# Challenges of Research Data Security

Scott Bradner

University Technology Security Officer
Harvard University
April 14, 2010

Security EDUCAUSE & Internet2 Security Professionals Conference

see resume on www.sobco.com for handouts

1

# Agenda

- context
- rules
  - laws & regulations
  - data use agreements
- what Harvard is doing
  - HEISP
  - HRDSP

Security EDUCAUSE & Internet2 Security Professionals Conference

2

# Context

- you need to (because you have to or because you should) protect research information
- researchers are not (generally) information security experts
- IRB members are not (generally) information security experts



# Human Subject Research

- research with human subjects
- research involving information about individual humans
- covers research that you might not expect
  - web-based surveys
  - data network usage
  - external data sets
  - etc

# Other Research

- information security may be required even if no people are identifiable
  - licenses that require confidentiality
  - research that will be patented
  - research with national security issues



# IRB

- Institutional Review Board
  - a.k.a. independent ethics committee
  - a.k.a. ethical review board
- mandated by federal law (Title 45 CFR Part 46) to review and approve human subject research
- some specific exceptions to approval requirement
  - e.g., research on the effectiveness of instructional techniques

# IRB, contd.

- makeup
  - at least 5 people
  - varying backgrounds & professions
  - at least one "community member"
  - at least one non-scientist
  - men & women
  - representative of any vulnerable population subject to research reviewed by the IRB (can be ad hoc member)
  - no requirement to include an IT or security expert



# IRB, contd.

- all human subject researchers should know about the IRB & get their research reviews & approved by the IRB
- human subject research generally requires annual reports
- thus, human subject research is about the only area in a University where there is a reliable information conduit to individuals who need to know & follow security rules

# Classifying Information

- a general information classification
  - not just for research information
  - 1/ information that requires notification in case of a breach
    - e.g., SSNs, student record information
  - 2/ other confidential information
  - 3/ non-confidential information



# Another Classification

- 1/ information subject to specific externally defined protection requirements
  - e.g., medical records in a HIPAA covered entity, SSNs in Massachusetts
- 2/ other information that must be secured
  - e.g., SSNs not in Massachusetts

# Rules: Laws & Regulations, 1

- state breach notification laws (45)
  - cover name + financial identifiers of state residents
    - California also covers medical records
  - note that the definition of a resident is a legal one, not just a postal address
  - require disclosure of a "breach"
    - definition of "breach" varies
      - most exempt encrypted information (w/o key)
  - some disclosure requirements contradictory



# Rules: Laws & Regulations, 2

- federal breach notification laws & regulations
  - no federal breach notification laws for financial information
    - but multiple are in process
    - likely to override state breach notification laws if passed
  - multiple federal breach notification laws & regulations for HIPAA & student record information

# Rules: Laws & Regulations, 3

- state protection laws & regulations
  - e.g. Massachusetts 201 CMR 17
  - specific protection requirements
- California information privacy laws
  - > 80 laws (right to privacy in CA constitution)



# Rules: Laws & Regulations, 4

- federal human subject laws & regulations
  - 45 CFR 46, 21 CFR 50, 21 CFR 56, FDA E6 good clinical practice (GCP) guidance
  - require Institutional Review Boards (IRBs)
    - require human subject research to be reviewed by IRB
  - require protection of identifiable research information
    - but no specific rules

# Rules: Laws & Regulations, 5

- Health Insurance Portability and Accountability Act (HIPAA)
  - directly applies to "covered entities"
    - health care providers, health plan, health care clearinghouse
  - deals with Protected Health Information (PHI)
    - medical record & payment history
- Privacy Rule
  - regulates & controls use and distribution of PHI
- Security Rule
  - detailed specific security requirements
- Transactions & Code Set and Enforcement Rules



# Rules: Laws & Regulations, 5b

- HIPAA Security Rule
  - administrative safeguards
    - written security policy, rules on who can access PHI, training, dealing with vendors, etc
  - physical safeguards
    - secure facility, PHI access controlled & monitored, etc
  - technical safeguards
    - encryption on open networks, partner authentication, data integrity, etc

# Rules: Laws & Regulations, 5c

- HIPAA term: "deidentified health information"
  - FERPA regulations point to HIPAA definition
  - remove personal identifiers: e.g., name, ssn, geo tag smaller than a state, all date other than a year, phone #s, email addresses, medical record #s, license #s, device and vehicle IDs & serial #s, IP addresses, biometric IDs (finger & eye print), photos of face
- deidentified health information is not confidential



# Rules: Laws & Regulations, 5d

- HIPAA term: "limited data set"
  - remove direct identifiers: e.g., name, SSN, address, license #, device and vehicle IDs & serial #s, IP addreses, medical record #s, biometric IDs (finger & eye print), photos of face
  - can include full date of birth, zip code, sex
    - (note: can re-identify most people using this info)
  - still confidential but less so than raw medical records

# Rules: Laws & Regulations, 6

- Health Information Technology for Economic and Clinical Health Act (HITECH Act)
- extend HIPAA security & privacy requirements to business associates
- added breach notification requirements
  - detailed in rules by Dept. of Health & Human Services and the Federal Trade Commission
- HHS also published guidance for making "unsecured PHI" unusable, unreadable or indecipherable
  - including de-identification, encryption & destruction



# Rules: Laws & Regulations, 7

- Family Education Rights and Privacy Act (FERPA) & regulations
  - student can control distribution of student record information
    - multiple exceptions, including for health or safety & de-identified information
- Genetic Information Nondiscrimination Act (GINA) & regulations
  - defines that genetic information is covered medical information under HIPAA

# Rules: Laws & Regulations, 8

- Federal Information Security Management Act (FISMA)
- law requiring NIST to define security requirements for government agencies
- requirements: NIST 800-53 rev3
- starting to show up in grants, contracts & data use agreements from US government agencies
- 3-levels of risk
  low, moderate & high
- many rules



# Rules: Laws & Regulations, 8b

**NIST 800-53 Rev. 3**

| control group | total | low | mod | high |
|---|---|---|---|---|
| access control | 20 | 11 | 16 | 18 |
| awareness and training | 5 | 4 | 4 | 4 |
| audit and accountability | 14 | 10 | 11 | 10 |
| security assessment & authorization | 6 | 6 | 6 | 6 |
| configuration management | 9 | 6 | 9 | 9 |
| contingency planning | 9 | 6 | 9 | 9 |
| identification & authentication | 8 | 7 | 8 | 8 |
| incident response | 8 | 7 | 8 | 8 |
| maintence | 6 | 4 | 6 | 6 |
| media protection | 6 | 3 | 6 | 6 |
| physical & environmental protection | 19 | 11 | 18 | 18 |
| planning | 5 | 4 | 5 | 5 |
| personnel security | 8 | 8 | 8 | 8 |
| risk assessment | 4 | 4 | 4 | 4 |
| system & services acquisition | 14 | 8 | 11 | 13 |
| systems & communications protection | 34 | 8 | 20 | 23 |
| systems & information integrity | 13 | 5 | 11 | 12 |
| totals | 188 | 112 | 160 | 167 |

# Rules: Laws & Regulations, 8c

- FISMA low could be met in a university data center with some work
  - e.g., 2 automated tools, lots of process, < $1M
- FISMA moderate could be met in a university data center with a lot of work
  - e.g., 5 automated tools, more process, $1-3M
- FISMA high, forget it
  - e.g., 14 automated tools, endless process



# Rules: Laws & Regulations, 8d

- FISMA issues
  - incomplete specification of requirement:
    - requirements in grants, contracts & data use agreements do not specify what level is required -- just say "must meet FISMA"
    - can self categorize with NIST 800-60 but the agency may not agree with categorization
  - audit:
    - may require formal certification and accreditation
      - see NIST 800-37

# Rules: Laws & Regulations, 9

- Children's Online Privacy Protection Act (COPPA)
  - websites dealing with children under 13 must comply with COPPA
    - must have accurate privacy statement with specific content
    - must get verifiable parental consent to collect info from child
  - fines up to $1M have been issued for violations
- Freedom of Information Act (FOIA)
  - data from federally funded research projects can be requested under FOIA
    - request must come via the granting agency



# Rules: Laws & Regulations, 10

- Federal Trade Commission Act section 5
  - FTC considers inaccurate privacy statements an unfair business practice
- electronic signatures
  - e.g., for subject's agreement to terms & conditions
  - Uniform Electronic Transactions Act (UETA) (state based)
  - Electronic Signatures in Global and National Commerce Act (eSIGN)

# Data Use Agreements

- a DUA comprises terms & conditions placed on information ...
  - received from an external source
  - developed under grant or contract
- a DUA can require protecting the confidentiality of information
- a DUA can require protecting the economic value of information



# DUA, contd.

- wide range of requirements
- types:
  - requesting
  - cursory
  - referential
  - detailed

# DUA, Example 1

- requesting - tell us what you will do and we will say if that is OK

*How will you maintain the confidentiality of the data obtained? Include an explanation of how and where such data will be stored as well as how and when you plan to dispose of the data after your study is completed. Also describe the safeguards that exist (or will be implemented) to ensure that the data will be used solely for the purpose of this research project.*



# DUA, Example 2

- cursory - just do the right things

*Applicants, contractors, or sub-contractors handling PHC4 data shall use appropriate safeguards to prevent use or disclosure of data other than as permitted by this agreement.*

# DUA, Example 3

- referential - follow these rules

*The User agrees to establish appropriate administrative, technical, and physical safeguards to protect the confidentiality of the data and to prevent unauthorized use or access to it. The safeguards shall provide a level and scope of security that is not less than the level and scope of security requirements established by the Office of Management and Budget (OMB) in **OMB Circular No. A-130, Appendix III**--Security of Federal Automated Information Systems (http://www.whitehouse.gov/omb/circulars/a130/a130.html) as well as **Federal Information***



# DUA, Example 3b

***Processing Standard 200** entitled "Minimum Security Requirements for Federal Information and Information Systems" (http://csrc.nist.gov/publications/fips/fips200/FIPS-200-final-march.pdf); and, **Special Publication 800-53** "Recommended Security Controls for Federal Information Systems" (http://csrc.nist.gov/publications/nistpubs/800-53-Rev2/sp800-53-rev2-final.pdf). The User acknowledges that the use of unsecured telecommunications, including the Internet, to transmit individually identifiable,*

# DUA, Example 3c

*bidder identifiable or deducible information derived from the file(s) specified in section 5 is prohibited. Further, the User agrees that the data must not be physically moved, transmitted or disclosed in any way from or by the site indicated in section 17 without written approval from XXXXXX unless such movement, transmission or disclosure is required by a law.*



# DUA, Example 3, Details

- "*security that is not less than the level and scope*" of the requirements in the following:
  - OMB Circular No. A-130, Appendix III
    - 85 pages (at 12 point type), 28,900 words
    - covers need to use Internet to disseminate info, required reports, adequate security, international travel, and many other topics
  - Federal Information Processing Standard 200
    - 17 pages
    - introduction to NIST Special Publication 800-53
  - NIST Special Publication 800-53
    - FISMA

# DUA, Example 4

- detailed requirements

  *The information may only be processed on a isolated computer in a secure facility. Any wireless networking must be disabled and the Ethernet cable must be removed from the computer when the data is present on the computer.* (paraphrased)

  2 additional pages of requirements



# DUAs, Penalties

- DUAs can include specific penalties for failure to meet the requirements in the DUA - e.g.:

  required destruction of information

  fines

  return of grant money

  criminal charges (including jail time)

# DUA, Legal Issue

- many DUAs require a signature "for the university"
  - signing officer may be personally subject to penalties if DUA requirements not met
  - in almost all cases researchers are not authorized to sign for the university
    - but many do
- signing officer needs a way to know that the researcher will meet whatever requirements there are in the DUA



# Signing DUAs at Harvard

- the Office of Sponsored Projects (OSP) is authorized to sign for the university
  - even if no money involved
- OSP gets signoff from school CIO and IRB before signing
  - signoff is that all security requirements are understood and the researcher can meet them
  - description of process & forms on security site

process: resources -> forms -> Office for Sponsored Programs Research Data Protection - Process

form: resources -> forms -> Office for Sponsored Programs Research Data Protection - Form

# Harvard OSP Process

- researcher fills out request form
  - attach any security requirements
- send form to school CIO and to IRB
- CIO works with researcher & IRB to be sure required protections are understood and are in place
- CIO forwards form to OSP if CIO is satisfied
  - OSP can then sign for the university
- IRB can ask researcher for copy of CIO OK
  - for IRB records

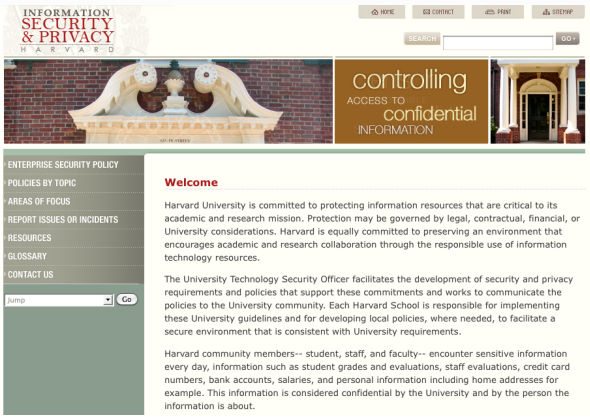


# Stealth Requirements

- too often the researcher signs a DUA on their own & does not tell IT
  - legal problem if something goes wrong - university may not legally be able to support researcher
- bigger issue with data security requirements in grants and contracts
  - too often not noticed by anyone
    - but still binding
  - need to make grant/contract review for requirements a part of acceptance process

# Harvard Security Web Site

- www.security.harvard.edu
  - also security.harvard.edu & privacy.harvard.edu






# HEISP

- Harvard Enterprise Security Policy
- applies to all confidential information at Harvard
  - to date, focused on administrative information
  - cover here because HEISP provides the environment for the research data security policy
- annual assessment process
  - use self assessment questionnaire (also used by internal audit)

# HEISP Sections

1. High Risk Confidential Information (HRCI)
2. Confidential Information
3. Student Information
4. Credit Card Information
5. Building Access & Physical Environment
6. Working With Vendors
7. Computers & Servers
8. Other IT Policies
9. Federal & Regulatory
10. Web Based Surveys



# HEISP Section 1

- Financial High-Risk Confidential Information
  not store on a user computer
  secure paper records
  get written permission to use HRCI
- Human Subject Information
  get IRB approval, including for data security
  updated by HRDSP
- Personally Identifiable Medical Information
  follow HIPAA if a covered entity
  else treat as HRCI

# HRCI

*High-Risk Confidential Information includes a person's name (or other identifier) in conjunction with the person's Social Security, credit or debit card, individual financial account, driver's license, state ID, or passport number, or a name in conjunction with biometric information about the named individual. High-risk confidential information also includes some human subject information and much personally identifiable medical information*



# Permission to Use HRCI

- formal process to request access to HRCI
- signoff by school or university CIO
  - after review by school or university security officer
- must justify
  - the need for the specific HRCI
  - who will have access to the HRCI
  - length of retention of HRCI
- must show compliance with HEISP
- aim: reduce the use & locations of HRCI

# Rest of HEISP

- specific rules on information security
- general rules for confidential information
- additional rules for HRCI



# HEISP as Environment

- school IT groups understand & (mostly) meet requirements in HEISP
- thus, systems and processes are in place
- so they are better able to assist researchers in protecting research information
- supporting researchers is now a requirement for school IT groups
  - does not mean that the researchers know who the school IT people are

# HRDSP

- Harvard Research Data Security Policy
- widely reviewed draft policy
  - IRBs, OGC, Social Science Committee, OSP, Provost, CIOs (school & university), VP Research, researchers,...
  - Provost will take final version to University "Joint Committee on Inspection" soon
- "owner": Vice Provost for Research



# HRDSP, Sections

- sections:
  - Research Information from Non-Harvard Sources
  - Research Information from Harvard Sources
  - Information Security Categories
  - Legal Requests for Research Information
- includes specific protection requirements

# Data From Non-Harvard Sources

- if data has a use agreement
  - protection must meet requirements in use agreement
- if research done in a non-Harvard facility
  - facility owner may define data protection requirements
- otherwise
  - treat as if data is from a Harvard source



# Data From Harvard Source

- human subjects research
  - research must be reviewed by a IRB
  - research proposals must include "acceptable, effective, and documented procedure" to protect personally identifiable research information
  - researchers should work with IRBs to determine data categories
- other sensitive research
  - e.g. research with national security implications
  - researchers should work with school CIOs to determine data categories

# Information Security Categories

level 5 - extremely sensitive information about individually identifiable people

level 4 - very sensitive information about individually identifiable people

level 3 - sensitive information about individually identifiable people

level 2 - benign information about individually identifiable people

level 1 - de-identified research information about people and other non-confidential research information



# Why 5 Levels?

- started with HEISP - 3 levels
  - high risk confidential information (HRDSP level 4)
  - other confidential information (HRDSP level 3)
  - non-confidential information (HRDSP level 1)
- added level 5
  - because non-network-connected requirement is in some use agreements and some research data deserves this level of protection
- added level 2
  - to cover truly minimal risk information

# Bright Lines?

- unless directed by a data use agreement, or, for example, a identity theft law, categorizing research data will be subjective
- IRB categorizing skills will evolve over time
- there will be tussles with researchers



# Protections

- 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 this policy
    - must be determined by IRB - changing understanding

# Level 5

*Level 5 information includes individually identifiable information that 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.*



# Level 4

*Level 4 information includes individually identifiable information that includes High Risk Confidential Information (HRCI) as defined by the Harvard Enterprise Information Security Policy. This includes Social Security numbers as well as other individually identifiable financial information. Medical records that are not categorized as extremely sensitive and other individually identifiable research information that, if disclosed, could reasonably be expected to present a non-minimal risk of civil liability, moderate psychological harm, or material social harm to individuals or groups should also be classified as Level 4 information.*

# Level 3

*Level 3 information includes individually identifiable information that, if disclosed, could reasonably be expected to be damaging to a person's reputation or to cause embarrassment. Student record information protected by FERPA also generally falls under Level 3.*



# Level 2

*Level 2 information includes individually identifiable information, disclosure of which would not ordinarily be expected to result in material harm, but as to which a subject has been promised confidentiality.*

# Level 1

*Research information in which all information that could be used, directly or indirectly, to identify an individual has been removed or modified is referred to as "de-identified research information." There are no specific University requirements for the protection of de-identified research information or for other non-confidential research information, but researchers may want to protect such data for their own reasons, i.e., keeping data private until a paper about the data is published.*



# Protection Requirements

- detailed protections requirements for each level information
  designed to be auditable
- protections for levels 2-4 taken directly from HEISP
- protections for level 5 includes relevant protections for level 4, 3 & 2 plus protections relating to requirement for no network connectivity

# Exceptions

*Except where there are legal protection requirements, the IRB or the UTSO in consultation with the IRB, have the authority to approve a variance of the following security requirements, in consultation with appropriate Harvard technical experts (such as the School CIO or Security Officer), if the requirements would otherwise inappropriately affect the conduct of the research and if alternate methods will still provide adequate protection of confidential information.*



# Data Collection

- processes & protections provided for field collection of information at each level
- general requirement - get the data off the data collection device ASAP
  - using a secure transfer process or device
    - e.g., VPN or encrypted thumb drive

# Legal Requests for Research Info.

- if researcher receives a subpoena, national security request or court order forward it to the OGC
  - researcher not authorized to respond
- if researcher receives a FOIA request notify OSP
  - researcher not authorized to respond
- consider obtaining a Certificate of Confidentiality
  - IRB can help



# Open Issues

- training IRBs, researchers & IT staff
- predictable categorizations
- getting researcher compliance
  - or even knowledge of signing requirements
  - or getting out of "it's IT's problem" mode
- resources - e.g. for device encryption
- getting non-IT buy in & ownership
- . . .