

Data Security in Research: Is the IRB Responsible?

1

MARIAN HUGHLETT
UNIVERSITY OF LOUISVILLE

RUSSELL OPLAND
UNIVERSITY OF CALIFORNIA

2

DATA BREACH TRENDS

Privacy Rights Clearinghouse Breaches

3

When: Sept 29, 2006

Where: University of Iowa, Dept of Psychology

What: A computer containing SSNs of psychology department research study subjects was the object of an automated attack designed to store pirated video files for subsequent distribution.

Records: 14,500

Privacy Rights Clearinghouse Breaches

4

When: Oct 24, 2006

Where: Jacobs Neurological Institute

What: The laptop of a research doctor was stolen from her locked office at the Institute. It included records of patients and her research data.

Records: Unknown

Privacy Rights Clearinghouse Breaches

5

When: Feb 2, 2007

Where: University of Missouri

What: A hacker broke into a UM computer server mid-January and might have accessed personal information, including SSNs, of 1,220 researchers on 4 campuses. The passwords of 2,579 individuals might also have been exposed.

Records: 3,799

Privacy Rights Clearinghouse Breaches

6

When: Aug 21, 2007

Where: Walter Reed Army Institute of Research

What: Boxes of documents containing personal information were supposed to be shredded but instead turned up in an off-base trash bin. Police do not believe anyone had access to the information other than the person who found the records.

Records: Unknown

Privacy Rights Clearinghouse Breaches

7

When: Mar 24, 2008

Where: National Institutes of Health

What: A laptop was stolen from the trunk of a car. It contained information about heart disease patients, including their names, dates of birth and diagnoses of their medical conditions. Ongoing review of the computer's last-known contents, performed on data backed up from the laptop before it was stolen, found a file that, unbeknownst to the lead researcher, had been loaded onto the laptop by a research associate. That file included SSNs for at least 1,281 of the 3,078 patients enrolled in the multi-year study, which is sponsored by the NIH's National Heart, Lung and Blood Institute.

Records: 3,078

Privacy Rights Clearinghouse Breaches

8

When: May 20, 2008

Where: Duke University

What: The Fuqua School of Business notified former New York University students that some of their personal information was inadvertently accessible by targeted Internet searches. The personal data included names and Social Security numbers and was contained in the faculty member's research records. The information could have been accessed only if searched by specific student names, along with a search code for SSNs.

Records: 273

Privacy Rights Clearinghouse Breaches

9

When: Aug 20, 2009

Where: California State University (Los Angeles)

What: The theft of two desktop and 12 laptop computers from an office caused identity theft concerns for students and faculty members. Someone broke a window in the office of the university's Minority Opportunities in Research program to steal the computer. The computers stolen contained individual names, SSNs and addresses, according to campus.

Records: >600

Privacy Rights Clearinghouse Breaches

10

When: Sept 25, 2009

Where: University of North Carolina (Chapel Hill)

What: A hacker infiltrated a computer server housing the personal data of women enrolled in a UNC Chapel Hill research study. Among the information exposed: the SSNs of 163,000 participants. The data is part of the Carolina Mammography Registry, a 14-year-old project that compiles and analyzes mammography data submitted by radiologists across North Carolina.

Records: 236,000

Privacy Rights Clearinghouse Breaches

11

When: Jan 1, 2010

Where: Netflix

What: A class action suit was filed against Netflix, Inc., in the United States District Court for the Northern District of California. Plaintiffs in the suit are claiming that Netflix has “perpetrated the largest voluntary privacy breach to date.” According to the complaint, Netflix knowingly and voluntarily disclosed the sensitive and personal information of Netflix subscribers when Netflix provided participants in a contest initiated to improve Netflix’s movie recommendation systems with data sets containing over 100 million subscriber movie ratings and preferences. Netflix has claimed that the data sets provided to the contest participants were anonymized and that the subscribers’ movie ratings were accompanied only by “a numeric identifier unique to the subscriber” (as opposed to the subscriber’s name or other personal information). However, the complaint cites the results of several researchers who, in fact, were able to crack Netflix’s anonymization process and identify individual subscribers.

Records: 480,000

Privacy Rights Clearinghouse Breaches

12

When: Mar 10, 2010

Where: Veterans’ Affairs Medical Center (Atlanta)

What: An assistant allegedly recorded two sets of patient data on to a personal laptop for research purposes. One set included 3 years’ worth of patient data and another held 18 years of medical information. The physician assistant’s laptop was never connected to the VA network and any data she recorded on her laptop was hand entered. The department has not disclosed the number of patients involved in the incident, what kind of personal data was copied, or whether it plans to notify the veterans whose records were downloaded.

Records: Unknown

Common Characteristics of These Breaches?

13

- All research-related
- Every year since record-keeping began in 2005
- Electronic and paper
- Public and private sectors
- Hundreds to hundreds of thousands of records
- Varied threats and risks
- Subjects', faculty, and students' information
- Administrative and research information

Vulnerability Trends

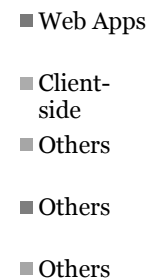
14

In 2009, IBM X-Force® documented 6,601 new vulnerabilities:

49% of vulnerabilities affect web applications (usually server-based)

20% of vulnerabilities affect the client-side (desktop, laptop, etc.)

Vulnerabilities



2009 Web Application Vulnerabilities

15

- Of these, 19% affect the web application itself, and 81% affect the plug-ins

Vendor Patching:

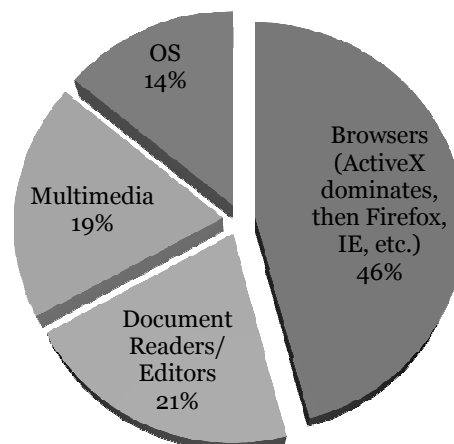
- 5 to 23% of web applications were not patched
- 13 to 86% of plug-ins were not patched
- Vendor averages: 52% not patched in 2009
 - 60% of Critical and High not patched in 2009

Attack Vectors:

- Cross-Site Scripting and Injection attacks vastly dominate

2009 Client-Side Vulnerabilities

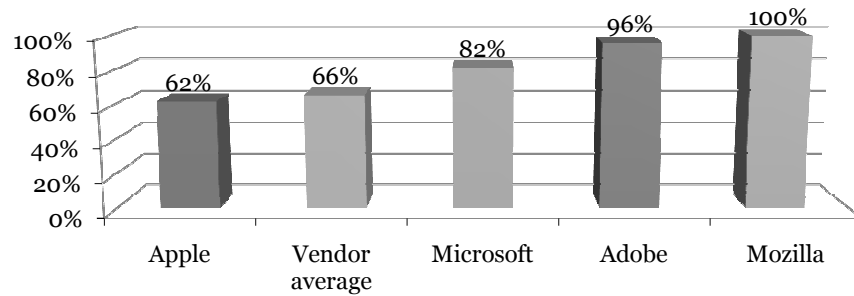
16



2009 Client-Side Patching

17

Percentage of “Critical” and “High” Vulnerabilities Patched in 2009



18

IRB AND RESEARCHER OBLIGATIONS TO SAFEGUARD DATA

Who is responsible for protecting research data?

19

Common Rule

Applies to IRBs,
Researchers

Privacy Rule

Applies to Covered
Entities

Research Regulatory Obligations: Common Rule

20

HHS and FDA protections of human subjects regulations include protections to help ensure the privacy of subjects and the confidentiality of information

Research Regulatory Obligations: Privacy Rule

21

Privacy Rule *supplements* these (HHS and FDA) protections by requiring covered entities to implement specific measures to safeguard the privacy of protected health information (PHI)

—from *Institutional Review Boards and the HIPAA Privacy Rule*,
NIH Publication Number 03-5428 (August 2003)

Regulatory Permissions for Research

22

Common Rule

- Waiver of written ICF
- Waiver of informed consent
- Informed Consent Form (ICF)

Privacy Rule

- Authorization OR
- Exceptions:
 - Waiver of authorization
 - Activities preparatory to research
 - Research on decedents
 - Limited Data Set
 - Deidentified Data

Regulatory Permissions for Research	
23	
Common Rule	Privacy Rule
<ul style="list-style-type: none"> • Waiver of written ICF • Waiver of informed consent • Informed Consent Form (ICF) 	<ul style="list-style-type: none"> • Authorization OR • <u>Exceptions:</u> <ul style="list-style-type: none"> ○ Waiver of authorization ○ Activities preparatory to research ○ Research on decedents ○ Limited Data Set ○ Deidentified Data

24	
<p>Common Rule: Criteria for IRB Approval of Research</p> <ul style="list-style-type: none"> ➤ Waiver of informed consent ➤ Informed Consent Form (ICF) 	<p>“In order to approve research... the IRB shall determine that <u>all</u> of the following requirements are satisfied:</p> <p style="text-align: center;">...there are adequate provisions to <u>protect the privacy of subjects</u> and to <u>maintain the confidentiality of data.</u>”</p> <p style="text-align: right;">—45 CFR 46.111(a)(7) <i>(emphasis added)</i></p>

How do IRBs meet the criteria for approving research?

25

“...adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data...”

- Are there adequate provisions? (What are they?)
- Are they documented? (e.g. protocol submission)
- Does the IRB deliberate on this? Who determines what is adequate?
- Are these deliberations or determinations of adequacy documented in the review process?

26

Common Rule: Requirements for informed consent

- Informed Consent Form (ICF)

“...in seeking informed consent the following information shall be provided to each subject:

... A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained; ...”

—45 CFR 46.116(a)(5)

How do IRBs meet these criteria?

27

Informed Consent Form:

- Is there a statement regarding confidentiality of information?
- What does it tell the subject? Is the language consistent with the Authorization (if separate) and the protocol?
- Does anyone monitor the researcher to verify the statement's accuracy throughout the life of the research study?

28

Privacy Rule: Waiver of authorization criteria

➤ Waiver of authorization

1. Use/disclosure involves no more than minimal risk to subject's privacy based on
 - a) plan to protect identifiers
 - b) plan to destroy identifiers
 - c) written assurances PHI will not be reused, disclosed
2. Research could not be conducted without the waiver
3. Research could not be conducted without access/use of PHI

—45 CFR 164.512(i)(2)(ii)(A)

How do IRBs/PBs meet these criteria?

29

Waiver of Authorization:

- Did the researcher submit a plan to protect and destroy identifiers? (Did anyone read the plan?)
- Is the plan consistent with the protocol?
- Does anyone monitor the researcher to verify the plan's accuracy throughout the life of the research study?

30

STRATEGIES, TOOLS FOR REDUCING RISK OF DATA BREACH

31

If you don't
have it,
you can't
lose it!

Researcher and IRB Considerations

32

- What data are essential for the research?
- What is the “minimum necessary”?
- Must it be identifiable?
- Would a Limited Data Set suffice?
- Can Authorization practicably be obtained?

➤ i.e., Waiver is a last resort

De-Identification Criteria

33

A covered entity may determine that health information is not individually identifiable health information only if:

- (1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:
 - (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
 - (ii) Documents the methods and results of the analysis that justify such determination; **OR** [emphasis added]
- (2)(i) The following identifiers of the individual or of relatives, employers, or household members of the individual, are removed ...

—45 CFR §164.514(b)

De-identified data omits all 18 HIPAA identifiers

34

- (1) Names (including initials);
- (2) Street address, city, county, precinct, zip code, and equivalent geo-codes
- (3) ALL elements of dates (except year) for dates directly related to an individual and all ages over 89 (this would include procedure dates, date of admission, date of lab work, etc.)
- (4) Telephone numbers;
- (5) Fax numbers;
- (6) Electronic mail addresses;
- (7) Social security numbers;
- (8) Medical record numbers;
- (9) Health plan ID numbers;
- (10) Account numbers;
- (11) Certificate/license numbers;
- (12) Vehicle identifiers and serial numbers, including license plate numbers;
- (13) Device identifiers/serial numbers;
- (14) Web addresses (URLs);
- (15) Internet IP addresses;
- (16) Biometric identifiers, incl. finger and voice prints;
- (17) Full face photographic images and any comparable images; and
- (18) Any other unique identifying number, characteristic, or code
- AND....

De-identified data elements (cont'd)

35

“...the covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information.”

—45 CFR 164.514(2)(ii)

Limited Data Sets

36

- ✦ Limited to specified data elements
- ✦ Requires Data Use Agreement signed by Covered Entity and researcher
- ✦ Excluded from Accounting for Disclosures requirement

- ✦ Often not practical for clinical drug and device trials because sponsors want more than LDS allows

Limited Data Set?

37

- A) Names;
- (B) Street address, **town or city, county, precinct, zip code, and equivalent geo-codes**
- (C) **All elements of dates (except year) for dates directly related to an individual and all ages over 89**
- (D) Telephone numbers;
- (E) Fax numbers;
- (F) Electronic mail addresses;
- (G) Social security numbers;
- (H) Medical record numbers;
- (I) Health plan ID numbers;
- (J) Account numbers;
- (K) Certificate/license numbers;
- (L) Vehicle identifiers and serial numbers, including license plate numbers;
- (M) Device identifiers/serial numbers;
- (N) Web addresses (URLs);
- (O) Internet IP addresses;
- (P) Biometric identifiers, incl. finger and voice prints;
- (Q) Full face photographic images and any comparable images; and
- (R) **Any other unique identifying number, characteristic, or code.**

Data Flows

38

- How are the data collected?
 - Paper?
 - Verbal?
 - Electronic?
 - How are the data stored and processed?
 - How are the data transmitted or moved?
 - How are the data disposed?
- **In each case: how will we protect the data??**

Greatest Risks

39

- Network-accessible data
- Server with web applications
- Mobile data
 - Laptops
 - Thumb/USB drives
 - Smart phones
 - Desktops
 - Paper records

Security Basics

40

- Patch, patch, patch! (OS and apps and plug-ins)
 - Scan
- Harden (remove/disable unnecessary apps, plug-ins, services, ports, etc.)
 - Scan
- Application security (secure programming, admin passwords, etc.)
 - Scan
- End-point protection (anti-virus, encryption, etc.)
- Firewalls
- Network security

What can the IRB do to reduce risk?

41

- Evaluate PI's plan for securing data (document!!)
 - Develop checklists to identify security parameters for consideration
 - READ and EVALUATE the PI's plan
- Conduct Risk assessment for each project in its proper context
 - Reference: OMB circular

	c.	Describe your plan to protect identifiers in paper and/or electronic format from improper use and/or disclosure by completing the applicable questions below.			
	c.1.	Are you storing PHI in paper form? Yes <input type="checkbox"/> No <input type="checkbox"/> If No, please proceed to 6.c.2.			
		Please describe the permanent location of the paper form.			
		Please describe the security measures that you will put in place for stored data.			
		Will the data be kept in a locked file cabinet?	Yes <input type="checkbox"/>	No <input type="checkbox"/>	
		Will the cabinet be kept in a locked office or store room?	Yes <input type="checkbox"/>	No <input type="checkbox"/>	
		Will the area be a locked/limited access area?	Yes <input type="checkbox"/>	No <input type="checkbox"/>	
		Describe any additional security measures.			
	c.2.	Are you storing PHI on any electronic media (server, desktop computer, laptop, PDA/Smart phone, USB [flash, thumb, jump drive] drive/card, external hard drive, DVD/CD or any other electronic storage media)?			
		If no, proceed to "Attestation of Investigator". If yes, please indicate the number of each type of device(s) utilized. If none, put zero.			
		Device Type	Number	Device Type	Number
		Server		USB drive/card (flash, thumb, jump)	
		Desktop Computer		External hard drive	
		Laptop		DVD/CD	
		PDA/Smart phone		Other electronic storage media:(Describe)	
		Complete Attachment A for each device.			

42

Device Type					
Location and Security of Device					
Device Owner (Institution, Business, Individual)					
If mobile device, name of primary custodian					
Is Device...	Password Protected?	Encrypted?	Temporary Storage?	Permanent Storage?	Back -Up Storage?
YES					
NO					
Back Up / Recovery Plan		<i>[Not required for temporary or back up storage device]</i>			

43

What can the CE do to reduce risk?

44

- Educate researchers and IRBs on data security
- Consider creating dedicated IT resources for researchers
- Monitor and Audit
- Enforce policies

Questions?

45

Marian Hughlett, CHC, CHRC

Deputy Privacy Officer
University of Louisville
marian.hughlett@louisville.edu

Russell Opland, MPH, CIPP, CISM

Systemwide Privacy Officer & HIPAA Privacy and Security Officer
University of California
russell.opland@ucop.edu