Data Collection and Privacy Considerations

In research, a formal data collection process is necessary to ensure integrity of data and to protect against the risks of unauthorized use of research data. The primary rationale for preserving data integrity is to support the detection of errors, whether made intentionally (deliberate falsifications) or not (systemic or random errors). Equally important is the process to address the privacy and security of a research participant’s identity throughout the data’s lifecycle from collection and transmission, to sharing, storage, and destruction.

Risks of unauthorized use of research data are already associated with traditional data collection means and methods (i.e. via paper or local computers). However, in today’s interconnected world, data is not just locked in a cabinet. Data is held by and analyzed with various methods, and transmitted across institutions over wired or wireless networks (e.g. Cellular, Bluetooth, and telephone networks) that present the risk of intercepted transmissions. These networks are susceptible to eavesdropping and wireless carrier security holes, allowing unauthorized users access to the accounts and usage data. Hackers can use hardware and technologies to intercept and decrypt calls.\(^1\) Use of the Internet adds another complexity to security risks as investigators use cloud services and other offerings that are dependent on a multitude of third-party services such as apps or cloud providers. Such services may enhance or decrease risk, depending on several factors including the nature of the research and the method in which the data will be processed. Research teams must first be aware of these risks, then assess and mitigate them in collaboration with their institutions and IRBs.

All investigators and research staff should become familiar with their institution and/or department information security policies and procedures. Investigators should work with information security experts to review their data collection, transmission, sharing, storage, and destruction procedures to minimize the risk of unauthorized access to, or exposure of, sensitive information.\(^2\) In IRB review, the process of protecting the privacy and security of data should be documented and methods relating to data collection, transmission, sharing, storage, and destruction should be clearly described.

To learn more about online privacy, see Glenn Greenwald’s TED Talk on why privacy matters.

Below are model statements investigators may adapt to describe technology-specific risks to confidentiality or privacy.

Sample: Research Data Collection – No Guarantees Information Will Remain Confidential

There is no guarantee your information will remain confidential during collection. Your confidentiality can be protected by the protections in place on the technology being used, as well as through additional precautions suggested by the research team and other steps you can take personally. While efforts are made to protect your data, confidentiality of your data cannot be guaranteed.

---


Sample: Research Data Collection – Risks to Loss of Privacy
We collect information, including personal information that you voluntarily provide to us when you choose to participate in [name activity: questionnaire, survey, etc.]. When you use this interactive tool accessed through the Internet, there may be some risk(s) to your privacy. For example, we may have outside companies perform services relating to the development, operation and maintenance of this research website or relating to other services. These third party service providers may have access to your personal information, as is reasonably necessary, for their services.

We take great care to protect your information; however there is a slight risk of loss of privacy. This is a low risk because we code your data by separating your personal information (information that can directly identify you (such as your name or phone number) from the research study data. Only a few members of the research team are allowed to see your identifiable information. All others will only be able to see your coded information. The information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permissions or as required by law. The information collected about you will be coded using a fake name (pseudonym) for initials and numbers, for example ABC–123, etc., and the information which has you’re identifiable information will be kept separately from the rest of your data.3 However, even with removal of this information, experts in re-identification may be able to reverse our processes and /or attempt to re-identify an individual given enough cross-reference information about him or her.

Accidental public disclosure may occur such as unintended data breaches by hacking or other activities outside of the procedures authorized by the study. In such a case, your data may be misused or used for unauthorized purposes.

Sample: Research Data Collection – Protections Utilized by Study Team
The following procedures will be used to collect and protect the confidentiality of your study records [Please describe all types of electronic data]. All electronic files [include all the types of electronic files that are used, such as databases, spreadsheets, etc.] containing identifiable information will be password-protected. Such files will also have password protection to prevent access by unauthorized users. Only the members of the research staff will have access to the passwords. Back-up data may be kept on server logs even after this research has been completed. At the end of this study, when investigators publish their findings, the information will be presented in summary format: you will not be identified in any publications or presentations. Data will be kept for the length of the study [list out the length of study in months or years]. After that time all identifiable data about you will be destroyed or de-identified, meaning we may retain and share certain elements of the study records for future research, but we will replace your identifying information with a code that does not directly identify you.