Changing Concepts of Anonymity, Confidentiality, and Privacy in SBER

Scott Bradner & Dean Gallant

PRIM&R's 2015 AER Conference | November 12-15 Boston, MA

1

Agenda

- Anonymity
- Privacy
- Confidentiality
- Private information under regulations
- Rule sets
- Best practices
- Establishing responsibilities
- Certificates of confidentiality
- NPRM

2

Anonymity

- Traditional definition: "*being unnamed*"
- Wallace: "*Noncoordinatability of traits in a given respect*"

  "*Anonymity*," in Ethics and Information Technology, 1999, Kathleen Wallace, Chairperson, Department of Philosophy, Hofstra University.
- Most data represents traits of some kind – anonymity is the inability to coordinate those traits to a person

3

Anonymity, permanence

- Noncoordinatability may not be forever

  Example: Unabomber
- He had anonymity when he was "just" the "sender of bombs to computer scientists"
- That changed when he got his writings published in the New York Times

  His brother saw traits in the writings that identified him
- New information from an additional venue changed things

4

Anonymity in data

- Data with anonymity: data recorded without any link to a particular subject
  - E.g., no IP address, photo, etc.
- May still contain traits
  - E.g., patterns of responses
- Getting harder to ensure permanent anonymity as the use of “big data” expands
  - Data from many sources about many people

5

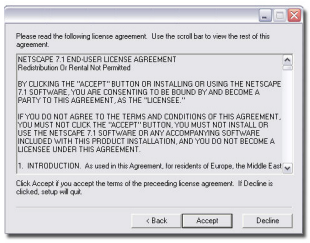
De-identified data

?

- Question: is de-identified data the same as data that was collected without any identifiers? (data with anonymity)

6

Do contracts help?



- Contractual agreements to not attempt (re)identification may help protect anonymity, but are not guarantees
- Contracts apply to those who sign or otherwise actually agree to them.
  - E.g., not the recipient of data exposed in a data breach

7

Aside about public data

- Is it time to reassess the need for legal controls on the analysis of public data?
  - Big data tools can find patterns in public data that can reveal information a subject would rather not have revealed
- E.g., security cameras, license plate readers, social media, ad trackers

8

Anonymity in data, biological

?

- Question: can a biological sample ever have permanent anonymity?
  Considering genetic analysis?

9

Privacy

Merriam-Webster

- The state of being alone: the state of being away from other people, the state of being away from public attention
- More specifically in the context of information, privacy is the ability to control the distribution of your personal information

10

Exposed personal information

?

- Question: if the personal information becomes known, but cannot be tied to an individual, has privacy been violated?

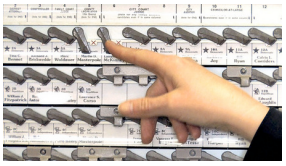
11

Privacy?

"Oh, look . . . they're reading '1984' in Ms. Smith's English class."

12

Privacy ≠ Anonymity



- In anonymity the information itself does not identify a subject
  - No one knows
- In privacy, information that identifies a subject is not known to people who should not know it
  - Authorized people know
- An example – voting is an act where your privacy is protected, but not your anonymity
  - The fact that *you* voted is known, but not who you voted for

13

Privacy, history

- Not much privacy in small towns/ settlements
  - Everyone knows what everyone is doing
- Not much privacy from kings, etc.
  - e.g., Magna Carta required due process but not privacy
- Some privacy in early English law
  - 'home is castle' (1499)
  - Eaves-droppers that spread *"mischievous tales" "are a common nuisance"* & can be fined
    - Blackstone: book 4, chap 13 (1769)

14

Privacy, history, contd.

HARVARD

LAW REVIEW.

VOL. IV. DECEMBER 15, 1890. NO. 5.

THE RIGHT TO PRIVACY.

"It could be done only on principles of private justice, moral fitness, and public convenience, which, when applied to a new subject, make common law without a precedent; much more when received and approved by usage."

WILLES, J., in Millar v. Taylor, 4 Burr. 2303, 2312.

THAT the individual shall have full protection in person and in property is a principle as old as the common law; but it has been found necessary from time to time to define anew the exact nature and extent of such protection. Political, social, and economic changes entail the recognition of new rights, and the common law, in its eternal youth, grows to meet the new demands of society. Thus, in very early times, the law gave a remedy only for physical interference with life and property, for trespasses *vi et armis*. Then the "right to life" served only to protect the subject from battery in its various forms; liberty meant freedom from actual restraint; and the right to property secured to the individual his lands and his cattle. Later, there came a recognition of man's spiritual nature, of his feelings and his intellect. Gradually the scope of these legal rights broadened; and now the right to life has come to mean the right to enjoy life,—the right to be let alone; the right to liberty secures the exercise of extensive civil privileges; and the term "property" has grown to comprise every form of possession—intangible, as well as tangible.

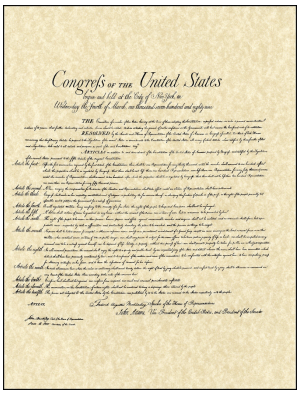
Thus, with the recognition of the legal value of sensations, the protection against actual bodily injury was extended to prohibit mere attempts to do such injury; that is, the putting another in

"That the individual shall have full protection in person and in property is a principle as old as the common law; but it has been found necessary from time to time to define anew the exact nature and extent of such protection. Political, social, and economic changes entail the recognition of new rights, and the common law, in its eternal youth, grows to meet the new demands of society."

Warren & Brandeis - The Right to Privacy, 1890

15

US Constitution & Privacy



Congress of the United States

- The word "privacy" is not in the U.S. Constitution
- But the Supreme Court has found that a right of privacy is implied by a number of the amendments in the Bill of Rights:
  - 4th amendment – protection against unreasonable searches and seizures
  - 5th amendment – self incrimination, due process, etc.
  - 9th amendment – not all rights enumerated

16

US Constitution & Privacy, contd.

- Privacy was called a "penumbra right" in Griswold v. Connecticut (1965)
  - Penumbra: "*a body of rights held to be guaranteed by implication in a civil constitution*"
- The Supreme Court has found that the Constitution protects a "zone of privacy" in two areas
  - Independence in making certain types of decisions
  - Avoiding disclosure of personal matters

17

Supreme Court & Privacy, contd.

- Key case: Katz v. U.S. (1967)
  - Government wiretapped a phone booth w/o warrant
  - Court found that government violated the Fourth Amendment (unreasonable searches and seizures)
  - In doing so, moved the right to privacy from a place (e.g., home) to a person
  - Added the 'reasonable assumption of privacy' test
    - 1: Did person exhibit personal expectation of privacy?
    - 2: Does society recognize the expectation as reasonable?

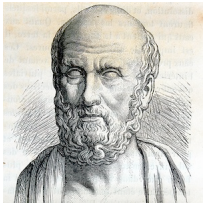
18

Prosser on privacy

- William Prosser survey of privacy cases (1960)
- Showed 4 classes of charges
  1. Intrusion on the seclusion or private affairs of another
  2. Appropriation of name or likeness of another
  3. Public disclosure of private facts
  4. False light (presenting a false impression of subject)
- Note - privacy applies to 'natural persons'

19

Confidentiality



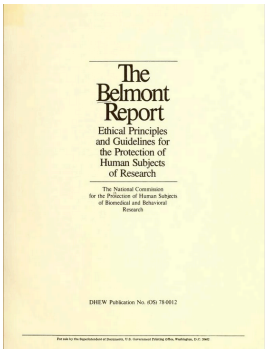
- Informational privacy
  Public disclosure of private facts
- Confidential data: data kept within certain boundaries (whether specified or understood) by agreement

Whatever, in the course of my practice, I may see or hear (even when not invited), whatever I may happen to obtain knowledge of, if it be not proper to repeat it, I will keep sacred and secret within my own breast.

Hippocratic Oath

20

Belmont principles



- Respect for Persons
  - Includes respect for personal privacy
  - Adherence to pledges of confidentiality
- Beneficence
  - Privacy violation = dignitary harm
  - Breach of confidentiality can be multi-dimensional harm
- Justice

21

Common Rule references

Human subject means a living individual about whom an investigator … obtains ... identifiable private information.

Private information includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record).

Private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects. --45 CFR 46.102(f)

22

Common Rule references

[Criteria for IRB approval include] when appropriate, there are adequate provisions to protect the **privacy** of subjects and to maintain the **confidentiality** of data. *--45 CFR 46.111(a)(7)*

[Informed consent must include] a statement describing the extent, if any, to which **confidentiality** of records identifying the subject will be maintained. *--45 CFR 46.116(a)(5)*

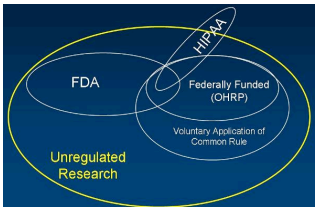
23

Privacy and Confidentiality

- Information privacy depends on good (in context) security
- The quality of security required depends on the ability of those who are after the information
  - Protecting against a nation-state is close to impossible
    - Either one with a subpoena or one willing to use extralegal means
  - Protecting against a local identity thief is feasible
- There are many rule sets relating to private/confidential information

24

Other rule sets for Confidential Data



- In addition to the Common Rule, some research information is defined as private and subject to protection requirements under other sets of rules or standards; some of these include:
  - U.S. federal and state laws
  - EU directive & laws
  - Institutional rules
  - HIPAA

25

U.S. Laws: HIPAA

HIPAA
Health Insurance Portability and Accountability Act

- Health Insurance Portability and Accountability Act (HIPAA)
  - Applies to records created by medial service providers (e.g., hospitals, insurance and billing services) and others that receive such records
  - HIPAA Security Rule defines security safeguards
    - Administrative, physical, and technology safeguards
  - Business Associate Agreement (BAA) often required with entities that receive HIPAA data
    - Contract-based protection

26

HIPAA identifiers

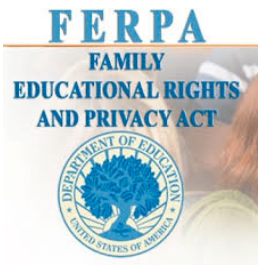
"De-identification" involves the removal of 18 identifiers of the individual and relatives, employers, household members, plus "no actual knowledge" that the remaining information could be used to identify the individual.

Address, incl. most ZIP info
Relevant dates (except year)
Phone numbers
Fax numbers
SSN
Medical record numbers
Health plan beneficiary numbers
Account numbers
Certificate/license numbers

Vehicle ID, S/N, license numbers
Device ID, serial numbers
URLs
IP addresses
Biometric ID (incl. finger & voiceprint)
Full face photo & comparable images
Any other unique number, code, or characteristic

27

U.S. Laws: FERPA



- The Family Educational Rights and Privacy Act (FERPA)
- Must protect the confidentiality of student records
  - Conditional exception for directory information
- No specific administrative, physical and technology requirements

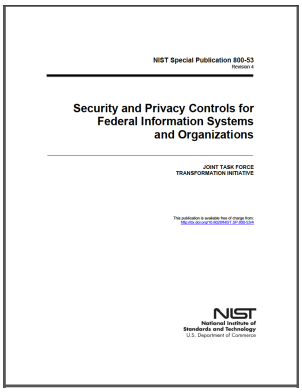
28

U.S. Laws: FISMA

- The Federal Information Security Management Act (FISMA)
- Requires U.S. federal agencies to develop and execute information security protection plans
- Large (endless) set of administrative, physical and technology safeguards
- FISMA compliance requirements sometimes show up in grants or data use agreements

29

U.S. Laws: FISMA, contd.


NIST Special Publication 800-53
Revision 4

Security and Privacy Controls for Federal Information Systems and Organizations

JOINT TASK FORCE TRANSFORMATION INITIATIVE

This publication is available free of charge from:

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

- Three levels:
  - Low – hard to meet
  - Moderate – very hard to meet
  - High – You will not meet

30

U.S. Laws: COPPA

- The Children's Online Privacy Protection Act (COPPA)
- Special requirements for websites collecting information about users younger than 13
  - Parental permission required
  - Required to have reasonable procedures to protect the confidentiality, security, and integrity of personal information collected from children.
- Does not apply to nonprofits
  - But would be very bad PR to violate

31

U.S. State Laws

- California Constitution Article 1, Section 1:
  - *All people are by nature free and independent and have inalienable rights. Among these are enjoying and defending life and liberty, acquiring, possessing, and protecting property, and pursuing and obtaining safety, happiness, and privacy.*
- California has adopted many privacy-related laws
- Some other states have as well
  - E.g., Massachusetts 201 CMR 17
    - *Specific protection requirements*

32

Massachusetts 201 CMR 17

- Targets financial information
  - E.g. credit card #s, SSNs
  - No matter what they are collected for
- Regulations include specific technical and process requirements to protect the information
  - E.g., lockout of bad password guesses, validate vendor ability to meet law & contractually require them to do so
    - Even out of state vendors

33

Secretary's Advisory Committee on Human Research Protections

SACHRP

- "Considerations and Recommendations Concerning Internet Research and Human Subjects Research Regulations" (2013):

If individuals intentionally post or otherwise provide information on the Internet, such information should be considered public unless existing law and the privacy policies and/or terms of service of the entity/entities receiving or hosting the information indicate that the information should be considered "private."

34

Other applicable rules, codes, etc.

- Best practices
- Professional organization codes of ethics (APA, ASA, AAA, etc.)
- Institutional policies
- European privacy laws

35

Best Practices

- Follow applicable laws and regulations
- Do not collect more information than necessary for the purposes of the research
- Remove identifiers (or assign opaque codes) as soon as possible if data are sensitive; separate and protect code key

36

Best Practices, contd.

- Consider strategies for grouping data (but be aware of limitations)
- Practice good data hygiene and data security

37

Codes of Ethics

American Psychological Association

ASA
American Sociological Association

American Anthropological Association

- Many professional organizations have adopted codes of ethics
- Help separate “right” from “wrong” in the context of the organization

38

European privacy laws

- Basic concept of information privacy is different in US & EU
  - In US, point solutions to specific issues
  - In EU, broad principle-based regulations
- Restrictions on information collection, analysis and retention
- If you think you want to do human subject research in the EU, check with a local lawyer first

39

De-identified data, issues

- Even assuming that de-identified data can not be re-identified
  - An unproven premise
- Analyzing de-identifying data can produce different results than analyzing the original data
- See papers posted to conference site

40

Establishing responsibilities

- It is important to define who is responsible for what when it comes to protecting confidential information
- First step, develop an overall research data security policy and program
- Includes categorizing the types of information and establishing process and technical protection requirements for each category

http://vpr.harvard.edu/pages/harvard-research-data-security-policy

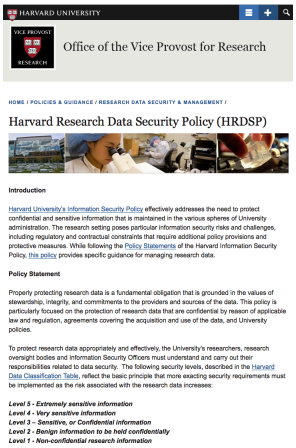
41

Harvard Categories

LEVEL 1	Public information
LEVEL 2	Information the University has chosen to keep confidential, but the disclosure of which would not cause material harm
LEVEL 3	Information that, if disclosed, could cause risk of material harm to individuals or the University
LEVEL 4	Information that, if disclosed, would likely cause serious harm to individuals or the University
LEVEL 5	Information that, if disclosed, would cause severe harm to individuals or the University

42

Harvard Research Data Security Policy (HRDSP)


HARVARD UNIVERSITY

Office of the Vice Provost for Research

HOME / POLICIES & GUIDANCE / RESEARCH DATA SECURITY & MANAGEMENT /

Harvard Research Data Security Policy (HRDSP)

Introduction

Harvard University's Information Security Policy effectively addresses the need to protect confidential and sensitive information that is maintained in the various spheres of University administration. The research setting poses particular information security risks and challenges, including regulatory and contractual constraints that require additional policy provisions and protective measures. While following the Policy Statements of the Harvard Information Security Policy, this policy provides specific guidance for managing research data.

Policy Statement

Properly protecting research data is a fundamental obligation that is grounded in the values of stewardship, integrity, and commitments to the providers and sources of the data. This policy is particularly focused on the protection of research data that are confidential by reason of applicable law and regulation, agreements covering the acquisition and use of the data, and University policies.

To protect research data appropriately and effectively, the University's researchers, research oversight bodies and Information Security Officers must understand and carry out their responsibilities related to data security. The following security levels, described in the Harvard Data Classification Table, reflect the basic principle that more exacting security requirements must be implemented as the risk associated with the research data increases:

Level 5 - Extremely sensitive information
Level 4 - Very sensitive information
Level 3 – Sensitive, or Confidential information
Level 2 - Benign information to be held confidentially
Level 1 - Non-confidential research information

- Builds on the Harvard Information Security Policy
  - www://policy.security.harvard.edu
  - Covers all confidential information at the university & at university vendors
- HRDSP Establishes roles and responsibilities for:
  - Researchers
  - Information Security Officers
  - Research Oversight Bodies
  - Office of the Vice Provost for Research

43

HRDSP: Researchers

Researchers have these responsibilities:
1. Identifying confidentiality and data security obligations, based on laws, regulations, policies, and binding commitments such as data use agreements and participant consent agreements.
2. Except in cases where it is the responsibility of a research oversight body, (see Definitions) it is the responsibility of researchers to identify the appropriate data security level for research data. (See procedures (link) for how to get assistance in setting a data security level.)
3. When the data security level has been established, researchers are responsible for creating and maintaining data documentation, implementing the security controls corresponding to the requirements of the data security level and developing and following a data security plan and procedures over the course of their projects.

1. Identify any confidentiality obligations (e.g. laws, data use agreements, etc.)
2. Identify appropriate security level for data (except when IRB does it)
3. Develop & follow security plan, that implements appropriate security controls, maintain data documentation

44

HRDSP: Info Security Officers

Local or School Information Security and HUIT Information Security are responsible for assisting researchers with implementation of appropriate security controls in accordance with the level assigned by the Research Oversight Body or specific controls outlined in a DUA. Information Security Officers may be asked to review DUAs for information security controls specified by a data provider. …

1. ***Variances:*** The Information Security Officer and the Researcher may apply compensating controls for the assigned data security level, if certain controls prescribed for the security level are not feasible. These compensating controls will be documented and attested to by the researcher and the ISO(s), and the ISO will inform the IRB if the project is under IRB review.
2. ***Signature:*** A checklist will not be considered complete without the researcher attestation via signature.
3. ***Facility Certification:*** A research facility may be certified at a certain data security level, enabling projects classified as up to and including that data security level to be exempt from separate review and approval.

- Assist researchers in implementing appropriate security controls, may be asked to see if requirements in data use agreements are being met

1. May grant variances if appropriate compensating controls in place
2. Must sign off on security plan
3. May certify a facility for a particular data level – remove requirement for individual reviews

45

HRDSP: Research Oversight Bodies

Research oversight bodies are responsible for:
1. Assessing data security risks associated with the research within their purview and assigning data security levels for the research.
2. Establishing procedures to set security levels, either on a project by project basis, or by category of research data, and
3. Informing researchers about data security risks and working with them to set appropriate data security levels.

While all research oversight bodies share these same basic roles and responsibilities with respect to their engagement with researchers and information security officers, the procedures will vary, depending on the particular research and the oversight body or bodies that may be involved.

1. Assessing data security risks & assigning data security levels for some research
2. Establish procedures to assign data security levels
3. Inform researchers about data security risks & collaboratively set data security levels

46

HRDSP: VP Research

The Office of the Vice Provost for Research is responsible for:
Implementing this policy

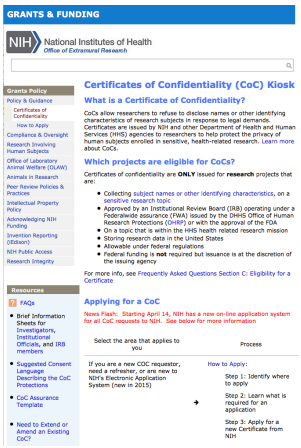
1. Working with research oversight bodies to identify data security risks and set data security levels, and
2. Working with researchers and IT and HUIT as appropriate, to foster awareness and understanding of the policy.
3. Periodically reviewing adherence to the policy

- Implement the HRDSP

1. Work with research oversight bodies to identify data security risks & set data security levels
2. Work with researchers and IT groups to foster awareness and understanding of the HRDSP
3. Periodically review adherence to the HRDSP

47

Certificates of confidentiality



- What they do: protect against compelled disclosure of certain types of confidential information
- What they don't do: prohibit voluntary disclosure of the same information; override mandated reporting or Tarasoff obligations

48

NPRM

- The NPRM includes, directly or by implication, several changes to the concept of privacy and/or how identifiable information will be handled and protected.
- New privacy safeguards are proposed for biospecimens and identifiable private information.
- HHS will develop standards (not available yet)

49

NPRM, contd.

- [from the discussion in the preamble of exclusion 101(b)(2)(i)]: *In the case of observation of public behavior, even if the subject does not know that an investigator is watching his or her actions, the subject's behavior is public and could be observed by others and thus the research observation is not inappropriately intrusive.*

50

References

- SACHRP "Considerations and Recommendations Concerning Internet Research and Human Subjects Research Regulations," 2013. http://www.hhs.gov/ohrp/sachrp/mtgings/2013%20March%20Mtg/internet_research.pdf
- How to de-identify your data. O. Angiuli et al. ACMqueue, 25 Oct 2015. http://queue.acm.org/detail.cfm?id=2838930
- Is de-identification sufficient to protect health privacy in research? M. Rothstein American Journal of Bioethics, Sept 2010. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3032399/pdf/nihms-264889.pdf
- American Anthropological Association Statement on Ethics, 2012.
- http://ethics.aaanet.org/category/statement/
- American Psychological Association Ethics Code, 2010. http://www.apa.org/ethics/code/index.aspx
- American Sociological Association Code of Ethics, 1999. http://www.asanet.org/images/asa/docs/pdf/CodeofEthics.pdf
- COPPA FAQ: https://www.ftc.gov/tips-advice/business-center/guidance/complying-coppa-frequently-asked-questions

51

Image Credits

Slide source
3 - http://www.merriam-webster.com/dictionary/anonymity
http://www.kathwallace.com/
4 - both pictures: https://pl.wikipedia.org/wiki/Theodore_Kaczynski
5 - http://www.gardenista.com/products/unfinished-diamond-willow-hiking-staff
7 - http://www.brightlabs.com.au/page/Blog/Click_wrap_agreements-_What_Are_They_and_Are_They_Enforceable/
8 - cameras https://en.wikipedia.org/wiki/Closed-circuit_television
9 - http://www.duimiami.com/blood-tests/
12 - Bruce Beattie cartoon 2005
13 - voting machine - http://blog.syracuse.com/opinion/2009/03/forget_lever_voting.html
14 - magna carta- http://www.bl.uk/magna-carta
15 - Warren & Brandeis http://heinonline.org/HOL/LandingPage?collection=journals&handle=hein.journals/hlr4&div=31&id=&page=
16 - http://www.historicdocumentsofamerica.com/images/BillofRights001.jpg
17 - Supreme court seal https://commons.wikimedia.org/wiki/File:Seal_of_the_United_States_Supreme_Court.svg
18 - Katz case - http://www.nydailynews.com/new-york/city-expand-free-wifi-transforming-pay-phones-hotspots-article-1.1788374
19 - prosser - http://heinonline.org/HOL/AuthorProfile?search_name=Prosser%2C+William+L.&collection=journals&base=js
20 - http://afflictor.com/2010/01/07/medical-world-the-importance-of-daydreaming
21 - https://archive.org/details/belmontreporteth00unit

52

Image Credits

Slide source

24 IRA http://www.dailymail.co.uk/news/article-2619994/Former-IRA-member-taped-interviews-led-arrest-Gerry-Adams-fear-life-tensions-grow-Northern-Ireland.html

25 – common rule http://www.hhs.gov/ohrp/sachrp/mtgings/mtg07-08/present/nelson.html

26 - hipaa http://www.limswiki.org/index.php/Health_Insurance_Portability_and_Accountability_Act

28 – ferpa http://www.campussafetymagazine.com/article/schools_beware_some_see_ferpa_update_as_inevitable

29 – fisma http://www.fediq.com/fisma-defined/

30 - http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf

31 – coppa https://wiki.uiowa.edu/pages/viewpage.action?pageId=60267257

FTC logo - https://commons.wikimedia.org/wiki/File:US-FederalTradeCommission-Seal.svg

32 - https://en.wikipedia.org/wiki/California

33 - https://en.wikipedia.org/wiki/Seal_of_Massachusetts

34 - http://www.histalkpractice.com/2009/12/

35 - http://kelleherpatentlaw.com/news/supreme-court-hear-arguments-i4i-ltd-v-microsoft-corp

36, 37 - https://www.youtube.com/watch?v=FIlvssCuLuM

38 - APA logo - apa - http://www.apa.org/

ASA logo - http://www.asanet.org/

AAA logo - http://politicalandlegalanthro.org/about/

53

Image Credits

Slide source

39 - http://europa.eu/about-eu/basic-information/symbols/images/flag_yellow_low.jpg

40 - sweeney - http://www.gov.harvard.edu/people/faculty/latanya-sweeney

41 - harvard logo - https://en.wikipedia.org/wiki/Harvard_University

43 - Hrdsp - http://vpr.harvard.edu/pages/harvard-research-data-security-policy

48 - https://grants.nih.gov/grants/policy/coc/index.htm

49, 50 - http://www.hhs.gov/ohrp/humansubjects/regulations/nprmhome.html

54