

---

**Technical Issues in Data Security**

Scott Bradner  
University Technology Security Officer  
Harvard University

PRIM&R  
April 29, 2011

1

---

---

---

---

---

---

---

---

**Issues in Research Data Security**

---

- ◆ communication
- ◆ mindset
- ◆ communication
- ◆ understanding
- ◆ ‘does this apply to ME?’
- ◆ acceptance
- ◆ data categorization
- ◆ technology
- ◆ communication

2

---

---

---

---

---

---

---

---

**A Process & Technology Example**

---

- ◆ Harvard as a test case
- ◆ security policy at Harvard
- ◆ addressing research data security (or not)
- ◆ finally done
- ◆ now what?

3

---

---

---

---

---

---

---

---

## Harvard

- ◆ general university structure is distributed
  - “cloud education” (maybe ‘quantum education’)
  - informal associations among Schools
  - long history of local management
- ◆ recent (in the context of Harvard) push to change
  - previous president ‘do not accept’ ‘this is the way we have always done it’
  - e.g., forced unified calendar
  - current president continuing to push
    - e.g., new (September) unified CIO for central admin & FAS
    - mostly through design phase of new IT organization
    - covers central + FAS but will offer university-wide services

4

---

---

---

---

---

---

---

---

## Harvard Risk Management

- ◆ new (2010) university risk management structure
- ◆ university risk management team
  - chaired by University Executive Vice President
  - but as of yet, no chief risk officer
- ◆ central Risk Management and Audit Services
  - includes university auditor & insurance office
- ◆ risk management team in each school
  - generally chaired by school Administrative Dean
  - includes all major administrative groups
  - i.e., IT is only a member of team

5

---

---

---

---

---

---

---

---

## Harvard IT Security

- ◆ central policy, local implementation
- ◆ information security policy & compliance process
  - has evolved to now be university-wide
  - University Technology Security Officer (UTSO)
  - Harvard Enterprise Information Security Policy (HEISP)
  - HEISP compliance process
  - Harvard Research Data Security Policy (HRDSP)
- ◆ new CISO named but role still being defined
- ◆ how policy & compliance fits in new IT organization still under development

6

---

---

---

---

---

---

---

---

## Harvard Research

---

- ◆ research oversight is slightly less distributed
  - e.g., 3 Institutional Review Boards
  - fewer than at times in the past
  - e.g., 3 Offices of Sponsored Projects
- ◆ Vice Provost for Research
  - <http://vpr.harvard.edu/>
  - research policy, conflict of interest policy, IPR policy, etc.
  - Chief Research Compliance Officer

7

---

---

---

---

---

---

---

---

## HEISP

---

- ◆ Harvard Enterprise Information Security Policy (HEISP)
  - a set of University-wide policies to protect confidential information
  - annual training, etc
  - annual compliance assessment process
  - checked by Risk Management (Internal Audit) during audits
- ◆ collaboratively developed & updated
  - UTSO & CIOs

8

---

---

---

---

---

---

---

---

## HEISP Information Categories (3)

---

- ◆ High Risk Confidential Information (HRCI)
  - financial identifiers (SSN, credit card, bank account)
  - government identifiers (drivers license, passport)
  - health information & biometric identifiers
  - most also covered by Mass disclosure reporting law
- ◆ other confidential information
  - student & employment information
  - university-designated confidential information
- ◆ non-confidential information

9

---

---

---

---

---

---

---

---

## HEISP, contd.

- ◆ detailed requirements for each type of confidential information  
<http://www.security.harvard.edu/enterprise-security-policy>
- ◆ detailed self assessment worksheet  
<http://www.security.harvard.edu/files/resources/forms/EnterpriseSecurityComplianceWorksheetFinal.xls>
- ◆ annual compliance process uses worksheet & in person visits  
each school & central administration group

10

---

---

---

---

---

---

---

---

## Research Data Policies - #1

- ◆ prodded by Patriot Act requirements - draft policies were developed to protect research data reviewed by IRBs presented at PRIM&R provided to VP for research
- ◆ but...  
draft policies went nowhere  
VP for research left  
no one owned the problem or the solution

11

---

---

---

---

---

---

---

---

## Data Use Agreements

- ◆ researcher received a data use agreement that threatened jail time if the data was not protected
- ◆ resulted in formal process for signing  
use agreement signed by OSP if school CIO says researcher can meet protection requirements even if no money involved
- ◆ note - OGC says that the university must not support a researcher that signs on their own if agreement required signing "for the university"
- ◆ same issue for grants & contracts  
can include stealth security requirements

12

---

---

---

---

---

---

---

---

## Grant & Contract Requirements

- ◆ data protection requirements are appearing in grants and contracts.  
potential increase in FISMA requirements; e.g., research grants with VA data require FISMA
- ◆ researchers and Sponsored Projects groups must be warned to look for these requirements; it is unlikely that researchers will notice however...  
requirements are binding even if they were not noticed

13

---

---

---

---

---

---

---

---

## Data Use/Protection Requirements

- ◆ becoming quite common to get 3rd party data and in grants and contracts  
not just in government g&c
- ◆ can include very specific requirements
- ◆ can just say 'protect the data'
- ◆ potentially significant penalties for non-compliance  
e.g., can be required to return already spent grant money and in a few cases, criminal charges

14

---

---

---

---

---

---

---

---

## FISMA

- ◆ Federal Information Security Management Act  
mixed view of effectiveness
- ◆ some push in federal agencies to include FISMA security requirements in grants & contracts  
grant agent may add requirement w/o understanding
- ◆ 3-level requirements  
low - could be met by well run university data centers with some effort  
moderate - possible to be met by well run university data centers with a lot of effort & expense  
high - unlikely to ever be met by a university data center

15

---

---

---

---

---

---

---

---

## FISMA, contd.

- ◆ NIST 800-53rev3, July 2009, errata to June 2010
- ◆ 237 page document
- ◆ more than 200 requirements
  - high level generally requires automated mechanisms to meet requirements
  - moderate level sometimes requires automated mechanisms to meet requirements
- ◆ too often “FISMA” is required w/o classifying data
- ◆ some pushback against FISMA application
- ◆ new guidance document - NIST 800-39

16

---

---

---

---

---

---

---

---

## Other Data Protection Requirements

- ◆ most states also have data protection requirements
  - e.g., Mass 201 CMR 17
- ◆ federal requirements for medical & student records (HIPPA, FERPA)
  - e.g., Mass Gen agreed to pay a \$1M penalty for misplacing medical records concerning 192 people
- ◆ VA requires FISMA protections
  - university researcher locked out of research lab for failure to meet FISMA requirements
- ◆ local penalties can be harsh
  - UNC researcher demoted & pay cut after breach
  - <http://chronicle.com/article/Chapel-Hill-Researcher-Fights/124821>

17

---

---

---

---

---

---

---

---

## Research Data Policies - #2

- ◆ this time process driven by chair of Social Science Committee, Provost and new Vice Provost for Research
  - policy “owned” by VP for Research
- ◆ draft reviewed by IRBs, School CIOs, OGC, Social Science Committee, Provost, University Joint Committees on Inspection, ...
- ◆ multi-year process
- ◆ (finally) approved October 2010
  - <http://www.security.harvard.edu/research-data-security-policy>

18

---

---

---

---

---

---

---

---

## HRDSP, Sections

---

- ◆ Introduction
- ◆ Research Information from Non-Harvard Sources
- ◆ Research Information from Harvard Sources
- ◆ Information Security Categories
- ◆ Legal Requests for Research Information

19

---

---

---

---

---

---

---

---

## Introduction

---

- ◆ responsibilities: investigators:
  - disclose nature of data
  - prepare data security plans & procedures
  - implement plans & procedures
- ◆ responsibilities: IRB
  - ensure adequacy of investigators plans & procedures
- ◆ responsibilities: IT
  - assist investigators in determining proper levels
  - assist investigators in implementing security

20

---

---

---

---

---

---

---

---

## Data From Non-Harvard Sources

---

- ◆ if data has a use agreement (DUA)
  - protection must meet requirements in DUA agreement
    - note: researchers can not sign DUAs for the University - OSP is the designated signer (even if no money involved)
  - IRB can determine that DUA requirements are too weak
    - if so, treat as if data is from a Harvard source
- ◆ if research done in non-Harvard facility
  - facility owner may define protection requirements
- ◆ otherwise
  - treat as if data is from a Harvard source

21

---

---

---

---

---

---

---

---

## Data From Harvard Source

- ◆ human subjects research
  - research must be reviewed by a IRB
  - information used in research must be protected against inadvertent or inappropriate disclosure
  - IRB will confirm security level categorization
- ◆ other sensitive research
  - e.g. research with national security implications
  - researchers should work with school IT groups to determine data categories

22

---

---

---

---

---

---

---

---

## Data Categories

- ◆ five research data Levels were created by augmenting the HEISP.
  - Level 5 - extremely sensitive information about individually identifiable people
  - Level 4 - very sensitive information about individually identifiable people (same as HEISP HRCI)
  - Level 3 - sensitive information about individually identifiable people (same as HEISP other confidential information)
  - Level 2 - benign information about individually identifiable people
  - Level 1 - de-identified research information about people and other non-confidential research information

23

---

---

---

---

---

---

---

---

## Why 5 Levels?

- ◆ started with HEISP - 3 levels
  - high risk confidential information (level 4)
  - other confidential information (level 3)
  - non-confidential information (level 1)
- ◆ added level 5
  - because non-network connected requirement is in some use agreements and is the right thing for some data
- ◆ added level 2
  - pragmatic - researchers are not willing to be significantly inconvenienced just to protect information they do not see as risky

24

---

---

---

---

---

---

---

---

## De-Identification Key

- ◆ key for coded de-identified research information must be protected at the level that would have been applicable to the non-de-identified data
- ◆ what constitutes de-identification is not addressed in policy

25

---

---

---

---

---

---

---

---

## Level 5

- ◆ description:  
*Disclosure of Level 5 information could cause significant harm to an individual if exposed, including, but not limited to, serious risk of criminal liability, serious psychological harm or other significant injury, loss of insurability or employability, or significant social harm to an individual or group*
- ◆ examples  
currently mostly requirement from data use agreements  
raw census data, some mental health records

26

---

---

---

---

---

---

---

---

## Level 5: Protections

- ◆ stored in physically secure rooms in university space
  - not on janitor's key or building master key
  - need accessible fireman's key
- ◆ computers must not be connected to a network that extends outside the room

27

---

---

---

---

---

---

---

---

## Level 4

### ◆ description

*Disclosure of Level 4 information could reasonably be expected to present a non-minimal risk of civil liability, moderate psychological harm, or material social harm to individuals or groups.*

### ◆ examples

HEISP high risk confidential information (HRCI)  
e.g., subject's SSNs  
medical research records  
information with national security implications

28

---

---

---

---

---

---

---

---

## Level 4: Protections

- ◆ do not store on user computers or devices even if encrypted (too much risk of error)
- ◆ servers in physically secure Harvard environments  
card based access best - create access log
- ◆ local network-based firewalls
- ◆ access limited to IRB approved individuals
- ◆ media must be encrypted or stored in a locked safe
- ◆ separate networks using private addressing
- ◆ regular vulnerability testing
- ◆ backup tapes must be encrypted

29

---

---

---

---

---

---

---

---

## Level 3

### ◆ description

*Disclosure of Level 3 information would could reasonably be expected to be damaging to a person's reputation or to cause embarrassment.*

### ◆ examples

most non-de-identified human research information  
student record information (FERPA)  
some commercial data  
employment records

30

---

---

---

---

---

---

---

---

### Level 3: Protections

---

- ◆ encrypt laptops and portable devices
- ◆ use automatic patching
- ◆ virus protection
- ◆ encrypt all transfer over networks and on portable media
- ◆ limit access to those doing the research
- ◆ host-based firewalls
- ◆ lock up all non-electronic records

31

---

---

---

---

---

---

---

---

### Level 2

---

- ◆ description  
*Disclosure of Level 2 information would not ordinarily be expected to result in material harm, but as to which a subject has been promised confidentiality.*  
called “minimal risk” information under the common rule
- ◆ examples  
data from reaction time experiments  
customer satisfaction survey data

32

---

---

---

---

---

---

---

---

### Level 2: Protections

---

- ◆ good computer hygiene  
secret complex passwords  
not shared accounts  
regular patching  
avoid dangerous web sites  
don't respond to phishing

33

---

---

---

---

---

---

---

---

## Level 1

### ◆ description

de-identified research information about people and other non-confidential research information

### ◆ examples

de-identified research information

but might be private until publication

student directory information

except for students with 'FERPA blocks'

research information where no anonymity promised

34

---

---

---

---

---

---

---

---

## Legal Requests for Research Info.

◆ forward any legal request of information (e.g., a subpoena, national security request or court order demanding disclosure of information in researcher possession) to OGC

◆ researchers not authorized to provide the information

◆ consider obtaining a Certificate of Confidentiality  
allow refusal to disclose

35

---

---

---

---

---

---

---

---

## Other Information

◆ policies include specific guidance on how to do data collection in the field for each level data

◆ web site also includes:

requirements when working with vendors

process for responding to Freedom of Information Act (FOIA) requests (send to OGC)

classified work (can not do)

advice for travelers

<http://www.security.harvard.edu/advisory-travelers>

rules concerning paying subjects (i.e., tax requirements)

36

---

---

---

---

---

---

---

---

## How Much Detail?

- ◆ 1st version gave general directions
  - e.g., treat as HEISP Level 4
- ◆ pushback from Joint Committee on Inspection
  - wanted self contained requirements that could be audited
- ◆ now getting pushback that the requirements are blocking research
  - making things too hard
  - want “risk-based approach”
- ◆ going to be a common conflict
  - need to be detailed to meet detailed regulations, but too much detail is ‘too hard to meet’

37

---

---

---

---

---

---

---

---

## Implementation

- ◆ specific protection requirements for each level
  - existing HEISP level protection requirements well understood
  - Levels 5 and 2 will take some work
    - special facilities for Level 5
    - researcher cooperation for Level 2
- ◆ communications to researchers
  - annually by Deans
  - day-to-day by IRBs
    - a better path than for administrative information security
    - IRBs have created new forms
- ◆ enforcement is an open question

38

---

---

---

---

---

---

---

---

## Another Issue

- ◆ federal regulations require that the university “immediately” produce data from federally funded research
  - e.g., in case accusation of research fraud
- ◆ can be a problem if researcher runs their own systems or uses non-university resources
  - can you say “cloud computing”?
- ◆ not addressed in HRDSP

39

---

---

---

---

---

---

---

---

## Remaining Issues

---

- ◆ communication
- ◆ mindset
- ◆ communication
- ◆ understanding
- ◆ 'does this apply to me?'
- ◆ acceptance
- ◆ categorization of actual data
- ◆ technology
- ◆ communication

40

---

---

---

---

---

---

---

---

questions?

41

---

---

---

---

---

---

---

---