Health IT Policy Committee

A Public Advisory Body on Health Information Technology to the National Coordinator for Health IT

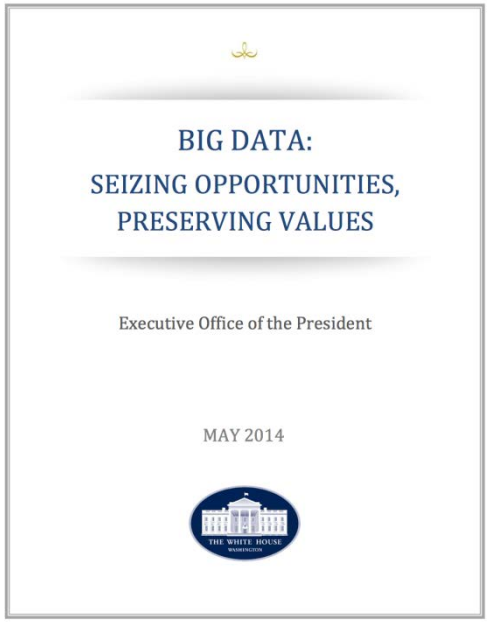
Privacy and Security Workgroup

Update on Health Big Data Deliberations

March 10, 2015

Context / Background

White House Big Data Report: Observations

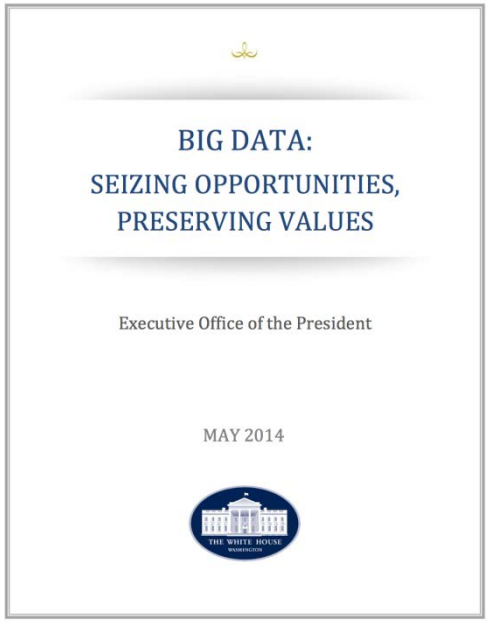


White House Report (May 2014)

- Big data is characterized by 3 Vs (Volume, Variety, Velocity)
- Other key observations:
  - De-identification is insufficient to protect privacy in big data analytics
  - Meta data raises significant privacy issues
    - Should not necessarily treat as less risky than content
  - Focus on assuring responsible uses, vs. trying to control collection; role of notice and consent should be re-examined

Source: http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

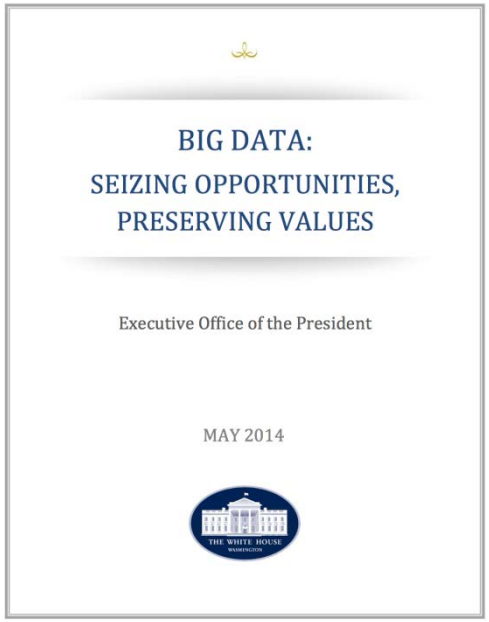
Why are we considering Big Data? continued



- "The government should lead a **consultative process** to **assess how** the Health Insurance Portability and Accountability Act (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."

Source: http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

Why are we considering Big Data?



- "The complexity of complying with numerous laws when data [is] combined from various sources raises the potential 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."

Source: http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

White House Big Data Recommendations

- Big data analytics is challenging current policy frameworks in the health space
  - Payment: fee-for-service → quality outcomes
  - Building a learning health system (LHS)
  - User generated data = identify lifestyle risks
  - Population health analysis and research
  - Predictive medicine
  - Myriad of federal, state, and organizational rules
  - "Health" data within and outside of HIPAA

Source: http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

Other Developments

- PCAST Report (Big Data and Privacy)
- White House Open Government Partnership
- 21st Century Cures Initiative
- Precision Medicine Initiative
- Big data business opportunities
  - Venture capital in data analytics
  - Mobile health
  - Maintain fairness and privacy for individuals
- Anthem breach

Source: http://energycommerce.house.gov/cures
Source: http://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative

Privacy Protections for Health Data

- Current federal protections
  - HIPAA Privacy and Security Rules
  - FTC Section 5 (Unfair and Deceptive Trade Practices) Authority
  - Common Rule (federally-supported research & research institutions)
  - Part 2 (federally supported substance abuse treatment programs)
  - HITECH breach notification (FTC & HIPAA)
  - Others (for example, FERPA)

- State law protections (see Interoperability report)
- HIPAA has limited applicability
  - HIPAA Applies to HIPAA Covered Entities (CE)
    - Health Plans
    - Health Care Clearinghouses
    - Health Care Providers
    - Business Associates (BA) acting on behalf of a CE

HIPAA – What is Covered and What is Not?

- Privacy Rule covers all paper, electronic, and oral PHI
- Security Rule applies to PHI in electronic form only
- HIPAA rules *do not* apply to:
  - Individually identifiable health information that is maintained by someone other than a covered entity or business associate
  - Information which has been de-identified in accordance with the HIPAA Privacy Rule
- HIPAA CEs have clear rules to abide by, ***but*** there are a growing number of organizations involved in health and wellness that maintain, transmit, or receive health information that are not CEs

Work Completed to Date

PSWG Meetings and Hearings

Date	Topic
Oct 27	PSWG Meeting: health big data topic introduction
Nov 10, Nov 24	PSWG Meetings: public hearing preparation
Dec 5	Public Hearing Day 1 Link: http://www.healthit.gov/facas/calendar/2014/12/05/policy-privacy-security-workgroup-virtual-hearing
Dec 8	Public Hearing Day 2 Link: http://www.healthit.gov/facas/calendar/2014/12/08/policy-privacy-security-workgroup-virtual-hearing
Jan 12	Reoriented and prioritized key themes; begin deliberations
Jan 28	De-identification; consent
Feb 9	Health big data security testimony; review consent Link: http://www.healthit.gov/facas/calendar/2015/02/09/policy-privacy-security-workgroup
May	Resume health big data deliberations
June	Current estimated to brief recommendations to HITPC

Public Hearing Panelists

Day 1 – December 5

Health Big Data Opportunities and the Learning Health System (LHS):
- Steve Downs, RWJF
- Richard Platt, Harvard Pilgrim
- Patricia Brennan, U. Wisconsin

Health Big Data Concerns:
- Michele DeMooy, CDT
- Mark Savage, NPWF
- Anna McCollister-Slipp, Galileo Analytics

Protections for Consumers:
- Khaled El Emam, U. of Ottawa
- Bob Gellman, Private Consultant
- Fred Cate, Indiana U.

Day 2 – December 8

Current Law:
- Melissa Bianchi, Hogan Lovells
- Kirk J. Nahra, Wiley Rein
- Deven McGraw, Manatt

Health Big Data Opportunities:
- Linda Avey, 23 and Me, Curios, Inc.
- Kald Abdallah, Project Data Sphere
- Ella Mihov, Ayasdi

Learning Health System:
- Paul Wallace, Optum Labs
- Josh Gray, AthenaHealth

Health Big Data Concerns:
- Leslie Francis, U. Utah
- Melissa Goldstein, George Washington U.

Scope

In Scope:

- Privacy and security issues – concerns and potential barriers to progress/innovation
- Potential harmful uses (related to privacy)

Out of Scope:

- Data quality/data standards
- Non-representativeness of data
  - Should not try to resolve this from the standpoint of increasing "representativeness" of data but should be considered in discussion of harmful uses

Post-Public Hearing: Topics

1. Concerns about tools commonly used to protect privacy
   - **A. De-identification**
   - **B. Patient consent v. norms of use**
   - **C. Security**
   - D. Transparency
   - E. Collection/use/purpose limitations
2. Preventing/Limiting/Redressing Harms
3. Legal Landscape
   - A. Gaps or “under” regulation
   - B. “Over-” or “mis-” regulation

De-Identification: Concerns

- De-identified data is useful, but not a panacea
- Re-identification risk → persistent concern
  - HIPAA safe harbor method inadequacies
  - Combining data sets (mosaic effect)
  - Sensitive information/attributes may be revealed, even if data is not fully re-identified
  - Overly restrictive de-identification may be unsustainable/stifle innovation
- HIPAA expert determination method
  - Concerns about inadequate transparency or objective scrutiny
  - No standards or certification of "experts"
- No explicit prohibitions on re-identification
- Lack of transparency re: de-identified data disclosures, uses

De-Identification: Possible Solutions

- Context should drive re-identification risk reduction measures
  - Public use data sets v. data in enclaves
- Consistent de-identification standards for all personal data
  - Regulator collaboration
  - Incentivize use of de-identified data
- Define standards & best practices for expert determination
  - OCR-led public-private collaboration
  - Vet statistician approaches; publish
- Certification or accreditation for de-identification experts/organizations
- Automate statistical expertise = easy and affordable alternative to safe harbor

De-Identification: Possible Solutions, continued

- Legislation to prohibit and establish penalties for re-identification
  - Public policy exceptions (health & safety; white hat)
- Regulation to to require re-assessment of re-identification risk when datasets are combined
- Regulation to impose security requirements (commensurate with risk) to protect de-identified data
- OCR: re-evaluate Safe Harbor

De-Identification: Possible Solutions, continued

- Reduced requirements for de-identification in certain validated research circumstances
  - Controlled environments (e.g. data enclaves)
  - Internal use only
  - Data use agreements (a la limited data sets)
    - Permitted uses and prohibiting re-identification
  - Patient-controlled research initiatives
  - Research approved by an IRB or Privacy Board

Consent: Concerns

Within the HIPAA Framework

- PSWG focused on research and whether HIPAA and Common Rule appropriately advance innovation and learning
  - Low-risk research
  - Role of transparency coupled with “appropriate use” definitions

Consent: Concerns, continued

Outside the HIPAA Framework

- Consent not required
  - FTC – "deceptive or unfair" trade practices
- Consent not scalable? (the 3 Vs)
- "Health"-related data constantly expanding
- Hard to list "expected uses" of non-HIPAA data
  - Consider a list of factors to determine reasonable/expected uses; also identify factors that should *not* be considered

Consent: Concerns, continued

Overarching concerns for both environments:

- All in or all out → limited choice
- Granularity and scalability of consent
  - Granular choice limited by policy environment and technological capabilities
  - Enable individual control, including broad sharing
- Persistent consents
  - Across environments (HIPAA → non HIPPA → HIPAA)
- Over-reliance on consent: patient burden / ethics

Consent: Possible Solutions

Research in HIPAA environment

- Refine ANPRM recommendation re: consent for research
  - Entity in control follows FIPPs
- IRB consent waiver for low-risk research?
- Transparency + appropriate use = no consent

Consent: Possible Solutions, continued

Non-HIPAA Environment

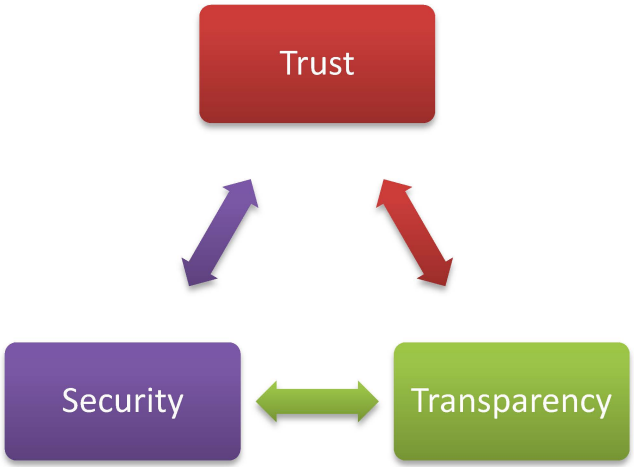
- Assess when consent is required
  - "Appropriate" or "expected use" of consumer-generated data; context
  - Middle ground = more difficult; edge cases = easy
  - Based on: privacy risk, identifiability, commercial use or profit
  - Disclosure outside initial collection environment
- Follow FIPPs
- Conditional consents: educate consumers

Security Testimony

- PSWG heard testimony on February 9, 2015
- 3x5 minute presentations; 45 minute discussion

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 Andersen	Milliman	Director of Software as a Service

Security: Key Themes



- Maintain a holistic, flexible approach to security
- Regulatory compliance can be impacted by organizational resources
  - Complex tech solutions are difficult and costly
- Many threats; greatest impact by focusing internally
- Embrace common security frameworks (e.g. HITRUST)

Remaining Work – Path Forward

Remaining Work

- Refine "possible solutions" into more specific recommendations on topics of:
  - De-identification
  - Consent
  - Security
- Dive more deeply into topics that we've only managed to discuss indirectly as a part of other topics:
  - Transparency
  - Collection/use/purpose limitations
  - Preventing/Limiting/Redressing Harms
  - Legal Landscape
    - Gaps or "under" regulation
    - "Over-" or "mis-" regulation

Question and Answer

Backup

Summary of Hearing Testimony

Health Big Data Opportunities & the Learning Health System Testimony

Beneficial opportunities for using data associated with the social determinants of health

- User generated data; e.g., track diet, steps, workout, sleep, mood, pain, and heart rate
- 3 characteristics: (1) breadth of variables captured, (2) near continuous nature of its collection, and (3) sheer numbers of people generating the data
- Personal benefits → predictive algorithms for risk of readmission in heart failure patients
- Community benefits → asthma inhaler data to identify hot spots; track aggregate behavior of runners
- Key issues: privacy, informed consent, access to the data and data quality
- Important to allow experimentation for the technology and methods to improve
- Important to allow institutions catch up to learn how best to take advantage of opportunities and realize potential benefits

"Care between the care" → patient defined data. May ultimately reveal a near total picture of an individual – merged clinical and patient data; data must flow back and forth

Data needs access, control and privacy mechanisms throughout its life cycle, at level of data use, not just data generation; data storage is not well thought through

Health Big Data Opportunities & the Learning Health System Testimony, con't

Must embed learning into care delivery; we still do not have answers for a large majority of health questions

Key points:

1. Sometimes there is a need to use fully identifiable data
2. It is not possible to get informed consent for all uses
3. Impossible to notify individuals personally about all uses
4. Can't do universal opt-out because answers could be unreliable
5. There is likely a standard that could be developed that determines "clearly good/appropriate uses" and "clearly bad/inappropriate uses"

Focus on:

1. Minimum necessary amount of identifiable data (but offset by future use needs)
2. Good processes for approval and oversight
3. Uses of data stated publicly (transparency)
4. Number of individuals who have accessed to data minimized (distributed systems help accomplish this)

When we use identifiable data, we must store it in highly protected locations – "data enclaves"

Health Big Data Opportunities & the Learning Health System Testimony, con't

- Shift in the way we look into data and its use
- Paradigm of looking into the data first and then beginning to understand different findings and correlations that you didn't think about in standard hypothesis-driven research, but you do when you're doing data driven research
- Focus on sharing, integrating, and analyzing cancer clinical trial data
- Use de-identified data; de-identification is the responsibility of the data provider); most data providers use expert determination method

- Data collected and used to conduct topological data analysis
- Mathematics concept that allows one to see the shape of their data
- Analysis can identify healthcare fraud, waste, and abuse, as well as reduce clinical variation and improve clinical outcomes
- Use de-identified data
- We have not been able to get a data set that shows a continuum of care for a patient
- While interoperability isn't exactly perfect in other industries, in healthcare we've seen that to be a unique issue

Health Big Data Opportunities & the Learning Health System Testimony, con't

- Partners drawn from academia, care delivery, industry, technology and patient and consumer interest
- Key asset is the database – 7.7 terabytes of de-identified data from administrative claims of over 100 million individuals over 20 years, clinical data from electronic health records of 25 million patients, and consumer data on 30 million Americans
- Data provided to researchers vie secure enclave
- Premise: combine the insights of multiple partners
- Key issue: systematically coordinating uses of de-identified techniques with subsequent uses of PHI

- Cloud-based, single instance software platform with 59,000 healthcare provider clients
- Products include EHR, practice management, and care coordination services
- Data immediately aggregated into databases; near real-time visibility into medical practice patterns
- Monitor visit data for diagnoses of influenza-like illness
- Tracking the impact of the ACA on community doctors; sentinel group of 15k doctors; measuring # patients seen, health status, and out-of-pocket payment requests

Health Big Data Concerns Testimony

A person's health footprint now include Web searches, social media posts, inputs to mobile devices, and clinical information such as downloads from implantable devices

Key issues include (1) notice and consent, (2) unanticipated/unexpected uses, and (3) security

HIPAA does not apply to most apps

Without clear ground rules and accountability for appropriately and effectively protecting user health data, data holders tend to become less transparent about their data practices

Patient perspective

- Frustration with "data dysfunction" - cannot access and combine his/her own data
- Privacy and security are cited as excuses/barriers that prevent access to personal data
- Health data is a social asset; there is a public need for data liquidity

Health Big Data Concerns Testimony, continued

Issues from conferences on big data and civil rights:

- The same piece of data can be used both to reduce health disparities and empower people and to violate privacy and cause harm
- All data can be health data
- Focus on uses and harms rather than costs and benefits. Focusing on C&B implies trade-offs. Instead, seek redress via civil rights laws.
- Universal design. Design the technology and services to meet the range of needs without barriers for some.
- Ensure privacy and security of health information via all the FIPPs, not just consent
- Principle of preventing misuse of patient data. There are many good uses of health information, but there must also be some prohibitions.

Consumer Protections Testimony

- Ease of re-identification narrative may be misleading
- If you de-identify data properly, success rate is very low for attacks. If you don't use existing methods or de-identify data at all, and if data is attacked, success rate is high
- De-identification is a powerful privacy protective tools
- Most attacks on health data have been done on datasets that were not de-identified at all or not properly de-identified
- De-identification standards are needed to continue to raise the bar. There are good de-identification methods and practices in use today, but no homogeneity.
- HIPAA works fairly well – but mounting evidence that Safe Harbor has important weaknesses
- De-identification doesn't resolve issues of harmful uses; may need other governance mechanisms, such as an ethics or data access committees
- Privacy architectures. Still need to de-identify the data that goes in to Save Havens
- Distributed computation. You push the computations out to the data sources and have the analysis done where the data is located

Consumer Protections Testimony, continued

- Cant' regulate something called "big data" because once you define it, people will find a way around it
- The people who think privacy protections don't apply to big data are likely the same people who have always been opposed to privacy protections
- No reason to think HIPAA's research rules need to be different because of big data. HIPAA at least sets a clear and consistent process that covered entities and business associates must follow

- Privacy laws today are overly focused on individual control
- Individual control is inadequate as both a definition and an aspiration. Impossible expectation to think a person can control his or her personal health data
- The effect of control is an impediment to availability. For most patients & families, the primary concern about data misuse was that they would be contacted
- Privacy is too critical and important a value to leave to a notion that individuals should police themselves
- We need to be thinking about how to make sure data is protected at the same time that it's available. We don't let the mechanisms of protection by themselves interfere with the responsible use of the data

Current Law Testimony

- HIPAA Safe Harbor de-identification requires removal of 18 fields
  - May not give researchers the data they need/want; but some researchers cited the value of de-identified data
- Limited data set is a bit more robust, but not a lot
- Definition of research same under HIPAA and Common Rule (generalizable knowledge)
- May receive a waiver to use data by an IRB or privacy board
- HITECH changes:
  - Authorization may now permit future research (must adequately describe it)
  - Some compound authorizations now permitted for research purposes

- HIPAA applies to covered entities and business associates; patient authorization/consent is not required for treatment, payment, or healthcare operations purposes
- Paradox in HIPAA
  - Two studies that use data for quality improvement purposes using the same data points done to address the same question or sets of questions and done by the same institution will be treated as operations (no consent required) if the results are not intended to contribute to generalizable knowledge (intended for internal quality improvement instead)

Current Law Testimony, continued

- HIPAA does not cover a large amount of healthcare data
- Past few years = explosion in amount of data that falls outside of HIPAA
  - Mobile applications, websites, personal health records, wellness programs
- FTC is default regulator of privacy and security → unfair or deceptive acts or practices
  - Very active on general enforcement of data security standards
  - Debate as to whether the FTC really has authority to do this; 2 pending cases
- Less FTC enforcement in privacy space, especially healthcare
  - Tough question is broader FTC ability to pursue unfair practices in area of data privacy (enforcement of deceptive practices is easier)
- Fair Credit Reporting Act (FCRA) governs how information is gathered, used, and what people must be told about contents of credit reports
  - Specific prohibitions → using medical data for credit purposes
- Many conflicting state laws, which are often confusing, outdated and seldom enforced
- Key issue: substantial gaps exist
  - More and more data that is health-related is falling outside the scope of HIPAA rules