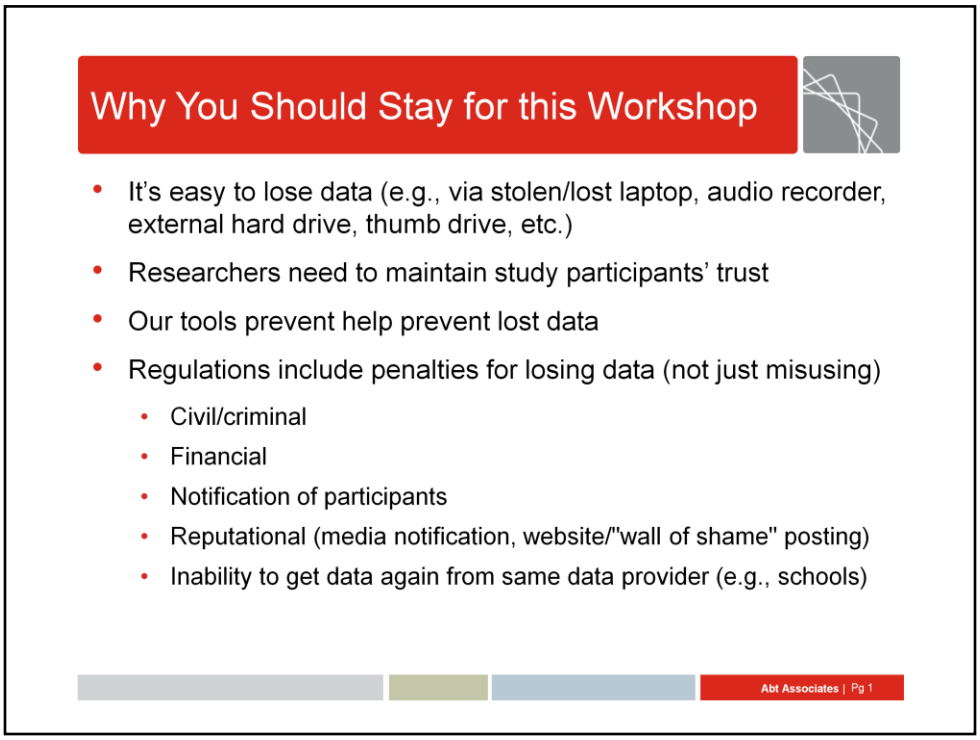
How to Protect Sensitive Data: Tools for Researchers and Programs

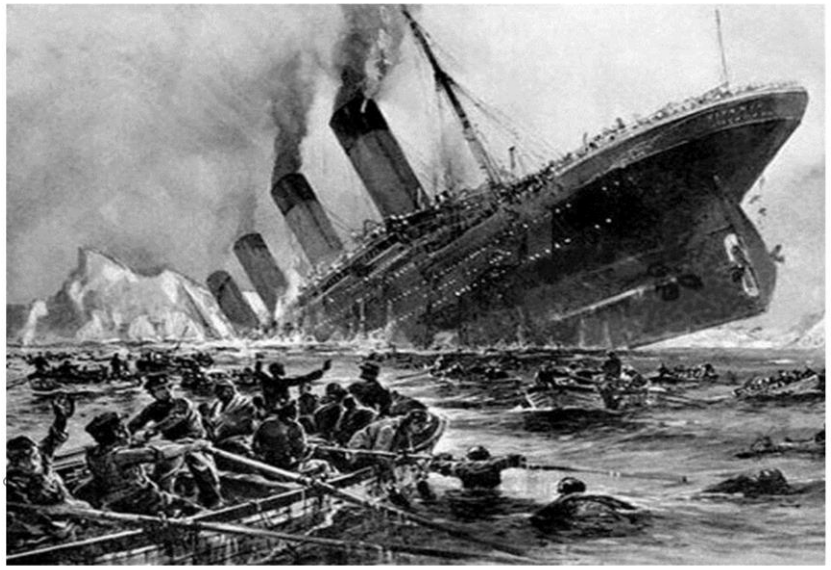
APPAM

Sean Owen, CISSP, CAP, CRISC
Teresa Doksum, PhD, MPH

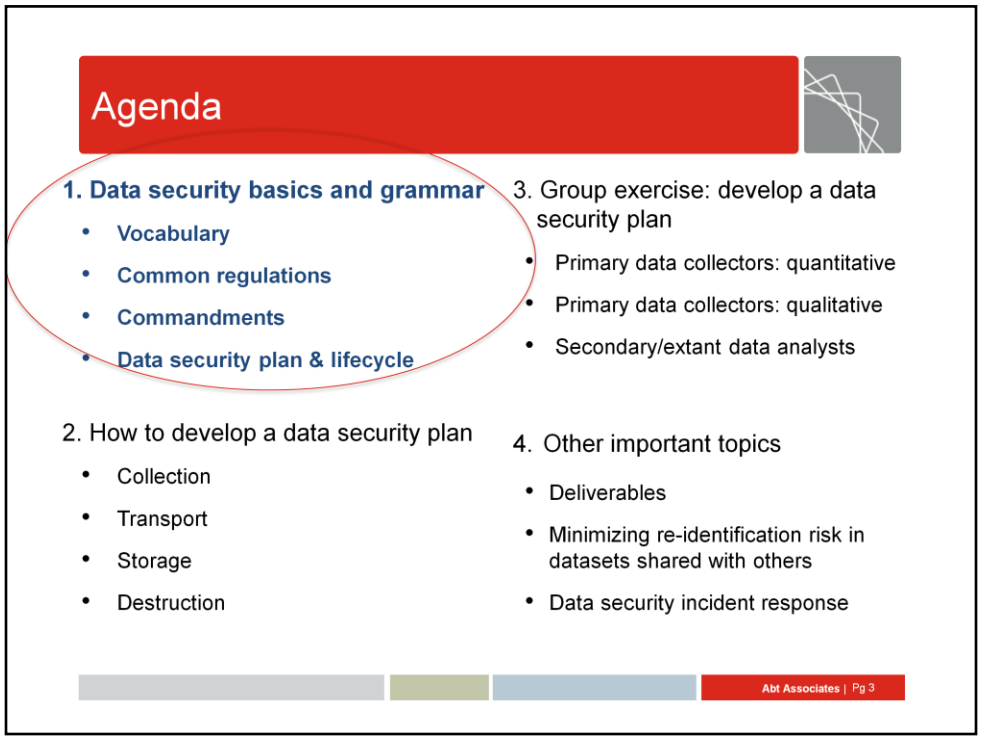
November 6, 2014



Handling any data introduces the possibility that data will be lost, misplaced, or stolen. It can happen to the best of us. There seems to be a new report every week about a well-respected hospital, university, or company that had something go wrong. These events have a detrimental effect on all of us because we need information from trusting participants to conduct research and make the world a better place. Aside from the loss of trust, there are other, more direct implications like fines.

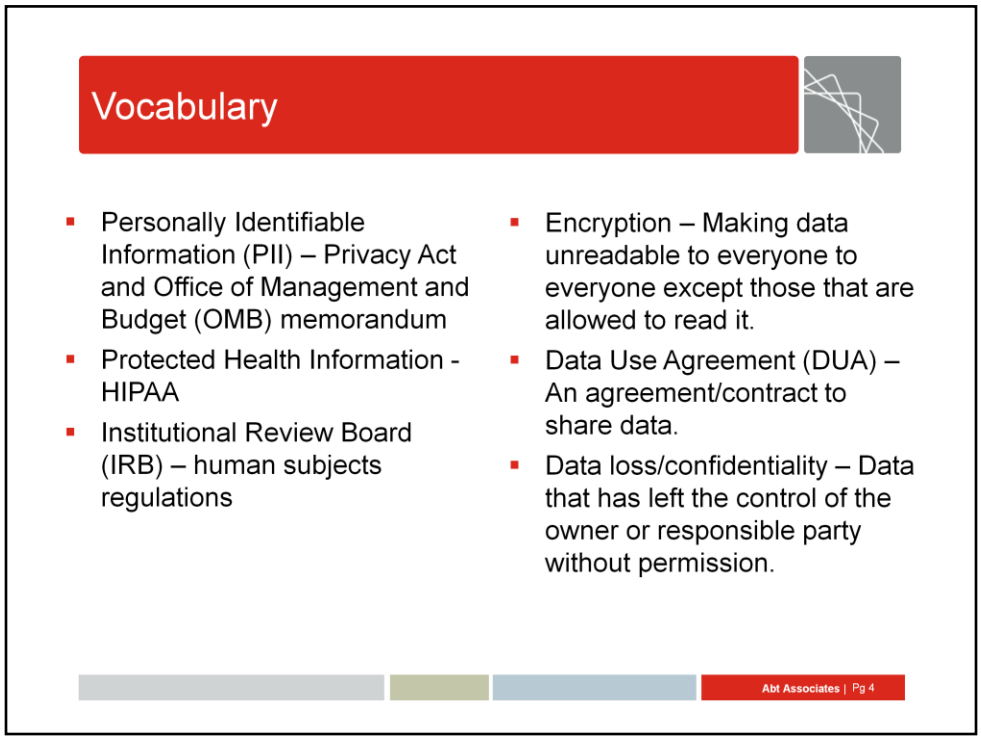


- Based on our experience, without a detailed plan with basic security measures, the likelihood of something going wrong is higher. We recently had a study partner call us in a panic because they are researching a potential problem with their data that they shared with us. They needed to know who had what data and what the current status of that data are. Within 1 hour, we pulled out the plan we worked with the team on and knew the answer to every question. The study partner is now in a much better position to avoid fines.



There are lots of topics we could cover;

Feel free to interrupt and ask questions. Let's start with some basics so we are all speaking the same language.



Encryption – Making data unreadable to everyone to everyone except those that are allowed to read it.

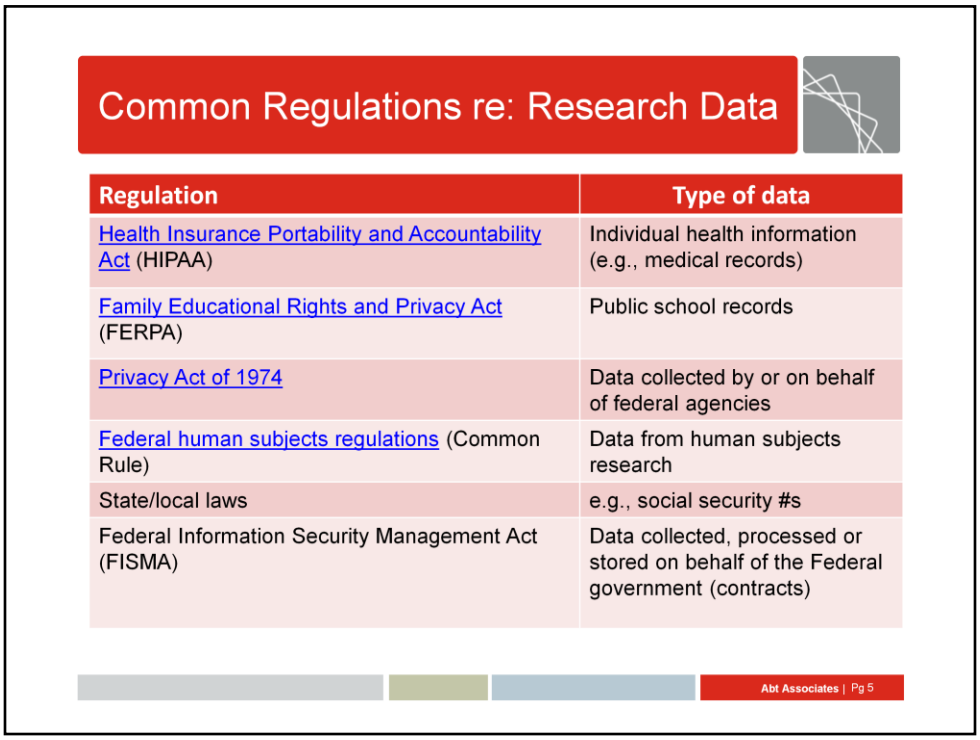
PII – personally identifiable information is any piece of data that individually or in combination with other data could be used to identify an individual

PHI – Information regarding an individual's health. See HIPAA for a complete definition.

DUA – DUA's are used by two parties to share data. They often include security requirements, allowed uses, and destruction requirements.

Data loss/confidentiality – Data loss is a common type of incident. This is often the most grevious that can happen to our community. The steepest penalties from these types of incidents. HIPAA, Federal reqs, and states all have varying (but similar enough) definitions of an incident.

Institutional Review Board – IRB's protect human subjects, often a source of data for us, and ensure that protections have been put in place. The requirements around data security for human subjects are not described in detail in the Common Rule enforced by OHRP.



- Bottom line—the data you use for your research may be subject to several of these regulations.
- Our IRB requires researchers to create a detailed data security plan (see the handout—it includes a template to help researchers protect data using procedures that comply with the requirements of the most common regulations.

Useful websites:

HIPAA: http://www.hhs.gov/ocr/privacy/hipaa/understanding/special/research/index.html

FERPA: http://www2.ed.gov/policy/gen/guid/fpco/index.html

Privacy Act: https://www.federalregister.gov/articles/2013/09/11/2013-22072/privacy-act-of-1974-systems-of-records

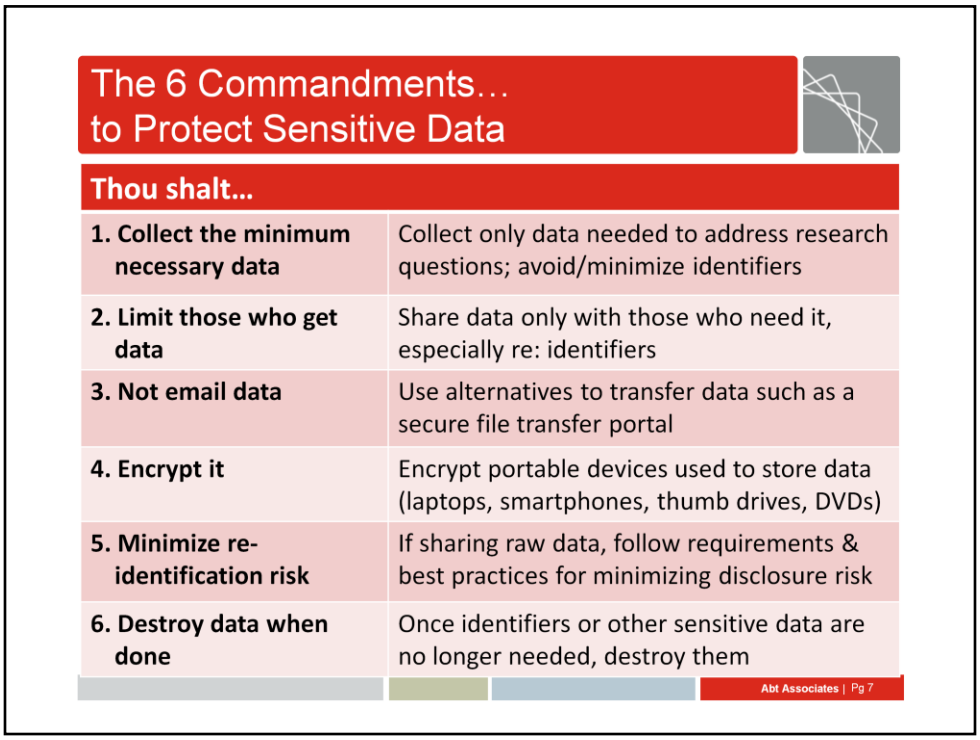
Common Rule: http://www.hhs.gov/ohrp/humansubjects/commonrule/index.html

FISMA: www.nist.gov

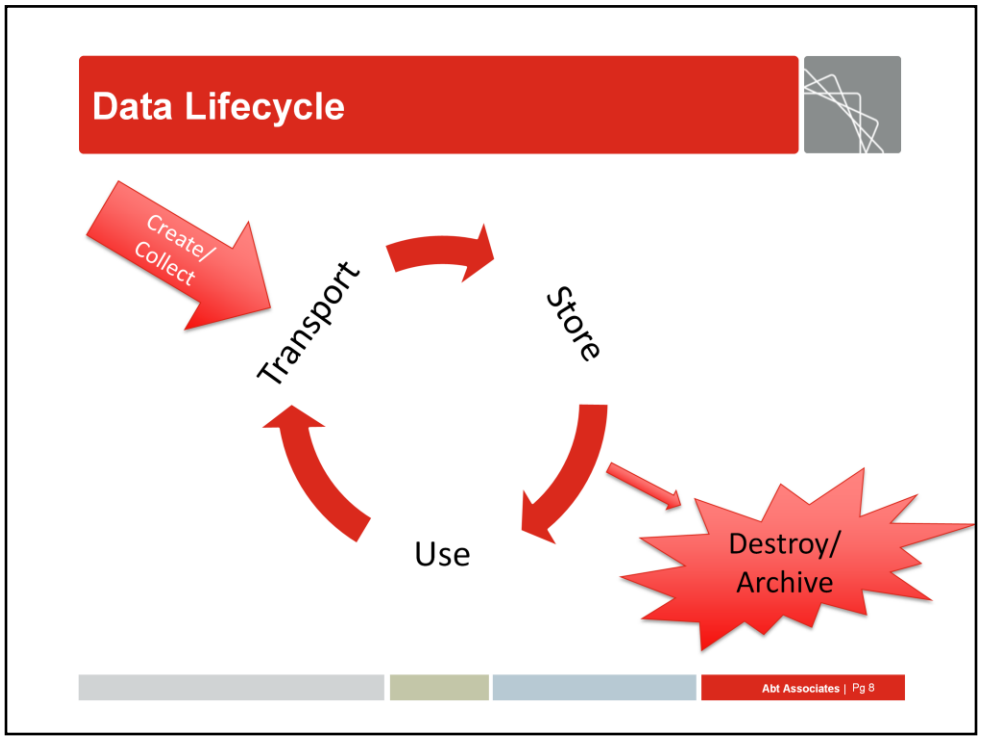
The Data Security Commandments

We had 9 Commandments but…

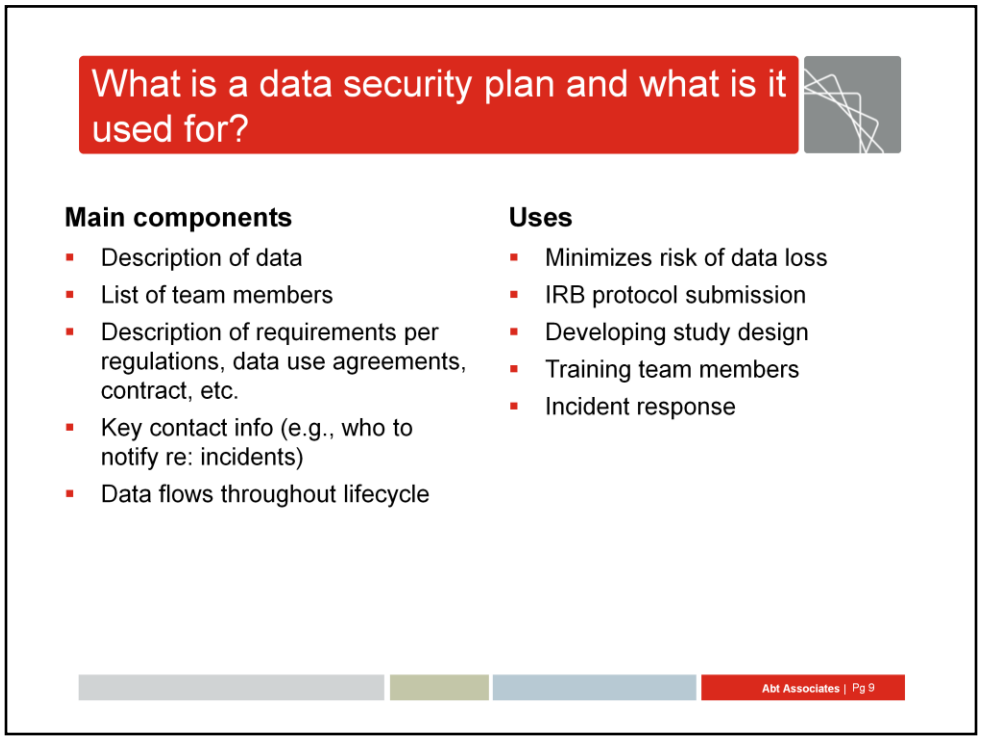
See handout for all 9 Data Security Commandments.



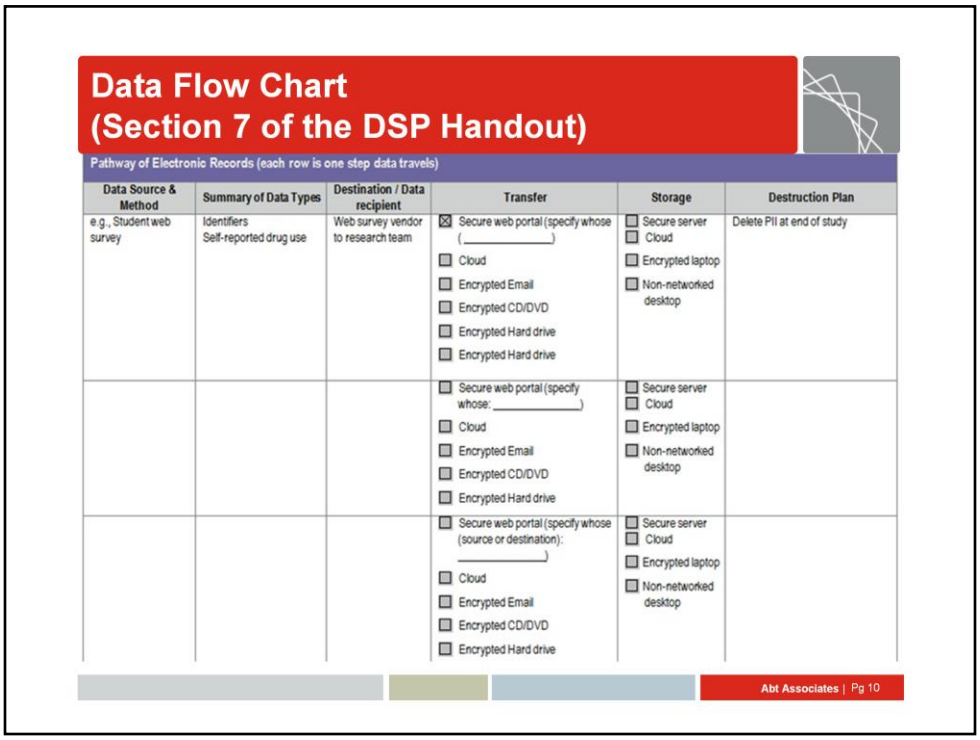
These 6 commandments are the minimum needed to comply with most regulations.



The data security plan is based on the lifecycle of the data during your study. Think of data as a physical object that explodes into existence (for you) at the point of creation or collection. You are now responsible for this object until it is destroyed by requirement or no longer needed, or it is archived for future use as allowed.



- Data security plan is a living document. It should be updated and refined as needed.
- Changes to the protocol should happen only after reviewing the data security plan to verify you won't introduce problems with your revision.
- So how do you translate these commandments into a plan for your data and study? DSP
- Your university and IRB may have own requirements & policies
- We require more detail like what's in this DSP because this is main risk in our studies



Data Flow Chart (Section 7 of the DSP Handout)

Pathway of Electronic Records (each row is one step data travels)

Data Source & Method	Summary of Data Types	Destination / Data recipient	Transfer	Storage	Destruction Plan
e.g., Student web survey	Identifiers Self-reported drug use	Web survey vendor to research team	☒ Secure web portal (specify whose (_____) ☐ Cloud ☐ Encrypted Email ☐ Encrypted CD/DVD ☐ Encrypted Hard drive ☐ Encrypted Hard drive	☐ Secure server ☐ Cloud ☐ Encrypted laptop ☐ Non-networked desktop	Delete PII at end of study
			☐ Secure web portal (specify whose: _____) ☐ Cloud ☐ Encrypted Email ☐ Encrypted CD/DVD ☐ Encrypted Hard drive	☐ Secure server ☐ Cloud ☐ Encrypted laptop ☐ Non-networked desktop	
			☐ Secure web portal (specify whose (source or destination): _____) ☐ Cloud ☐ Encrypted Email ☐ Encrypted CD/DVD ☐ Encrypted Hard drive	☐ Secure server ☐ Cloud ☐ Encrypted laptop ☐ Non-networked desktop	

Sean likes tables and linear thinking, so here is the lifecycle in table format.

This is a snapshot of the key part of the data security plan we will cover today in our exercise.

It shows where data come from and go, as security procedures that will be used to transfer it and store it.

Note it also includes when and how destruction will occur.

Agenda

1. Data security basics and grammar
   - Vocabulary
   - Common regulations
   - Commandments
   - Data security plan & lifecycle

2. **How to develop a data security plan**
   - **Collection**
   - **Transport**
   - **Storage**
   - **Destruction**

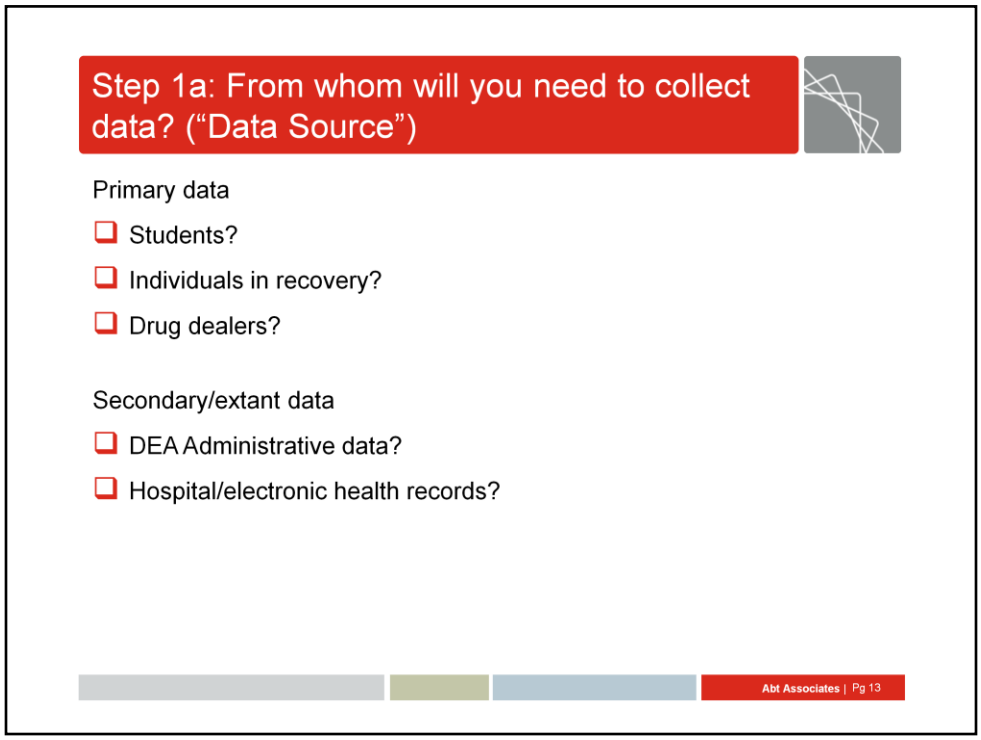
3. Group exercise: develop a data security plan
   - Primary data collectors: quantitative
   - Primary data collectors: qualitative
   - Secondary/extant data analysts

4. Other important topics
   - Deliverables
   - Minimizing re-identification risk in datasets shared with others
   - Data security incident response

How to Develop a Data Security Plan

Example research question:

"Has *Breaking Bad* caused more people to use crystal meth compared to before the series?"



We are going to walk you through Section 7 of the DSP which is the detailed data flows & steps to protect the data.

Answers the questions for each type of data, what data is going where, how will it be protected while in transit & in storage, and when will it be destroyed?

Covers all 6 commandments.

We’ll start with the easy study design basics that researchers are familiar with.

We’ll discuss security considerations & implications of the study design.

Your selection will have pros/cons re: security of data and risk of data loss.

Audience: can you think of pros/cons in terms of data security for paper vs. web survey? Paper vs. audiorecordings?

Web – Pros: reaches a large audience quickly, allows for rapid determination of results. Cons: Even if you don't collect names, identifiable information is collected by web systems (IP addresses), security is a large concern because the site is often at least partially available to anyone in the world.

Paper & pencil – Pros: straightforward management and reliable. Regulations are easy to satisfy with paper. Cons: identifiers are hard to separate from data and paper is easily lost.

Audio recordings – Pros: Higher quality of data and more rich. Cons: files are large, voices can be identified, you never know what someone will say. Often sent to a transcriptionist which increases something going wrong.

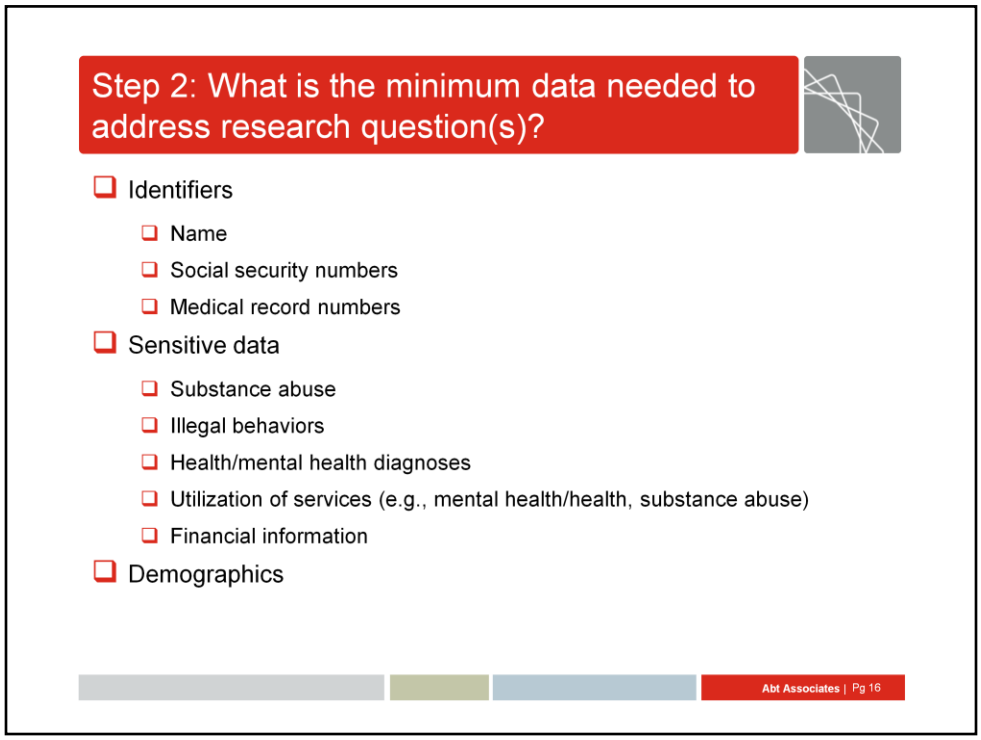
Video – Same as audio, but even more identifiable!

Paper notes – Often need to be transcribed and notes are often not destroyed in a systematic manner.

EMR – HIPAA almost always applies and with it comes risk.

Steps 1a-1b: Select Data Source and Method to Collect

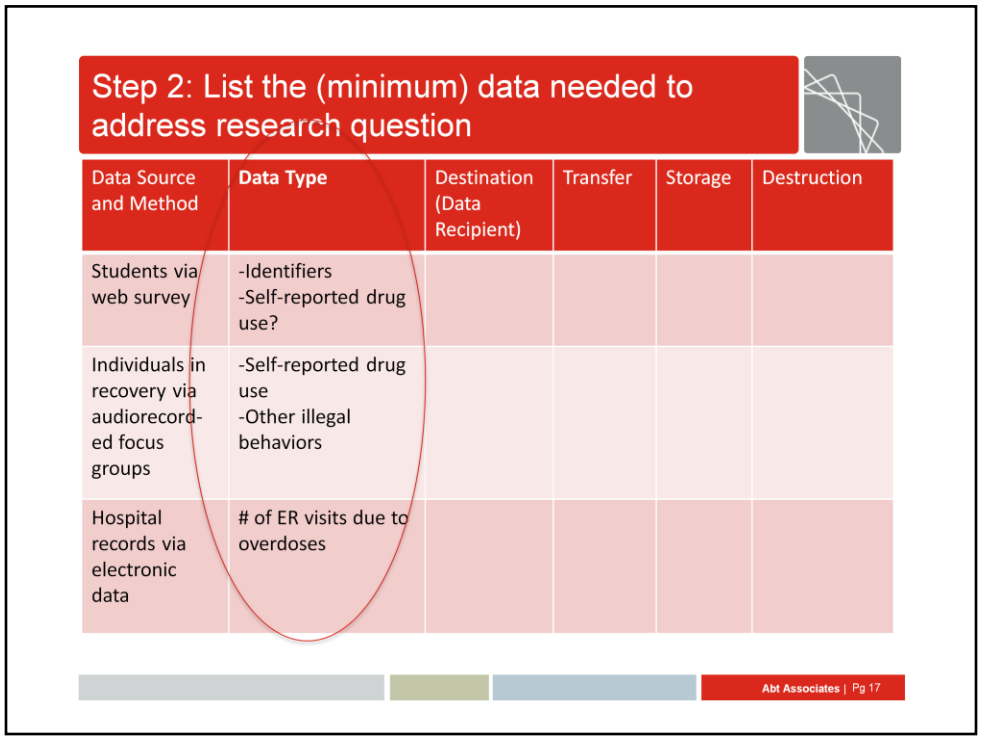
Data Source and Method	Data Type	Destination (Data Recipient)	Transfer	Storage	Destruction
Students via web survey					
Individuals in recovery via audiorecord-ed focus groups					
Hospital records via electronic data					



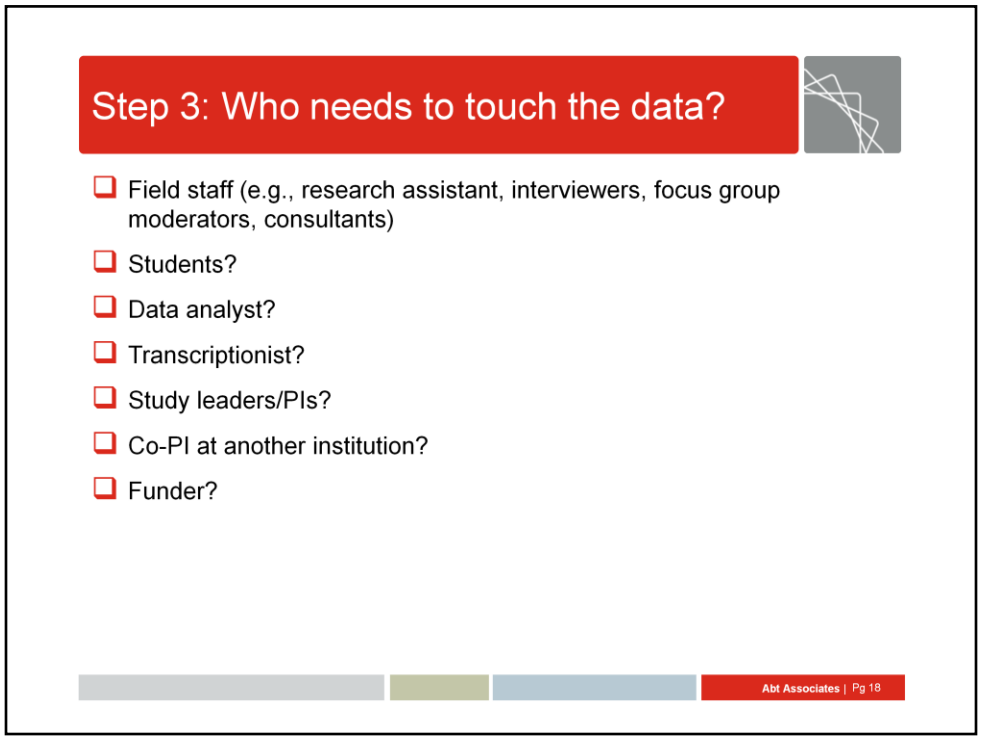
The need to get/use identifiers may depend on your design (pre/post/longitudinal) and need to link data from different sources (that would require identifiers like SSNs)

Don't forget that obtaining names for sampling frame means you will have identifiers!

More data = more risk.



This is commandment #1

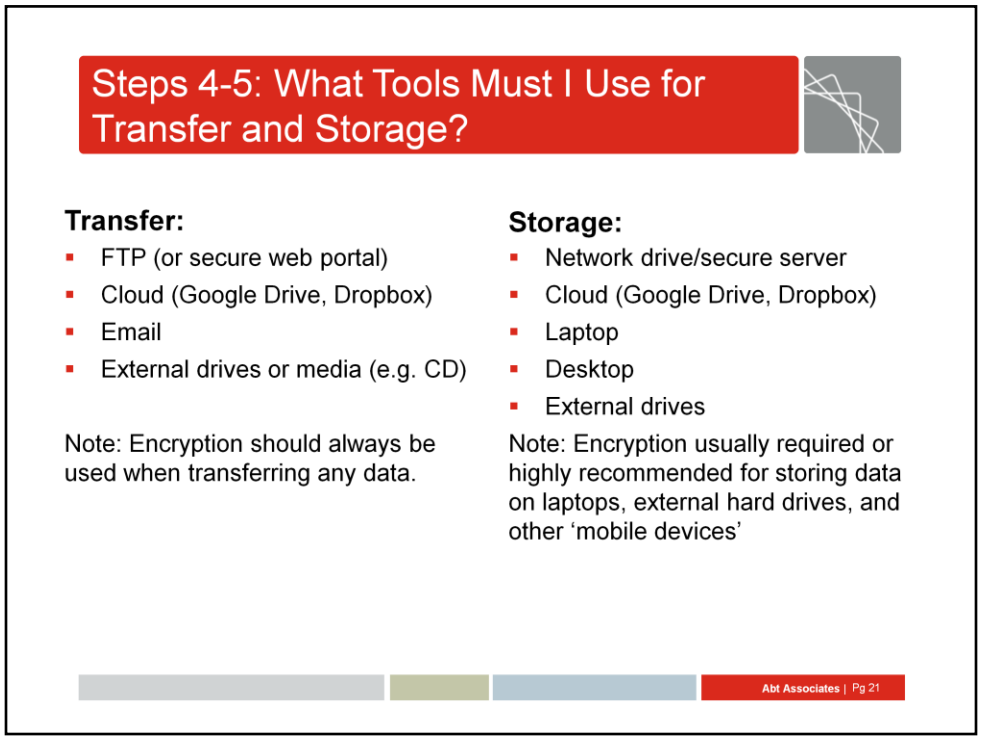


Think through who will have access…where will that data go. Every person that touches the data introduces another point of failure/increases risk.

Step 3: Identify minimum # of people who need to receive data ("Destination")

Data Source and Method	Data Type	**Destination (Data Recipient)**	Transfer	Storage	Destruction
Students via web survey	-Any identifiers -Self-reported drug use?	Web survey vendor			
Individuals in recovery via audiorecord-ed focus groups	-Self-reported drug use -Other illegal behaviors	Focus group moderator (consultant)			
Hospital records via electronic data	# of ER visits due to overdoses	Data analyst			

Now we switch to the IT side of things. Now that you have a great design, lets find what IT tools can be used to protect the data. Accomplishing your study is very important and to complete it and do the next great study means you need to protect the data you have been entrusted with.



Sensitive data will definitely require these tools; less sensitive data may not (but ask your IRB/IT Security Officer to confirm). Each one has pros and cons. All of them can be made secure and you should pick the one that matches your USERS and protocol NEEDS, not which one is the cheapest or coolest. The two most common are cloud and email.

Cloud – So cheap, so available, and so out of your control. When you put data on the cloud, you are often also signing up for that data being scanned and added to a "big data" dataset. Don't avoid cloud, but make sure you know what will be done with the data you upload, your rights and the cloud provider's rights, and the security in place.

Email – Most "oopses" happen through email. An accidental "reply-all" for "forward" can result in serious fines. Some regulations place clear restrictions on the use of email. There are email security tools provided in the notes that can encrypt the data. This often satisfies regs.

Vendors for encryption:

Symantec Encryption (formerly PGP)

Sophos SafeGuard Encryption

McAfee Complete Endpoint Protection
Microsoft BitLocker

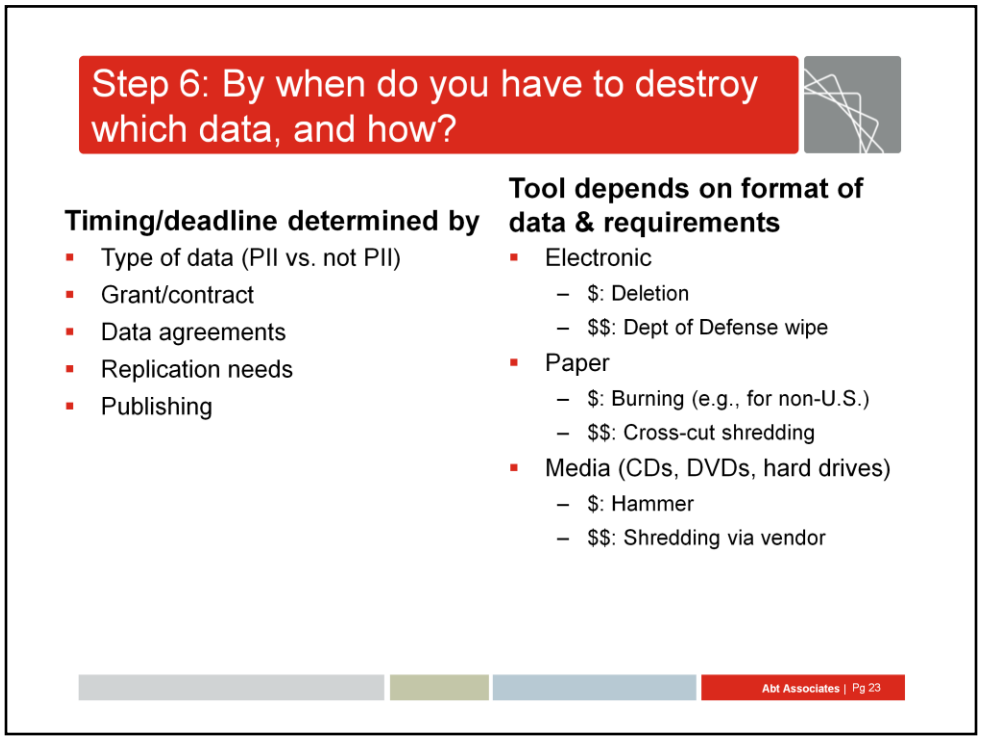
Vendors for file transfer:
FileZilla, WS_FTP

Vendors for cloud:
Fpweb, box.com, spideroak

Vendors for secure email:
DataMotion, GPGmail, Symantec Desktop Email Encryption

Steps 4-5: Determine transfer and storage tools

Data Source and Method	Data Type	Destination (Data Recipient)	**Transfer**	**Storage**	Destruction
Students via web survey	-Any identifiers -Self-reported drug use?	Web survey vendor to researcher	FTP	Secure server	
Individuals in recovery via audiorecord-ed focus groups	-Self-reported drug use -Other illegal behaviors	Focus group moderator (consultant)	FTP (via encrypted laptop)	Secure server	
Hospital records via electronic data	# of ER visits due to overdoses	Data analyst	Encrypted CD	Secure server	



Destruction can be creative. NIST SP800-88 has a guide on data destruction. The key is knowing what you must destroy, by when, and how thorough you must be. All destruction should be documented in a "certificate of destruction".

Step 6: Define destruction date and method

Data Source and Method	Data Type	Destination (Data Recipient)	Transfer	Storage	**Destruction**
Students via web survey	-Any identifiers -Self-reported drug use?	Web survey vendor to researcher	FTP	Secure server	Identifiers deleted at study end
Individuals in recovery via audiorecord-ed focus groups	-Self-reported drug use -Other illegal behaviors	Focus group moderator (consultant)	FTP (via encrypted laptop)	Secure server	Audiorecordings deleted at study end
Hospital records via electronic data	# of ER visits due to overdoses	Data analyst	Encrypted CD	Secure server	Deleted 6 yrs after publication

Agenda

1. Data security basics and grammar
   - Vocabulary
   - Common regulations
   - Commandments
   - Data security plan & lifecycle

2. How to develop a data security plan
   - Collection
   - Transport
   - Storage
   - Destruction

3. Group exercise: develop a data security plan
   - **Primary data collectors: quantitative**
   - **Primary data collectors: qualitative**
   - **Secondary/extant data analysts**

4. Other important topics
   - Deliverables
   - Minimizing re-identification risk in datasets shared with others
   - Data security incident response

Break into Groups for Exercise

Fill out Worksheet 7

- Pick one type of data source (e.g., student survey)
- Fill out table from start to finish for that type of data (where it comes from, all steps it has to go)
- Don't forget:
  - Identify then protect the weakest link(s)
  - Recruitment list/names
  - Focus group audiorecordings turn into transcripts, coded NVIVO files
  - Multiple study partners who might need access to data
- Pick someone to take notes & report out to group

Group Exercise

Example research question:

"Has *Breaking Bad* caused more people to use crystal meth compared to before the series?"

Breaking BadLib

- As the greatest researchers in the world, we feel that (source) are the most reliable source of information and (method) is the most effective way to collect data. We know that we have to provide the data to (destination) for analysis. We are going to send the data to them using (transfer). They are going to store it on (storage) each day and our partner will do the same. After the end of the study, (date) we will destroy it by (method).

Agenda

1. Data security basics and grammar
   - Vocabulary
   - Common regulations
   - Commandments
   - Data security plan & lifecycle

2. How to develop a data security plan
   - Collection
   - Transport
   - Storage
   - Destruction

3. Group exercise: develop a data security plan
   - Primary data collectors: quantitative
   - Primary data collectors: qualitative
   - Secondary/extant data analysts

4. **Other important topics**
   - **Deliverables**
   - **Minimizing re-identification risk in datasets shared with others**
   - **Data security incident response**

Products/Deliverables of Study

- Summary report with aggregated data only (no small cell sizes)
- Raw data (survey data with or without identifiers? transcripts?)
- Public use dataset
- Restricted use data set

Resources/Guidance re: Minimizing Disclosure Risk

- OMB's Federal Committee on Statistical Methodology (*Report on Statistical Disclosure Limitation Methodology*, *2005, Working Paper 22*)
- Department of Education's National Center for Education Statistics (*Statistical Standards: Processing and Editing Data* http://nces.ed.gov/statprog/2002/std4_2.asp)
- Department of Health and Human Services (*Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act Privacy Rule, 2012*)

Data Security Incidents: What is a data breach?

- **Don't panic!!!** Most data security incidents are NOT breaches!
- Incidents can be many things. If you are unsure to report or not, report it!

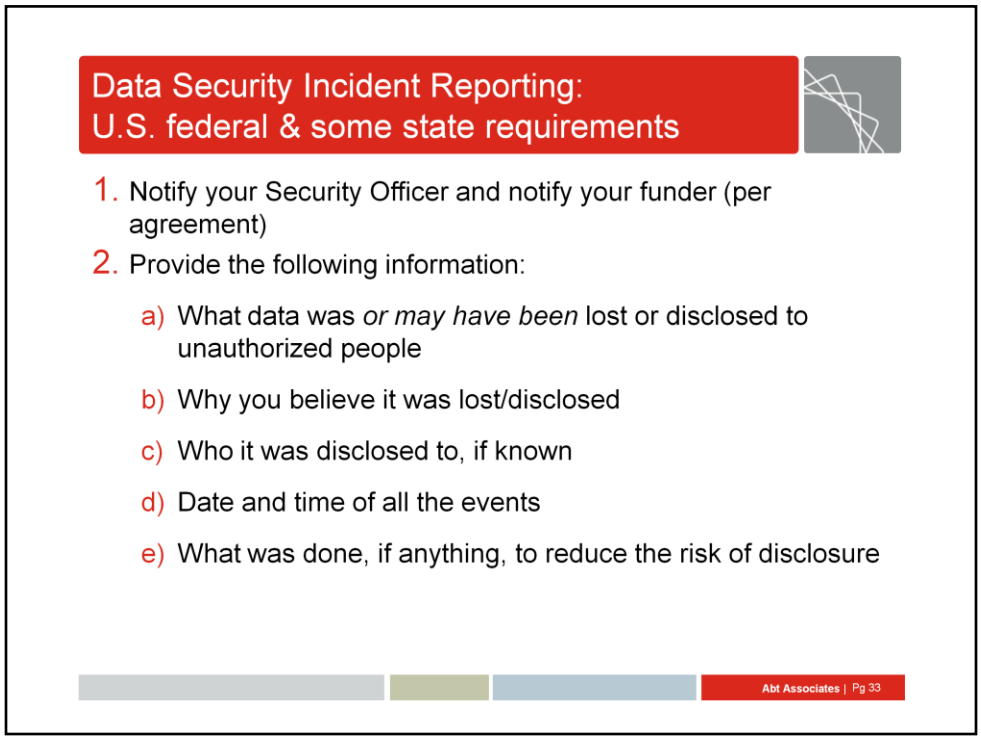
What qualifies as a breach?

- Lost paper with personally identifiable information (PII)
- PII on lost/stolen unencrypted portable device/media (flash drive, CD-ROM/DVD, etc.)

What is not a breach? (but still need to report)

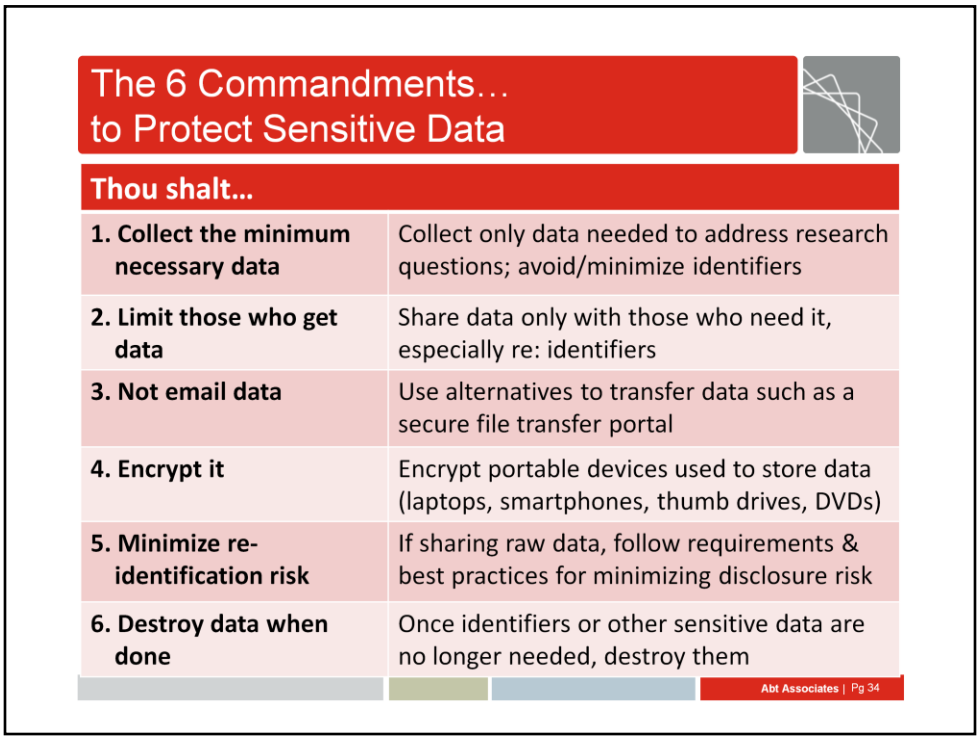
- Emailed PII between authorized study team members
- Giving/getting more data than requested and authorized

Even the best technology and plans can have hiccups. Review the laws and agreements/contracts/grants that are applicable to your data and review the incident/breach definitions and conditions. The data security plan should be the go-to document for determining the extent of the incident and the severity. How you react after the discovery of the incident is very important. Hiding the incident never turns out well and its an opportunity to prove that you are a trustworthy data custodian by your prompt reporting and resolution.



When something does go wrong; do not delay in responding to client requests/notifying.

Notify appropriate institution expert; they will help you decide who else needs notification



These 6 commandments are the minimum needed to comply with most regulations.

Contact Information

Sean Owen, Director of Abt's Client Cybersecurity Center

Sean_Owen@abtassoc.com

301-347-5734

Teresa Doksum, Abt IRB Chair and Director of Research Ethics

Teresa_Doksum@abtassoc.com

617-349-2896