Top Ten Research Data Security Tips
for Investigators using Personally Identifiable Information

INTRODUCTION

These Top Ten Research Data Security Tips are a non-exhaustive selection of data security measures investigators can take to help prevent unauthorized uses of personally identifiable information in electronic or non-electronic forms.

The list is informed by a holistic evaluation of regulatory agency guidance as well as a review of published breach incidents, which occurred at leading universities and academic medical centers.

The list applies broadly to research data with personal identifiers—personally identifiable information (“PII”), defined as information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record).

It is not intended to replace existing institutional guidance.

I. INITIAL CONCEPT AND RESEARCH DATA MANAGEMENT PLANNING

1. **Data safeguards:** At the initial concept and planning stage, consult with your institution and collaborators regarding safeguards for PII throughout the lifecycle of your research. Safeguards may differ depending on how data is collected, generated, stored or accessed.

2. **Third Party Recipients:** Consider the varying individuals or entities that may require access to your research data for varying purposes. Individuals or entities may, for example, de-identify, process or analyze data on your behalf. For studies where public dissemination of study results is anticipated, plan accordingly. For example, federal law requires registration of certain clinical trials and submission of certain aggregate research data elements to a publicly accessible database (ClinicalTrials.gov).

Researchers should engage the IRB and other appropriate administrative offices at their institution to plan for use of PII in the IRB application. For details, consult the IRB office overseeing your research. For contact information, see the [Harvard Catalyst Regulatory Atlas](http://connects.catalyst.harvard.edu/regulatoryatlas/?mode=c).

II. DATA COLLECTION AND USE

---

3. **Portable Devices**: Personally identifiable information should not be persistently stored on a portable device including thumb drives, hard drives, smart phones, or laptops unless the PII is encrypted on the device. Your institution may have policies regarding the use of portable devices. Consult your IRB to learn if your institution has a specific policy for research and the use of portable devices. Learn more about policies at your institution through the research data protection section of the Harvard Catalyst Regulatory Atlas. See also: “A Statement by the NIH Director, Elias Zerhouni on Encryption and Data Security.”

4. **Use a secure computer or server**: Where possible, PII from existing data sets or prospectively collected data sets should be stored on a secure server inaccessible to the public (i.e. internet search engines) or unsecured networks. Some institutions, for example, require PII to be held only on institutionally managed servers. To determine the secure solution appropriate for your project connect with your institution’s research data protection resources through the Harvard Catalyst Regulatory Atlas.

   **Password tips:**
   - Use a unique password that is not used for other accounts
   - At least 15 characters long using a pass phrase, no numbers at the beginning or end of the phrase
   - Does not contain your user name, real name, or company’s name
   - Does not contain a complete dictionary word
   - Contains characters from each of the following four groups: lowercase letters, uppercase letters, numerals, and special characters

5. **Use codes for personal identifiers or other appropriate de-identification**: Where possible, substitute codes for personal identifiers and store the key in a different locked physical location. Statistical experts or others specializing in standards for data de-identification may be needed. Consult your local research data protection resource through the Harvard Catalyst Regulatory Atlas to determine the secure solution appropriate for your project.

6. **Training and education**: Study team members such as research assistants, research coordinators, data managers and other study staff with research PII access should be trained and educated in the importance of confidentiality and the potential risks of harm to subjects. In some situations, staff could be asked to sign confidentiality agreements. Check with your local IRB on what training and education is required. Find your IRB contact on the Harvard Catalyst Regulatory Atlas.

7. **Track and document data transmissions**: Track and document any transmissions, transfers, or dissemination of research PII including copies or extracts made of your PII, even if mixed with data that is not PII. Especially document any transfers to third parties (i.e. statisticians who are de-identifying data, data disposal vendors, cloud computing service providers).

8. **Use and retain minimum necessary**: Use the minimum amount of PII necessary to conduct your research. And limit the number of individuals who have access to the data. If possible de-identify
research data as soon as practicable and store the file on a secure server with access limited to only those required.

9. Incident response and reporting: Know your institution’s incident reporting policy, if any, in the event of discovery of an unauthorized use or disclosure of PII. Ensure that third parties commit to following any applicable incident response policies. Consult your local research data protection resource through the Harvard Catalyst Regulatory Atlas to determine the secure solution appropriate for your project. If you discover unauthorized use or disclosure of PII during a research study, consult your IRB (find your IRB contact on the Harvard Catalyst Regulatory Atlas).  

III. PROJECT CONCLUSION, REPORTING, PUBLICATION, RETENTION AND DISPOSAL

10. At project conclusion, ensure that any final dissemination, public access, retention, archive or disposal plan is secure.

   - Work with your local IRB and IT offices to ensure that data is appropriately encrypted or secured when prepared for archiving, return to initial data source, or destruction by third party vendors.
   - Ensure that any dissemination plan that requires aggregate data does not contain any personal identifiers.
   - Consult your local institution on how to prepare your data for submission to any publicly accessible repositories. For example, certain studies involving genomic data require submission to dbGaP.  
   - If your data is being made available to the public (such as through ClinicalTrials.gov), obtain assurances that any publicly accessible website complies with the appropriate security controls.
   - For publications, ensure that data is sufficiently de-identified prior to submission. Determine the data content and format requirements of repositories to which data must be sent. You may need experts to de-identify your PII into an acceptable format for reporting and publication.

QUESTIONS?

If you have any questions about how to secure your data, consult the Harvard Catalyst Regulatory Atlas to contact the IRB or access your institution’s Research Data Protection resources.

5. IRB Contacts: http://connects.catalyst.harvard.edu/regulatoryatlas/?mode=c&id=3