<table>
<thead>
<tr>
<th>Content</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC State University Regulatory Information for Data Security</td>
<td>p.2</td>
</tr>
<tr>
<td>Sensitivity of Data and Data Protection (Controls)</td>
<td>p.3</td>
</tr>
<tr>
<td>Information to Include in IRB application</td>
<td>p.6</td>
</tr>
<tr>
<td>IRB Recommended Tools</td>
<td>p.7</td>
</tr>
<tr>
<td>Collaborative Projects and Data Management/Security</td>
<td>p.8</td>
</tr>
<tr>
<td>Issues Related to Identifiability of Data</td>
<td>p.9</td>
</tr>
<tr>
<td>Definitions and Resources</td>
<td>p.13</td>
</tr>
<tr>
<td>Example Language for IRB applications</td>
<td>p.15</td>
</tr>
</tbody>
</table>
Data and Privacy breaches of study participants is one of the largest risks to participants in research at NC State University. As a result, a description for managing your data must be part of your IRB application. This document is intended to provide researchers and their support staff guidance on what to include in an IRB application.

In accordance with the NC State University Data Security requirements, the IRB is requiring data security to be addressed in all approved protocols. The following bullets provide information on specific data security requirements at the University.

- **NC State University Data Management Procedures Regulation**
- **Rule 08.00.18 NC State University Endpoint Protection Standard**
- It is important to address research data protection and address what protection measures are needed to ensure adequate Confidentiality, Integrity, and Availability.
- It is important to address data protection thought the entire data life-cycle (collection/creation, processing/accessing/sharing, and storage/retention/disposition/destruction).

All software, hardware, equipment, and processes used during the course of an IRB approved research protocol needs to protect data at all levels. It is important to consult with your departmental IT services when determining solutions and data protection methods. This is particularly important when ensuring that required encryption and access controls are implemented.

It is imperative that the appropriate IT service providers for your department be identified and consulted to ensure data protection controls are implemented and verified. Unless specified, all standard services provided by OIT are provided at the sensitive/yellow classification level (see below for classification levels).

- Standard services provided by IT include data center services, storage, desktop support and management, email, web and application hosting.
- Please consult with the designated IT service provider or [LAN Tech](#) for your campus unit for guidance in complying with the encryption and data security requirements of the University.

Be aware of the requirements for [Compliant Printing/Copying](#). Printers, scanner, scanners, and fax machines should be in a secure location and identified as being able to print highly sensitive/red data. This is important when equipment is going to be discarded or not used anymore. You should work with the proper IT staff to ensure that no data is retained in any of these devices, and that, if the data is stored on these devices, that data is securely erased.
Sensitivity of Data as Defined by NC State

The University uses a five level classification system in order to protect data, based on many factors, including risk, litigation potential, regulatory, and contractual. Classification levels are Ultra Sensitive (purple), Highly Sensitive (red), Moderately Sensitive (yellow), Normal (green), and General (white). NC State is responsible for the Confidentiality, Integrity, and Availability of all classification levels but General/white. These classifications each have required and recommended controls for securing the data.

It is recommended that you consult with your departmental IT to determine the best data security plan for your research. There are many different avenues for protecting your data and these are called controls. Some examples include: limiting who has access to what data, use of authentication and authorization software and processes, use of encryption throughout the lifecycle of the data (generation, access, transfer, storage, disposition), monitoring data streams, management of hardware and software used at any point during the research process, formal agreements, de-identification, etc.

Each set of data has a different level of sensitivity (risk to participants) and as a result, a different level of required control that is determined by your specific project (ways to mitigate those risks to participants). These controls change over time as technology changes and evolves. Below you will find information about varying types of data, their sensitivity, and you will be linked to NC State’s requirements for protecting those types of data. Working with your departmental IT will make this easier to implement in your project as well as easier to articulate in your IRB application.

**Ultra Sensitive (PURPLE)** - Represents the most extreme risk to the University and unauthorized disclosure or loss may not be repairable. It is important to eliminate or minimize the collection, use, and storage of this type of data within a research project whenever possible. Ultra sensitive data elements are:

- Biometric Data
- Social Security Numbers/Tax IDs
- Primary Account Numbers and Credit Card Numbers
- Encryption Keys and other Privileged Access Control Information

Use of email and unapproved non-University devices for this classification level is prohibited. If a personal device will be used, an exception request needs to be submitted to provide authoritative review and approval to use a personal device for Highly Sensitive classified data.

In order to manage and protect your Purple Data, work with you departmental IT to meet the requirements listed here for controls for purple data. Describe the agreed upon protections in your IRB application. You may complete the IRB’s data management and security plan (See website) with your departmental IT and upload that with your study.
**Highly Sensitive (RED)** - most research data is considered to be classified at this level. Examples of this data include:

- Driver’s License Numbers
- VINs
- Passport Numbers,
- State Identification Card Numbers
- Mother’s Maiden Name
- Patented and Pre-Patentanted Information
- Controlled Unclassified Information
- Data Controlled by Non-disclosure Agreement

Highly sensitive data must be protected by encryption in transit and while at rest. University email and Google Drive may be used with proper security configuration. All removable media must be safeguarded and encrypted.

Use of unapproved non-University devices for this classification level is prohibited for storage. If a personal device will be used, an exception request needs to be submitted to provide authoritative review and approval to use a personal device for Highly Sensitive/Red classified data.

In order to manage and protect your Red Data, work with you departmental IT to meet with the requirements listed here for controls for red data. Describe the agreed upon protections in your IRB application. You may complete the IRB’s data management and security plan (See website) with your departmental IT and upload that with your study.

**Moderately Sensitive (YELLOW)** is generally data that is comprised of Personally Identifiable Information or Privacy protections are required. At the University, Family Educational Rights and Privacy Act (FERPA) data is usually classified at Moderately sensitive. Use of University email and most storage methods are permitted. Data security requirements include:

In order to manage and protect your Yellow Data, work with you departmental IT to meet with the requirements listed here for controls for yellow data. Describe the agreed upon protections in your IRB application. You may complete the IRB’s data management and security plan (See website) with your departmental IT and upload that with your study.

**Normal (GREEN)** – Is not considered sensitive data. Unauthorized disclosure or loss would not cause any adverse impact on the participant. NC State public sites and communications and materials provided to stakeholders is within this classification level.

In order to manage and protect your Green Data, we suggest you work with you departmental IT to meet with the requirements listed. You can also access information about controls for green data. Describe the agreed upon protections in your IRB application. You may complete the IRB’s data management and security plan (See website) with your departmental IT and upload that with your study.
**General (WHITE)** - This data is not sensitive and the University has no responsibility for Confidentiality, Integrity, and Availability for it. It is often open source data. Examples include Wikipedia, non-NCSU websites, open-source materials. You can access information regarding controls for white data.
The IRB Wants to know the Following Information

- Will University owned or non-University owned endpoints, hardware, software, and/or equipment be used? ([Rule 08.00.18 NC State University Endpoint Protection Standard](https://example.com))
  - During the course of an NC State research project, it will be important to ensure that all endpoint devices (e.g., computers, laptops, tablets, and smartphones) used to access, process, and store data meet the data protection requirements.
  - If a personal device will be used, an exception request needs to be submitted to provide authoritative review and approval to use a personal device for accessing Highly Sensitive/Red classified data. No personal devices can be used for ultrasensitive/purple data (social security numbers and biometric data).

- During the course of an NC State research project, it will be important to ensure that all endpoint devices (e.g., computers, laptops, tablets, and smartphones) used to access, process, and store data meet the data protection requirements.
- If a personal device will be used, an exception request needs to be submitted to provide authoritative review and approval to use a personal device for accessing Highly Sensitive/Red classified data. No personal devices can be used for ultrasensitive/purple data (social security numbers and biometric data).
- RUL 08.00.16 NC State University Security Standards for Sensitive Data and Systems
- Rule 08.00.18 NC State University Endpoint Protection Standard
- FAQs for Endpoint Protection
- Be aware of the requirements for Compliant Printing/Copying. Printers, scanners, scanners, and fax machines should be in a secure location and identified as being able to print highly sensitive/red data. This is important when equipment is going to be discarded or not used anymore. You should work with the proper IT staff to ensure that no data is retained in any of these devices, and that, if the data is stored on these devices, that data is securely erased.

- Who will be the user groups and what level of use is required for each?
- Who will provide IT and any applicable data related services?
  - It is imperative that the appropriate IT service providers for your department be identified and consulted to ensure data protection controls are verified and implemented. Unless specified, all standard services provided by OIT are provided at the sensitive/yellow classification level.
    - Standard services provided by IT include data center services, storage, desktop support and management, email, web and application hosting.
    - Please consult with the designated IT service provider or LAN Tech for your campus unit for guidance in complying with the encryption and data security requirements of the University.

- Data flow and protection methods through the entire data life-cycle
  - collection/creation
    - What data elements will be collected, stored, and used?
    - Who will collect, store, and use data and what equipment, tools, and services will be used?
    - What data formats will be used (hard copy/paper, digital, audio, video, etc.)
  - processing/accessing/sharing
    - Who and how will the data be accessed and/or shared?
  - storage/retention/disposition/destruction
    - What storage methods will be used?
    - What is required for retention and how long does the data need to be kept?
- When and how will the data be destroyed?
- You can also complete the Data Security, Access, and Management Plan found on the IRB website

The IRB Office Recommends:
- Using Zoom for video conferencing
- Using Qualtrics for survey data collection
- Using NC State Google Drive (associated with your unity ID) for cloud storage
- Using a secure VPN while using the internet, especially when accessing and transferring data
- Using NC State managed computers and devices
- Windows Defender is available at University for anti-malware/anti-virus
- For encrypting files using a strong password:
  - Windows: 7Zip
  - macOS:
    - For built-in, the macOS has UNIX command line zip -e ([howto](#))
    - If you like 7zip then Keka is a free implementation (with AES256 option) that is drag and drop on macOS
    - The Disk Utility.app that ships with every macOS device can create encrypted disk images of folders ([howto](#))
Collaboration and Data Security

- **Collaboration within NC State**
  - NC State researchers should work with their departmental IT on setting up secure methods of collaboration as proper access needs proper configurations that your IT folks can help you with.
  - Whatever you and your IT person decide, you must communicate in your IRB application.
  - What your Unity ID can help you access:
    - NC State Google Drives (not approved for Ultrasensitive/purple data but is approved for Normal, Moderately sensitive, and Highly sensitive classification levels for storage.)
    - NC State Network Shared Drives (password protected and with secure access controls)
    - Secured data transmission tools and methods such as Apricorn Aegis products (coordinate with your IT support when using), and ShareBase (data transfer solution). Contact NC State’s ISRA for more information (oit_isra@help.ncsu.edu).
    - Secure NC State VPN Access
    - Secured NC State Servers

- **Collaboration Outside of NC State**
  - NC State researchers should work with their departmental IT on setting up secure methods of collaboration with outside researchers as proper access needs proper configurations that your IT folks can help you with.
  - It is likely that the only difference between internal and external collaboration would be access to shared drives and Google Drives. As noted above, your departmental IT will help configure this.
  - Whatever you and your IT person decide, you must communicate in your IRB application.

**Contact Information for Help and Guidance:**

- OIT Security & Compliance Unit - security@ncsu.edu
- OIT Compliance Officer and NC State IRB Rep
  - Help Desk: 919.515.HELP or help@ncsu.edu
  - LAN Tech List for Campus Units
  - Elizabeth Cole-Walker eacolewa@ncsu.edu
- OIT Compliance Officer and NC State IRB Representative, Darren Fallis dfallis@ncsu.edu
- Complete this training: Data Security in Research Overview 2018 in REPORTER.
Appendix A
Issues Related to Identifiability of Data

Anonymous Data: Any information about a living individual that when collected was never associated with identifiers and that no one was ever able to identify from whom the information was collected. Subjects' identities are unknown to the investigator, not requested, not recorded and not given. There is no possible way that the researcher, research team or anyone else could possibly link the data to the participant.

- When you generate anonymous data – you as the researcher/research team should not collect any type of identifier that would ever allow you to be able to identify a participant/respondent. There should be no way that you or anyone on the research team can know who said/did what.
- When you receive anonymous datasets – the original dataset may have IDs on them, but when that dataset is shared with you – there are absolutely no identifiers on the dataset and no one on the research team receiving the dataset can identify an individual from the dataset (with current knowledge, access, or expertise).

Identifiable Data: Any information about a living individual that is linked, associated with, or contains the name or any details of the individual that would allow someone to be able to directly or indirectly identify a subject from the information collected.

Indirectly Identifiable: Indirectly identifiable datasets have two different meanings:

- Identifiable due to researcher expertise/access/role:
  - The identifiers are considered “readily ascertainable” to the researchers/PIs due to their expertise, access to related information and technologies, and their roles outside of the research.
  - When a respondent can be indirectly identified from a dataset due to the expertise of the PI/researcher, due to the role of the PI/researcher, and/or due to the access the PI/researcher has due to their position outside of the research.
  - For example, you are an admissions officer by profession and you want to use admissions data for research purposes. When you have the data-set, you have the access needed due to your role, to re-identify the data, even if it has been de-identified.

- Identifiable due to data content:
  - When the data shared can be indirectly identifiable to anyone with a modicum of effort.
    - For example, compiling multiple indirect identifiers to be able to directly identify someone. One could potentially use someone’s race, gender, years of experience, and rank to be able to identify a participant.
    - For example, when the content of the data reveals details of an experience that someone could identify the respondent based on details shared within the data.
  - Special Note: sometimes the nature of the data will require the IRB to consider if the data could be identifiable to others outside of the research team, and if so – what is the
likelihood of that data being re-identified and what is the magnitude of harm to the respondents should that data be re-identified. In this case we also take into consideration data handling procedures regarding data protection and security.

De-identified Data: De-identified data refers to a data set created by an NCSU investigator that has had all direct and indirect identifiers removed from the data set when appropriate. They at one time had access to identifiable/indirectly identifiable data, but they have processed the data in such a way that the data no longer has IDs associated with it and the NCSU investigator cannot identify or re-identify respondents from the new data-set.

- De-identified data with codes: Identifiers have been removed from the dataset but can readily be found through the use of a master list that is accessible to the investigator.
  - The link that cross-references the subject’s identity with the code should be stored in a separate location from the data and should be protected.

- Limited Dataset: A limited data set is a type of dataset specifically termed by HIPAA and only refers to HIPAA covered entities and their release of a data set.
  - A limited data set excludes 16 categories of direct identifiers and may be used or disclosed, for purposes of research, public health, or health care operations, without obtaining either an individual's Authorization or a waiver or an alteration of Authorization for its use and disclosure, with a data use agreement.
  - A limited data set allows retention of specific elements of identifying private information: geographic subdivisions, town, city, state, ZIP code, dates, age. Limited data sets are not considered to be de-identified information.
  - Usually when a Limited Data Set is shared, an agreement between institutions/organizations must be in place. A data use agreement is the means by which covered entities obtain satisfactory assurances that the recipient of the limited data set will use or disclose the PHI in the data set only for specified purposes.

The Nature of Re-identifiability and Questions for Consideration

- What access does the researcher and others have that would allow them to re-identify the data?
  - Do they have special access to data due to their professional role?
    - For example: Do they work for a university and have access to registration and records data as a part of their job? Could they possibly/reasonably re-identify an individual due to that access?
  - Do they have special access due to the technology available to them?
    - For example: Do they have access to technology that would allow them to run diagnostics or matching that would allow for re-identification?
  - Do they have special knowledge that others may not?
    - For example: Are they close in relation to the participant in some way and would they know of stories or ideas that would clearly link the data to a participant?
● Do they have expertise that would allow them to re-identify the data?
  ○ For example: Can they take medical records and blood samples and subsequently identify a person from the data?

● What is the likelihood of re-identification from the raw data?
  ○ For example: How likely is it that someone would be able to access the raw data and re-identify someone? The researcher? Someone else?
  ○ Is there information in the data that can be triangulated with other data to re-identify someone?
  ○ Is there information in the public domain that could be paired with the raw data in order to re-identify someone?
  ○ What is the statistical likelihood of re-identification of raw data occurring?
    ■ Of the researcher re-identifying them?
    ■ Of people who know the participant of re-identifying them? Of strangers re-identifying them?

● What is the likelihood of re-identification from the reported data?
  ○ For example: How likely is it that someone would be able to use the reported data and re-identify someone?
  ○ Is there information in the reported data that can be triangulated with other data to re-identify someone?
  ○ Is there information in the public domain that could be paired with the reported data in order to re-identify someone?
  ○ What is the statistical likelihood of re-identification of reported data occurring?
    ■ Of the researcher re-identifying them?
    ■ Of people who know the participant of re-identifying them?
    ■ Of strangers re-identifying them?

● What is the magnitude of harm that the participant could incur based on re-identification?
  ○ Breach of data confidentiality and participant privacy can result in many types of harm. Some harms are transient and some harms have lasting effects.
  ○ These harms are related to the type of risk the data could pose to the participant. These include risks such as employment, reputation, financial, legal, social, academic, psychological, medical, insurability, etc.
  ○ Think about what risk is related to the type of data you have and what that risk could actually do in harming the participant.
    ■ For example, if the data were to be accessed, could your participant be fired from their job? Could your participant be arrested? Could your participant have issues with insurability or with their medical benefits? Could your participant’s identity be stolen? Could your participant be retaliated against in their academic setting? Could your participant lose their social standing with their friends? Their religious organization? Could it negatively affect their relationships?
○ Think about what type of harm you could expose your participants to and what the lasting effects of that harm are - articulate this in your IRB application.
○ Once you have addressed the harms that your participants are exposed to, talk about the likelihood of those harms occurring. This is where your data protection plan comes into play as it decreases this likelihood of harm.
Appendix B
Definitions and Resources

Anti-Malware - Malware is a broad term used to describe all sorts of unwanted or malicious code and software. Although, malware can be viruses, the term anti-Malware refers to software that uses behavior-based and not pattern-based detection rules to identify, repair, and protect your devices and data against malicious code and viruses.

Anti-Virus - Viruses are a specific type of malware (designed to replicate and spread). Anti-Virus is software that uses a set of pattern-based detection rules to identify, repair, and protect your devices and data against malicious code and viruses.

Archiving Data - Data archiving is the process of moving data that is no longer actively used to a separate storage device or repository for long-term retention. Archive data consists of older data that is still important to the study and may be needed for future reference, as well as data that must be retained for regulatory compliance.

Data Disposition - Data disposition is the process of disposing of obsolete or unwanted data when retention is no longer required. All data should have a disposition date and should be in accordance with REG 01.25.12 – University Record Retention and Disposition Regulation and REG 07.40.01 – Disposal of University Property.

Computer Networks - Computer networks can be categorized in several different ways. One approach defines the type of network according to the geographic area it spans.

- Local area networks (LANs), for example, typically span a single home, school, or small office building
- Wide area networks (WANs), reach across cities, states, or even across the world.
- The Internet is the world’s largest public WAN.

Non-Networked Device - a device that is not connected to any type of network.

Firewall - A firewall is a security system that monitors and controls the incoming and outgoing network traffic based on predetermined rules about security. A firewall typically establishes a barrier between a trusted internal network and untrusted external network. Firewalls are often categorized as either network firewalls or host-based firewalls.

- Network firewalls filter traffic between two or more networks and run on network hardware.
- Host-based firewalls run on host computers and control network traffic in and out of those machines.

Password/Passphrase Protection - is a security process that protects information accessible via electronic devices and equipment that needs to be protected from certain users. Password/Passphrase protection is intended to only allow those with an authorized reason to gain access to certain information. Files, folders, drives, laptops, desktops, and servers can be protected.
with a password or passphrase. This protection method is a minimum protection standard and does not suffice for protecting your data (this does not include tools that derive an encryption key from a password).

**Encryption** - the process of encoding a message or information in such a way that only authorized parties can access it and those who are not authorized cannot. The software protects the data by using an algorithm. The algorithm uses a password to derive/generate a key and that key is used to encrypt the data. This password “unlocks” the data which can only be read by individuals with a password to unlock information about that data. You can encrypt files, folders, and hard drives.

- NC State suggests Encryption Software based on use and need. We suggest working with your LAN tech from your unit or any NC State IT professional to determine which software is best.

**Cloud Storage** - Saving data to an off-site storage system maintained by a non NC State affiliated third party.

- Examples Dropbox, iCloud, Google Drive
- For IRB purposes, you should only use the following cloud services:
  - NC State Google Drive (the google drive linked to your @ncsu.edu account) or a dedicated project Shared/Team Drive (note that all access and handling configuration must be used, i.e., read only, no print, share only with specific people, etc.)
  - If you need to use a Cloud Storage system other than Google Drive, work with your IT to determine which is most appropriate and secure for your work.

**Two-Factor Authentication (2FA)** - is a way to confirm a user’s identity where a computer grants a user access only after presenting something the user knows (password, phrase, PIN) and something the user has (e.g., login code, response to a prompt on a mobile phone) or something the user is (e.g., fingerprint, iris). NC State uses DUO for 2FA.

**Virtual Private Network (VPN)** - A VPN allows users to send and receive data across public networks as if they were directly connected to a private network. Here, you can access information about NC State University’s [VPN](#).

**Eduroam** - Encrypted wireless Internet service. Enrolling a wireless device in eduroam at NC State allows you to use that device at any of thousands of other educational and research institutions that participate in eduroam worldwide. [eduroam](#)

**Data Trustee** - Data Trustees are designated oversight authority and responsibility for the portion of University Data (referred to as a Data Element) that is related to the university functions managed and administered by the units and/or personnel who report to them, as delegated by the Chancellor. Each Data Trustees will assign one or more Data Stewards to be responsible for all University data elements managed within the Data Trustee’s domain. NC State Data Trustees are:

- Executive Vice Chancellor and Provost
Data Custodian - Data Custodians are persons who are assigned specific data management responsibilities by the Data Stewards because of their knowledge and position at the operational level. Data Custodians typically will manage access rights to data they oversee. Each Data Custodian may delegate specific custodial responsibilities for different subsets of data under their authority.

Data Steward - Data Stewards are responsible and accountable for the confidentiality, Integrity, and availability of University Data elements within their business or mission area. They shall ensure proper levels of security and protection measures are implemented.

Resources:

- Mobile Device Security Requirements and Recommendations
- Mobile Device Security - Doing it Yourself
- Sensitivity Levels for Data
Appendix C
To Address in IRB Application - Language Examples

● Regarding devices used, select one:
  ○ The data will be stored and handled on NC State University Devices only.
  ○ The data will be stored and handled on both NC State University Devices and Personal Devices.
  ○ The data will be stored and handled on Personal Devices.

● If personal devices used, complete the sentence below:
  ○ Personal Devices used for this project will be protected via _____.
  ○ *make sure the protection plan is in accordance with the sensitivity level of the data.

● If NC State devices used, use the sentence below and make sure you execute it in practice:
  ○ NC State University devices will be managed in accordance with NC State University’s data protection standards and regulations.

● The OIT person I have worked with is ______. They are helping in the following ways _____

● Data will/will not have identifying information on it and will be coded. I will/will not have a master list for this code. The master list will be kept _____ and will be protected by ____. I can/cannot re-identify the data due to ______ (discuss access, expertise, and triangulation of data). The data can/cannot be re-identified by others due to _____ (discuss access, expertise, and triangulation of data).

● The following people have access to all of the data ______. They have access to all of the data because ______. The following people have access to part of the data ______.

● When encryption is used, the following items will be encrypted ______.

● The data will be transferred via _____ and will be protected during transfer by _____.

● When the data is at rest, it will be protected via ______.

● The data will be destroyed in _____ amount of time. The data will be securely destroyed via _____.

If you do not want to use this language, you can complete the “Data Access and Security” form found on the NC State IRB Office Website.