# Executive Summary

This report addresses privacy and security concerns as well as recommendations related to the application of big data in the healthcare space. The Privacy and Security Workgroup (PSWG) of the Health Information Technology Policy Committee (HITPC) is charged with investigating and providing recommendations to the National Coordinator of the Department of Health and Human Services (HHS) on privacy and security issues related to the electronic exchange of health information. The application of big data in healthcare impacts one of the PSWG’s core values; specifically, that patient needs and expectations should be considered, and that “patients should not be surprised about or harmed by collections, uses or disclosures of their information.”[[1]](#footnote-2)

The collection, analysis, and use of large volumes of data will be a driver in the economy for the foreseeable future. Through the proliferation of software applications and mobile devices, the amount of health-related data is growing exponentially. As the volume, velocity, and variety of data continue to grow, so do the potential risks arising from unknown and inappropriate uses of protected health information (PHI).[[2]](#footnote-3) Many see the application of big data analytics in healthcare as an opportunity to improve the health of both individuals and their communities. Others are concerned about new risks to the privacy and security of personal information.

In response to a charge from the White House to consider the impacts of big data analyses, the PSWG invited relevant experts and interested stakeholders to testify on the opportunities and challenges of health big data, health big data concerns and harms, and the advantages and limits of current laws concerning the use of big data and emerging technologies. The PSWG held a total of three (3) hearings between December 2014 and February 2015, where 21 individuals from across the healthcare spectrum were invited to speak. The invited panelists are leading experts with a diverse perspective on issues related to big data in healthcare, and they represent a wide range of stakeholder groups, including consumer and privacy advocacy groups, consumer-facing enterprises, academia, big data analytics companies, and healthcare delivery systems.

## Opportunities

The collection and analysis of large data sets offers the promise of many improvements in the health environment. Big data is expected to improve our understanding of the efficacy and safety of medical treatments and improve outcomes for the treatment of common diseases.[[3]](#footnote-4) Additionally, big data can help us better understand the performance of medical devices like artificial joints.[[4]](#footnote-5)

The use of big data analytics in healthcare is anticipated to provide both personal and community-wide benefits. On an individual level, big data can advance personalized medicine by providing evidence for vaccine safety and providing insight into which treatments may work best for certain people.[[5]](#footnote-6) On a community level, big data is expected to advance population health and improve community-wide care[[6]](#footnote-7) by understanding how conditions like asthma, obesity, high blood pressure, and diabetes are concentrated in specific communities or grouped by age, gender, or other characteristics.[[7]](#footnote-8) Moreover, big data will continue to enable public officials to identify and track the spread of infectious diseases and respond in a timely manner.[[8]](#footnote-9)

Big data analytics can also support growth of a learning health system, which is “an environment that links the care delivery system with communities and societal supports in ‘closed loops’ of electronic health information flow, at many different levels, to enable continuous learning and improved health.”[[9]](#footnote-10) ONC recently released a draft roadmap for the interoperability of clinical data to support research and big data analyses on the path to achieving a nationwide learning health system.[[10]](#footnote-11)

Finally, data is driving rapid growth in the technology industry, particularly with regard to wearable devices and software applications that track personal health-related information. Although there are “exciting potential benefits” in wearables and other technologies, there are “cautions to be heard and challenges to be overcome.”[[11]](#footnote-12)

## Challenges

The application of big data in healthcare also faces challenges with regard to privacy and security. First, big data is not precisely defined, which makes it more difficult for legislators to enact laws that regulate the collection, use and analysis of it.[[12]](#footnote-13) Additionally, in a big data world, almost any kind of data can be health-related data,[[13]](#footnote-14)so laws or regulations that target health data or traditional health care providers, may not regulate certain uses of Big Data for health purposes, such as the use of non-health data to infer health conditions. Thus, appropriately scoping legislation or regulation is a significant challenge.

Defining privacy harm is a challenge because the concept of harm is often subjective and dependent on context.[[14]](#footnote-15) Defining acceptable uses of data faces similar challenges.[[15]](#footnote-16)

Challenges with Big Data are also found in the traditional privacy and security protection practice of data de-identification. While de-identification is a well-established and useful tool for traditional data protection, Big Data introduces new risks of re-identification due to the sheer volume of data and the broad variety of data sources in the Big Data ecosystem..[[16]](#footnote-17)

Data security is also a concern. Security risks have increased with the growth of health-related data and its value to both the health industry and to businesses. Security challenges include both internal threats and sophisticated external attacks, such as the recent Anthem breach.[[17]](#footnote-18)

Additionally, algorithms that rely on big data and that are applied to make decisions about individuals, which may shape their behaviors and opportunities, are not well understood by the public. This lack of transparency, which was repeatedly cited as a concern, creates the potential for undetected discrimination to occur, which could reduce public trust in data exchange.[[18]](#footnote-19)

Finally, the current data protection legal landscape is complex and confusing, which is partially caused by both under-regulation and over-regulation.[[19]](#footnote-20) Most health-related data generated today is not regulated by the Health Insurance Portability and Accountability Act (HIPAA), which provides baseline privacy and security rules for covered entities (health care providers, payers, and health care clearinghouses) and their business associates.[[20]](#footnote-21) On the other hand, regulations applicable to research, although well-intended, have posed challenges for researchers. In addition, patients experience barriers to gaining full access to their personal health information.

## Overarching Themes

During the course of its deliberations on big data, the PSWG identified three overarching themes. First, big data analytical methods are applying pressure on traditional fair information practice principles (FIPPs),[[21]](#footnote-22) which include transparency, individual participation (including consent), data minimization (including collection, use, and purpose limitation), and security. Regardless of these pressures, the FIPPs remain the most flexible, comprehensive, and resilient framework for ensuring the privacy of personal information.[[22]](#footnote-23)

Second, preventing harm in a big data ecosystem will be a challenge, especially when there is a lack of consensus on which uses are “harmful” or “acceptable.” While defining the ends of the spectrum on appropriate use would be fairly straight-forward (ie, uses that are clearly good and uses that are clearly harmful), defining the middle of the spectrum would be very difficult. In the middle of the spectrum, what one community would define as a harmful or acceptable use of data could be different from how another community would define it. Furthermore, the definition of harmful or acceptable use could change over time. The combination of the lack of definition of “harmful” use and inability to predict what future uses could be harmful creates challenges in developing policies to prohibit harmful use of data.

Third, efforts to appropriately address health big data confront a complex legal landscape, which continues to confuse patients, providers, health IT developers and other stakeholders in the Big Data ecosystem, including mobile app developers. Traditional health care entities and the data they collect and generate are governed by the Office for Civil Rights’ enforcement of the HIPAA privacy and security rules. However, a great deal of health-related data is now generated and consumed outside of this HIPAA-regulated space. Whereas covered entities and their business associates are bound by HIPAA’s Privacy and Security Rules, non-covered entities are subject to different legal obligations, which include the Federal Trade Commission’s (FTC) consumer protection authority to combat unfair or deceptive trade practices under Section 5 of the FTC Act. The exact same health-related data is regulated differently merely based on the entity processing the data. Additionally, data flowing between HIPAA and non-HIPAA environments may face both sets of laws and regulations. Finally, state Consumer Protection laws based on similar principles of deception, enforced by State Attorneys General, as well as State HIPAA-like laws, add an additional layer of complexity. Consequently, privacy and security risks of non-compliance are difficult to understand without painstaking mapping of data movement and a thorough knowledge of state and federal laws. This results in misapplication of existing rules which has resulted in both lost opportunity and increased risk..

This report is divided into the following six sections:

* Section 1: Executive Summary
* Section 2: Background (including the PSWG’s charge)
* Section 3: Scope (including summaries of other related initiatives)
* Section 4: Expert Testimony
* Section 5: Problem statements
* Section 6: Recommendations
* Section 7: Bibliography

[***To be Added****: for final draft, add high-level summaries of recommendations to executive summary*]

# Background

## Privacy and Security Workgroup Charge

In response to the White House report on big data[[23]](#footnote-24) and other complementary federal initiatives, the PSWG was charged to investigate privacy and security issues related to big data in the healthcare space and recommend actions to address critical challenges. This section briefly summarizes the reports that shaped the PSWG’s charge.

### White House Report on Big Data and the President’s Council on Advisors for Science and Technology Report

On May 1, 2014, the White House released a report on big data that highlights the pressure on traditional privacy-protective measures (i.e., FIPPs), such as de-identification, notice and consent.[[24]](#footnote-25) The report recommends that government “lead a consultative process to access how HIPAA and other relevant federal laws and regulations can best accommodate the advances in medical science and cost reduction in health care delivery enabled by big data.”[[25]](#footnote-26) The report acknowledges the complexity of the current federal and state legal landscape regarding patient data and privacy, and suggested the “need to carve out special data use authorities for the health care industry if it is to realize the potential health gains and cost reductions that could come from big data analytics.”[[26]](#footnote-27) Finally, the report highlights that neither HIPAA nor other privacy laws regulate many organizations that collect health-related data, and that consumer privacy expectations may not be met in the current ecosystem.

The White House report is complemented by another report released on the same day by the President’s Council of Advisors for Science & Technology (PCAST) report, which reinforced the pressure that big data places on the FIPPs.

### White House Open Government Partnership

The PSWG took note of the White House Open Government Partnership, which is a global initiative that began in 2011 to promote transparency and leverage new technologies, among other objectives. The Partnership calls for the use of big data to support greater openness and accountability, and highlights the need to “ensure privacy protection for big data analysis in health.”[[27]](#footnote-28) Specifically, it recommends that to “ensure that privacy is protected while capitalizing on new technologies and data, the Administration, led by the Department of Health and Human Services, will: (1) consult with stakeholders to assess how Federal laws and regulations can best accommodate big data analyses that promise to advance medical science and reduce health care costs; and (2) develop recommendations for ways to promote and facilitate research through access to data while safeguarding patient privacy and autonomy.”[[28]](#footnote-29)

## PSWG Plan of Action

Beginning in October 2014, the PSWG held several public meetings and hearings in which experts presented and discussed key issues. The Workgroup analyzed the testimony and began drafting and refining its recommendations.

PSWG held two days of public hearings on December 5 and December 8, 2014. The Workgroup invited panelists from industry, non-profit organizations, academia, and law to address the following issues as they relate to big data: (1) health big data opportunities, (2) health big data concerns, (3) the learning health system, (4) protections for consumers, and (5) current laws. Please see Appendix A for a list of public hearing topics and speakers.[[29]](#footnote-30)

Following these hearings, the PSWG began its deliberations. In February 2015, the PSWG heard additional testimony on health big data security issues, and in March, the PSWG updated the HITPC on the workgroup’s progress. Workgroup deliberations continued through June 2015.

# Scope

In identifying specific issues to address within its charge, the PSWG benefitted from the lessons and recommendations of other initiatives and activities. The following section summarizes several of these initiatives.

## Concurrent Initiatives and Efforts

The PSWG recognized the important big-data-related work that is being performed by both public and private stakeholders, often in partnership. The PSWG strove to complement these efforts and address gaps where they were identified. Below are brief descriptions of some of the efforts that are currently underway.

### Federal Trade Commission Internet of Things Report

In January 2015 the Federal Trade Commission (FTC) released a report[[30]](#footnote-31) on the Internet of Things (IoT). The report, which summarized the discussions from a workshop held by the FTC in 2013, focused on FIPPs-related issues such as security, data minimization, and notice and consent. One of the report’s recommendations is that Congress should enact general data security legislation.[[31]](#footnote-32) The Commission also reaffirmed its commitment to strengthen data security enforcement tools, enforce existing privacy laws, educate consumers and businesses, participate in multi-stakeholder groups, and advocate for consumers.

### Precision Medicine Initiative

Launched in 2015, the Precision Medicine Initiative aims to “generate the scientific evidence needed to move the concept of precision medicine into clinical practice.”[[32]](#footnote-33) Among the Initiative’s objectives is a commitment to protecting privacy. The White House intends to accomplish this via a “multi-stakeholder process with HHS and other Federal agencies to solicit input from patient groups, bioethicists, privacy, and civil liberties advocates, technologists, and other experts in order to identify and address any legal and technical issues related to the privacy and security of data in the context of precision medicine.” The initiative also seeks to modernize regulations by evaluating what changes are needed to support new research and care models, including a privacy and participant protection framework.[[33]](#footnote-34)

### 21st Century Cures

Launched in 2014, the 21st Century Cures initiative aims to “help accelerate the discovery, development, and delivery of promising new treatments and cures for patients and maintain the nation’s standing as the biomedical innovation capital of the world.”[[34]](#footnote-35) The House Energy and Commerce Committee released an initial discussion document on January 27, 2015,[[35]](#footnote-36) and on May 21, 2015, the Committee unanimously approved advancing the 21st Century Cure Act.[[36]](#footnote-37) The 21st Century Cures Act aims to advance interoperability among patients, researchers, providers and innovators, modernize and personalize health care, while encouraging greater innovation and supporting research.[[37]](#footnote-38)

### Federal Health IT Strategic Plan and the Shared Nationwide Interoperability Roadmap

The 2015-2020 Federal Health IT Strategic Plan builds on ONC’s previous strategy to advance the widespread adoption of health IT.[[38]](#footnote-39) Under the plan, ONC’s vision is that “health information is accessible when and where it is needed to improve and protect people health and well-being,” and its mission is to “improve health, health care, and reduce costs through the use of information technology.”[[39]](#footnote-40) Objective 5B - Accelerate the development and commercialization of innovative technologies and solutions - references big data. For this objective, ONC plans to adopt a strategy to “fund organization learning and research, and promote innovation for new health IT products and solutions” that incorporate “advances in big data, computation and analytic methods, and other scientific discoverers that use health IT securely to help resolve challenging health problems.”[[40]](#footnote-41)

ONC’s Shared Nationwide Interoperability Roadmap[[41]](#footnote-42) leverages the second goal of the Federal Health IT Strategic Plan; which is to advance secure and interoperable health information.[[42]](#footnote-43) This goal provides the foundation for achieving the balance of ONC’s goals.[[43]](#footnote-44) Big data is referenced under the learning health system (LHS) requirement for shared governance of policy and standards that enable interoperability across the health ecosystem. In the 2018-2020 timeframe, ONC plans to participate with stakeholders in a coordinated governance process to *define a policy framework* for the interoperability of clinical data that supports research and big data analyses.[[44]](#footnote-45) In the 2021-2014 timeframe, ONC and stakeholders will continue their coordinated governance process to *define criteria and implementation specifications* to support the interoperability of clinical data to support big data analysis nationwide.[[45]](#footnote-46)

### Patient-Centered Outcomes Research

The Patient-Centered Outcomes Research Institute (PCORI) is a nonprofit, nongovernmental organization established as part of the Patient Protection and Affordability Care Act of 2010. PCORI’s mandate is to “improve the quality and relevance of evidence available to help patients, caregivers, clinicians, employers, insurers, and policy makers make informed health decisions.”[[46]](#footnote-47)

PCORI funds comparative clinical effectiveness research (CER) that will provide evidence to help patients and their caregivers make better-informed decisions. To facilitate more efficient CER that could significantly increase the amount of information available to healthcare decision makers, PCORI has created PCORnet: The National Patient-Centered Clinical Research Network. PCORnet is a national patient-centered research network that seeks to leverage the power of large amounts of data, including from EHR and information reported by patients to help draw information from real-world clinical settings to conduct critical CER and other types of studies more quickly and cost-effectively.

### Secretary’s Advisory Committee on Human Research Protections (SACHRP)

The Secretary's Advisory Committee on Human Research Protections (SACHRP) provides expert advice and recommendations to the Secretary on issues and topics pertaining to the protection of human research subjects.[[47]](#footnote-48) SACHRP recently provided recommendations regarding Human Subjects Research Implications of “Big Data” Studies.[[48]](#footnote-49) Some of these recommendations called on the Office for Human Research Protections (OHRP) to provide guidance on: consent waiver standards for research, proposed changes to rules to account for an exemption category for research involving big data, and asked OCR to clarify the extent to which HIPAA applies to big data research.[[49]](#footnote-50)

## Scope of the Privacy and Security Workgroup’s Recommendations

Given the breadth of big data as a topic, the PSWG narrowed the scope of its discussions and recommendations to privacy and security concerns and potentially harmful uses of big data in healthcare. The PSWG also focused on prevailing legal frameworks and potential gaps in privacy and security protections, as well as the degree to which existing laws facilitate an environment that enables data to be “leveraged for good” while still protecting individual’s privacy interests.

The PSWG identified several issues that were out of scope. These included matters related to data quality, data standards, and the non-representativeness of data (e.g., data that does not accurately reflect the composition of the population, which has the potential to ignore under-served communities). Where possible, the PSWG sought to avoid discussing issues that have been addressed by other projects and initiatives, as summarized above, though some topics and themes were complementary.

# Public Testimony

This section provides a summary of testimony from the Workgroup’s public hearings and deliberations during the Workgroup’s regular public meetings.[[50]](#footnote-51) The hearings and meetings surfaced several key themes, which provide the following structure for this section: (1) concerns about tools commonly used to protect privacy; (2) preventing, limiting, and redressing privacy harms; and (3) the complex legal landscape, including issues of under- and over-regulation.

## Concerns About Tools Commonly Used to Protect Privacy

Big data is blurring the lines between traditional health data (e.g., clinical or billing data) and other information (e.g., user-generated data about diet, steps, workouts, sleep, and mood).[[51]](#footnote-52) Consequently, defining health data is becoming more difficult because almost all data has potential to, in some way, become health-related data, depending on how it is used.[[52]](#footnote-53) Big data analytical methods are putting pressure on traditional privacy principles (i.e., FIPPs), such as confidentiality, security, individual participation through meaningful patient consent, transparency and data minimization (including collection, use, and purpose limitation). Nevertheless, presenters defended the resiliency of the FIPPs, which continue to provide “a strong, standardized structure that promotes responsible and efficient use of data while allowing for innovations in analytics and application.”[[53]](#footnote-54)

### De-identification

De-identification refers to the data anonymization methods that obfuscate health information to keep it confidential. HIPAA provides two methods of de-identification – safe harbor and expert determination – that are widely used to facilitate health research and are considered a powerful privacy protective tool.[[54]](#footnote-55) Nevertheless, several presenters noted important weaknesses in the current HIPAA de-identification practices and offered specific solutions.

While panelists directly involved with research and big data analytics were all using de-identified data and did not require fully identifiable data for their specific research purposes, some panelists cautioned that HIPAA’s safe harbor de-identification method may be too extensive and resulted in data that did not satisfy the needs of some researchers.[[55]](#footnote-56) Conversely, other panelists cautioned that current de-identification methods may not be extensive enough to adequately protect privacy, since there is “accumulating evidence that the safe harbor method has some important weaknesses” that would allow a higher risk of re-identification.[[56]](#footnote-57) Because Safe Harbor is a technique that is being copied globally, it makes sense with respect to Big Data analytics to reevaluate when using such a simple standard can result in both beneficial use of Big Data and is also protective enough of privacy.[[57]](#footnote-58)

With respect to the expert determination method, HIPAA does not establish specific standards or require the use, or the vetting of, specific methodologies.[[58]](#footnote-59) Standards are needed to raise the bar in de-identification.[[59]](#footnote-60) Creating expert determination standards serves multiple purposes, which include (1) ensuring that methods are known, published, and scrutinized, and (2) creating a professional community of practice based on certification that could facilitate the development of more sophisticated methods and practices.[[60]](#footnote-61)

De-identification does not eliminate the risk of re-identification,[[61]](#footnote-62) but if it is done well, the risk of re-identification can be very low.[[62]](#footnote-63) De-identification can be enhanced by other controls, including accounting for threat models, contractual controls (e.g., prohibiting the joining of data sets), privacy and security controls at recipient sites, and good governance mechanisms, such as ethics committees or data access committees, which determine acceptable uses of data.[[63]](#footnote-64)

Finally, several presenters suggested the need for legal controls that prohibit and provide penalties for inappropriate re-identification, especially since de-identification cannot eliminate all risk of re-identification.[[64]](#footnote-65) They advanced a position that Congress should focus on addressing accountability for re-identification or negligent anonymization/de-identification[[65]](#footnote-66) in lieu of requiring broader use of de-identified data or mandating heightened standards for de-identification that could serve to interfere with beneficial research purposes.

### Patient consent

A patient’s meaningful consent to authorize the use and sharing of personal health information is a valuable tool for protecting privacy and individual autonomy.[[66]](#footnote-67) In a big data analytics world, it is becoming more difficult to obtain meaningful consent because secondary uses of data may not be contemplated or anticipated, as the data itself can generate the hypotheses.[[67]](#footnote-68) The Workgroup approached the consent issue by assessing how it works both within the HIPAA environment and outside the HIPAA environment, with a particular focus on consent for research.

Individual control of data through informed consent has both advantages and disadvantages.[[68]](#footnote-69) Consent empowers patients to control their information and take a more active role in their health, but consent also enables patients to withhold information, which can make data sets less valuable.[[69]](#footnote-70)

Presenters also disagreed over the degree to which people want control over their health information.[[70]](#footnote-71) Some highlighted that it may not even be possible to notify every person about how their data are used,[[71]](#footnote-72) and the complexity of privacy policies (which few people read) often makes consent meaningless.[[72]](#footnote-73) Others stated that privacy is too important to expect individuals to shoulder the burden of policing themselves.[[73]](#footnote-74) But others argued that new technologies can enable us to economically ask people for their consent, and more thought should be given to a person’s ability to opt-out or opt-in to research.[[74]](#footnote-75)

### Data security

Security, including data breaches, was highlighted as one of the most pressing issues for big data in healthcare.[[75]](#footnote-76) To build trust in a robust health big data ecosystem, one panelist recommended the “development and implementation comprehensive adaptable privacy and security policy and technology frameworks” that apply to health data, regardless of whether a HIPAA-covered hospital or a non-HIPAA-covered health app manufacturer collects the information.[[76]](#footnote-77) In any environment, data must be protected at the same time it is made available, and the methods of protection should not interfere with the responsible use of data.[[77]](#footnote-78)

Identifying the need for a larger discussion on data security, the PSWG held an additional public hearing on February 9, 2015.[[78]](#footnote-79) The main points from this hearing are as follows:

* First, risk can never be eliminated, and the security threat landscape is incredibly complex and varies over time.[[79]](#footnote-80)
* Second, the best security approach is a holistic approach to security that looks at operations end-to-end and applies a risk-based framework.[[80]](#footnote-81) HIPAA defines high-level objectives, but it falls short of establishing a risk-based framework that defines specific, contextual, and evolving controls is needed.[[81]](#footnote-82) HITRUST was mentioned as an example of a common security framework that the healthcare community may consider applying.[[82]](#footnote-83)
* Third, Moving to a common framework will be difficult for many organizations. HIPAA compliance varies significantly across hospitals based on their levels of resources.[[83]](#footnote-84) The resources go beyond IT sophistication to hospital infrastructure, and staffing.[[84]](#footnote-85) Consequently, any regulatory incentive or effort must acknowledge that compliance varies across hospitals and providers.[[85]](#footnote-86)
* Fourth, distributed data networks may augment good security practices by minimizing the need to aggregate and centralize data. For example, the Food and Drug Administration (FDA), the National Institutes of Health (NIH), and the PCORI have created or funded distributed data networks to support some of their needs.[[86]](#footnote-87)
* Finally, organizations can adopt privacy architectures, such as “safe havens” or data enclaves, and organizations can embrace distributed computation, which avoids risks associated with pooling data by performing analysis at the data sources.[[87]](#footnote-88)

### Transparency

One participant noted that “[t]he foundational principle of openness and transparency is perhaps the most important component of the FIPPs for all entities using big data.”[[88]](#footnote-89) Certainly, the need for transparency in the application of big data analytics in healthcare was echoed throughout the hearing testimony.

People are not fully aware of how their data are used and for what purpose; this extends to a common assumption that HIPAA covers all medical data when in reality it does not.[[89]](#footnote-90) This lack of understanding extends to the use of search algorithms that contextualize information and enable machines to make decisions for and about people.[[90]](#footnote-91) Besides raising concerns about the unanticipated or unexpected uses of data, poor transparency engenders a lack of trust, and trust is essential for future learning and the beneficial application of big data in healthcare.

In particular, notices often fail to foster transparency (and trust) because they are overly broad and vague.[[91]](#footnote-92) Nevertheless, entities should provide notice whenever individuals may think the usage or collection of data is unexpected or objectionable, and notice should be provided at a relevant time (e.g., contextual (just-in-time) notice).[[92]](#footnote-93)

As to increasing transparency for algorithms, it was acknowledged that because sophisticated algorithms are proprietary intellectual property, it is very difficult to determine their inputs and outputs, and how they make decisions about people.[[93]](#footnote-94) One participant suggested that transparency and disclosure should extend to “what [data] informs the algorithms, how … cohorts are defined, and how individuals are separated,” because if these are not known, trust cannot be achieved.[[94]](#footnote-95)

Finally, potential parallels were drawn to the transparency elements of the Fair Credit Reporting Act (FCRA). The FCRA not only frames acceptable uses of data, it also provides consumers with transparency if data has an adverse impact on them. Even so, some cautioned that FCRA is tailored to particular circumstances and it may not scale well in the health arena.[[95]](#footnote-96)

### Collection, use, and purpose limitation

Big data analytics and research begins with researchers examining trends and patterns in large data sets without first formulating a hypothesis.[[96]](#footnote-97) The need to collect as much data as possible before identifying a research purpose conflicts with longstanding FIPPs principles around limiting the use of personal information.

Regardless of how big data analytics is performed, hearing participants stated: “as a general principle, the minimum necessary amount of identifiable data should be used to answer a question.”[[97]](#footnote-98) Others suggested that organization should often ask themselves why they need the information they have collected, and they should avoid retaining data for some future, unnamed use simply because they think it might be valuable.[[98]](#footnote-99)

Concerning purpose limitation, participants struggled to clearly define acceptable and unacceptable uses of health information in big data analytics,[[99]](#footnote-100) save the obvious prohibition on using health data for discriminatory purposes (see more on harms, below).

In summary, the PSWG heard many concerns regarding the tools commonly used to protect privacy. The PSWG also heard several suggestions and recommendations. In particular, the PSWG heard following recommendations regarding the appropriate use of electronic health data:

* As a general principle, the minimum necessary amount of identifiable data should be used to answer a question;
* There should be good processes for approval and oversight; and
* The uses of data should be stated publicly and the number of individuals who have access to identifiable data should be minimized.[[100]](#footnote-101)

## Preventing, limiting, and redressing privacy harms

Defining privacy harm is very difficult, and panelists were not able to reach a consensus regarding a clear definition of harm.[[101]](#footnote-102) Defining an acceptable use of data is subjective because acceptable use is culturally specific and will change over time.[[102]](#footnote-103)

Moreover, data can be used for either positive or negative effect.[[103]](#footnote-104) For example, understanding demographic granularity can help address health disparities, but it can also increase the risk of harmful profiling.[[104]](#footnote-105)

To arrive at a consensus around harms or non-permitted abuses, one participant suggested an effort be made to identify laws that may already prohibit certain activities, identify gaps, and catalogue federal laws with non-discrimination provisions.[[105]](#footnote-106) Additionally, the FTC and other agencies may have a role to play. The FTC can help can help identify boundaries through the cases it pursues under its ability to combat unfairness and deception.[[106]](#footnote-107)

## The complex legal landscape

Complexity in the legal landscape, both within the HIPAA environment and outside of it, was a core theme during each public hearing and meeting. Testimony confirmed that there continues to be a lack of clarity and understanding of privacy and security laws and rules.[[107]](#footnote-108) Additionally, the legal landscape for protecting health information is uneven. For example, legal coverage is extensive and even contradictory in some areas (e.g., research under HIPAA and the Common Rule), while coverage is significantly lacking in other areas.[[108]](#footnote-109) Moreover, State law is confusing, often outdated and seldom enforced.[[109]](#footnote-110)

HIPAA applies only to covered entities (health plans, health care clearinghouses, health care providers) and business associates acting directly on their behalf. The bulk of health-related data being generated today falls outside of HIPAA regulation.[[110]](#footnote-111) Regardless of which laws apply – HIPAA for healthcare entities or the FTC for much of the private sector – both individuals and organizations continue to struggle to gain access to health information so that it can be used and studied meaningfully.

This section summarizes the testimony concerning the legal landscape including concerns about access to information, under-regulation, and over-regulation.

### Access to Information

Participants voiced a need for greater data liquidity.[[111]](#footnote-112) On a personal level, patients want to access and combine their data in meaningful ways, but they face significant impediments[[112]](#footnote-113) and privacy and security issues are frequently seen as one of the biggest barriers.[[113]](#footnote-114)

Public access to data is also very important. Health data can be viewed as a social asset; there is a social responsibility to give back to the community by making data available to researchers and to patient groups.[[114]](#footnote-115)

Within HIPAA, providers may access patient information without their consent for treatment, payment, or healthcare operations purposes.[[115]](#footnote-116) This concept of normal, routine uses was cited as one of HIPAA’s greatest strengths.[[116]](#footnote-117) Participants also cited some progress regarding data liquidity. The recently finalized HIPAA Omnibus Rule introduced important changes, including (1) authorization to permit future research and (2) the ability to permit compound authorizations for research purposes. [[117]](#footnote-118) Nevertheless, testimony generally indicated that access to data still needs to be improved for big data to benefit personal and public health.

### Under-Regulation

For the purposes of the PSWG’s investigation, “under-regulation” refers to the gaps in law in which health-related data is not afforded the same privacy and security protections that exist under a regime like HIPAA. Although the FTC has taken the lead to protect privacy and security in the commercial space, the FTC usually only acts if an organization violates its own privacy and security policies. To date, the FTC has refrained from pursuing cases involving “unfair” privacy practices, which can be difficult to define. Moreover, private industry is not subject to privacy and security rules and policies set forth for government entities (e.g., requirements under the E-Government Act of 2002 (which includes the Federal Information Security Management Act (FISMA).

A rapidly growing amount of health-related information is not regulated by the HIPAA.[[118]](#footnote-119) These include information in mobile applications, websites, and personal health records. Additionally, healthcare entities that are covered by HIPAA are using non-health data for healthcare purposes.[[119]](#footnote-120)

Although the FTC has recently become very active on the general enforcement of data security standards,[[120]](#footnote-121) it has been less active in the privacy sphere, particularly in healthcare. One reason is that while deception is relatively straightforward to litigate, it is not clear what may be considered a legally enforceable “unfair” privacy practice.[[121]](#footnote-122)

### Over-Regulation

The PSWG used the term “over-regulation” to refer to the multiplicity of laws addressing certain holders of health and health-related data and the extent to which those laws help leverage beneficial uses of health big data. PSWG discussions focused mainly on research, which is regulated by both HIPAA and the Common Rule.

One panelist stated that the HIPAA Privacy Rule does not protect privacy as well as it should, and in fact, HIPAA impedes the use of data for important health research.[[122]](#footnote-123) Others cited HIPAA’s strengths, noting that it establishes common rules that apply uniformly; which serves to improve access to information.[[123]](#footnote-124) Ideally, entities should not be penalized and disincentives should not be created when organizations contribute to the general knowledge base for healthcare.[[124]](#footnote-125)

For example, one presenter noted an apparent “paradox” in HIPAA. While the definition of “research” is the same under both HIPAA and Common Rule,[[125]](#footnote-126) different rules about patient consent are applied depending on whether the research results are shared for “generalizable knowledge” or are used for quality improvement purposes and kept within an organization (i.e., covered under “healthcare operations”). Stated another way, “two studies that use data for quality improvement purposes using the same data points done to address the same question … by the same institution will be treated as operations if the results are *not* intended to contribute to generalizable knowledge, … but treated as research [requiring consent] if you intend to share the results with others so that learning may occur.”[[126]](#footnote-127)

Finally, the PSWG’s predecessor, the Privacy and Security Tiger Team, previously provided recommendations[[127]](#footnote-128) approved by the HITPC on the topic of modernizing the Common Rule and creating more consistency with in HIPAA.[[128]](#footnote-129) The Tiger Team recommended that uses of EHR data to evaluate the safety, quality, and effectiveness of prevention and treatment activities should not require consent or IRB approval. Consequently, such investigations should not be labeled as research – even if the results are used for generalizable knowledge – because doing so would pose an obstacle to learning.[[129]](#footnote-130)

## General Suggestions

In summary, the PSWG received many hours of helpful testimony over the course of several days of public hearings and meetings. In assessing the concerns raised about protecting privacy and security in health big data analytics, presenters offered several general suggestions. Some of these suggestions follow:

* It is important to allow experimentation for technology and methods to improve. It is also important that organizations, that are initially slow to move, learn how best to take advantage of big data opportunities and realize potential benefits.
* The best approach for protecting privacy is to start with the FIPPs. The FIPPs are flexible yet structured, and can apply to the traditional healthcare sector as well as the emerging consumer applications market.[[130]](#footnote-131)
* Finally, the PSWG might consider three options to address legal gaps:
	+ Develop a specific set of principles applicable only to “non-HIPAA health care data” (with an obvious ambiguity about what “health care data” would mean);
	+ Develop a set of principles (through an amendment to the scope of HIPAA or otherwise) that would apply to all health care data; or
	+ Develop a broader general privacy law that would apply to all personal data (with or without a carve-out for data currently covered by the HIPAA rules.[[131]](#footnote-132)

# Detailed Problem Statements

As electronic health IT adoption and use has advanced, large data sets of health information have been formed within electronic health records, health applications and personal health records. Data mining and the application of big data analytics across these data sets offer promising opportunities for learning and for improving patient outcomes, but they also uncover additional challenges to maintaining privacy and security. The diversity of health data, the structure of the healthcare environment, and policy gaps further complicate the beneficial impact of big data. Analysis and discussions by the PSWG yielded high priority areas for discussion and solution development with a focus on the big data analytics. This section outlines the key problem areas for focus and provides greater detail about the specific problem that needs to be addressed.

## Potential for Harmful or Discriminatory Practices

During our hearings, among the most oft cited concerns about health “big data” is the potential for health data to be collected and used in a way that harms individuals or groups. Discrimination is just one example of a harm that can result from certain analytic uses of health big data. U.S. laws prohibit some discriminatory uses of health data – for example, use of health data to make decisions about health insurance coverage – but other discriminatory uses of health data are either not prohibited or are expressly permitted (for example, use of health information in life and disability insurance decisions).

Beyond discrimination, some see other uses of health data as being “harmful” (for example, marketing and other “commercial” uses). However, there is a lack of consensus on which uses are “harmful,” particularly with respect to health big data analytics, as well as an inability to predict which future uses could be harmful and which beneficial, which creates challenges to enacting policies to prohibit or place additional constraints on such uses. During our hearings, some presenters expressed concern about the use of algorithms to make decisions about people or communities, and the lack of “transparency” about both the data used to inform these algorithms and precisely how the algorithms are used. Unfair practices resulting from the use of algorithms are more insidious when further obscured by a lack of transparency since the harm itself may be difficult if not impossible to detect. Thus, it is only with an understanding of the process used to arrive at a decision that can reveal the existence of harmful bias or practices.

Failing to pay attention to these issues undermines trust in health big data analytics, which could create obstacles to leveraging health big data to achieve gains in health and well-being.

## Two different domains of regulation (HIPAA and “Other”) Yields contradictions and unpredictability

HIPAA covers many sources of health big data – but not all. Consequently, we lack comprehensive, FIPPS-based protections for health data analytics (and analytics leveraging data that on its face is not “health” but is used for health purposes or to infer a health status) in many domains, which is confusing for consumers and imperils trust in health big data. In addition, even when health data analytics is regulated, those rules may not have been written in a way that maximizes our ability to learn from health data while still protecting it from risks to privacy, confidentiality and security. Three concerns in particular were surfaced by the hearings:

* Access – Individuals often lack the ability to access, use, and share their own data, including for research and learning health system (LHS) activities. Even with respect to HIPAA covered entities, which are required to provide this right to individuals, the right is often difficult for individuals to exercise.
* Transparency – There is a lack of transparency regarding how holders of personal information use that information and how information is exchanged, especially in the Big Data ecosystem outside of traditional healthcare. This lack of transparency erodes trust and exacerbates the fear of harm or discrimination.
* Research – when it is regulated, the rules do not necessarily regulate based on privacy risk and, as a result, create higher hurdles for uses of data for “research” purposes that intend to contribute to “generalizable knowledge” (i.e., the greater good).

## Lack of Confidence in De-identification Methodologies and the Risk of Re-identification

De-identification is a useful tool for protecting privacy in big data research – but we over-rely on it as a matter of policy and do not have ways to hold people accountable for unauthorized re-identification of data or negligently failing to protect data that is vulnerable to re-identification. In addition, de-identification does not address the potential for harmful uses of health big data.

HIPAA has regulatory requirements for de-identification – but there are no such requirements for de-identification of health data outside of HIPAA. HIPAA standards for de-identification are often voluntarily used by entities not subject to HIPAA, but it is not required.

Concerns have been raised about both methodologies currently used for de-identification under HIPAA – safe harbor and expert determination. The former may not be sufficiently protective in all contexts (particularly given increases in publicly available data); the expert methodology is required to take “context” into account but there are no objective criteria governing it.

There is increased risk of re-identification when data sets are combined (the mosaic effect). A mosaic effect occurs when disparate threads can be pieced together in a way that yields information that is supposed to be private.[[132]](#footnote-133)

In addition, de-identification – even under HIPAA – has never meant zero risk, but de-identified data is not subject to regulation (so the residual risk that remains is unregulated). We do not have consistent mechanisms for punishing people/entities who re-identify or who negligently leave datasets vulnerable to easy re-identification.

Conversely, de-identification is also not the panacea for enabling valuable uses of data. Emphasizing (or favoring, through reduced regulatory requirements) data de-identified pursuant to HIPAA as the enabling mechanism for data use often significantly reduces the potential for valuable uses of data even where the risk associated with the use of more identifiable data is very low. In addition, de-identification using the expert methodology, which is generally believed to be both more effective at reducing re-identification risk (because it accounts for context) and more valuable for researchers (because it doesn’t per se require the removal of certain data fields) is perceived by many research entities to be too expensive and time intensive.

## Security Threats and Gaps

The lack of an end-to-end secure environment for health data was a problem mentioned by many who presented – but no entity (or federal agency) is responsible for assuring those end-to-end protections. Instead we have silos of protections. For example, HIPAA coverage applies in some places, FTC and FDA in others, Gramm-Leach-Bliley in financial contexts; state law may govern; some may be covered by multiple laws, and some may be covered by none. The lack of baseline security requirements was broadly seen as a significant risk for deteriorating patient and consumer trust in the healthcare system and in entities involved in health big data analytics both inside and outside of healthcare. The call for such end-to-end security requirements was referenced as one of the highest priorities.

In addition, the laws that do exist do not necessarily provide incentives for adopting privacy-enhancing technical architectures for big data analytics (for example, data enclaves). (In other words, the privacy rules governing analytic uses of data arguably are the same regardless of the technical architecture used to analyze the data.)

Congress is the only policy-making body equipped to authorize national security and/or cybersecurity requirements that would facilitate the requirement to provide a consistent baseline level of security for health data, regardless of the entity that holds that data, in an end-to-end environment that is desirable for building trust. But the workgroup did not recommend specifically directing Congress to address this issue at this time.

# Solutions and Recommendations

## Addressing Harm, Including Discrimination Concerns

To address discriminatory practices: without a national consensus on what constitutes harm with regard to health big data analytics, the Workgroup encourages ONC and other federal stakeholders to conduct more public inquiry and pursue or promote initiatives or projects that could yield greater understanding of the scope of the problem and the potential for harm - both harm to individuals and harm to communities or subpopulations.[[133]](#footnote-134) Federal stakeholders should continue to focus on identifying gaps in legal protections against what are likely to be an evolving set of harms from big data analytics.

Additionally, policymakers should consider adopting measures (for example, spending conditions, regulations, and guidance) that could increase transparency about actual health data uses. Greater education and knowledge about actual health data uses could help spur greater public dialogue about which uses are harmful, and as a result advance a national consensus around harms and the best ways to prevent or hold entities accountable for them.

With respect to addressing distrust in big data algorithms, the workgroup expressed a desire to have greater transparency about algorithms – for example, what data informs them, how the data are collected, how those data are weighted or used in the algorithm, and whether (and if so, how) the algorithm is evaluated with respect to the accuracy and fairness of its outcome. At the same time, the workgroup recognizes that many algorithms are considered to be proprietary and frequently are machine-generated, so there is less than complete understanding of the inputs and the processes even among those using the algorithms. Additionally, the workgroup recognized that detailing all of the data inputs for a given algorithm, may, in many cases, be a near impossible task given the ephemeral nature of the data input and the volume of data utilized. Nevertheless, the workgroup recommends policymakers explore how to increase the transparency around the use of algorithms, perhaps with an approach similar to that used in the Fair Credit Reporting Act (FCRA). The FCRA is a federal law that regulates consumer reporting agencies (CRAs) and empowers people by providing transparency about the use of consumer credit information where the credit information is used in algorithms to create a credit score. The FCRA offers the following protections:

* A consumer must be told if his/her information has been used to deny them service or opportunity.
* A consumer has the right to know what information is in their credit file
* A consumer has the right to ask for the credit score that has been computed and where it falls on the scale of other credit scores
* A consumer has the right to dispute incomplete or inaccurate information in their credit file
* Consumer reporting agencies must correct or delete inaccurate, incomplete, or unverifiable information.
* Consumer reporting agencies may not report outdated negative information and may not use such information in computing credit scores
* Access to and the use of a consumer’s file must be limited.
* A consumer must give consent for reports to be provided to employers.
* A consumer may limit “prescreened” offers of credit and insurance based on information in their credit report.
* A consumer may seek damages from violators.[[134]](#footnote-135)

Any such regulation or best practices governing algorithms should aim to maximize transparency (to the extent possible), validity and fairness. Although this may be desirable for algorithms used in a range of contexts, it is particularly important where algorithms are used to evaluate and/or make decisions that have an impact on the health of individuals and communities.

## Address Uneven Policy Environment

The Health IT Policy Committee has issued previous recommendations urging that holders of health data (and personal data being used for health purposes) implement protections based on the Fair Information Practice Principles (FIPPs) to protect the privacy, confidentiality and security of that data. FIPPs include provisions to enable individuals to make reasonable choices about the collection, use and disclosure of their health information – but the FIPPs do not focus just on consent as the primary mechanism. FIPPs are principles of responsible data stewardship and obligate data holders to adopt reasonable limits and safeguards regardless of whether an individual’s consent is sought. HIPAA and other privacy laws are based on FIPPs – but we lack FIPPs-based protections for health data outside of the HIPAA environment.

Congress could address this through legislation, but such protections could be achieved through voluntarily adopted codes of conduct, which can be enforced under Section 5 of the FTC Act by the FTC for entities subject to their jurisdiction. A number of efforts are under way to develop such codes – those efforts should be encouraged and HHS, FTC and other relevant federal agencies should offer to review and provide suggestions for such efforts in order to more quickly establish dependable “rules of the road” that help build trust in the use of health big data (Of note: the Health IT Policy Committee has already asked the Consumer Workgroup to consider an evaluation effort for consumer-facing health data tools like health mobile apps.)[[135]](#footnote-136) Such codes of conduct should emphasize transparency (regarding data collection, transfer and use), individual access, accountability, and use limitations, at a minimum. They could also reward/promote the use of privacy enhancing architectures for big data analytics, such as data enclaves. A data enclave is a controlled, secure environment in which eligible researchers can perform analyses using restricted data resources.[[136]](#footnote-137)

Policymakers also should evaluate existing rules governing uses of data that could contribute to a learning health system to assure those rules promote the responsible re-use of data to contribute to generalizable knowledge. The Policy Committee had previously recommended treating certain research uses of data conducted under the management and control of a HIPAA covered entity as operations (not requiring consent or IRB review), and we reiterate that recommendation.

Policymakers also should consider modifying rules around research uses of data so that they provide incentives for entities to use more privacy protecting architectures (for example, entities using secure data enclaves for research would not need to undertake as significant a level of de-identification).

Existing rules giving individuals access rights to health information should, over time, be strengthened to bring them into the digital age, so that individuals can access, download, and transmit their health information (both within HIPAA and as part of any rules or voluntary codes covering the non-HIPAA space) as easily as they can access their financial information, for their own use or in order to facilitate research into diseases that impact them or in any area of learning that they seek to support.

In the meantime, education of consumers, health care providers, technology vendors and other stakeholders about the limits of legal protections, and about best practices to protect the privacy, confidentiality and security of sensitive health information is critical, particularly given the patchwork of regulation and the lack of comprehensively adopted, robust codes of conduct. The Health IT Policy Committee recently endorsed recommendations from this Workgroup with respect to providing guidance and educating stakeholders on these topics; we reinforce those recommendations again.[[137]](#footnote-138)

## Protect Health Information by Improving Trust in De-identification Methodologies and Reducing the Risk of Re-identification

We ask OCR to be a more active “steward” of HIPAA de-identification standards and conduct ongoing review of the methodologies to determine robustness and recommend updates to the methodologies and policies. The analysis could be performed by an outside expert, such as NIST, but would be vetted and ultimately endorsed by OCR.[[138]](#footnote-139)

We further urge OCR to carefully consider the following additional recommendations that came out of the hearing testimony:

* Limit use of safe harbor only to circumstances where data represent a random sample of a population. “The Safe Harbor standard for de-identification is being copied and used by uncovered entities and is actually being used globally, it’s been copied globally as well. So, we need to revisit the value and the risks from using such simple standards for de-identifying data and maybe additional guidance is needed to limit the situations under which such a simple standard...such a simple method should be used."[[139]](#footnote-140)
* Consider whether de-identification status of a dataset should be required to be re-evaluated when context changes (such as when data set is combined with other data).
* Develop or encourage the development of programs to objectively evaluate statistical methodologies; consider granting safe harbor status to methodologies proven to be effective in particular contexts.

Consideration should be given to risk-based de-identification requirements and re-identification risk when data is held by entities or in environments where re-identification risk remains low (for example, data enclaves, data havens or data repositories voluntarily adopting HIPAA security rules).

Establishing accountability for re-identification or negligent de-identification also was of interest to the Workgroup. This is another issue that Congress could address – however, the workgroup did not believe specifically asking Congress to address this at this time was the advisable.

## Supporting Secure Use of Data for Learning

The PSWG seeks a widely-accepted security framework that assures accountability for security at all endpoints; It is yet another issue Congress could address – but consistent with prior recommendations, the workgroup urges the development of voluntary codes of conduct that also address robust security provisions. The FTC has previously recommended the enactment of strong, flexible, and technology-neutral legislation to strengthen the Commission’s existing data security enforcement tools.[[140]](#footnote-141) In addition, education of stakeholders about cybersecurity risks and recommended precautions is critical, and both the public and private sectors have a role to play in this effort.

Federal policymakers, through regulations, spending conditions, and guidance, should provide incentives for entities to use privacy-enhancing technologies and privacy-protecting technical architectures, such as secure data enclaves, secure distributed data systems, and distributed computation.

The workgroup also reiterates recommendations made by the Tiger Team and endorsed by the Health IT Policy Committee in 2011[[141]](#footnote-142) with respect to the HIPAA Security Rule. Specifically:

* Security policies for entities collecting, storing and sharing electronic health information needs to be responsive to innovation and changes in the marketplace.
* It also needs to be flexible and scalable to reflect differences in size and resources; at the same time a solid baseline of security policies needs to be established and consistently implemented across all entities.
* Providers will continue to need education and specific guidance on how to comply with the security rule.

HHS should have a consistent and dynamic process for updating security policies and the rapid dissemination of new rules and guidance to all affected. As part of this process, HHS should look to other security frameworks to assure the Security Rule keeps up with the latest threats and innovations in security protections. NIST had previously issued guidance on HIPAA security compliance that many entities have found helpful; NIST should continue to update this guidance and keep it current and relevant for a changing risk environment.

# Bibliography

# Appendix A – Health Big Data Public Hearing Topics and Speakers

## Health Big Data Public Hearings, December 5 and 8, 2014

|  |  |
| --- | --- |
| Day 1 – Friday, December 5, 2014 | Day 2 – Monday, December 8, 2014 |
| Panel 1: Health Big Data Opportunities and the Learning Health System (LHS) | **Panel 1: Current Law** |
| * Steve Downs, RWJF
* Richard Platt, Harvard Pilgrim
* Patricia Brennan, U. Wisconsin
 | * Melissa Bianchi, Hogan Lovells
* Kirk J. Nahra, Wiley Rein
* Deven McGraw, Manatt, Phelps & Philips, LLC
 |
| Panel 2: Health Big Data Concerns | **Panel 2: Health Big Data Opportunities** |
| * Michele DeMooy, CDT
* Mark Savage, NPWF
* Anna McCollister-Slipp, Galileo Analytics
 | * Linda Avey, 23 and Me, Curios, Inc. (invited but could not attend)
* Kald Abdallah, Project Data Sphere
* Ella Mihov, Ayasdi
 |
| Panel 3: Protections for Consumers | **Panel 3: Learning Health System** |
| * Khaled El Emam, U. of Ottawa
* Bob Gellman, Private Consultant
* Fred Cate, Indiana U.
 | * Paul Wallace, Optum Labs
* Josh Gray, AthenaHealth
 |
|  | **Panel 4: Health Big Data Concerns** |
|  | * Leslie Francis, U. Utah
* Melissa Goldstein, George Washington U.
 |

## Data Security in Health Big Data Hearing, February 9, 2014

|  |  |  |
| --- | --- | --- |
| Panelist | Organization | Position |
| Andrei Stoica | IMS Health | VP of Global Systems Development and Security |
| Denise Anthony | Dartmouth College | Vice Provost for Academic Initiatives, Professor of Sociology; SHARPS contributor |
| Ryan Anderson | Milliman | Director of Software as a Service |

# Appendix B – Supporting Testimony

| Topic | Testimony |
| --- | --- |
| Concerns About Tools Used to Protect Privacy |  |
|  | * All data can be health data, or data from which inferences about health are drawn or correlations with health are made.[[142]](#footnote-143)
* Some say the FIPPs are unsuited for the era of big data[[143]](#footnote-144) (e.g., analytical methods are putting pressure on traditional principles such as confidentiality, security, individual participation through meaningful patient consent, transparency and data minimization (including collection, use, and purpose limitation).
* Nevertheless, the FIPPs continue to provide “a strong, standardized structure that promotes responsible and efficient use of data while allowing for innovations in analytics and application.”[[144]](#footnote-145)
 |
| De-identification |  |
|  | * De-identification does not eliminate the risk of re-identification.[[145]](#footnote-146) Nevertheless, if de-identification is done well, the risk of re-identification can be very low.[[146]](#footnote-147)
* While some research organizations indicated their satisfaction with de-identified data sets,[[147]](#footnote-148) others stated that sometimes it is necessary to use fully identified data (e.g. when electronic health data must be matched to an external source like the National Death Index).[[148]](#footnote-149) Some stated that HIPAA’s safe harbor de-identification method may not give researchers the data they need or want. Limited data sets are slightly more robust, but still may or may not be sufficient for research needs.[[149]](#footnote-150)
* The group agreed that when identifiable information is used, “it should be stored in highly protected locations like data enclaves.”[[150]](#footnote-151)
* There is “accumulating evidence that the safe harbor method has some important weaknesses” that would allow data to be shared with a higher risk of re-identification.[[151]](#footnote-152) These risks include a reduction in the data utility and the consequence that under certain conditions, Safe Harbor allows data to be shared with a higher risk of re-identification.[[152]](#footnote-153) Safe harbor is being copied and used globally, so HHS should reexamine the value of such simple standards and provide additional guidance to limit situations when simple standards are applied.[[153]](#footnote-154)
* There are no widely accepted standards for expert determination method and there is “ no homogeneity in how de-identification is actually done.”[[154]](#footnote-155) Standards are needed to raise the bar in de-identification.[[155]](#footnote-156) Creating standards for the expert determination serves multiple purposes. These include (1) ensuring that methods are known, published, and scrutinized, and (2) creating a professional community of practice based on certification that could facilitate the development of more sophisticated methods and practices.[[156]](#footnote-157)
* Participants echoed the rise in demand for standards, as a lack of guidance is inhibiting willingness to share data and IRBs are uncomfortable evaluating privacy issues in the face of conflicting advice.[[157]](#footnote-158) The HITRUST alliance is already working on a general health standard for de-identification.[[158]](#footnote-159)
* Additionally, as experience with expert determination grows, one could account for “the value of information in different settings” and balance whether to use experts to de-identify data or mimic or replace their processes with a degree of automation.[[159]](#footnote-160) This works when experts can anonymize in a fixed space with known data elements, but the process may require change when external data elements are introduced (as mosaicking may increase re-identification risks).[[160]](#footnote-161)
* De-identification can be enhanced by other controls. These include contractual controls (e.g., prohibiting the joining of data sets), privacy and security controls at recipient sites, and good governance mechanisms, such as ethics committees or data access committees, which determine acceptable uses of data.[[161]](#footnote-162) Additionally, organizations can adopt privacy architectures, such as “safe havens” or data enclaves, and organizations can embrace distributed computation, which avoids risks associated with pooling data by performing analysis at the data sources.[[162]](#footnote-163)
* Several presenters suggested the need for legal controls that prohibit and provide penalties for re-identification, especially since de-identification cannot be eliminate all risk of re-identification.[[163]](#footnote-164) They thought that the Congress would need to address accountability for re-identification or negligent anonymization/de-identification[[164]](#footnote-165).
 |
| Patient Consent |  |
|  | * Consent is a privacy and security issue for both providers in the HIPAA environment as well as for app developers and wearable device manufacturers outside the HIPAA space.[[165]](#footnote-166) While HIPAA provides for certain expected uses of data that do not require consent (e.g. sharing for treatment, payment, and healthcare operations among covered entities), rules for consent outside the HIPAA space are less structured and rely on the FTC to protect consumers by working to prevent unfair or deceptive acts or practices.[[166]](#footnote-167)
* Individual control of data through informed consent has both advantages and disadvantages.[[167]](#footnote-168) Consent empowers patients to control their information and take a more active role in their health, but consent also enables patients to withhold information, which can make data sets less valuable.[[168]](#footnote-169)
* Presenters disagreed over the degree to which people want control over their health information.[[169]](#footnote-170)
* One presenter stated, “it is not possible to obtain individual consent for all uses of an individual’s data, and it may be impossible to notify every person about all the uses of their data.[[170]](#footnote-171) Additionally, the length and complexity of privacy policies (which few people read) often makes consent meaningless.[[171]](#footnote-172) The sentiment was shared by other presenters who thought that current privacy laws are overly focused on individual control.[[172]](#footnote-173) They urged that it is nearly impossible to expect that people will be able to control their own data.[[173]](#footnote-174) Privacy is too valuable and important to expect individuals to shoulder the burden of policing themselves.[[174]](#footnote-175) However, others argued that new technologies can enable us to economically ask people for their consent, and more thought should be given to a person’s ability to opt-out or opt-in to research.[[175]](#footnote-176)
* Another presenter argued that society has a collective right, expressed through law and regulation, to automatically include people in important research for the greater public good without asking for consent.[[176]](#footnote-177) A study was cited, which revealed that people are often willing to contribute their data to research as long as their identity is protected.[[177]](#footnote-178) Consequently, transparency may be a preferable strategy to engage individuals rather than consent.[[178]](#footnote-179)
 |
| Data Security |  |
|  | * There is no such thing as zero risk, and the security threat landscape varies over time.[[179]](#footnote-180) The security threats are driven by vulnerabilities that arise from designing and deploying highly complex software and hardware.[[180]](#footnote-181)
* HIPAA defines high-level objectives, but what is needed is a risk-based framework that will define very specific, contextual, and evolving controls that are applied to reduce risk … to an acceptable level.[[181]](#footnote-182)
* In response to this complexity, organizations should adopt a balanced, holistic approach to security that looks at operations end-to-end and applies a risk-based framework.[[182]](#footnote-183) This holistic approach should consider things like physical security.
* “The only pragmatic way to secure data in healthcare and in any other domain is to consistently follow an industry developed risk-based framework.”[[183]](#footnote-184) HITRUST is an example of a common security framework that the healthcare community may consider applying.[[184]](#footnote-185)
* HIPAA defines high-level objectives, but what is needed is a risk-based framework that will define very specific, contextual, and evolving controls that will be applied to reduce risk … to an acceptable level.”[[185]](#footnote-186) “The security objective should be based on outcomes, not the means, because the means (e.g., the hardware, the software, the attack mitigation) change constantly.”[[186]](#footnote-187)
* Moving to a common framework will be difficult for many organizations. Specifically, it will be challenging for organizations that do not have an IT department and rely on outsourcing, but it will be easier for organizations with sophisticated IT operations.[[187]](#footnote-188) If an organization employs a good computer science approach, which involves backing up machines, firewalls, and antivirus software on desktops, then it should be a medium effort to achieve good security.[[188]](#footnote-189)
* HIPAA compliance varies significantly across hospitals based on their levels of resources.[[189]](#footnote-190) The resources go beyond IT sophistication to hospital infrastructure, and staffing.[[190]](#footnote-191) Consequently, any regulatory incentive or effort must acknowledge that compliance varies across hospitals and providers.[[191]](#footnote-192)
* Distributed data networks may augment good security practices. One participant testified that “[d]istributed data networks minimize the need to aggregate individual data[,] are increasingly powerful[,] and should be considered when they are appropriate. These methods move the analyses to the data systems that already possess the data and return results that can be combined across multiple sites. The Food and Drug Administration (FDA), the National Institutes of Health (NIH), and the PCORI have created distributed data networks to support some of their needs.”[[192]](#footnote-193) Others cautioned that some distributed networks do not have security evaluations or security proofs and it would be important to perform such proofs and evaluate security protocols before leveraging distributed computation systems.[[193]](#footnote-194)
 |
| Transparency |  |
|  | * Notices are overly broad and vague, and they do not end up fostering transparency.[[194]](#footnote-195)

Notices are drafted in highly technical language and are so vague that people do not read or understand them, so transparency is rarely achieved.[[195]](#footnote-196)* As a matter of ethics, transparency is crucially important if data is used without a person’s explicit consent.[[196]](#footnote-197)
* One participant explained that entities should provide notice whenever individuals may think the usage or collection of data is unexpected or objectionable, and notice should be provided at a relevant time. Contextual (just-in-time) notice helps clarify consumer expectations. Such notice should explain what type of data is collected, when it is collected, what it is used for, the secondary uses contemplated, how long it will be retained, and what security measures are in place.[[197]](#footnote-198)
* Without clear ground rules in the non-HIPAA space, organizations “tend to become less transparent about their data practices,” which extends to the use of algorithms, which are “crucial decision-making mechanisms.”[[198]](#footnote-199)
* “Algorithmic transparency is crucial.... Many companies have entered the health data space and they consider their models proprietary and refuse to reveal them, which leaves a gaping hole where our understanding of these decision-making mechanisms should be.”[[199]](#footnote-200) Because sophisticated algorithms are proprietary intellectual property, it is very difficult to determine their inputs and outputs, and how they make decisions about people.[[200]](#footnote-201) Moreover, “[a]lgorithms have become extremely sophisticated and nuanced to the point where they are [replacing the] human decision-making processes.”[[201]](#footnote-202)
* One participant suggested that transparency and disclosure should extend to “what [data] informs the algorithms, how … cohorts are defined, and how individuals are separated. If that’s opaque, … then nobody will ever trust the system.”[[202]](#footnote-203)
* Another participant drew parallels to transparency provisions in the Fair Credit Reporting Act (FCRA). When the FCRA was introduced, Credit Rating Agencies said that people were not asking for their credit information; nevertheless, access rights were put in the law. Today, 91% of people surveyed by the participant stated that it was important to find out to whom their personal information had been disclosed.[[203]](#footnote-204) Although the FCRA is a statute that frames acceptable uses of data, it provides consumers with transparency if data has an adverse impact on them. Some cautioned that FCRA is tailored to particular circumstances and it may not scale well in the health arena.[[204]](#footnote-205)
 |
| Collection, Use, and Purpose Limitation |  |
|  | * Big data driven research enables researchers to examine trends and patterns in large data sets without first formulating a hypothesis.[[205]](#footnote-206)
* “… as a general principle, the minimum necessary amount of identifiable data should be used to answer a question.”[[206]](#footnote-207)
* Organizations should ask themselves why they need the information they have collected, and they should avoid retaining data for some future, unnamed use simply because they think it might be valuable.[[207]](#footnote-208)
* With regard to health data, there should be a requirement to delimit the collection and use of data, and it is not acceptable to retain data for an unknown purpose.[[208]](#footnote-209)
* Concerning purpose limitation, participants struggled to clearly define acceptable and unacceptable uses of health information in big data analytics.[[209]](#footnote-210)
 |
| Privacy Harms |  |
|  | * It is very difficult to define or put a frame around what is harm.[[210]](#footnote-211)
* Defining an acceptable use of data is subjective because acceptable use is culturally specific and will change over time.[[211]](#footnote-212)
* Some current rules, such as HIPAA and the Common Rule, provide for permitted uses, but they do not enumerate “non-permitted abuses.”[[212]](#footnote-213)
* Commercial use of personal information without a clear disclosure could be viewed as harmful.[[213]](#footnote-214)
* Any sort of discrimination or denial of opportunity, such as the loss of employment or insurance, or any public embarrassment would be classified as harmful.[[214]](#footnote-215)
* When people were asked about their greatest concern regarding the use of their health information, one survey revealed that the top concern was future contact at a later date.[[215]](#footnote-216)
* To arrive at a consensus around harms or non-permitted abuses, an effort could be made to identify laws that may already prohibit certain activities, identify gaps, and catalogue federal laws with non-discrimination provisions.[[216]](#footnote-217)
* Additionally, the FTC can help can help identify boundaries through the cases it pursues under its ability to combat unfairness and deception.[[217]](#footnote-218)
 |
| Complex Legal Landscape |  |
|  | * There continues to be a lack of clarity and understanding of privacy and security laws and rules.[[218]](#footnote-219)
* Legal coverage is extensive and even contradictory in some areas (e.g., research under HIPAA and the Common Rule), while coverage is significantly lacking in other areas.[[219]](#footnote-220)
* State law is confusing, often outdated and seldom enforced.[[220]](#footnote-221)
* HIPAA applies only to covered entities (health plans, health care clearinghouses, health care providers) and business associates acting directly on their behalf. The bulk of health-related data being generated today falls outside of HIPAA regulation.[[221]](#footnote-222)
 |
| Access to Information |  |
|  | * Generally, there is a need for greater data liquidity.[[222]](#footnote-223) On a personal level, patients want to access and combine their data in meaningful ways, but they face significant impediments.[[223]](#footnote-224) Privacy and security are seen as some of the biggest burdens/barriers.[[224]](#footnote-225)
* Public access to data is also very important. Health data can be viewed as a social asset; there is a social responsibility to give back to the community by making data available to researchers and to patient groups.[[225]](#footnote-226)
* Within HIPAA, patient consent is not required for the use and exchange of protected health information (PHI) for treatment, payment, or healthcare operations purposes.[[226]](#footnote-227) On presenter cited this concept of normal, routine uses as one of HIPAA’s greatest strengths.[[227]](#footnote-228)
* Some progress has been made to improve data liquidity. HITECH introduced important changes, including (1) authorization to permit future research and (2) the ability to permit compound authorizations for research purposes. [[228]](#footnote-229)
 |
| Under-Regulation |  |
|  | * A rapidly growing amount of health-related information is not regulated by the HIPAA.[[229]](#footnote-230) These include mobile applications, websites, and personal health records. Additionally, healthcare entities that are covered by HIPAA are using non-health data for healthcare purposes.[[230]](#footnote-231)
* The FTC has general authority to prevent unfair or deceptive acts or practices, and it has recently become very active on the general enforcement of data security standards, though several have challenged this authority. [[231]](#footnote-232) However, the FTC has been less active in the privacy sphere, particularly in healthcare. One reason is that while deception is relatively straightforward to litigate, it is not clear what may be considered a legally enforceable “unfair” privacy practice.[[232]](#footnote-233)
* The HIPAA model could be extended to define covered entities in a broader way. For example, in Texas, anyone who touches healthcare data is considered to be covered entity. This, however, alters the legal approach by shifting the analysis from which entities process personal information (i.e., whether the entity is a covered entities) to what kind of personal information is being processed (i.e., whether the data is health data).[[233]](#footnote-234) This change would be difficult in a big data world in which health data is not clearly defined and information flows through many different people who don’t necessarily have a direct relationship with the individual. [[234]](#footnote-235)
 |
| Over-Regulation |  |
|  | * One panelist stated that the HIPAA Privacy Rule does not protect privacy as well as it should, and in fact, HIPAA impedes the use of data for important health research.[[235]](#footnote-236)
* Others cited HIPAA’s strengths, noting that it establishes common rules that apply uniformly; which serves to improve access to information.[[236]](#footnote-237)
* There is an apparent paradox in HIPAA. While the definition of “research” is the same under both HIPAA and Common Rule,[[237]](#footnote-238) different rules about patient consent are applied depending on whether the research results are shared for “generalizable knowledge” or are used for quality improvement purposes and kept within an organization (i.e., covered under “healthcare operations”). “Two studies that use data for quality improvement purposes using the same data points done to address the same question … by the same institution will be treated as operations if the results are *not* intended to contribute to generalizable knowledge, … but treated as research [requiring consent] if you intend to share the results with others so that learning may occur.”[[238]](#footnote-239)
* A panelist asked how the learning health system (and data protection) can be advanced if a rule is based on what an entity intends to do with the results, not how data is safeguarded and treated in the underlying research project.[[239]](#footnote-240) It was also stated that people should not be penalized and that disincentives should not be created when organizations contribute to the general knowledge base for healthcare.[[240]](#footnote-241)
* It was noted that the PSWG’s predecessor, the Privacy and Security Tiger Team, provided recommendations in the past[[241]](#footnote-242) on the topic of modernizing the Common Rule and creating more consistency with in HIPAA, and these were approved by the HITPC.[[242]](#footnote-243)
* Acknowledged that the learning health system requires more widespread dissemination of information, the Tiger Team recommended that uses of EHR data to evaluate the safety, quality, and effectiveness of prevention and treatment activities should not require consent or IRB approval. Thus, such investigations should not be labeled as research – even if the results are used for generalizable knowledge – because doing so would pose an obstacle to learning. This exemption should be granted when the provider entity retains oversight and control over EHR data.[[243]](#footnote-244)
 |

1. HITPC Transmittal Letter, September 1, 2010, p. 4, <http://www.healthit.gov/sites/faca/files/hitpc_transmittal_p_s_tt_9_1_10_0.pdf>. [↑](#footnote-ref-2)
2. See 45 CFR § 160.103. [↑](#footnote-ref-3)
3. Public Hearing Responses of Richard Platt, p. 3, <http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Richard_Platt_Reply_to_Questions_for_Panelists_2014-12-05.pdf> [hereinafter "Richard Platt Responses”]. [↑](#footnote-ref-4)
4. Richard Platt Responses, p. 3. [↑](#footnote-ref-5)
5. Richard Platt Responses, p. 3. [↑](#footnote-ref-6)
6. Toolkit for Communities Using Health Data: How to collect, use, protect, and share data responsibly, NCVHS, Draft 8: November 6, 2014, <http://www.healthit.gov/facas/sites/faca/files/PSWG_Presentation_Leslie_Francis_2014-12-08.pdf/> [↑](#footnote-ref-7)
7. Richard Platt Responses, p.3. [↑](#footnote-ref-8)
8. Richard Platt Responses, p.3. [↑](#footnote-ref-9)
9. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap, Draft Version 1.0, p. 8, <http://www.healthit.gov/sites/default/files/nationwide-interoperability-roadmap-draft-version-1.0.pdf>. See also, Testimony of Richard Platt, p. 1 (“The term “Learning Health System” connotes a commitment to improve care, both by learning from all patients’ experiences and by implementing the results of the learning activities.”). [↑](#footnote-ref-10)
10. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap, Draft Version 1.0, p. 35, <http://www.healthit.gov/sites/default/files/nationwide-interoperability-roadmap-draft-version-1.0.pdf>. [↑](#footnote-ref-11)
11. Stephen J. Downs, PSWG Transcript 2014-12-05, p. 7-9 [hereinafter “December 5”]. [↑](#footnote-ref-12)
12. Robert Gellman, December 5, p. 51. [↑](#footnote-ref-13)
13. Stephen Downs, December 5, p. 20; Mark Savage, December 5, p. 30; David McCallie, Jr., PSWG Transcript 2014-12-08, p. 21 [hereinafter “December 8”]. [↑](#footnote-ref-14)
14. Michelle De Mooy, December 5, p. 44. [↑](#footnote-ref-15)
15. Khaled El Emam, December 5, p. 49-50. [↑](#footnote-ref-16)
16. Michelle De Mooy, December 5, p. 30. [↑](#footnote-ref-17)
17. Andrew Rosenthal, *Anthem is warning consumers about its huge data breach. Here’s a translation*, L.A. Times, March 5, 2015, available at: <http://www.latimes.com/business/hiltzik/la-fi-mh-anthem-is-warning-consumers-20150306-column.html#page=1>. [↑](#footnote-ref-18)
18. Michelle De Mooy, December 5, p. 34. [↑](#footnote-ref-19)
19. Deven McGraw, HITPC Transcript, March 10, 2015, p. 19. [↑](#footnote-ref-20)
20. For information on covered entities and business associates, see <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/>; for definitions, see 45 CFR §160.103. [↑](#footnote-ref-21)
21. There is no definitive version of the FIPPs, which are recognized worldwide as the foundational principles for data privacy. Appropriate sources include the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, [http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm - part2](http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm#part2), the Markle Connecting for Health Common Framework, <http://www.markle.org/sites/default/files/CF-Consumers-Full.pdf>, the White House’s 2012 Consumer Bill of Rights, <https://www.whitehouse.gov/sites/default/files/privacy-final.pdf>, and the NIST National Strategy for Trusted Identities in Cyberspace, <http://www.nist.gov/nstic/NSTIC-FIPPs.pdf>. [↑](#footnote-ref-22)
22. Stanley Crosley, December 8, p. 14. [↑](#footnote-ref-23)
23. Big Data: Seizing Opportunities, Preserving Values, May 2014, <https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf> [hereinafter “White House Big Data Report”]. [↑](#footnote-ref-24)
24. White House Big Data Report, p. 54, (stating that “re-identification is becoming more powerful than de-identification,” and “focusing on controlling the collection and retention of personal data . . . may no longer be sufficient to protect personal privacy.”). [↑](#footnote-ref-25)
25. White House Big Data Report, p. 62. [↑](#footnote-ref-26)
26. White House Big Data Report, p. 23. [↑](#footnote-ref-27)
27. FACT SHEET: Announcing New U.S. Open Government Commitments on the Third Anniversary of the Open Government Partnership, <https://www.whitehouse.gov/the-press-office/2014/09/24/fact-sheet-announcing-new-us-open-government-commitments-third-anniversa>. (hereinafter Open Government Partnership). [↑](#footnote-ref-28)
28. Open Government Partnership. [↑](#footnote-ref-29)
29. Appendix A, Health Big Data Public Hearing Topics and Speakers. [↑](#footnote-ref-30)
30. Internet of things: Privacy & Security in a Connected World, FTC Staff Report, January 2015, p. i, <https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf>. [↑](#footnote-ref-31)
31. Id. [↑](#footnote-ref-32)
32. About the Precision Medicine Initiative, National Institutes of Health, available at: <http://www.nih.gov/precisionmedicine/>. [↑](#footnote-ref-33)
33. FACT SHEET: President Obama’s Precision Medicine Initiative, <https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative>. [↑](#footnote-ref-34)
34. 21st Century Cures Discussion Document, <http://energycommerce.house.gov/sites/republicans.energycommerce.house.gov/files/114/Analysis/Cures/20150127-Cures-Discussion-Document-One-Pager.pdf> [↑](#footnote-ref-35)
35. 21st Century Cures Discussion Document. [↑](#footnote-ref-36)
36. Energy and Commerce Cures, <http://energycommerce.house.gov/cures> [↑](#footnote-ref-37)
37. Id. [↑](#footnote-ref-38)
38. Federal Health IT Strategic Plan: 2015-2020, <http://www.healthit.gov/sites/default/files/federal-healthIT-strategic-plan-2014.pdf> [hereinafter “Federal Health IT Strategic Plan 2015-2020”]. [↑](#footnote-ref-39)
39. Federal Health IT Strategic Plan: 2015-2020, p. 3. [↑](#footnote-ref-40)
40. Federal Health IT Strategic Plan: 2015-2020, p. 26. [↑](#footnote-ref-41)
41. See Connecting Health and Care for the National: A Shared Nationwide Interoperability Roadmap, DRAFT Version 1.0, available at: <http://www.healthit.gov/sites/default/files/nationwide-interoperability-roadmap-draft-version-1.0.pdf>. [↑](#footnote-ref-42)
42. See Federal Health IT Strategic Plan: 2015-2020, p. 5. [↑](#footnote-ref-43)
43. Goal 3: strengthen health care delivery; goal 4: advance the heath and well being of individuals and communities; and goal 5: advance research, scientific knowledge, and innovation. See Federal Health IT Strategic Plan: 2015-2020, p. 5. [↑](#footnote-ref-44)
44. Connecting Health and Care for the National: A Shared Nationwide Interoperability Roadmap, DRAFT Version 1.0, p, 35, <http://www.healthit.gov/sites/default/files/nationwide-interoperability-roadmap-draft-version-1.0.pdf>. [↑](#footnote-ref-45)
45. Connecting Health and Care for the National: A Shared Nationwide Interoperability Roadmap, DRAFT Version 1.0, p, 35. [↑](#footnote-ref-46)
46. Patient-Centered Outcomes Research Institute, <http://www.pcori.org/about-us>. [↑](#footnote-ref-47)
47. http://www.hhs.gov/ohrp/sachrp/ [↑](#footnote-ref-48)
48. Human Subjects Research Implications of “Big Data” Studies, US Department of Health and Human Services, <http://www.hhs.gov/ohrp/sachrp/commsec/hsrimplicationsofbig_datastudies.html>. [↑](#footnote-ref-49)
49. Id. [↑](#footnote-ref-50)
50. Information about the hearing agendas and testifiers is provided in Appendix A. [↑](#footnote-ref-51)
51. Stephen J. Downs, December 5, p. 8. [↑](#footnote-ref-52)
52. Stephen Downs, December 5, p. 20; Mark Savage, December 5, p. 30; D. McCallie, JR., December 8, p. 21. [↑](#footnote-ref-53)
53. Testimony of CDT for the HIT Privacy and Security Work Group Virtual Hearing (December 5, 2013), p. 3, <http://www.healthit.gov/facas/sites/faca/files/PSWG_Testimony_Michelle_DeMooy_CDT_2014-12-05_0.pdf>. [↑](#footnote-ref-54)
54. Khaled El Emam, December 5, p. 48. [↑](#footnote-ref-55)
55. Melissa Bianchi, December 8, p. 6 [↑](#footnote-ref-56)
56. Khaled El Emam, December 5, p. 49. [↑](#footnote-ref-57)
57. [↑](#footnote-ref-58)
58. Khaled El Eman, December 5, p. 49 Khaled El Emam, December 5, p. 49. [↑](#footnote-ref-59)
59. Khaled El Emam, December 5, p. 49. [↑](#footnote-ref-60)
60. Khaled El Emam, December 5, p. 49. [↑](#footnote-ref-61)
61. Michelle De Mooy, December 5, p. 30. [↑](#footnote-ref-62)
62. Khaled El Emam, December 5, p. 48. [↑](#footnote-ref-63)
63. Khaled El Emam, December 5, p. 51-52. [↑](#footnote-ref-64)
64. Michelle De Mooy, December 5, p. 29; Mark Savage, December 5, p. 38, 41; Fred Cate, December 5, p. 63. [↑](#footnote-ref-65)
65. Fred Cate, December 5, p. 63. [↑](#footnote-ref-66)
66. Stanley Crosley, PSWG Transcript 2015-01-26, p. 17 [hereinafter “January 26”]. [↑](#footnote-ref-67)
67. Stanley Crosley, January 26, p17. [↑](#footnote-ref-68)
68. Richard Platt, December 5, p. 26. [↑](#footnote-ref-69)
69. Mark Savage, December 5, p. 32. [↑](#footnote-ref-70)
70. Fred Cate, December 5, p. 64; Robert Gellman, December 5, p. 64. [↑](#footnote-ref-71)
71. Richard Platt, December 5, p. 10 (Additionally, offering a universal opt-out may be undesirable because it would introduce bias in research results, and thereby create unreliable answers from the data, which is a data quality concern). [↑](#footnote-ref-72)
72. Fred Cate, December 5, p. 65. [↑](#footnote-ref-73)
73. Fred Cate, December 5, p. 53/2513-14 (Throughout Workgroup discussions, people commented that consent places a significant burden for privacy on the individual; see Stanley Crosley, January 26, at 17; see also Deven McGraw, February 9, at 26). [↑](#footnote-ref-74)
74. Robert Gellman, December 5, p. 65. [↑](#footnote-ref-75)
75. Michelle De Mooy, December 5, p. 28. [↑](#footnote-ref-76)
76. Michelle De Mooy, December 5, p. 28. [↑](#footnote-ref-77)
77. Fred Cate, December 5, p. 53. [↑](#footnote-ref-78)
78. See generally, <http://www.healthit.gov/facas/calendar/2015/02/09/policy-privacy-security-workgroup>. [↑](#footnote-ref-79)
79. Transcript of PSWG Meeting (February 9, 2015), Andrei Stoica, p. 6 [hereinafter “February 9”]. [↑](#footnote-ref-80)
80. Andrei Stoica, February 9, p. 6. [↑](#footnote-ref-81)
81. Andrei Stoica, February 9, p. 6. [↑](#footnote-ref-82)
82. Andrei Stoica, February 9, p. 14. [↑](#footnote-ref-83)
83. Denise Anthony, February 9, p. 10. [↑](#footnote-ref-84)
84. Denise Anthony, February 9, p. 16. [↑](#footnote-ref-85)
85. Denise Anthony, February 9, p. 10. [↑](#footnote-ref-86)
86. Written testimony of Richard Platt, p. 4, <http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Richard_Platt_Reply_to_Questions_for_Panelists_2014-12-05.pdf>. [↑](#footnote-ref-87)
87. Khaled El Emam, December 5, p. 51-52. [↑](#footnote-ref-88)
88. Testimony of CDT for the HIT Privacy and Security Work Group Virtual Hearing (December 5, 2013), p. 3, <http://www.healthit.gov/facas/sites/faca/files/PSWG_Testimony_Michelle_DeMooy_CDT_2014-12-05_0.pdf>. [↑](#footnote-ref-89)
89. Michelle De Mooy, December 5, p. 35. [↑](#footnote-ref-90)
90. Michelle De Mooy, December 5, p. 34/1593, 1597-98 [↑](#footnote-ref-91)
91. [↑](#footnote-ref-92)
92. Testimony of CDT for the HIT Privacy and Security Work Group Virtual Hearing (December 5, 2013), p. 3, <http://www.healthit.gov/facas/sites/faca/files/PSWG_Testimony_Michelle_DeMooy_CDT_2014-12-05_0.pdf> Testimony of CDT for the HIT Privacy and Security Work Group Virtual Hearing (December 5, 2013), p. 3, <http://www.healthit.gov/facas/sites/faca/files/PSWG_Testimony_Michelle_DeMooy_CDT_2014-12-05_0.pdf>. [↑](#footnote-ref-93)
93. Michelle De Mooy, December 5, p. 29. [↑](#footnote-ref-94)
94. Anna McCollister-Slipp, December 5, p. 35. [↑](#footnote-ref-95)
95. Kirk Nahra, December 8, p. 16, 17. [↑](#footnote-ref-96)
96. Ella Mihov, December 8, p. 30/138401387; 1413-1414 [↑](#footnote-ref-97)
97. Richard Platt, December 5, p. 10. [↑](#footnote-ref-98)
98. Michelle De Mooy, December 5, p. 42. [↑](#footnote-ref-99)
99. Khaled El Emam, December 5, p. 49. [↑](#footnote-ref-100)
100. Richard Platt, December 5, p. 10h. [↑](#footnote-ref-101)
101. Deven McGraw, December 8, p. 19/871 [↑](#footnote-ref-102)
102. Khaled El Emam, December 5, p. 49/2235-36 [↑](#footnote-ref-103)
103. Mark Savage, December 5, p. 30. [↑](#footnote-ref-104)
104. Mark Savage, December 5, p. 30/1416-17 [↑](#footnote-ref-105)
105. Melissa Bianchi, December 8, p. 21. [↑](#footnote-ref-106)
106. Michelle De Mooy, December 5, p. 46. [↑](#footnote-ref-107)
107. See Anna McCollister-Slipp, December 5, p. 33; Michelle De Mooy, December 5, p. 37 (… people think … HIPAA really covers all medical data and have … no idea or understanding that it doesn’t in a lot of circumstances”); Michelle De Mooy, December 5, p. 37 (outreach is needed to app developers to ensure they understand their ethical responsibilities). [↑](#footnote-ref-108)
108. Deven McGraw, HITPC Transcript, March 10, 2015, p. 19. [↑](#footnote-ref-109)
109. Kirk Nahra, Slide Presentation, p. 6, available at: <http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Kirk_Nahra_Health_Care_Privacy_Paradigm_2014-12-08.pdf> [↑](#footnote-ref-110)
110. Kirk Nahra, December 8, p. 11 (noting the explosion of data created by mobile applications, websites, personal health records, and wellness programs that are not subject to HIPAA). [↑](#footnote-ref-111)
111. Anna McCollister-Slipp, December 5, p. 32. [↑](#footnote-ref-112)
112. Anna McCollister-Slipp, December 5, p. 32. [↑](#footnote-ref-113)
113. Anna McCollister-Slipp, December 5, p. 32. [↑](#footnote-ref-114)
114. Anna McCollister-Slipp, December 5, p. 32. [↑](#footnote-ref-115)
115. Melissa Bianchi, December 8, p. 5. [↑](#footnote-ref-116)
116. Kirk Nahra, December 8, p. 19. [↑](#footnote-ref-117)
117. Melissa Bianchi, December 8, p. 6. [↑](#footnote-ref-118)
118. Kirk Nahra, December 8, p. 11. [↑](#footnote-ref-119)
119. Kirk Nahra, December 8, p. 11. [↑](#footnote-ref-120)
120. Kirk Nahra, December 8, p. 11. [↑](#footnote-ref-121)
121. Kirk Nahra, December 8, p. 11, 12 (see also the discussion regarding the inability to define privacy harm, supra). [↑](#footnote-ref-122)
122. Fred Cate, December 5, p. 53. [↑](#footnote-ref-123)
123. [↑](#footnote-ref-124)
124. Robert Gellman, December 5, p. 51. Deven McGraw, December 8, p. 9. [↑](#footnote-ref-125)
125. Melissa Bianchi, December 8, p. 5; see also 45 CFR 164.501 (defining research as “a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.”); see also <http://www.hhs.gov/ocr/privacy/hipaa/understanding/special/research/>. [↑](#footnote-ref-126)
126. Deven McGraw, December 8, p. 8, 9. [↑](#footnote-ref-127)
127. [↑](#footnote-ref-128)
128. [↑](#footnote-ref-129)
129. [↑](#footnote-ref-130)
130. HITPC Transmittal Letter, October 18, 2011, available at: <http://www.healthit.gov/sites/default/files/pdf/HITPC_Privacy_and_Security_Transmittal_Letter_10_18_11.pdf> [hereinafter “October 18 HITPC Recommendations”] Deven McGraw, December 8, p. 8; see October 18 HITPC Recommendations.October 18 HITPC Recommendations. Michelle De Mooy, December 5, p. 28; Deven McGraw, December 8, p. 19. [↑](#footnote-ref-131)
131. Kirk Nahra, Slide Presentation, p. 7, available at: <http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Kirk_Nahra_Health_Care_Privacy_Paradigm_2014-12-08.pdf> [↑](#footnote-ref-132)
132. The Business of Federal Technology. <http://fcw.com/articles/2014/05/13/fose-mosaic.aspx> [↑](#footnote-ref-133)
133. Harms identified in RWJF Report [↑](#footnote-ref-134)
134. A Summary of Your Rights Under the Fair Credit Reporting Act, <https://www.consumer.ftc.gov/articles/pdf-0096-fair-credit-reporting-act.pdf>. [↑](#footnote-ref-135)
135. May 22, 2015 HITPC meeting. <http://www.healthit.gov/facas/sites/faca/files/HITPC_PSWG_Meeting_Slides_2015-05-22_Final.pdf> [↑](#footnote-ref-136)
136. National Institute of Health. NIH Data Sharing Policy and Implementation Guidance <http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#enclave> [↑](#footnote-ref-137)
137. 8/16/2011 HITPC Transmittal Letter. <http://www.healthit.gov/sites/faca/files/HITPC_PSTT_Transmit_8162011.pdf> and May 22, 2015 HITPC meeting. <http://www.healthit.gov/facas/sites/faca/files/HITPC_PSWG_Meeting_Slides_2015-05-22_Final.pdf> [↑](#footnote-ref-138)
138. Reference NIST evaluation of efficacy of de-identification effort. [↑](#footnote-ref-139)
139. Khaled El Emam, December 5, p. 51 [↑](#footnote-ref-140)
140. See FTC Internet of Things Report, January 2015, p. 49. [↑](#footnote-ref-141)
141. 8/16/2011 HITPC Transmittal Letter. <http://www.healthit.gov/sites/faca/files/HITPC_PSTT_Transmit_8162011.pdf> [↑](#footnote-ref-142)
142. Stephen Downs, December 5, p. 20; Mark Savage, December 5, p. 31; D. McCallie, JR., December 8, p. 21. [↑](#footnote-ref-143)
143. Testimony of CDT for the HIT Privacy and Security Work Group Virtual Hearing (December 5, 2013), p. 3, <http://www.healthit.gov/facas/sites/faca/files/PSWG_Testimony_Michelle_DeMooy_CDT_2014-12-05_0.pdf>. [↑](#footnote-ref-144)
144. Testimony of CDT for the HIT Privacy and Security Work Group Virtual Hearing, p. 3. [↑](#footnote-ref-145)
145. Michelle De Mooy, December 5, p. 30. [↑](#footnote-ref-146)
146. Khaled El Emam, December 5, p. 48. [↑](#footnote-ref-147)
147. See testimony of Ella Mihov (Ayasdi), December 8, p. 32, 36; see also, Kald Abdallah (Project Data Sphere, LLC), December 8, p. 31, 36. [↑](#footnote-ref-148)
148. Richard Platt, December 5, p. 10. [↑](#footnote-ref-149)
149. Melissa Bianchi, December 8, p. 6 [↑](#footnote-ref-150)
150. Richard Platt, December 5, p. 10. [↑](#footnote-ref-151)
151. Khaled El Emam, December 5, p. 49. [↑](#footnote-ref-152)
152. [↑](#footnote-ref-153)
153. [↑](#footnote-ref-154)
154. Khaled El Eman, December 5, p. 49 Khaled El Eman, December 5, p. 49 Khaled El Emam, December 5, p. 49. [↑](#footnote-ref-155)
155. Khaled El Emam, December 5, p. 49. [↑](#footnote-ref-156)
156. Khaled El Emam, December 5, p. 49. [↑](#footnote-ref-157)
157. Khaled El Emam, December 5, p. 67 and Fred Cate, December 5, p. 67. [↑](#footnote-ref-158)
158. Khaled El Emam, December 5, p. 66. [↑](#footnote-ref-159)
159. Paul Wallace, December 8, p. 55. [↑](#footnote-ref-160)
160. McCallie, Jr., December 8, p. 56. [↑](#footnote-ref-161)
161. Khaled El Emam, December 5, p. 51-52. [↑](#footnote-ref-162)
162. Khaled El Emam, December 5, p. 51-52. [↑](#footnote-ref-163)
163. Michelle De Mooy, December 5, p. 29; Mark Savage, December 5, p. 38, 41; Fred Cate, December 5, p. 63. [↑](#footnote-ref-164)
164. Fred Cate, December 5, p. 63. [↑](#footnote-ref-165)
165. Michelle De Mooy, December 5, p. 28. [↑](#footnote-ref-166)
166. Kirk Nahra, December 8, p. 11. [↑](#footnote-ref-167)
167. Richard Platt, December 5, p. 26. [↑](#footnote-ref-168)
168. Mark Savage, December 5, p. 32. [↑](#footnote-ref-169)
169. Fred Cate, December 5, p. 64; Robert Gellman, December 5, p. 64. [↑](#footnote-ref-170)
170. Richard Platt, December 5, p. 10 (Additionally, offering a universal opt-out may be undesirable because it would create unreliable answers from the data, which is a data quality concern). [↑](#footnote-ref-171)
171. Fred Cate, December 5, p. 65. [↑](#footnote-ref-172)
172. Fred Cate, December 5, p. 52 [↑](#footnote-ref-173)
173. Fred Cate, December 5, p. 52 [↑](#footnote-ref-174)
174. Fred Cate, December 5, p. 53 (Throughout Workgroup discussions, people commented that consent places a significant burden for privacy on the individual; see Stanley Crosley, January 26, at 17; see also Deven McGraw, February 9, at 26). [↑](#footnote-ref-175)
175. Robert Gellman, December 5, p. 65. [↑](#footnote-ref-176)
176. Robert Gellman, December 5, p. 66 [↑](#footnote-ref-177)
177. Stephen J. Downs, December 5, p. 15. [↑](#footnote-ref-178)
178. See PSWG Meeting Slides, January 26, 2015, at 11, <http://www.healthit.gov/facas/sites/faca/files/PSWG_Meeting_Slides_2015-01-26_v9.pptx>. [↑](#footnote-ref-179)
179. Andrei Stoica, February 9, p. 6. [↑](#footnote-ref-180)
180. Andrei Stoica, February 9, p. 6. [↑](#footnote-ref-181)
181. Andrei Stoica, February 9, p. 6. [↑](#footnote-ref-182)
182. Andrei Stoica, February 9, p. 6. [↑](#footnote-ref-183)
183. Andrei Stoica, February 9, p. 6. [↑](#footnote-ref-184)
184. Andrei Stoica, February 9, p. 14. [↑](#footnote-ref-185)
185. Andrei Stoica, February 9, p. 6. [↑](#footnote-ref-186)
186. Andrei Stoica, February 9, p. 6. [↑](#footnote-ref-187)
187. Andrei Stoica, February 9, p. 15, 21 (stating that “there is a huge impediment and huge cost differential for security, but as you go [down] a path to critical mass and you have a decent sized IT operation … then it becomes easier and easier….”). [↑](#footnote-ref-188)
188. [↑](#footnote-ref-189)
189. Andrei Stoica, February 9, p. 15. Denise Anthony, February 9, p. 10. [↑](#footnote-ref-190)
190. Denise Anthony, February 9, p. 16. [↑](#footnote-ref-191)
191. Denise Anthony, February 9, p. 10. [↑](#footnote-ref-192)
192. Written testimony of Richard Platt, p. 4, <http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Richard_Platt_Reply_to_Questions_for_Panelists_2014-12-05.pdf>. [↑](#footnote-ref-193)
193. Khaled El Emam, December 5, p. 54. [↑](#footnote-ref-194)
194. [↑](#footnote-ref-195)
195. [↑](#footnote-ref-196)
196. Testimony of CDT for the HIT Privacy and Security Work Group Virtual Hearing (December 5, 2013), p. 3, <http://www.healthit.gov/facas/sites/faca/files/PSWG_Testimony_Michelle_DeMooy_CDT_2014-12-05_0.pdf> Fred Cate, December 5, p. 69. Fred Cate, December 5, p. 69. [↑](#footnote-ref-197)
197. Testimony of CDT for the HIT Privacy and Security Work Group Virtual Hearing, p. 3. [↑](#footnote-ref-198)
198. Michelle De Mooy, December 5, p. 28. [↑](#footnote-ref-199)
199. Michelle De Mooy, December 5, p. 29. [↑](#footnote-ref-200)
200. Michelle De Mooy, December 5, p. 29. [↑](#footnote-ref-201)
201. Michelle De Mooy, December 5, p. 29. [↑](#footnote-ref-202)
202. Anna McCollister-Slipp, December 5, p. 35. [↑](#footnote-ref-203)
203. Denise Anthony, February 9, p. 13. [↑](#footnote-ref-204)
204. Kirk Nahra, December 8, p. 16, 17. [↑](#footnote-ref-205)
205. Ella Mihov, December 8, p. 30. [↑](#footnote-ref-206)
206. Richard Platt, December 5, p. 10. [↑](#footnote-ref-207)
207. Michelle De Mooy, December 5, p. 42. [↑](#footnote-ref-208)
208. Michelle De Mooy, December 5, p. 42. [↑](#footnote-ref-209)
209. Khaled El Emam, December 5, p. 49. [↑](#footnote-ref-210)
210. Michelle De Mooy, December 5, p. 44. [↑](#footnote-ref-211)
211. Khaled El Emam, December 5, p. 49. [↑](#footnote-ref-212)
212. David McCallie, December 8, p. 21. [↑](#footnote-ref-213)
213. Stephen Downs, December 5, p. 21. [↑](#footnote-ref-214)
214. Stephen Downs, December 5, p. 21. [↑](#footnote-ref-215)
215. Fred Cate, December 5, p. 53. [↑](#footnote-ref-216)
216. Melissa Bianchi, December 8, p. 21. [↑](#footnote-ref-217)
217. Michelle De Mooy, December 5, p. 46. [↑](#footnote-ref-218)
218. See Anna McCollister-Slipp, December 5, p. 33; Michelle De Mooy, December 5, p. 37 (… people think … HIPAA really covers all medical data and have … no idea or understanding that it doesn’t in a lot of circumstances”); Michelle De Mooy, December 5, p. 37 (outreach is needed to app developers to ensure they understand their ethical responsibilities). [↑](#footnote-ref-219)
219. Deven McGraw, HITPC Transcript, March 10, 2015, p. 19. [↑](#footnote-ref-220)
220. Kirk Nahra, Slide Presentation, p. 6, available at: <http://www.healthit.gov/facas/sites/faca/files/PSWG_Background_Kirk_Nahra_Health_Care_Privacy_Paradigm_2014-12-08.pdf> [↑](#footnote-ref-221)
221. Kirk Nahra, December 8, p. 11 (noting the explosion of data created by mobile applications, websites, personal health records, and wellness programs that are not subject to HIPAA). [↑](#footnote-ref-222)
222. Anna McCollister-Slipp, December 5, p. 32. [↑](#footnote-ref-223)
223. Anna McCollister-Slipp, December 5, p. 32. [↑](#footnote-ref-224)
224. Anna McCollister-Slipp, December 5, p. 32. [↑](#footnote-ref-225)
225. Anna McCollister-Slipp, December 5, p. 32. [↑](#footnote-ref-226)
226. Melissa Bianchi, December 8, p. 5. [↑](#footnote-ref-227)
227. Kirk Nahra, December 8, p. 19. [↑](#footnote-ref-228)
228. Melissa Bianchi, December 8, p. 5. [↑](#footnote-ref-229)
229. Kirk Nahra, December 8, p. 11. [↑](#footnote-ref-230)
230. Kirk Nahra, December 8, p. 11. [↑](#footnote-ref-231)
231. Kirk Nahra, December 8, p. 11. [↑](#footnote-ref-232)
232. Kirk Nahra, December 8, p. 11, 12 (see also the discussion regarding the inability to define privacy harm, supra). [↑](#footnote-ref-233)
233. Kirk Nahra, December 8, p. 18. [↑](#footnote-ref-234)
234. Kirk Nahra, December 8, p. 17, 18. [↑](#footnote-ref-235)
235. Fred Cate, December 5, p. 53. [↑](#footnote-ref-236)
236. [↑](#footnote-ref-237)
237. Robert Gellman, December 5, p. 51. Melissa Bianchi, December 8, p. 5; see also 45 CFR 164.501 (defining research as “a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.”); see also <http://www.hhs.gov/ocr/privacy/hipaa/understanding/special/research/>. [↑](#footnote-ref-238)
238. Deven McGraw, December 8, p. 8, 9. [↑](#footnote-ref-239)
239. Deven McGraw, December 8, p. 9. [↑](#footnote-ref-240)
240. Deven McGraw, December 8, p. 9. [↑](#footnote-ref-241)
241. [↑](#footnote-ref-242)
242. [↑](#footnote-ref-243)
243. HITPC Transmittal Letter, October 18, 2011, available at: <http://www.healthit.gov/sites/default/files/pdf/HITPC_Privacy_and_Security_Transmittal_Letter_10_18_11.pdf>. Deven McGraw, December 8, p. 8; see October 18 HITPC Recommendations.October 18 HITPC Recommendations. [↑](#footnote-ref-244)