Data Security Plan Development Guide for Researchers

November 2014

Prepared for:
Association for Public Policy Analysis and Management Fall Research Conference

Submitted by:
Sean Owen, CISSP, CAP and Teresa Doksum, Ph.D., M.P.H.
Abt Associates Inc.
4550 Montgomery Avenue
Suite 800 North
Bethesda, MD 20814

Data Security Plan Development Guide

Table of Contents

1. Introduction

This is a 'how-to' guide for researchers to develop a data security plan. It can be used for all types of studies and data—large and small-scale studies, quantitative and qualitative data, U.S. and non-U.S. based studies. It is organized into checklists and/or worksheets for each section of the data security plan, beginning with key questions for the researcher. After completing the checklists and worksheets in this guide, a researcher should have a comprehensive plan to keep study data secure. Section 1 provides an overview of data security plans and key documents needed to get started. Section 2 summarizes the key requirements common to most security regulations for safeguarding data. Appendix 1 is an example of a completed data security plan for secondary/extant data. It is designed to be completed in sequential order and skipping steps can lead to an incomplete data security plan.

1.1 Purpose of a Data Security Plan

A data security plan is the recipe that study teams should develop and follow to protect the data from the beginning to the end of the study. It is organized around the lifecycle of the study and the data—from beginning to end.

Data security plans are often required by the regulations governing the data, the funder, and/or any Institutional Review Boards reviewing the study. Even if it is not required, it is a valuable document to document the security around a study. It can be a stand-alone document or incorporated into the study plan.

The data security plan describes the technical, physical and administrative safeguards for the protection of data. The procedures to safeguard the data should be commensurate with the level of sensitivity of the data and in accordance with requirements from relevant regulations.

As a living document, the data security plan should be updated as needed throughout the study and continue to match the lifecycle of the data.

1.2 Checklist of Items You Will Need to Develop Data Security Plan

The data security plan should include the requirements for keeping the data secure per any contractual or other documents that describe these requirements. Before you develop the data security plan for your study, gather the following documents, if relevant:

- ☐ Research grant or contract (look for clauses re: information security, confidentiality)
- ☐ Data use agreements
- ☐ Study design including data collection instruments, consent language
- ☐ Protocol approved by Institutional Review Boards
- ☐ OMB package—confidentiality section
- ☐ Privacy Act System of Record Notice

What is a data use agreement?

- A form of contract between the source of data such as a school or hospital ("data provider") and a data user (e.g., researcher).
- Contains permission to use the data for a certain purpose and a promise to keep the data secure.
- Other names include data sharing agreements, data transfer agreements, information transfer agreements, restricted use agreements, data licenses

2. Key Data Security Principles Common to Security Regulations

Below are some of the key principles to keep data secure per most security regulations. Researchers should check for additional requirements in the documents listed in Section 1.2 to develop the data security plan. As each section of the data security plan is completed, keep in mind these principles to comply with the regulations and to minimize the risks of a data security incident or breach.

Study Design

1. "Minimum necessary:" Plan to collect only the data needed to address the research questions; if identifiers are not needed, don't collect them.

2. "Need to know:" Share only the data that your study partners, inside and outside your institution, need to do their job.

3. "Ounce of prevention…:" Map out, in the form of a data security plan, which study partners will have which data, how it will get to them, and where/how it will be stored.

Data Collection

4. "Keep 'em separated:" Use study ID numbers and keep real identifiers separate from sensitive data while in transit and in storage.

5. "Encrypt it:" Encryption addresses lots of security challenges so use it for laptops, smartphones, thumb drives. However, it's not a silver bullet, so still use our other tips.

6. "Email oopses:" Avoid email to transmit sensitive data—use alternatives such as a secure file transfer portal.

Reporting and Close Out

7. "Protect your sources:" Ensure research reports do not include information identifiable to an individual and minimize risk of re-identification.

8. "Avoid data sprawl:" Collect all data from all study team members and securely archive it in one place.

9. "Don't need it? Destroy it:" Once identifiers are no longer needed, destroy them.

Worksheet 1 for Section 1: Data Security Contact Information

This section documents all the contact information for key project and client staff. This information is crucial in responding quickly to any security incident. Some regulations require notification in 60 minutes!

Instructions: In this section, provide name, title, telephone and email for lead study team members, including those from other institutions, reviewers of the data security plan, and the funder.

Study Team Information			
Role in Study	**Name**	**Organization**	**Phone/Email**
Lead Researcher			
Project Manager			
Lead Data Manager			
Other key team members (consultants, vendors, subcontractors)			

Reviewers of the Data Security Plan			
Role	**Organization**	**Phone**	**Email**
Director of Information Security			
Institutional Review Board Chair			
Other reviewer			

Research Funder (Client) Information			
Role	**Organization**	**Phone**	**Email**
Project Officer			
Contract Officer			
Information Security Incident Response			

Worksheet 2 for Section 2: Study Information

In worksheet 2, you should be pulling your information out of Section 2 earlier in the guide. This information serves as a reference guide to anyone reviewing the document to understand what contracts or agreements were used in the development of the data security plan. The dates are especially important because they help you determine when you should be considering destroying identifiers. These dates should all fall in your period of performance for the grant or contract.

Instructions: Provide a narrative description of the contractual relationships among the funder and study team members (e.g., study funded by X agency, which contracted with a study firm which engaged additional consultants/vendors:

_____

_____

_____

_____

Basic Study Information	
Study Title and Nickname	
Contract number:	
IRB Number & Status	
Funding Client Organization name	

Key Period of Performance Dates		
	Start Date	End Date
Study Contract:		
Data Collection:		

Study Subcontractors, Consultants, and Vendors Information			
Point of Contact Name & email/phone	**Organization**	**Contract Number**	**Data Agreement #, if any**

Worksheet 3 for Section 3: Description of Study Data and Study Security Procedures

3.1 Types of Study Data

The types of study data help you evaluate the sensitivity of the information. Understanding what data were collected by what instrument allows the study team determine what protections must apply to that data as it moves through its' data lifecycle.

Instructions: This section should describe what data will be collected (with special attention to identifiers), and how. The data being collected should be the minimum necessary to address the research questions (i.e., minimize collection of sensitive data including identifiers). The following supportive documentation should be included as appendices to the data security plan:

- **Data Dictionaries/list of variables**
- **Data collection instruments**
- **Data use agreements, non-disclosure agreements, research grant or contract, and other documents that indicate which laws govern the data**

Researchers can use either a table or narrative format for this section.

Description of Data		
Data Source	**Identifiers Needed**	**Type of Data**
EXAMPLE: Primary data collection (e.g., survey, interviews, focus groups)	• E.g., Student first/last name	• Satisfaction with program (see attached survey)
EXAMPLE: Secondary/extant data (e.g., administrative data)	• E.g., Student first/last names	• School records (grades for 2013)
	•	•
	•	•
	•	•
	•	•

Narrative format example:
The study will rely on primary (e.g., surveys, focus groups, interviews) and secondary (i.e., existing, extant) sources of data to inform the overall study:

(1) Primary data collection (briefly describe here, including identifiers, & attach data collection instruments)

a) _____

b) _____

(2) Secondary/existing datasets (attach variable list if available)

a) _____

b) _____

In summary, as part of our data collection effort:

- Individual level data will be collected;
- Personally identifiable `information (PII) will be collected on X types of participants (e.g., patients, stakeholders).

HINT: In the event of a data security incident, this section of the data security plan provides a quick way to assess the level of sensitivity of the data

3.2. Security Procedures

Instructions: This section provides a narrative of description of how each type of data will be collected and by which study team members and with whom the data will be shared, from beginning to end of the data lifecycle. The procedures should reflect any requirements specified in the contract, data agreement, etc.

Example procedures to maintain the confidentiality of data: Of the four data sources collected:

1. **X secondary/existing data** will be kept on a secure site managed by X study partner, as part of their contract with the funder for maintaining the X data system. De-identified data sets will be provided to Abt quarterly through a secure FTP site for analysis. X data sets will be maintained at Abt Associates on the secure server;
2. **X Data Collection Forms** begin as a hardcopies provided to Abt Associates by three organizations. The hard copy information is then processed by study partner X (see Appendix X – Study Partner Data Security Plan) into readable electronic files that are provided to Study Partner X through a secure FTP site. X then conducts an analysis of the data file and produces statistical tables for quarterly reporting to funder. The final file is then transferred to Abt via a secure FTP website for integration into the quarterly reporting. Original paper files will be returned to Abt once a year for secure storage. Electronic datafiles will be maintained by Study Partner X (master files) and Study Partner Y (analytical files) until the end of the contract. Report data will be maintained on Abt Associates common drive project folder.

X Interviews will be conducted by Abt Associates staff with grantees. Interviews will be conducted with grantee staff and focus on program implementation, data collection, clinical services, and program administration. PII will not be collected as part of this effort nor will consumers of services be interviewed. Notes taken during the site visit and reporting will be kept on Abt Associates common drive with access restricted to staff associated with the project.

Worksheet 4 for Section 4: Staff Training on Data Security and Monitoring

When completing the training section, keep in mind that many regulations or funders require specific training. If you are handling HIPAA covered data, the team should have the appropriate HIPAA training. Don't forget your study partners, consider what data they will be accessing and what is the appropriate training to safely handle that data.

Instructions: list study-specific training, any funder-required training, and general training required by the researchers' organization(s).

Key staff from the lead study organization and our partner organizations have 1) completed study-specific training that incorporates this data security plan, 2) received a copy of this data security plan, and 3) completed the following general trainings to promote data security and compliance. Management of trainings is handled within each organization that is part of the study and occurs annually or biannually. A list of some of the trainings completed by the lead researcher and their partner organizations is provided in the exhibit below.

Trainings	
Study Team Member	**Training**
Lead researcher	• General Security Awareness Training • CITI Human Subjects Training • HIPAA or relevant regulation governing the data
Study Partner X	• General Security Awareness Training
Study Partner Y	•

Study staff will be trained on security and supervised by:_____

Monitoring and supervision of the staff who are handling data and/or are interviewing program staff at the lead study organization and our partner organizations will allow for additional opportunities to identify and correct any security or procedural issues. All study staff will be made aware of the study-specific data regulations and best practices associated with handling data for the study. These practices will be incorporated in the study design and will be detailed in training plans for interviewers, as well as for support and data analytic staff. All staff who will have access to the ___ data have signed a confidentiality agreement per the requirements of the data use agreement or contract.

Worksheet 5 for Section 5: Deliverables

Instructions: List the key deliverables associated with the study. The deliverables listed here need to be consistent with the grant or contract, consent language (especially with regard to identifiers), IRB protocol, OMB package, and the data use agreements. Any restrictions from these documents need to be noted.

Deliverables		
Data Sources	**Deliverable**	**Any restrictions from data agreements?**
X	• E.g., Quarterly reports highlighting key process and outcome trends associated with grantee's efforts to provide services to clients; • E.g., Annual reports detailing key trends with grantees grantee service progress and outcomes.	e.g., draft needs to be shown to school districts first
Survey Data	• E.g., data sets like restricted use or public use dataset	e.g., Re-identification risk must be minimized per industry standards

- In all reports no PII will be included and all reported data will be aggregated at either the site or multi-site level.
- All data and reporting systems will be transitioned to the client at the conclusion of the project (including _____ dataset per data use agreement).
- Raw qualitative data (e.g., interview notes, transcripts) will not be shared or transitioned to the client during or at the conclusion of the project.

HINT: Before collecting data, ensure that you confirm with the funder how they will want study data reported, whether just aggregated or any raw data they will want and in what format. Then your design will include consent language and data agreements that are consistent with the funder's needs.

Worksheet 6 for Section 6: Physical Record Lifecycle

The lifecycle sections are the most complicated sections in this guide, but they are also the most valuable for identifying the appropriate security procedures to protect the data. The lifecycle should track the creation (or acquisition) at the earliest possible point in the study all the way until it is delivered to the client or destroyed. Your first row probably starts with participants and your last row is you providing a dataset/report to your client. Often the rows will feed into each other, where the destination from the previous row becomes the source for the next row.

Instructions for Section 6 and 7: These sections outline the path the data must travel from beginning to end and who will need/have access, how it will be transported, where it will be stored, and when it will be destroyed. Use these questions to guide developing these sections. Use the table format provided, or alternatively, use diagrams and narrative.

Questions re: Sharing the Data

1.1. Who is authorized to have access to the data (within the study team)? Ensure the contract, data agreement, consent allow access to various team members.

1.2. Which data are we providing to anyone outside the study team? (any identifiers?)

1.2.1. How do we plan to provide that data (e.g., via secure web portal, encrypted CD)?

HINT: Focus security efforts on the identifiers and most sensitive data that have stringent requirements, as well as the point(s) at which the data is most at risk of getting lost or shared with unauthorized individuals.

Questions re: Data Transport and Storage

1.3. Transport

1.3.1.How will the data be collected, especially the identifiers?

1.3.2.Do study team members need to receive or send data to each other? How?

1.3.3. Does the data need to be encrypted (check data agreements, contract, etc.)?

1.3.3.1. What method/product will be used to encrypt the data?

1.4. Storage

1.4.1. After receiving data, where will study team members store it?

1.4.2. Will the data be stored on any mobile devices like laptops, CDs, thumb drives?

1.4.2.1. What encryption will be used on those devices?

Questions re: Data Destruction

1.5. Which data must be archived vs. returned vs. destroyed? By when?

1.6. How will/must the data be destroyed?

Pathway of Physical Records (each row is one step data travels)						
Source of Data	**Summary of Data Types**	**Destination**	**Transport**		**Storage (Destination)**	**Return or Destruction Plan**
E.g., program staff	Interview data	Researcher	☒ Paper (interview notes)	☐ USPS (Registered) ☐ UPS ☐ FedEx from site ☐ Licensed/bonded carrier ☒ Hand-delivery/carry by study member	Researcher organization locked cabinet	Shred 3 yrs after end of study
e.g., program participant	Survey data (no PII)	Researcher	☒ Paper	☐ USPS (Registered) ☐ UPS ☒ FedEx ☐ Licensed/bonded carrier ☐ Hand-delivery by study member	Researcher organization locked cabinet	Return to funder at end of study via Fed-Ex
	I		☐ Paper ☐ Encrypted CD/DVD ☐ Tape ☐ Encrypted thumb drive ☐ Encrypted Hard drive	☐ USPS (Registered) ☐ UPS ☐ FedEx ☐ Licensed/bonded carrier ☐ Hand-delivery by study member		

Worksheet 7 for Section 7: Electronic Record Lifecycle

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	☐ 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	

NOTES