claims databases on patients aged 13 to 25 years old with attention-deficit hyperactivity disorder (ADHD) who started taking amphetamines or methylphenidate. The researchers evaluated these two most common ADHD treatments and found that, although the risk of psychosis is low, it is greater for patients who are taking amphetamines than for those taking methylphenidates.
- In India, one of the nation's largest private healthcare companies is using artificial intelligence (AI) to improve detection of cardiac illnesses that cause more than three million heart attacks in that country every year. Until now, it's been difficult for doctors to identify patients who are at risk for coronary disease because most prediction models are based on studies conducted in Europe and North America and don't apply well to Indian populations.
- In the U.S., a map of all prescription take-back locations across the country, constructed using data from HHS, is helping states and private companies begin to address the inappropriate use of opioids.

The July 2019 Roundtable described in this report addressed the challenge: How should we balance the need for privacy with access to health data that can make insights like these possible? That discussion is part of a national effort to address data privacy at different levels of government.

In the absence of definitive federal action, states have been actively passing new and expanded requirements for privacy and cybersecurity. While laws like the California Consumer Privacy Act are getting the most attention, many states are actively amending their breach notification laws. Most recently, Illinois, Maine, Maryland, Massachusetts, New Jersey, New York, Oregon, Texas, and Washington have all amended their breach notification laws to either expand their definitions of personal information, or to include new reporting requirements.

There has also been significant movement on the Hill to begin to establish federal protections that would take effect nationwide:

- In November 2018, Senator Ron Wyden (D-OR) released a draft Consumer Data Protection Act, designed to expand the FTC's regulatory and enforcement powers. Among other things, the draft Act would establish minimum national data privacy and cybersecurity standards. The draft Act would also create a system that would allow consumers to stop third parties from tracking their online activity and sharing their data.

- Shortly thereafter, Senator Brian Schatz (D-HI) released the draft Data Care Act, which would require websites, applications, and other online providers to establish practices to reasonably secure individual identifying data and promptly inform users of data breaches that involve sensitive information.

- Earlier this year, Senator Marco Rubio (R-FL) proposed a new privacy bill, the American Data Dissemination Act (ADDA). In contrast to the Consumer Data Protection Act and the Data Care Act, ADDA would not expand FTC authority to create and implement laws. Instead, ADDA would require Congress to pass applicable laws presented by the FTC, with the FTC ultimately gaining rule-making power if Congress is unable to pass a law within two years of the ADDA going into effect.

- In June 2019, Senators Amy Klobuchar (D-Minn.) and Lisa Murkowski (R-Alaska) introduced legislation aimed at safeguarding the privacy of consumer health data, specifically the data involved in DNA testing kits and health tracking apps. The Protecting Personal Health Data Act would require the Secretary of the U.S. Department of Health and Human Services (HHS) to create regulations for health data tracking apps, including wearable devices such as FitBits, and for genetic testing kits. The regulations would include a clause to enable consumers to review, change, and delete any health data collected by companies. The legislation would also create a National Task Force on Health Data Protection to evaluate and provide input on any potential cybersecurity and privacy risks of consumer products that use customer health data.

At HHS, we are at the center of this national conversation. As leaders of governments, as decision-makers in business, and as citizens, we must ask ourselves a basic question: What kind of world do we want to live in? My belief, regardless of the topic, is that lack of transparency in any system creates a fundamental crisis of trust and a stunting of the potential for progress. It is easy to use the inherent complexity of systems to delay decisions. Those of us who believe in technology's potential for good must lean into this conversation and embrace that it will be messy, incremental, and iterative. But at the end of the day, the voice of the consumer and the voice of the patient needs to be loudest.

That is why the HHS Office of the Chief Technology Officer (CTO), together with the nonprofit Center for Open Data Enterprise (CODE), convened an outstanding group of leaders for the Roundtable described in this report from CODE. The Roundtable brought chief privacy officers and national experts together with patients and patient advocates for a candid, in-depth, action-focused discussion. As HHS is grappling with how we share data across our organization internally to make better decisions, their perspectives continue to be essential in informing those discussions. I also want to acknowledge that the only way this work can be sustained long term is with tremendous leadership support. We hope you find CODE's summary report about health data privacy a useful resource as we begin the work ahead.

Mona Siddiqui
Chief Data Officer
Office of the Chief Technology Officer
U.S. Department of Health and Human Services

Executive Summary

The independent nonprofit Center for Open Data Enterprise (CODE) and the Office of the Chief Technology Officer (CTO) at the U.S. Department of Health and Human Services (HHS) are co-hosting a series of three Roundtables to find ways to improve how health data is shared and utilized for the public good.

As the second Roundtable in this series, CODE and the HHS Office of the CTO convened a *Roundtable on Balancing Privacy with Health Data Access*. This Roundtable brought together approximately 70 different expert stakeholders from industry, academia, law, government, and civil society to discuss issues of data privacy. It also included patient advocates from a variety of organizations who provided additional insight into health data privacy from a patient's perspective. The purpose of this Roundtable was to empower data providers and users to maximize the utility of sensitive health data while providing necessary privacy measures and addressing risk. Participants discussed privacy risks and current issues in health data use, shared approaches to managing risk, and identified actionable opportunities for HHS.

This Report summarizes the findings of the Roundtable in the following sections:

Introduction: the benefits and risks associated with health data use, and some of the major goals that HHS has set out for a new privacy paradigm.

Risk and Benefit Landscape of Health Data Use: findings from the opening Roundtable exercise, common and specific risks of different kinds of health data, and harms and benefits to key stakeholders.

Technical Approaches to Reducing Risk: the various technical approaches used to safeguard privacy and their limitations.

Data Governance Successes and Challenges: the current governance model for privacy, how that model is working, and what needs to be improved.

Recommendations and Solutions: the recommendations proposed by Roundtable participants to improve the existing governance framework and address types of data that may fall outside of those rules.

Conclusion: The Report concludes with an overview of possible paths forward to embrace these recommendations and other relevant updates since the Roundtable was hosted.

Introduction

The increasing availability of health data is transforming the health sector. Researchers are using clinical and surveillance data to better prevent, diagnose, and treat disease. Technology companies are using patient-generated data from mobile phones and wearable devices to help individuals track their medical conditions and customize their treatment plans. And healthcare providers are using administrative and claims data in combination with data on the social determinants of health to better understand risk factors for health conditions and improve healthcare delivery. This transformation is also taking place at the community level. By using new data sources, epidemiologists can better track the spread of disease and health epidemics, and public health agencies can leverage population data to drive better policy decisions to address health inequities.

At the same time, complex questions are emerging around data privacy and the harms that can impact individuals and communities when using health data without adequate privacy safeguards. Health data privacy is the protection of personal health information, such as an individual's medical conditions, health insurance records, genetic information, and fitness activities, with appropriate provisions for sharing and utilizing this information in ways that the subject of the data is aware of and has consented to. If privacy is not managed well, individuals may not understand how their data is being shared or the limits of what privacy laws cover. Communities may also face exclusion of services or discrimination based on the use of population-level data.

In the United States, the Health Insurance and Portability Accountability Act (HIPAA) and a patchwork of other laws seek to establish a number of approaches to safeguard the privacy of personal health information. But while these laws protect data collected by healthcare providers and health plans, they are not well designed to handle the many other kinds of health data produced and collected today. In particular, HIPAA only protects data collected by specific healthcare entities, including healthcare plans, healthcare providers, and healthcare clearinghouses. As a result, this can omit new kinds of data collected by fitness trackers, genetic analyses, or other commercial processes and devices.

Current regulations have also not accounted for the full range of possible harms that may arise from health data use. As the volume of health data increases, patients, providers, and private companies alike are increasingly uncertain about what data is and is not covered under federal statute. This has led to confusion over privacy-related issues including informed consent for data use, data access, and appropriate use of health data.

Patients and patient advocates are critical stakeholders in the conversation around balancing privacy with appropriate health data access and use, particularly in the context of individual-level health data. For example, patients can benefit from research that uses individual-level data to better diagnose disease and find new treatments. But if protected health information (PHI) is misused, patients may be at risk of discrimination, financial exploitation, or other harms. These harms can also compound at the community level as the analysis of large-scale health datasets may benefit one group at the expense of another.

The *Roundtable on Balancing Privacy with Health Data Access* was designed to bring together HHS leaders, patients, and health data experts in federal and state government agencies, industry, law, and patient-advocacy organizations to discuss strategies for appropriately accessing sensitive health data while safeguarding its privacy. The opening remarks from HHS Chief Data Officer Mona Siddiqui and Assistant

Deputy Secretary of HHS Charles Keckler underscored the Department's commitment to establishing HHS as a leading force in promoting the appropriate use of health data.

The Roundtable then featured lightning talks on challenges and issues in the privacy landscape from representatives of Omada Health, Ciitizen, Facing Our Risk of Cancer Empowered (FORCE), and Sage Bionetworks. These talks highlighted the tools being used and challenges faced by companies seeking to access data, strategies to improve data access for researchers, and issues created by data sharing with wellness programs. They highlighted areas that are working well within the current governance framework and areas in need of improvement.

Throughout the day, participants engaged in three breakout sessions: 1) The risks and rewards of accessing different types of data, 2) Effective strategies for balancing privacy with health data access, and 3) Actionable next steps. The day concluded with a presentation of highlights including specific recommendations that HHS could use to improve the health data privacy paradigm.

As policymakers update privacy rules to keep pace with the explosion of new types of data, this Report is designed to inform the discussion in two ways. First, it analyzes the potential benefits and harms different stakeholders face from the use of health data. Second, it reviews how the current governance framework could better manage these risks and improve the benefits of data use more broadly.

This Report, written by the staff of CODE, summarizes insights and recommendations provided by individual Roundtable participants in the course of the day. It is not meant to represent a consensus of the participants, and does not represent the views and opinions of HHS or its leadership or staff. CODE hopes that it will be of value to all stakeholders, inside and outside of government, as they continue to address these important issues.

The Risk/Benefit Landscape of Health Data

Appropriate access to health data, facilitated by technological advances such as cloud computing and artificial intelligence (AI), can greatly benefit patients, communities, and other stakeholders across the healthcare system. These benefits range from improving diagnostic accuracy to increasing the understanding of complex genetic conditions.[1] If personal health data is misused, however, then patients may be at risk of financial discrimination, reputational damage, or other harms from their loss of privacy. This kind of misuse can also harm communities, who may face discrimination or a lack of services if they are thought to be at high risk for illness.

The current portfolio of technical approaches and governance frameworks attempt to address the different risks and benefits associated with using health data. While regulations generally address only risks to the individual, both risks and benefits can impact a variety of stakeholders. This section describes the ways risks and benefits of health data are perceived by different stakeholder groups, the specific risks posed by different types of health data, and the ways different stakeholders may benefit or be harmed by the use of health data.

Group Perceptions of Risk and Benefit

CODE began the Roundtable with an informal data gathering activity that asked participants to rate the level of risk and benefit for six types of health data.[i] For each data type, CODE asked every participant to place a dot on a wall chart showing a matrix of risk and benefit, using dots color-coded to their stakeholder group: Civil Society and Academia (yellow), Patient Advocacy and Engagement (red), Government (green), or Private Sector (blue). They did this for six charts representing six types of high-value health data: administrative, clinical, genomic, patient-generated, social, and surveillance data.[2] It is important to note that the 51 participants who did this exercise were participants in a Roundtable of carefully selected experts, and this exercise was not meant to represent a scientific random sample. CODE also did not design the exercise to test any formal hypotheses about risk perception. Still, the results of the exercise are a useful starting point for understanding the risk/benefit landscape of health data.

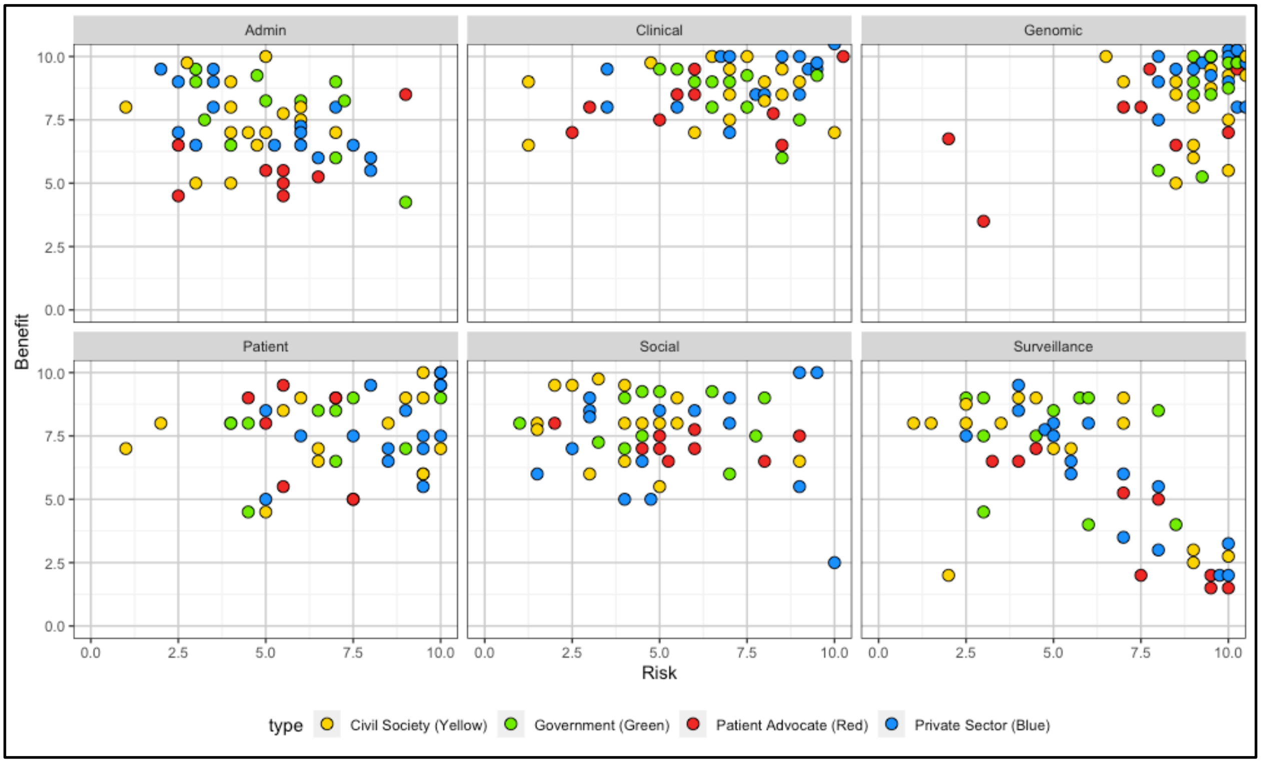
As the results below show, the exercise demonstrated that most stakeholders value the benefit of health data but disagree widely on the level of risk in using that data. For administrative, clinical, patient-generated, and social data, participants consistently rated the benefit highly - from 5 to 10 on a scale of 10 - but varied significantly in how they assessed risk, with ratings from 1 to 10. There were two exceptions to this pattern: genomic data, which participants agreed carries a high level of risk, and surveillance data, which showed less agreement on the level of benefit. Participants may have been unclear on the meaning of "surveillance data"; while CODE meant this to refer to surveys of population health, some participants seemed to think it meant individual surveillance through cell phones or other means.

The different stakeholder groups at the Roundtable also came to different conclusions about the risks and benefits of health data. Both representatives of civil society and patient advocates ranked risk lower than their private sector and federal government counterparts. This is noteworthy since civil society and patient

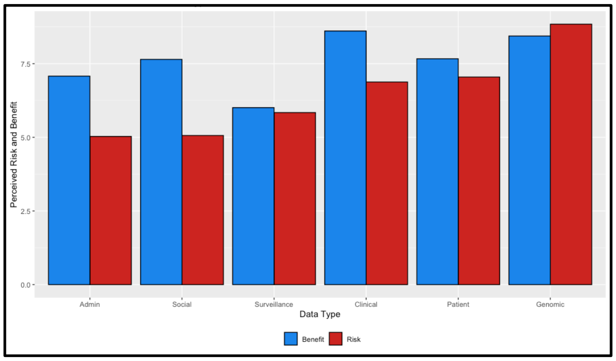
[i] This informal icebreaker activity distributed colored stickers to participants at the Roundtable by stakeholder type. CODE created matrices with the level of Risk along the x-axis and the level of Benefit along the y-axis. Participants were instructed to place the sticker somewhere along the plane based on their perceptions of risk and benefit. A one page overview of this activity is included in the appendix of the report.

advocacy groups represent the individuals who have the most to lose if their privacy is not protected. This finding also underscores the importance of including patients and civil society in discussions of health data privacy issues, since they may have a different perspective than other stakeholders.

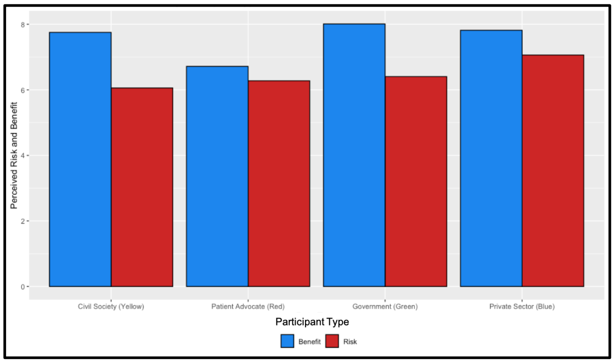
Overall Distribution of Points



By Data Type



By Participant



<u>Note:</u> Charts created from data points gathered at July 15 Roundtable.

What are the common risks across data types?

Many of the benefits of accessing datasets that include sensitive PHI are well documented, including improved preventive care, clinical outcomes, and coordination of care. Similarly, many of the major health data types pose shared challenges and issues that can negatively impact patients and communities.

Incomplete Data

Healthcare providers frequently make clinical decisions based on data that may not be complete or that lacks additional context from complementary data. While many physicians make decisions purely based on available patient electronic health record (EHR) data or claims data, they may miss potential insights and context from genomic data or data on the social determinants of health. Moreover, technical restrictions can lead to incomplete or inaccurate data. For example, because EHR clinical data is not standardized, the lack of interoperability may create medical gaps in building longitudinal patient profiles and providing comprehensive care. Moreover, non-interoperable EHR data and handwritten notes may not be machine readable and may have to be manually processed, making its analysis slower and more cumbersome.

Possibility of Re-Identification Through Other Datasets

De-identification has been used as a key strategy to preserve the anonymity of individuals and safeguard sensitive PHI. Data scientists and researchers have made advances in de-identifying data, which involves removing key identifiers from data, or anonymizing data, which entails changing the way variables are coded in a dataset. However, these technical fixes may not be enough to completely safeguard data privacy. Recent research has demonstrated how companies and researchers can take anonymized datasets and re-identify individuals with a high degree of accuracy when the anonymized data is combined with other third party data, a process called the "mosaic effect." A 2018 study described how researchers were able to use machine learning techniques to re-identify anonymized physical activity data, such as running patterns and heart rate, collected from wearable devices.[3] A study in the journal *Nature Communications* further confirmed the possibility of re-identification of heavily anonymized datasets after a machine learning method showed that almost any American can be re-identified by using 15 demographic attributes.[4] Since HIPAA does not place any restrictions on the use of de-identified data, it's especially important to learn more about the limitations of these approaches.

Inappropriate Data Sharing and Use

Personal health data may be misused by third party providers that violate informed consent or by data brokers that illegally obtain PHI without a patient's permission or knowledge. While data flows between covered entities are common, especially for treatment, payment, and healthcare operations (TPO), patients are not often informed about when their data moves from one entity to another.[ii] HIPAA defines six approved categories of use for sharing an individual's PHI. However, many patients are unaware of the limits of approved data use, and may not immediately realize when their right to privacy is being breached. For example, Yale University was recently sued after 5,400 Yale employees were required to share their personal PHI with a mandatory workplace wellness program that was not a covered entity under HIPAA.[5]

[ii] HIPAA outlines six broad categories of permitted uses and disclosures of PHI that do not need patient consent. Those include: 1) Disclosure to the individual, 2) Treatment, Payment, and Healthcare Operations, 3) Uses and Disclosures with the Opportunity to Agree or Object, 4) Incidental Uses and Disclosures, 5) Public Interest and Benefit Activities, and 6) As a limited dataset. Each of these categories features sub-categories of specific uses which are outlined in HIPAA's privacy rule. For more information about uses, please visit: https://www.hhs.gov/sites/default/files/privacysummary.pdf

Confusion About Data Safeguards and Protection

Many Roundtable participants stressed that both patients and providers are frequently confused or misunderstand how PHI is protected under HIPAA. Many patients are unaware that data-gathering mobile applications, medical devices, and voice assistants are sold and managed by entities not covered under HIPAA. A recent survey showed that nearly one third of healthcare providers do not have a HIPAA compliance plan and the same number are uncertain about security safeguards for personal data.[6] Moreover, many patients are unaware of the specific situations in which a healthcare entity covered by HIPAA's rules is allowed to access PHI. Lastly, patients and providers both may be misled by the privacy policies of third-party companies that do not have to abide by data safeguards. A 2016 ONC Report from the HHS Office of National Coordination (ONC) noted that many healthcare entities do not have adequate safeguards in place or comprehend the scope of security and privacy risk assessments.[7]

What are unique challenges for new data types?

In addition to these challenges, emerging types of health data pose unique challenges that policymakers and practitioners must consider. Understanding these risks can help establish better privacy protection strategies and inform how the healthcare system manages the growing use of non-traditional data types.

Genomic Data

The rapid rise of genomic data in personalized healthcare decision-making has been enabled by companies like 23andme, Ancestry.com, and MyHeritage. More widespread clinical genomics testing has also increased the availability of genetic data. The largest four genomics companies alone had received DNA samples from more than 26 million consumers as of January 2019.[8] Much of this data still falls outside of the purview of HIPAA and is not regulated by research-driven data use arrangements that place limits on how data is used and disclosed. This situation poses several risks, which include:

- **Disclosure and its Impacts on Relatives:** While the disclosure of most PHI can impact the individual, genomic data can also impact family members related to the individual whose genetic makeup is analyzed. Federal rules and regulations, including HIPAA, do not address risks posed to relatives. Genomic data has already been used in criminal cases to identify relatives of people who have provided DNA samples to commercial companies.[9] Similarly, genomic data can be used to trace blood relatives of people whose DNA has been tested without those relatives' consent.
- **Confusion Over Value and Use:** Genomics companies routinely emphasize the value of their diagnostic capabilities and ability to identify rare conditions in patients. These marketing tools may make patients more willing to provide their DNA to these companies. However, consumers may find that genomics companies provide less information and cover fewer genetic risks than they anticipated, or provide information without the full context.[10] For example, a consumer may believe that she will never develop breast cancer because she has tested negative for a specific mutation, even though she may still be at risk for breast cancer from other causes.[11]

Consumer-Generated Data from Non-Covered Entities

Consumer-generated data is health-related data collected from products and devices used by consumers, including data from the Internet of Things, and social media data. Consumer-generated data may fall outside of the purview of HIPAA if it is collected by technology companies that are not covered entities, are not regulated through their relationship with covered entities, and are not subject to clear guidelines from the FTC. HIPAA does specify that "business associates" of covered entities such as healthcare providers are subject to the same regulations on their use of health data as covered entities. However, if a company independently collects consumer-generated data, it may legally be able to use the data or sell it for commercial third party use in various ways. Some companies that are not covered entities or business associates under HIPAA have already released consumer-generated data to companies like Facebook and Google.[12] Additional privacy concerns include:

- **Lack of Data Minimization:** A general principle for ensuring privacy is data minimization - the goal of using the minimal amount of data for a particular purpose. While this principle is not applicable to treatment information due to research needs, it is an appropriate goal for consumer-generated data. However, consumer-generated health data can include extraneous information with unnecessary personal details. Although some of this data may have implications for an individual's health, additional data points may not be essential to the patient, but may be sent to third party providers for their own purposes.
- **Location and Consent:** Roundtable participants noted that smartphone apps, in particular, often have location-sharing built into their functionality, which could potentially gather information without the individual's knowledge of it being collected or about how it will be used.
- **Unregulated Technology Companies:** There are a number of routes by which consumer-generated data could be regulated: HIPAA, the FTC, and the FDA's rules on medical devices. As discussed above, HIPAA applies only to healthcare companies and providers that are considered covered entities or business associates, and may not cover other companies that gather consumer-generated data. The FTC has taken actions that penalize improper uses of consumer-generated health data but has not established clear guidelines about when to apply these penalties.[13] Lastly, the FDA has released guidelines on mobile medical health applications for industry but places less priority on "low-risk devices," so companies that gather consumer-generated data may not have to register their products with the FDA.

Social Determinants of Health

The social determinants of health (SDOH), including such factors as income, education, and housing, are a promising area of population data attracting increased interest from researchers, providers, and patients. For example, some research suggests that a person's ZIP code is actually more predictive of adverse health outcomes than that person's genetic code.[14] HHS Secretary Alex Azar noted that the social determinants of health "would be important to HHS even if all we did was healthcare services…but in our very name and structure, we are set up to think about all the needs of vulnerable Americans, not just their healthcare needs."[15] While the benefits of SDOH data hold enormous potential in reducing costs and improving patient care, several Roundtable participants noted a potential risk in the use of this data.

- **Profiling and Redlining:** As healthcare companies increasingly use the social determinants to better support diagnosis and treatment, it's also possible that these same companies could engage in healthcare "redlining" and exclude or profile communities that they identify as high-risk areas.[16] Alternatively, individuals may be directly profiled for residing in a high-risk ZIP Code, which could affect the quality of their treatment. These generalized assumptions could lead to unequal distribution of care and limit some groups' medical options.

Data Use: Potential Benefits and Harms by Stakeholder Group

Given the range of potential risks across the data spectrum, not all harms and benefits are equally distributed among different stakeholders when health data is accessed and used. Roundtable participants emphasized the need to analyze how the benefits and risks of health data could potentially be distributed to individuals, communities, and other stakeholders. Several participants noted that patients could be harmed by the misuse of health data in ways that would benefit healthcare providers and insurers, such as increasing premiums based on factors in their data. But others pointed out that additional stakeholder groups also stand to benefit or be harmed by health data in different ways. The table below assesses the potential harms and benefits that could occur to communities, individuals, and organizations based on access and use of health data.

Stakeholder Group	How Data Use Potentially *Benefits* Them	How Data Use Potentially *Harms* Them
Patients	● Improved coordination of care ● Fuller, contextual information about health risks ● Better selection of health plans ● Increased knowledge and empowerment using genetic results on risks of rare diseases ● Advancing research for rare conditions	● Data sold to data brokers without patient's knowledge or consent, reducing patient trust in institutions ● Sensitive information about patient's conditions being made available to third parties ● The possibility of identity theft relating to PHI ● Stereotyping or discrimination based on disclosed healthcare information ● Negative impacts on relatives due to sharing of genomic data by a family member
Communities	● Addressing health disparities by tying social services into clinical care ● Increased provision of social services based on high-priority geographic areas ● Increasing federal grant money for specific needs	● Communities redlined out of key services by healthcare providers and private companies based on cost and risk determinations ● Fewer resources allocated to uninsured communities ● Generalizations made about communities that may not apply to individuals
Providers	● Improved ability to administer coordinated care ● Reduction of costs through better risk management ● Increased revenue from better allocations of services ● Better understanding of patient's needs and treatment options through relevant health	● Decreased patient trust in providers if information continues to be widely shared, especially with non-covered entities

	information such as social determinants and genetic data	
Private Non-Covered Entities	● Employers informed of improved patient care ● Increased ability to tailor new programs based on rich health data	● Decreased consumer trust with improper data use ● Patient lawsuits increase costs and risks
Federal Government	● Enhanced research around surveillance data and clinical trials ● Improved ability to serve Medicare and Medicaid beneficiaries ● Reducing costs of programs through more effective interventions	● Additional staff and human resources needed to manage larger amounts of HIPAA-compliant data ● Federal agencies like the NIH and CDC are not covered by HIPAA ● Lawsuits and patient complaints due to data breaches
Public Health Agencies	● Enhanced understanding of health outbreaks using patient-generated data ● Improved administration of services at the national level ● Flagging rare diseases and potential epidemics	● Increasing health disparities from the possibility of redlining as agencies use data for precision medicine and predictive analytics ● Possibility of privacy concerns such as re-identification due to hyper-local data gathering points

Technical Approaches to Reducing Risk

Given the range of possible benefits and harms, researchers, insurers, and healthcare providers now use a variety of approaches to avoid unnecessary data collection, de-identify or anonymize sensitive PHI, and regulate access to health data. The table below presents an overview of these technical approaches and their limitations. While technical strategies are not a complete solution, they are an important part of any privacy-protection program.

Approach and Its Value	Limitations
The data minimization principle emphasizes that data collection and the amount of data used for any particular project is only what is necessary to accomplish the needed tasks. This reduces the possibility of unnecessarily gathering potentially sensitive information about an individual.	● Complex analysis for research using machine learning and AI may require larger amounts of data to see what is most meaningful ● May complicate patient consent for data use as data needs become broader
De-identification and anonymization strategies seek to remove sensitive personally identifiable information (PII) from individual-level and population-level data or otherwise make it difficult to identify the source of the data. These can include: 1. **Providing Anonymized Identifiers**: These identifiers allow researchers to connect disparate datasets while preserving the privacy of individuals. 2. **Removing Non-Critical Information**: Researchers can remove key variables such as ZIP code digits, social security numbers, account information, and other identifying information. 3. **Leveraging Synthetic Data**: Synthetic data is produced by "a complex statistical model that generates a simulated population that has the same general features as the original data."[17] 4. **Applying Differential Privacy**: Differential privacy places constraints on algorithms that rely on inputs from a database of information. This masks the personal information so an external user cannot determine if an individual's information was used in the computation process. **5. Generalized Statistical Approaches:** Statistical approaches often include adding "noise" to the data to obscure specific variables such as age range or location.[18]	● De-identified or anonymized data can often be re-identified using other datasets ● May not be optimal for important use cases where individual identity is important, e.g. providing disaster relief or addressing health epidemics ● Private sector companies outside of HIPAA use different de-identification strategies based on their own business choices, making it difficult to establish best practices across industry

Approach and Its Value	Limitations
Institutional Differential Access assumes that PHI can be made accessible to institutions under controlled conditions when release to the public is not appropriate or could negatively impact a patient's privacy. It grants access to datasets only under specific circumstances, to specific organizations and individuals (such as medical researchers), and for specific purposes. Approaches to differential access can include a federated data cloud model that grants trusted users credentials to access the data, and multiple levels of access for different types of users. Differential access may also include different options for data download in machine readable formats.	● Possibility of reduced care coordination if certain aspects of a patient's PHI are restricted ● May impede interoperability when different databases follow different standards, making data sharing between databases difficult
Patient-Based Differential Access enables individuals to grant access to their personal data for the benefits of public research. Patients may opt-in and provide consent to use their personal data for a specific purpose such as studying a rare disease or identifying genetic trends. Researchers are allowed to access this data based on the parameters of the patient's original consent.	● Requires providing clear, detailed information to patients ● Initial consent may not cover future legitimate uses of data

These technical approaches are employed by private and public data holders to ensure that valuable PHI is made accessible and usable to physicians, researchers, and other stakeholders in appropriate ways. The current range of approaches may be applied based on a variety of factors, including the type of data protected, the organization that collects this data, and how the stakeholder perceives benefit versus risk. Yet technical approaches alone are not foolproof; they cannot completely eliminate the risk of re-identification without decreasing the value of the data by removing, changing, or obscuring large amounts of information. And they don't prevent the misuse of data by approved users, who may use the data for purposes including discrimination and financial exclusion. The following section describes how the current governance framework is designed to offer additional privacy protections, and the ways in which that framework is still incomplete.

Health Data Governance: Successes and Challenges

While technical approaches represent the "what" and "how" of safeguarding health data privacy, governance frameworks provide the "why" and "when" for using these approaches. The U.S. health privacy framework has been built on a foundation of federal rules and regulations, with the additional support of state-level laws and legislation. This section describes the current U.S. health data privacy framework and examines the successes and shortcomings of this framework in addressing key privacy issues. It also describes some current efforts, such as the proposed *Protecting Personal Health Data Act*, that are designed to improve health data governance in important ways.

Current Regulations

The primary governance framework to manage the privacy of U.S. healthcare data is HIPAA. HIPAA was designed to create a federal floor for the privacy and security of personal health information, which HIPAA defines as data that "includes the individual's past, present, or future mental or physical condition, the provision of healthcare to an individual, and any past, present, or future payment for the provision of healthcare to the individual."

HIPAA, which was passed in 1996, sets the standards for how entities covered by the law must transmit this personal health information, which includes claims, enrollment, eligibility, payment, and coordination of benefits. The law defines "covered entities" as qualified healthcare providers, healthcare clearinghouses, and health plans. HIPAA also requires covered entities to have clear contractual arrangements with any "business associates" that manage their data to ensure that they follow HIPAA's rules. (Business associates include persons or businesses that perform certain functions on behalf of a covered entity, or provide services to a covered entity that involves the use or disclosure of a patient's PHI.)[19] These provisions are enforced by the HHS Office of Civil Rights (OCR), which can administer financial penalties for rule violations.

- **The HIPAA Privacy Rule:** Sets standards for individually identifiable health information and defines when and how use and disclosure of PHI is permitted. Those broad instances include release of data on behalf of the individual, for healthcare treatment or payment or operations giving an individual rights to their data, to give the individual an opportunity to agree to or correct the data, for the public interest and benefit, and for limited purposes of research, public health or health care operations.[20]
- **The HIPAA Security Rule:** Focuses on safeguarding electronic PHI (EPHI). It dictates that healthcare providers that create, receive, maintain, and transmit EPHI must institute measures to protect this EPHI from anticipated threats, hazards, and impermissible uses. The rule aims to ensure the confidentiality, integrity, and availability of EPHI.

Other relevant regulations that govern health data privacy include:

- **The Privacy Act of 1974:** Establishes a code of fair information practices for the collection, maintenance, use, and dissemination of information about individuals when that information is maintained in the federal system.[21]
- **Family Educational Rights and Privacy Act (FERPA):** Passed in 1974, FERPA establishes specific guidelines for protecting the privacy of personal health information in students' educational records, such as vaccinations and nurse visits.
- **Human Subjects Common Rule:** The 1979 Belmont Report outlines basic ethical principles that should underpin biomedical and behavioral research for human subjects. This report set the

foundation for the later 1991 Common Rule, which outlines specific protections for at-risk groups like prisoners, children, and pregnant women.[22]

- **The Confidential Information Protection and Statistical Efficiency Act (CIPSEA):** This 2002 legislation establishes laws to govern confidentiality protections for data collected by U.S. statistical agencies and units. The National Center for Health Statistics and the Center for Behavioral Health Statistics and Quality are the two HHS entities covered under CIPSEA.[23]
- **Genetic Information Nondiscrimination Act (GINA):** Passed in 2008, GINA aims to prevent discrimination based on a person's genetic information by employers and health insurers.[24] Companies are not allowed to make decisions related to eligibility, premium costs, or coverage based on this information.
- **State-Level Laws:** Maine's 2019 Act to Protect the Privacy of Online Consumer Information and California's Consumer Privacy Act seek to regulate the privacy of consumer-generated data and allow for robust opt-in and opt-out clauses for health and mobile data.[25]

Addressing Privacy Issues: Successes and Challenges

The current governance frameworks seek to balance the need for patient privacy with the need for appropriate health data access. HIPAA and other regulations govern several aspects of privacy protection, including the use of de-identification, patients' access to their information, patient consent for use of their data, and nondiscrimination. The following observations from Roundtable participants and further research by CODE describe where the governance structure is working well and where it needs improvement to manage these issues.

De-identification, Anonymization, and Technical Protections for Health Data

De-identification and anonymization of health data, combined with data security, serve to protect patients from having third parties access their information and tie the data to them individually.

What's Working. HIPAA outlines specific strategies to manage de-identification of sensitive PHI and sets technical requirements for covered entities that manage PHI. To ensure de-identified data, HIPAA mandates that all entities that share PHI either utilize "Safe Harbor" guidelines or follow "expert determination" to remove PII from a patient's PHI. HIPAA's Safe Harbor outlines a comprehensive list of variables that must be removed from a person's PHI including address, medical record numbers, email addresses, and other PII. Expert determination involves convening a panel of experts with statistical and scientific knowledge to evaluate the risks of re-identification from a person's PHI. Moreover, HIPAA's technical requirements ensure that covered entities institute protective measures and safeguards for their data management systems to prevent security breaches and other possible threats.

What Needs Improvement. HIPAA could be improved to manage two issues with its technical requirements on data management.

- <u>Excessive Costs and Technical Requirements:</u> The Health Information Technology for Economic and Clinical Health Act (HITECH), passed in 2009, enforces the HIPAA Privacy Rule by mandating compliance audits of covered healthcare providers, clearinghouses, and plans. These audits evaluate a covered entity's compliance with HIPAA which includes security risks, assets and devices, physical environment, and their policies and practices to ensure that patients can access their own data safely. These policies can cost thousands of dollars to implement and create barriers to entry for small companies working to manage sensitive PHI. For example, one Roundtable participant estimated that the HITRUST certification, which is a framework for HITECH compliance which is not required by law or endorsed by the OCR, would cost their company tens of thousands of dollars and include steep monthly costs to maintain the platform.[iii]

- Non-Covered Entities: The lack of oversight for non-HIPAA entities leaves many organizations and companies that manage and use health data outside of the rules for de-identification and technical protections. For example, many wellness programs do not fall into the covered entities category for HIPAA. The Federal Trade Commission (FTC) has broad authority to ensure privacy protections, but has a relatively small staff and budget to cover the many different kinds of privacy violations that may fall under its purview.

Patients' Access to Their Personal Health Information

Patients' access to their PHI ensures that they can monitor their own chronic conditions, adhere to treatment plans, identify any errors in their health data, and directly contribute this information to research programs.[26]

What's Working. HIPAA's Right of Access ensures that patients may access their PHI from covered entities at any time. The Right of Access specifically states that a patient has the right to inspect or obtain a copy of their PHI in a designated record set. A designated record set is a group of health records that includes the medical and billing records about a patient, the enrollment, payment, claims adjudication, case or medical management record systems maintained by or for a health plan, and a set of records used to make decisions about a patient.[27] The Right of Access is critical to patient advocacy groups and healthcare companies that rely on the right of access to ensure effective treatment plans for patients and members of healthcare plans.

What Needs Improvement. The current frameworks don't fully meet the necessary requirements:

- Inconsistent Compliance by Covered Entities: Although the Right of Access provides legal access to a patient's PHI, this rule is not always followed by covered entities. A 2018 assessment of US Hospital compliance with regulations for patients' access to their PHI found that nearly half of the 83 hospitals in the study did not comply with the patient's request to obtain their medical records.[28] This discrepancy between the law and practice has increased confusion among patients and providers alike.
- Non-Covered Entities and Patients' Rights of Access: Health data is increasingly generated from an individual's smartphone, wearable, or voice assistant. These devices are manufactured by companies that do not fall under FDA guidelines or HIPAA's covered entities and business associates. Patients may face difficulties in accessing this data since company privacy policies do not need to comply with HIPAA's Right of Access. The Protecting Personal Health Data Act, introduced by Senators Lisa Murkowski and Amy Klobuchar, seeks to amend existing regulations and improve an individual's ability to access, amend, and delete a copy of their personal health data.[29]
- Merging Disparate Sources of Data: As the volume and variety of health data increases, consumers, companies, and providers are increasingly seeking to merge and aggregate data from different sources. Some of this data stems from traditional HIPAA-covered entities with sources such as EHRs or claims data, while other health data comes from entities that are not covered by HIPAA, such as creators of social determinants or genomics data from home kits. The Security Rule dictates that data outside of HIPAA, such as housing or nutrition data, becomes subject to HIPAA rules when a HIPAA-

[iii] Other participants noted that the OCR created the security rule to be scalable to the size and resources of the company, and that the OCR has different expectations for a small company when compared with larger health plans.

covered entity obtains it. But this rule is often unclear to patients, especially when social determinant data is gathered at the population rather than clinical level. It also does not regulate cases when the reverse occurs and non-covered entities obtain access to HIPAA-compliant datasets.

Patient Consent and Opt-in

A patient's ability to grant or prevent access to their PHI is an important safeguard of health data privacy.

What's Working. Under the Right to an Accounting of Disclosures section of HIPAA, patients are entitled to request information about when and why their healthcare records were accessed and this applies to a limited set of permitted disclosures. While covered entities are allowed to release data for routine reasons like treatment, payment, and healthcare operations, covered entities must receive written consent from patients for other uses of their data such as marketing communications and research.

Additionally, HIPAA in many instances aims to ensure that research subjects must grant informed consent for use of their data and must be aware of how their health data will be used. Programs like the HHS All of Us Research program and the Million Veteran Program of the Department of Veterans Affairs are precision medicine initiatives that rely strongly on a patient's willingness to provide their data for research purposes. The All of Us Research program employs a *dynamic consent* model that allows patients to adjust their opt-in and opt-out preferences as the study is carried out.

What Needs Improvement. Patients may find it difficult to grant truly informed consent because of unclear and inconsistent definitions and rules.

- Limited Range of Permitted Disclosures: The HITECH requirement to include Treatment, Payment, and Healthcare Operations (TPO) in the Accounting of Disclosures section of HIPAA has not been implemented yet by the OCR. Additionally, entities are not required to share incidental disclosures to individuals during the Accounting Disclosures process.
- Unclear Definitions of Research: HIPAA defines research as "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge."[30] Federal regulation 45 CFR Part 46 provides the framework for informed consent as an ethical principle of human subjects research. However, research is being increasingly carried out in settings that generate data outside the rules required of HIPAA-covered entities. Pharmaceutical clinical trials data, for example, falls outside of HIPAA and may not be appropriately regulated.
- Inconsistent Opt-In Rules and Regulations: Varying types of sensitive health data, such as mental health or drug addiction information, has created a fragmented approach to what data is shareable and what data is protected. Moreover, patient opt-in and opt-out rules vary widely by state and across healthcare providers for health information exchange. For example, Florida, Nevada, California, New York, Vermont, Rhode Island, and Massachusetts maintain opt-in policies that require patient consent to share data with a qualified Health Information Exchange, but many other states have no such policies.[31]
- Confusing Terms of Service Agreements: Health-related data that is managed by an entity not covered by HIPAA is often subject to that company's privacy policy and terms of service agreements. These agreements can be overly complex or obscure how the company plans to use a patient's data. Many companies continue to use complex or misleading provisions in their End User License Agreements (EULAs) such as changing the terms of conditions without notification or they fail to describe how the application will monitor individuals.[32]

Nondiscrimination and Appropriate Data Use

The possibility of community redlining and individual financial discrimination are real concerns if health data is misused. As new kinds of data analysis are developed, comprehensive guidelines and measures are needed toreduce the possibility of data misuse and resulting discrimination.

What's Working. The current governance approach aims to reduce discrimination where possible and minimize the amount of data collected by covered entities. This fits into a "Privacy by Design" approach that encourages organizations to think about the possible adverse effects of using sensitive data during their initial design phases of a health-related application or program.

- Nondiscrimination measures: HIPAA includes several key nondiscrimination measures to ensure that insurance companies cannot increase premiums or exclude members based on their health status. Health status includes a series of factors such as medical conditions, claims experience, receipt of healthcare, disability, and evidence of insurability.[33] This provision is complemented by the Genetic Information Nondiscrimination Act (GINA) which similarly aims to prohibit employer- or insurance-based discrimination based on an individual's genetic information.
- Minimum Necessary Clause: HIPAA also builds on the data minimization principle with its "Minimum Necessary" use clause, which requires covered entities not to gather noncritical information about a person. Specifically, a covered entity "must make reasonable efforts to use, disclose, and request only the minimum amount of protected health information needed to accomplish the intended purpose of the use, disclosure, or request."[34]
- Privacy Impact Assessments: The E-Government Act of 2002 mandates that any agency that collects PII must evaluate the security of its systems to ensure adequate data protection. Most federal agencies achieve this by conducting a privacy impact assessment (PIA) of their operational and developmental systems. HHS publishes all of the PIAs from its various operating divisions and also shares the PIAs from its third party websites. The PIA follows a standard template and describes the systems of data collection, the technical security measures, and approaches to addressing any individual's concern that their data may have been inappropriately used.[35]

What Needs Improvement. While HIPAA sets rules for how health data should be managed, privacy-protected, and used, it does not set standards for data collection and also does not regulate de-identified data use. Moreover, although HIPAA covers a wide array of entities that manage data, it does not account for new types of data collected by entities not covered by HIPAA.

- Unregulated Data Use: While HIPAA limits the permitted uses for data that includes PII, it sets no rules for data use and disclosure of de-identified data. There is the growing possibility that de-identified data when combined with other big data (such as retail purchases or location information) could be employed by insurance companies to restrict coverage or raise premiums for certain communities. Additionally, the risk of re-identification suggests that de-identified data shared with third parties could be used to discriminate against individuals.
- Emerging Types of Data: HIPAA does not govern uncovered entities that are gathering emerging data types like consumer-generated data when that data is not collected in or through the financial sponsorship of a covered entity. For example, an exercise tracker handed out by your doctor or health insurance company is governed by HIPAA, but when you buy it in a department store, HIPAA does not apply. Agencies like the FTC have taken a more active role in safeguarding consumer-generated health data through its health breach notification rule. Despite this advancement, the rule applies

primarily to vendors of personal health records or related entities rather than companies that manage health-related mobile applications and wearables.[36]

- Incidental Use and Disclosure: HIPAA permits certain incidental uses and disclosures that may occur as a by-product of another, permissible use of data. They are allowed as long as the covered entity has instituted a reasonable set of technical, administrative, and physical safeguards. However, poor definitions of incidental and secondary use can create confusion and hinder accountability for inappropriate uses of health data.[37]

Recommendations and Solutions

Roundtable participants proposed a series of solutions and recommendations to address the shortcomings in current health data governance. The following recommendations reflect suggestions by individual Roundtable participants and further research by CODE and do not represent a consensus of Roundtable attendees. They fall into three categories: supporting and enforcing existing regulations, regulating data that is not covered by HIPAA, and empowering patients through new tools and resources.

Supporting and Enforcing Existing Regulations

While HIPAA and other existing regulations have some limitations, they provide a framework that is familiar to healthcare providers and can be applied more strongly. Participants proposed several measures for supporting and enforcing the provisions that HIPAA already includes.

Improve Individual Access to Health Data

The Right of Access is an important tool for patient empowerment and accessing a disparate array of PHI safeguarded under HIPAA. The OCR is currently responsible for enforcing this rule and ensuring that providers and payers comply when patients request this data. Despite this, many patients are denied access to their data or unaware that they are allowed to request this data from providers.

- **Solution:** Increase OCR enforcement of the HIPAA Right to Access and better monitor compliance
- **Impact:** Increased patient awareness of the ability to access their data enhances medical transparency and can improve treatment and self-care.
- **Resources Needed:** Additional government spending on awareness campaigns, increased staffing, and adding the capacity for public tracking on the OCR website.
- **Stakeholders:** HHS, particularly the OCR, patient advocates, vendors/developers, researchers, providers, and civil society organizations
- **Policy Changes:** Enforce existing policy to maximize patients' awareness of their rights.
- **Immediate Actions:**
  - Dedicate OCR staff to manage a widespread patient awareness campaign through PSAs to increase knowledge of the Right of Access.
  - Conduct outreach to payers and providers about their compliance responsibilities.
- **Long-Term Goals:** Fulfill the promise of the law from the consumer perspective through empowerment, education, and enforcement.

Hold Health-Related "Business Associates" Accountable

The passage of the HITECH Act required that business associates comply with the security and privacy rules of HIPAA just like covered clearinghouses, providers, and payers. Despite this advancement, these business associates are not directly regulated: the covered entities they work with are responsible for ensuring that their business associates follow the rules. HIPAA should improve its direct monitoring, and enforcement of business associates, to ensure that these business associates adhere to the standards of data de-identification, limited data collection, and the range of accepted and incidental data uses.

- **Solution:** Directly monitor business associate adherence to the privacy and security rules.

- **Impact:** Improved measures to increase data access, protection, and use in line with HIPAA regulations.
- **Resources Needed:** Additional HIPAA resources, special guidance and support for business associates navigating new requirements
- **Stakeholders:** HHS, particularly the OCR; experts in privacy law; business associates such as third-party health plan administrators, accounting firms, and consultants
- **Policy Changes:** HIPAA regulations and the HITECH Act may need to be amended to include business associates, or HHS may determine that the ability to regulate business associates falls under existing laws.
- **Immediate Actions:**
  - HHS or another interested party should convene privacy experts who can provide input on challenges and opportunities for implementation.
  - Take next steps, which may include drafting legislation, depending on judgment of HHS informed by privacy experts.
  - Increase HHS staff resources implementation.
- **Long-Term Goals:** Ensure that business associates face the same level of regulation when they manage a patient's PHI as covered entities do.

Help Start-Ups Comply with HIPAA's Requirements

HIPAA sets a high standard for covered entities that gather data with PHI, and costs for HIPAA compliance can be prohibitive. Obtaining the HITRUST certification to confirm compliance can cost tens of thousands of dollars. As one solution to help companies avoid these high startup costs, the CMS has established a Virtual Research Data Center (VRDC) to provide a secure portal to efficiently use de-identified CMS data that is approved for wider use.[38] This "containerized" approach creates a HIPAA-compliant virtual sandbox where small companies can submit and run their tech applications on the portal without ever having to download the data in ways that would require them to be HIPAA-compliant.

- **Solution:** Continue to build the CMS VRDC and develop similar data "containers" for other sensitive HHS data.
- **Impact:** Ensures patient privacy with a moderated virtual portal and expands access for researchers to de-identified PHI.
- **Resources Needed:** Funding to update the current cloud infrastructure and hire staff to manage access to sensitive PHI
- **Stakeholders:** CMS, OCR, and potentially other operating divisions in HHS; private companies like Amazon and Google that provide cloud services
- **Policy Changes:** To be determined; may not be necessary.
- **Immediate Actions:**
  - Find partners to help expand the CMS VRDC and build other similar resources.
  - HHS CTO to release a Request for Information, which may be followed by a Request for Proposals to pilot potential portals.
- **Long-Term Goals:** Pilot and scale selected proposals for large-scale adoption.

Regulating Data Not Covered by HIPAA

Despite the various improvements that can be made to HIPAA, patients are now accessing data falls outside its purview. Technology companies that manufacture and sell fitness wearables, mobile applications, and home assistants are not considered "covered entities" under HIPAA. Moreover, despite its increasing use, data on the Social Determinants of Health is also not regulated by HIPAA. These recommendations outline strategies for HHS and other partners to extend the kinds of protections provided by HIPAA to other data types.

Adopt Legislation to Broaden Data Privacy Rights

Participants noted that there is currently no federal oversight of consumer-generated health data. Many participants stated that this problem should not be left to industry self-regulation, but that the House and Senate should pass comprehensive legislation to properly regulate the appropriate use of patient-generated data. One possible route is the *Protecting Personal Health Data Act* introduced by Senators Murkowski and Klobuchar, which would create a comprehensive set of policies to regulate the use and sharing of consumer-generated health data.[39]

- **Solution:** Congress passes legislation to regulate consumer-generated health data.
- **Impact:** Adopt uniform standards for consumer-generated data and ensure de-identification, consent, and sharing rules are adopted.
- **Resources:** Additional funds for staff resources in the FTC, FDA, HHS, or other federal agencies as determined by legislation
- **Stakeholders:** FTC, HHS, Congress, patient and consumer advocates
- **Policy Changes:** New policies to be established by new legislation.
- **Immediate Actions:**
  - Provide expert input from HHS and stakeholders, as appropriate, in hearings on proposed legislation.
  - HHS and stakeholders to participate in expert task force (which may be required by legislation) on specific actions to address privacy of consumer-generated data.
- **Long Term Goals:** Create a flexible and effective legal framework to protect and regulate consumer-generated data.

Create Industry-wide Ethical Guidelines for Consumer-Generated Health Data

Even with legislation in place, the private sector will need to coordinate its efforts to adopt best practices for preventing individual discrimination or group harm from misuse of health data. Companies should collaborate to produce a set of ethical guidelines that govern the use of patient-generated data. This framework could build on existing models such as the Future of Privacy Forum, Consumer Technology Association, or the CARIN Alliance. For example, the MITRE Framework for the Use of Consumer-Generated Data in Healthcare outlines a set of Principles, Values, and Guidelines for companies using consumer generated data.[40] Moreover, companies should inform their consumers about these guidelines and publicly commit to following them.

- **Solution:** Develop industry-wide ethics guidelines and best practices for managing consumer-generated health data.
- **Impact:** Increased consumer trust, and reduced risk of harm.
- **Resources Needed:** Resources for industry-wide convenings and working groups to develop best practices for managing consumer-generated health data

- **Stakeholders:** Healthcare providers, private technology companies, FTC, HHS, and civil society groups such as the CARIN Alliance and MITRE
- **Policy Changes:** None required.
- **Immediate Actions:**
  - HHS can convene a working group of companies collecting consumer-generated data to identify guidelines and best practices needed to minimize harm.
  - Work to streamline and improve user agreements for consumer literacy.
  - Draft an initial set of guidelines based on both consumer and industry feedback.
  - Adopt an awareness strategy to inform consumers of these changes.
- **Long Term Goals:** Create a flexible overview of guidelines that can be iterative and change as new forms of consumer-generated health data become prevalent.

Increase Access to Data on Social Determinants of Health – With Legal Protections

The social determinants of health have emerged as a key priority for health providers across the country. But data on determinants like economic stability and education can be difficult to access, and when it is available, it falls outside of the purview of HIPAA. This solution seeks to increase access to SDOH data while simultaneously providing legal protections needed to prevent discrimination based on the SDOH.

- **Solution:** Identify ways to increase access and use of data on social determinants of health and establish protections to prevent its misuse.
- **Impact:** Increased benefits from the use of SDOH data with appropriate safeguards to reduce the risk of harm.
- **Resources Needed:** Funding to create innovative ways to access and use data (e.g. sandboxes); funding for legal analysis of ways to mitigate risks of data misuse
- **Stakeholders:** SDOH stakeholders (e.g. housing, transportation and education agencies), community members, private sector
- **Policy Changes:** Adopting federal levers to incentivize analysis and collection of SDOH
- **Immediate Actions:**
  - Convene different sectors relevant to SDOH such as social services, community representatives, housing experts, and others.
  - Identify the low hanging fruit among easily accessible data versus more difficult data.
  - Develop engagement and feedback loops between government, private sector, and local communities for use of SDOH data.
- **Long Term Goals:** Establish defined SDOH and create an approved SDOH repository for public use.

Empowering Patients Through New Tools and Resources

Whatever regulations and protections are established, individual patients will need to be informed and involved in the management of their own data to ensure that it is used in ways they approve of. Roundtable participants suggested two paths to this kind of patient engagement.

Use Technology to Improve Patient Consent for Data Sharing

Patients face confusing choices if they are interested in granting informed consent for the use of their personal health data. Technology platforms may provide new methods for creating “dynamic consent”, whereby patients electronically "control consent through time and receive information about the uses of their data."[41] This approach could provide a transparent, flexible, and user-friendly means to make more data available for use in a way that patients can trust. This model would take a similar approach to the All of Us Research program. The expansion of smartphones and other mobile devices enable greater user control over their records and the ability to quickly update their consent preferences. This could apply to End User License Agreements (EULAs) and Terms of Service agreements as well.

- **Solution:** Use technology to create a better, more dynamic system for informed consent and a spectrum of user preferences rather than a binary opt-in and opt-out.
- **Impact:** Facilitates data sharing while preserving patient control over data use.
- **Resources Needed:** Technological capacity, political will, and monetary resources for new technology adoption
- **Stakeholders:** Patients, clinicians, software development and tech companies, researchers, government agencies, healthcare institutions, and state governments
- **Policy Changes:** Providing incentives to use the technology and adopting new regulations for implementation.
- **Immediate Actions:**
  - Identification of pilot projects for technology application.
  - Dissemination campaign for stakeholders and patients.
- **Long Term Goals:** Generating better protections for patient data and increasing trust in research.

Create Patient-Centered Outreach and Engagement Programs

Patients often feel confused and resentful in the current data landscape due to the emergence of new forms of health data, inconsistent rules and regulations, and a lack of awareness around how their health data may be used. HHS and its partners should undertake a comprehensive outreach strategy that would increase awareness of the right to access their data, the regulations designed to prevent harm from data misuse, and the resources patients have at their disposal to protect their health data. Moreover, HHS can help provide recommendations to industry to improve consumer literacy for EULAs and terms of service agreements.

- **Solution:** Create an outreach and engagement program to inform and empower patients about the uses of their data.
- **Impact:** Empowers patients and consumers with knowledge that helps them better access, use, and control their health data.
- **Resources Needed:** Funding and staff resources, marketing campaign support, patient input
- **Stakeholders:** Patients, government agencies, private healthcare companies, healthcare institutions, and state governments
- **Policy Changes:** None required.
- **Immediate Actions:**
  - HHS convenes a patient stakeholder group to advise on major health data issues.
  - Develop a campaign with videos, webinars, and visual materials to improve awareness of health data privacy issues.
  - Provide guidelines to industry that recommend language and terminology for consumer awareness.
- **Long Term Goals:** Improved patient empowerment and sense of comfort and confidence navigating health data.

Conclusion

Today's privacy landscape has been adapted since HIPAA was passed in 1996 to address a rapidly changing environment of health data. The adoption of the Privacy Rule in 2000 and the Security Rule in 2003 represent attempts to keep pace with the ever-changing nature of health data privacy. But the privacy paradigm needs to be updated further to fully address the possibility of harm from misuse of health data. The *Roundtable on Balancing Privacy with Health Data Access* addressed this issue in detail. The Roundtable found that HHS can reinforce its existing privacy governance structure while building new paths that both protect sensitive health data and enable its use for appropriate applications.

HHS has two approaches it can use to improve the governance of data privacy. First, it can draw upon its existing legal frameworks to more effectively govern the harms and risks posed to both individuals and communities. Second, HHS can advise other agencies and the federal government about the most effective approaches to balancing privacy with health data access. In both of these strategies, HHS should make recommendations with the benefit and safety of patients and communities in mind. With the right protections and safeguards, patients will benefit from using health data to improve preventive care, diagnosis, and treatment.

HHS has moved forward in this area since the Roundtable was hosted. In August 2019, HHS proposed revising the privacy rule known as 42 CFR Part 2 to facilitate research on opioid addiction in the hope of improving prevention and treatment. Participants at a Roundtable that CODE co-hosted with HHS in July 2018, which focused on data sharing to address the opioid crisis, specifically identified this rule as a major roadblock to research that restricts access to sensitive data without providing significantly more patient protection than HIPAA does. HHS has proposed additional measures to make data appropriately usable while protecting privacy. In December 2018, HHS released a Request for Information (RFI) with the objective of improving HIPAA regulations for coordination of care and improving patient-centered outcomes.[42] Other HHS rules would require providers to share a patient's health information with third party mobile applications should the patient request it. While some providers caution that this proposed rule could open people up to serious data abuses, the HHS Coordinator for Health Information Technology has stated that the rule will grant patients more ownership over their health data and allow them to make better health decisions.[43]

These rule changes demonstrate the commitment at HHS to balance privacy with health data access for research and better health outcomes. CODE hopes that this report will provide context, perspective, and elements of a framework for the important work to come.

Acknowledgements

The *Roundtable on Balancing Privacy with Health Data Access* was funded through a Patient-Centered Outcomes Research Institute (PCORI) Engagement Award Initiative (12667-CODE). This Roundtable is part of the Open Data Roundtable Series: Sharing and Utilizing Data to Enhance and Protect Health and Well-Being funded through this award.

CODE would like to thank the HHS Office of the Chief Technology Officer for their partnership in co-hosting this Roundtable series.

This report was compiled by Paul Kuhne, Roundtables Program Manager. For additional questions and follow-up, he can be contacted at paul@odenterprise.org.

We also thank the Multi-Stakeholder Advisory Committee for this series:

Lisa Bari, CMS Innovation Center, Centers for Medicare and Medicaid Services
Sohini Chowdhury, Michael J. Fox Foundation
James Craver, National Center for Health Statistics, Centers for Disease Control and Prevention
Gwen Darien, National Patient Advocate Foundation
Stephanie Devaney, All of Us Research Program, National Institutes of Health
Natalie Evans-Harris, BrightHive
Jason Gerson, Patient-Centered Outcomes Research Institute
Joel Gurin, Center for Open Data Enterprise (CODE)
William Hoffman, World Economic Forum
Charles Keckler, Associate Deputy Secretary, HHS
Lisa Khorey, Allscripts Healthcare Solutions
Michael Seres, 11 Health
Mona Siddiqui, Chief Data Officer, HHS
Paul Tarini, Robert Wood Johnson Foundation
John Wilbanks, Sage Bionetworks

CODE would like to extend additional thanks to Deven McCraw, Lucia Savage, and Mark Savage for playing an instrumental role in the development and publication of this report.

The following resources are included as appendices to this report:

- Icebreaker Activity Overview
- High Value Health Data Types One Pager
- List of Participating Organizations
- Roundtable Agenda
- References

Appendices

Health Data Icebreaker Activity

Before the Roundtable begins, we're asking you share your perspective on the risks and benefits of accessing, sharing, and using health data. Please review the instructions below and ask event staff for additional help if needed.

Purpose of the group exercise: To start the day by seeing how different groups of Roundtable participants view the level of benefit and risk for accessing different types of health data.

Instructions:

1. You will find six colored stickers in the back of your nametag. The color of your stickers reflects the stakeholder group you are part of:

   - **Red Dot:** Patient Advocacy and Engagement
   - **Green Dot:** Government
   - **Blue Dot:** Private Sector
   - **Yellow Dot:** Civil Society and Academia

   As you enter the Great Hall, you will see six posters around the room that each correspond with a different type of health data (e.g. administrative and claims data, genomic data, patient-generated data, etc.).

   Each of the six posters has a graph for potential risks and benefits of accessing different types of health data.

   - ***Potential Risks:*** The X-Axis represents the level of risk that accessing, sharing, and using this data may pose to patients and other stakeholders. 1 represents low risk while 10 represents high risk.

   - ***Potential Benefits:*** The Y-Axis represents the possible benefits of accessing, sharing, and using this data for patients and other stakeholders. 1 represents low benefit while 10 represents high benefit.

2. From your perspective, please place one colored dot sticker somewhere along the plane for each type of health data. When you have done this for all six posters, please help yourself to some light refreshments and meet your colleagues. Thank you and enjoy the day!

High-Value Health Data Types

Administrative and Claims Data

Clinical Data

- Clinical Trials Data
- EHR Data

Genomic Data

Consumer-Generated Data

- IoT Data
- Social Media Data

Social Determinants of Health Data

Surveillance Data

- Registry Data
- Survey Data
- Vitals Data

Participating Organizations

Civil Society and Academia

AcademyHealth is a leading organization for health services researchers, policymakers, and health care practitioners and stakeholders. AcademyHealth supports and conducts high-level health services research designed to improve the public's understanding of the U.S. healthcare system.

ACT | The App Association represents more than 5,000 app companies and information technology firms across the mobile economy. ACT advocates for an environment that inspires and rewards innovation, while providing the necessary resources to help its members leverage their intellectual assets to raise capital and create jobs.

American Medical Association (AMC) promotes the art and science of medicine and the betterment of public health. AMC provides timely, essential resources to empower physicians, residents and medical students to succeed at every phase of their medical lives.

Berkman Klein Center for Internet and Society, Harvard University was established with a mission to explore and understand cyberspace; to study its development, dynamics, norms, and standards; and to assess the need or lack thereof for laws and sanctions.

Bipartisan Policy Center is a Washington, D.C. based think tank that actively fosters bipartisanship by combining the best ideas from both parties to promote health, security, and opportunity for all Americans. Its policy solutions are the product of informed deliberations by former elected and appointed officials, business and labor leaders, and academics and advocates who represent both sides of the political spectrum.

Center for Digital Health Innovation supports the discovery, innovation and implementation of digital health technologies by providing developmental resources and leveraging external partnerships. CDHI focuses on data science, product management, software development, EHR integration, and project management.

Center for Open Data Enterprise (CODE) is an independent nonprofit organization based in Washington, D.C. whose mission is to maximize the value of open government data for the public good.

Columbia University Mailman School of Public Health is an international thought leader addressing critical public health issues ranging from emerging infectious diseases and urban health disparities to the implications of health policy decisions. The Mailman School is committed to creating knowledge, translating science for impact, and educating the next generation of public health leaders.

John Hopkins Bloomberg School of Public Health is dedicated to the improvement of health for all people through the discovery, dissemination, and translation of knowledge, and the education of a diverse global community of research scientists, public health professionals, and others in positions to advance the public's health.

Michael J. Fox Foundation is dedicated to finding a cure for Parkinson's disease through an aggressively funded research agenda and to ensuring the development of improved therapies for those living with Parkinson's today.

MITRE is a non-profit organization that works in the public interest across federal, state and local governments, as well as industry and academia. We bring innovative ideas into existence in areas as varied as artificial intelligence, intuitive data science, quantum information science, health informatics, space security, policy and economic expertise, trustworthy autonomy, cyber threat sharing, and cyber resilience.

Robert Wood Johnson Foundation (RWJF) is the nation's largest philanthropy dedicated solely to health. RWJD supports research and programs targeting some of America's most pressing health issues—from substance abuse to improving access to quality health care.

Sage Bionetworks is a nonprofit biomedical research and technology development organization that was founded in Seattle in 2009. Its focus is to develop and apply open practices to data-driven research for the advancement of human health. Sage believes open practices can help improve the role of data in biomedicine.

School for the Future of Innovation in Society, Arizona State University is a transdisciplinary unit at the vanguard of ASU's commitment to linking innovation to public value. It is pursuing a vision of responsible innovation that anticipates challenges and opportunities, integrates diverse knowledge and perspectives, and engages broad audiences.

Government

The U.S. Department of Health and Human Services is a cabinet-level department of the U.S. federal government with the goal of protecting the health of all Americans and providing essential human services.

The Centers for Disease Control and Prevention Center (CDC) works to protect America from health, safety and security threats, both foreign and in the U.S. Whether diseases start at home or abroad, are chronic or acute, curable or preventable, human error or deliberate attack, CDC fights disease and supports communities and citizens to do the same.

The National Center for Health Statistics (NCHS), part of the CDC, compiles statistical information to help guide policies to improve the health of Americans. Holds a biennial data user conference; consult the NCHS website for date and location. NCHS disseminates data and statistics online and in print.

Center for Medicare and Medicaid Services (CMS) administers the Medicare program and works in partnership with state governments to administer Medicaid, the Children's Health Insurance Program (CHIP), and health insurance portability standards.

The Innovation Center with CMS supports the development and testing of innovative health care payment and service delivery models.

The Immediate Office of the Secretary (IOS) is responsible for operations and coordination of the work of the Secretary.

The National Institutes of Health (NIH) seeks fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability.

National Human Genome Research Institute offers access to reliable and timely information about genomics research and the human genome. Its resources and partnerships help spark scientific curiosity, improve genomic literacy, and foster engagement among learners in different communities.

The Office of the Assistant Secretary for Preparedness and Response leads the nation's medical and public health preparedness for, response to, and recovery from disasters and public health emergencies.

The Office of the Chief Technology Officer (CTO) provides leadership and direction on data, technology, innovation and strategy across the HHS. Areas of focus include promoting open data and its use to create value, driving more efficient operations through technology utilization, and coordinating innovation strategy across the Department to improve the lives of the American people and the performance of the Department.

The Office of the National Coordinator (ONC) works to improve the health and well-being of individuals and communities through the use of technology and health information that is accessible when and where it matters most.

U.S. Department of Veterans Affairs seeks to provide veterans the world-class benefits and services they have earned - and to do so by adhering to the highest standards of compassion, commitment, excellence, professionalism, integrity, accountability, and stewardship.

U.S. Government Accountability Office (GAO) examines how taxpayer dollars are spent and provides Congress and federal agencies with objective, reliable information to help the government save money and work more efficiently.

Healthcare Providers

Beth Israel Deaconess Medical Center is a new health care system that brings together academic medical centers and teaching hospitals, community and specialty hospitals, more than 4,000 physicians and 35,000 employees in a shared mission to expand access to great care and advance the science and practice of medicine through groundbreaking research and education. BIDMC is a world-class teaching hospital of Harvard Medical School and is located in the heart of Boston.

Partners Healthcare is a not-for-profit health care system that is committed to patient care, research, teaching, and service to the community locally and globally. Collaboration among its institutions and health care professionals is central to its efforts to advance our mission.

Law and Ethics

Crowell & Moring is an international law firm representing clients in litigation and arbitration, regulatory and transactional matters. They are internationally recognized for our representation of Fortune 500 companies in high-stakes litigation, as well as an ongoing commitment to pro bono service and diversity.

Hall Center for Law and Health, Indiana University Robert H. McKinney School of Law was established in 1987 to expand the curriculum and teaching of health law and provide opportunities for students.

The Harlow Group is a healthcare consulting organization that works with healthcare providers, vendors, and payers to help them navigate the maze of regulatory and business issues facing them on a daily basis. The Harlow Group works with both acute and non-acute inpatient and outpatient facilities.

Harris, Wiltshire & Grannis LLP is a boutique firm that focuses on solving serious legal problems that call for seasoned judgment and experience. They have a first-class reputation for excellence in telecom and technology regulation, trial and appellate litigation, legal and governmental ethics, energy, national security, and privacy.

Harvard Law School, Petrie-Flom Center was established with a founding mission to promote interdisciplinary analysis and legal scholarship in the fields of Health Law Policy, Biotechnology, and Bioethics.

Polsinelli is an AmLaw 100 firm with more than 800 attorneys in 20 offices. Their attorneys build enduring relationships by providing legal counsel informed by business insight to help clients achieve their objectives.

University of Houston Law Center is a law school with particular specialty in Health Law and Policy. The Law Center was established in 1947 and enrolls more than 800 students in its degree programs. The Law Center is truly a “global” school and is well-connected with the international legal and education communities.

University of Michigan Center for Bioethics and Social Sciences in Medicine (CBSSM) is a multidisciplinary unit integrating bioethics with key social science disciplines. CBSSM acts as a "home" for anyone interested in using empirical social science methods to improve health care decisions and the ethical practice of medicine.

Verrill Danna LLP is a full-service law firm based in New England that helps individuals and businesses achieve their goals in a manner that suits their unique legal needs and preferred work methods.

Patient Advocacy

National Patient Advocate Foundation (NPAF), the advocacy affiliate of the Patient Advocate Foundation, represents the patient voice, both the powerful stories of individuals and the collective needs of the community. The NPAF’s primary objective is to prioritize the patient voice in health system delivery reform to achieve person-centered care.

Patient-Centered Outcomes Research Institute (PCORI) helps people make informed healthcare decisions, and improves healthcare delivery and outcomes, by producing and promoting high-integrity, evidence-based information that comes from research guided by patients, caregivers, and the broader healthcare community.

Patient Privacy Rights’ purpose is to honor and empower the individual’s right to privacy through personal control of health information wherever such information is collected and used. They educate, collaborate and partner with people to ensure privacy in law, policy, technology, and maximize the benefits from the use of personal health information with consent.

FORCE: Facing Our Risk of Cancer Empowered’s mission is to improve the lives of individuals and families affected by hereditary breast, ovarian, and related cancers. FORCE accomplishes this mission by creating awareness, supplying information and support to our community, advocating for and supporting research and working with the research and medical communities to help people dealing with hereditary breast, ovarian, and related cancers.

The Open Artificial Pancreas System project (#OpenAPS) is an open and transparent effort to make safe and effective basic Artificial Pancreas System (APS) technology widely available to more quickly improve and save as many lives as possible and reduce the burden of Type 1 diabetes. OpenAPS means basic overnight closed loop APS technology is more widely available to anyone with compatible medical devices who is willing to build their own system.

PatientsLikeMe is the world's largest personalized health network. 650,000+ people living with 2,900 conditions have generated more than 43 million data points, creating an unprecedented source of real-world evidence and opportunities for continuous learning. Everything members have shared empowers the community with personal agency, establishing PatientsLikeMe as a clinically robust resource that has published more than 100 research studies.

Celiac Disease Foundation has funded and executed international initiatives in three principal areas to bring an end to the suffering caused by celiac disease: medical research, patient and healthcare provider education, and public policy advocacy.

Open Medicine Institute is an organization built from the ground up to put people/patients first in driving actionable knowledge and engagement for patients. It has created a neutral, safe place for patients to collect, curate and share their own medical data and more while helping to contribute (if they desire) to a community of science.

National Committee on Quality Assurance is an independent nonprofit designed to 'turn on the lights' of healthcare. They use measurement, transparency and accountability to highlight top performers and drive improvement in the field of healthcare. They work with government and private sector clients.

Private Sector

Amazon Web Services is a subsidiary of Amazon that provides on-demand cloud computing platforms to individuals, companies and governments, on a metered pay-as-you-go basis.

Booz Allen Hamilton provides management and technology consulting and engineering services to leading Fortune 500 corporations, governments, and not-for-profits across the globe. Booz Allen partners with public and private sector clients to solve their most difficult challenges through a combination of consulting, analytics, mission operations, technology, systems delivery, cybersecurity, engineering, and innovation expertise.

Ciitizen Corporation provides a platform that helps end users collect, summarize, and share your medical records digitally. It can be used to get a second opinion, coordinate with caregivers, or donate to research.

IBM Research is a community of thinkers dedicated to addressing some of the world's most complex problems and challenging opportunities for the benefit of all. They are one of the world's largest and most influential corporate research labs, with more than 3,000 researchers in 12 labs located across six

Mathematica Policy Research is dedicated to improving public well-being by bringing the highest standards of quality, objectivity, and excellence to bear on public policy. It advances its mission through objective, evidence-based standards, superior data, and collaboration.

Microsoft is an American multinational technology company with headquarters in Redmond, Washington. It develops, manufactures, licenses, supports and sells computer software, consumer electronics, personal computers, and related services.

Mpirica Health Analytics is a digital health company that uses machine learning, backed by a robust methodology, to provide quality scores for hospitals and surgeons based on objective clinical outcomes. Its cloud-based platform and API helps patients and payers, especially self-insured employers, avoid surgery risks and costs.

Novo Nordisk is a global healthcare company with 95 years of innovation and leadership in diabetes care. This heritage has given us experience and capabilities that also enable us to help people defeat other serious chronic conditions: rare bleeding disorders, growth hormone related disorders and obesity.

Omada Health is an innovative program designed to help individuals lose weight and brings together the individualized attention of professional health coaches with a researched curriculum and manageable but powerful goals.

Virtru is a cybersecurity organization that works hard to protect user data from cyberthreats. Virtru was founded on the core belief that privacy-preserving data protection is both a fundamental right and a force multiplier for organizations.

ZeOmega's mission is to deliver proven population health management software solutions that enable our clients to enhance the value of healthcare and bend the cost curve. We deliver integrated informatics and business process management solutions so actionable information can be delivered in real-time, at the right time, and to all stakeholders in the care management continuum.

Agenda for Roundtable on Balancing Privacy with Health Data Access

U.S. Department of Health and Human Services | Monday, July 15, 2019

***Purpose**: Empower data providers and users to maximize the utility of sensitive health data while providing necessary privacy measures and addressing risk.*

10:00	**Registration and Networking** *(Coffee and light refreshments will be provided)*
10:40	**Welcome** Mona Siddiqui, Chief Data Officer, U.S. Department of Health and Human Services (HHS)
10:50	**Opening Remarks** Charles Keckler, Associate Deputy Secretary, HHS
11:00	**Structure of the Day** Joel Gurin, President, Center for Open Data Enterprise (CODE)
11:05	**Lightning Talks: Opportunities and Challenges in Advancing Health Data Privacy** Deven McGraw, General Counsel and Chief Regulatory Officer, Ciitizen John Wilbanks, Chief Commons Officer, Sage Bionetworks Lucia Savage, Chief Privacy and Regulatory Officer, Omada Health Lisa Schlager, Vice President of Public Policy, FORCE: Facing Our Risk of Cancer Empowered
11:30	**Breakout Session 1: Risks and Rewards of Accessing Different Types of Data**
12:30	**Lunch Break** (*Lunch will be provided)*
1:30	**Breakout Session 2: Effective Strategies for Balancing Privacy with Health Data Access**
2:30	**Networking Break**
2:45	**Breakout Session 3: Actionable Next Steps**
3:30	**Presentation of Highlights**
4:15	**Closing Remarks & Next Steps** Mona Siddiqui, Chief Data Officer, HHS Joel Gurin, President, CODE
4:30	**Adjourn for Reception**

To ensure openness of discussion, the Roundtable will be held under the Chatham House Rule:
Any participant is free to use information from the day but is not allowed to reveal who made any comment.

References

[1] Miner, Luke. "Opinion | For a Longer, Healthier Life, Share Your Data." The New York Times, May 27, 2019, sec. Opinion. https://www.nytimes.com/2019/05/22/opinion/health-care-privacy-hipaa.html.

[2] An overview of CODE's analysis of high-value data types is available in the appendices. For the purposes of this report, CODE has renamed "patient-generated data" to "consumer-generated data."

[3] Na, Liangyuan, Cong Yang, Chi-Cheng Lo, Fangyuan Zhao, Yoshimi Fukuoka, and Anil Aswani. "Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning Feasibility of Reidentifying Individuals by Their Protected Health Information; Feasibility of Reidentifying Individuals by Their Protected Health Information." JAMA Network Open 1, no. 8 (December 21, 2018).

[4] Rocher, Luc, Julien M. Hendrickx, and Yves-Alexandre de Montjoye. "Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models." Nature Communications 10, no. 1 (July 23, 2019): 3069. https://doi.org/10.1038/s41467-019-10933-3.

[5] "Yale University Faces Lawsuit over Employee Wellness Program." *WTNH.Com* (blog), July 19, 2019. https://www.wtnh.com/news/connecticut/new-haven/yale-university-faces-lawsuit-over-employee-wellness-program/.

[6] Group, Compliancy. "HIPAA Compliant Survey 2016 Findings." Compliancy Group, February 7, 2017. https://compliancy-group.com/hipaa-compliance-survey-2017/.

[7] "Examining Oversight of the Privacy & Security of Health Data Collected by Entities Not Regulated by HIPAA." Washington, D.C: U.S. Department of Health and Human Services, June 17, 2016. https://www.healthit.gov/sites/default/files/non-covered_entities_report_june_17_2016.pdf.

[8] Evans, Garrett. "DNA Testing Companies Launch New Privacy Coalition." *The Hill*, June 25, 2019. https://thehill.com/regulation/lobbying/450124-dna-testing-companies-launch-new-privacy-coalition.

[9] Zhang, Sarah. "How a Genealogy Website Led to the Alleged Golden State Killer." *The Atlantic*, April 27, 2018. https://www.theatlantic.com/science/archive/2018/04/golden-state-killer-east-area-rapist-dna-genealogy/559070/.

[10] Hu, Jane. "Genetic Tests Like 23andMe Promise the Moon and Stars—but What Can They Actually Tell Us?" *Pacific Standard*. Accessed June 25, 2019. https://psmag.com/social-justice/what-can-genetic-tests-really-tell-us.

[11] Board, The Editorial. "Opinion | Why You Should Be Careful About 23andMe's Health Test." *The New York Times*, February 1, 2019, sec. Opinion. https://www.nytimes.com/interactive/2019/02/01/opinion/23andme-cancer-dna-test-brca.html.

[12] Huckvale, Kit, John Torous, and Mark E. Larsen. "Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation." JAMA Network Open 2, no. 4 (April 19, 2019).

[13] "Using Consumer Health Data?" Federal Trade Commission, April 27, 2015. https://www.ftc.gov/news-events/blogs/business-blog/2015/04/using-consumer-health-data.

[14] Sokol, Emily. "How Geographic Data Can Help Address Social Determinants of Health." *HealthITAnalytics*, July 12, 2019. https://healthitanalytics.com/features/how-geographic-data-can-help-address-social-determinants-of-health.

[15] U.S. Department of Health and Human Services Secretary Alex M. Azar II, "The Root of the Problem: America's Social Determinants of Health" (speech, Hatch Foundation for Civility and Solutions, Washington D.C., November 14, 2018), https://www.hhs.gov/about/leadership/secretary/speeches/2018-speeches/the-root-of-the-problem-americas-social-determinants-of-health.html

[16] Gottlieb, Laura M., and Hugh Alderwick. "Integrating Social and Medical Care: Could It Worsen Health and Increase Inequity?" The Annals of Family Medicine 17, no. 1 (January 2019): 77–81.

[17] Callier, Viviane. "How Fake Data Protects Real People's Privacy." The Atlantic, July 30, 2015. https://www.theatlantic.com/technology/archive/2015/07/fake-data-privacy-census/399974/.

[18] Trial Data, Committee on Strategies for Responsible Sharing of Clinical, Board on Health Sciences Policy, and Institute of Medicine. Concepts and Methods for De-Identifying Clinical Trial Data. National Academies Press (US), 2015. https://www.ncbi.nlm.nih.gov/books/NBK285994/.

[19] Rights (OCR), Office for Civil. "Methods for De-Identification of PHI." Text. HHS.gov, September 7, 2012. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html.

[20] "Summary of the HIPAA Privacy Rule." Text. HHS.gov, May 7, 2008. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.

[21] "Privacy Act of 1974," June 16, 2014. https://www.justice.gov/opcl/privacy-act-1974.

[22] "Federal Policy for the Protection of Human Subjects ('Common Rule." Text. HHS.gov, June 23, 2009. https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html.

[23] U.S. Department of Health and Human Services Office of the Chief Technology Officer, "The State of Data Sharing at the U.S. Department of Health and Human Services," September 2018, Retrieved from .https://www.hhs.gov/sites/default/files/HHS_StateofDataSharing_0915.pdf

[24] "Genetic Discrimination." Genome.gov. Accessed June 7, 2019. https://www.genome.gov/about-genomics/policy-issues/Genetic-Discrimination.

[25] Tolliver, Sandy. "Maine's New Privacy Law Means Well, but Goes Wrong." *The Hill*, August 13, 2019. https://thehill.com/opinion/technology/456920-maines-new-privacy-law-means-well-but-goes-wrong.

[26] Health Information Privacy Division, Office for Civil Rights (OCR). "Individuals' Right under HIPAA to Access Their Health Information." Text. HHS.gov, January 5, 2016. https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/index.html.

[27] "45 CFR § 164.501 - Definitions." LII / Legal Information Institute. Accessed August 23, 2019. https://www.law.cornell.edu/cfr/text/45/164.501.

[28] Lye, Carolyn T., Howard P. Forman, Ruiyi Gao, Jodi G. Daniel, Allen L. Hsiao, Marilyn K. Mann, Dave deBronkart, Hugo O. Campos, and Harlan M. Krumholz. "Assessment of US Hospital Compliance With Regulations for Patients' Requests for Medical Records." JAMA Network Open 1, no. 6 (October 5, 2018): e183014.

[29] "Klobuchar, Murkowski Introduce Bill to Protect Consumer Health Data Privacy – MeriTalk." Accessed June 17, 2019. https://www.meritalk.com/articles/klobuchar-murkowski-introduce-bill-to-protect-consumer-health-data-privacy/.

[30] Rights (OCR), Office for Civil. "Research." Text. HHS.gov, May 7, 2008. https://www.hhs.gov/hipaa/for-professionals/special-topics/research/index.html.

[31] "State Health IT Privacy and Consent Laws and Policies." Accessed May 14, 2019. /apps/state-health-it-privacy-consent-law-policy.php.

[32] "Dangerous Terms: A User's Guide to EULAs." Electronic Frontier Foundation, February 17, 2005. https://www.eff.org/wp/dangerous-terms-users-guide-eulas.

[33] "HIPAA Nondiscrimination — Compliancedashboard: Interactive Web-Based Compliance Tool." Accessed August 14, 2019. https://complianceadministrators.com/hipaa-nondiscrimination/.

[34] "Summary of the HIPAA Privacy Rule." Text. HHS.gov, May 7, 2008. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.

[35] Affairs (ASPA), Assistant Secretary for Public. "Privacy Impact Assessments." Text. HHS.gov, February 13, 2009. https://www.hhs.gov/pia/index.html.

[36] "Complying with the FTC's Health Breach Notification Rule." Federal Trade Commission, April 2, 2010. https://www.ftc.gov/tips-advice/business-center/guidance/complying-ftcs-health-breach-notification-rule.

[37] Rights (OCR), Office for Civil. "Incidental Uses and Disclosures." Text. HHS.gov, January 7, 2009. https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/incidental-uses-and-disclosures/index.html.

[38] "CMS Virtual Research Data Center (VRDC) | ResDAC." Accessed August 8, 2019. https://www.resdac.org/cms-virtual-research-data-center-vrdc.

[39] "Klobuchar, Murkowski Introduce Bill to Protect Consumer Health Data Privacy – MeriTalk." Accessed June 17, 2019. https://www.meritalk.com/articles/klobuchar-murkowski-introduce-bill-to-protect-consumer-health-data-privacy/.

[40] Granger, Eldesia, Jessica Skopac, Susan Mbawuike, Anna Levin, Susan Dwyer, Arnon Rosenthal, and Jillian Humphreys. “An Ethical Framework for the Use of Consumer-Generated Data in Healthcare.” MITRE, July 2019. https://www.mitre.org/sites/default/files/publications/A%20CGD%20Ethical%20Framework%20in%20Health%20Care%20final.pdf.

[41] Williams, Hawys, Karen Spencer, Caroline Sanders, David Lund, Edgar A. Whitley, Jane Kaye, and William G. Dixon. “Dynamic Consent: A Possible Solution to Improve Patient Confidence and Trust in How Electronic Patient Records Are Used in Medical Research.” *JMIR Medical Informatics* 3, no. 1 (January 13, 2015): e3.

[42] Heath, Sara. “HHS Issues RFI for HIPAA Improvements, Care Coordination.” *PatientEngagementHIT*, December 13, 2018. https://patientengagementhit.com/news/hhs-issues-rfi-for-hipaa-improvements-care-coordination.

[43] Singer, Natasha. “When Apps Get Your Medical Data, Your Privacy May Go With It.” *The New York Times*, September 3, 2019, sec. Technology. https://www.nytimes.com/2019/09/03/technology/smartphone-medical-records.html.