Data Security Procedures

Data security procedures are the formalized, internal processes and standards that the data agencies can employ to protect the administrative data that they collect. Data security requires adequate planning, development of procedures, and training and supervision to ensure that data are stored, archived, or disposed off in a safe and secure manner that preserves the integrity of data.

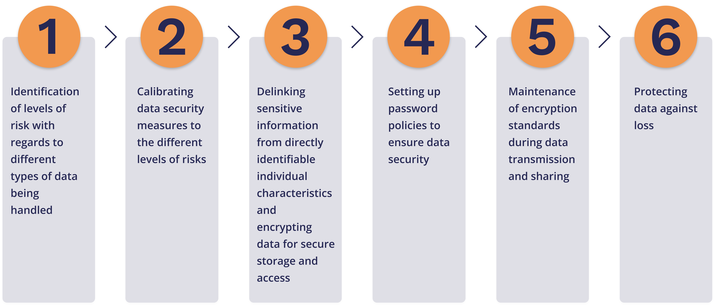
4.1: Rationale behind setting-up data security procedures

Data security is essential for safeguarding private information, maintaining privacy of the subjects, and complying with requirements and regulations. Ensuring data security involves the "interaction of legal, technical, statistical and, above all, human components".

The Five Safes Framework outlines a set of considerations to be kept in mind to check for secure data procedures:
Safe projects: Is the use of the data appropriate?
Safe people: Can the users be trusted to use it in an appropriate manner?
Safe data: Is there a disclosure risk of personal information in the data itself?
Safe settings: Does the access limit unauthorized use?
Safe outputs: Are the statistical results or outputs meeting confidentiality requirements or are released without required data owner's approval?

4.2: Setting-up data security procedures to protect administrative data

Flowchart: Setting-up data security procedures



Step 1: Identification of levels of risk regarding the different types of data being handled

In order to enable the access and usage of administrative data in a secure manner, it is essential to identify safe data flows while considering applicable regulations, confidentiality, data security capabilities, and the potential need to maintain the ability to add follow-up data or new data sources. The first step here is to determine what type of information is stored and processed in your environment, such as medical or financial data, as defined by your organization or by law. The next step is to classify the data by the level of risk associated with handling it; using a standard classification table (that details out the risks involved for each level) For example; Level 1 - Information intended and released for public use (de-identified data which can be published) or Level 2 - Low Risk confidential information that may be shared only within the smaller group (unpublished intellectual property) or Level 3 - Medium risk confidential information intended only for those with a "need to know." or Level 4 - High risk confidential information that requires strict controls (contains personally identifiable sensitive information).

Step 2: Calibrating data security measures to correspond to the different levels of risks

Data security measures must be calibrated to respond to these levels of risk, and corresponding requirements for security can be imposed. We must use information security policies that classify data into levels based on confidentiality and sets standards for use and sharing such that the higher the data level accorded to a type of data, the greater is the required protection. For example, according to the policy, a level 1 classification of data are "information that is considered public", such as research data that has been de-identified in accordance with applicable rule.

Step 3: Delinking sensitive information from directly identifiable individual characteristics and encrypting data for secure storage and access

Data poses the most risk when sensitive information is linked directly to identifiable individuals. Once delinked, the data must be handled separately, and the identifiers should remain encrypted at all times.

In order to ensure safety, the selected sensitive data must be first transformed into an encrypted code that needs a password or pair of "keys" to decipher it. This encryption of data may take place at multiple levels (such as at the device, folder, or file levels), at various phases of the data lifecycle, and using a range of software and hardware packages as well as methods to balance privacy and usability. Once separated, the "identifiers" data set and the "analysis" data set should be stored separately, analyzed separately, and transmitted separately. Once separated, the identifiers should remain encrypted at all times, and the two data sets should only be linked again if necessary to adjust the data matching technique. Access to the "identifiers" data should be limited only to a few personnel.

Step 4: Setting up password policies to ensure data security

To guarantee data security, strong passwords are necessary. Each high-value account should have a unique password. For instance, the passwords for institutional servers, email, and encrypted files should all be unique. Furthermore, passwords must not be shared using file sharing mechanisms or over the phone. It is important to rely on password storage systems and other safer alternatives for password sharing.

The following mechanisms must be incorporated to ensure secure data transactions:
- Methods to handle inactivity or timeouts during remote access,
- Processes to handle non retrievable passwords (i.e. if a user forgets his or her password, the password is reset by the
- System, rather than the original password being returned),
- Restriction on the number of password guesses permitted before account lockout, and,
- Storing access logs that describe who signed in, from where, and when.

Step 5: Maintenance of encryption standards during data transmission and sharing

Data must be safeguarded both at rest and while being transferred between the data provider, data recipients and their collaborators. "Even though using encryption may decrease convenience (a password or a hardware key needs to be used each time decryption occurs), utilizing encryption for data and devices should be mandated as a minimum-security feature as part of any data access mechanism."
A whole-disk encrypted laptop or a secure server that stores encrypted data may or may not provide protection for data in transit. In order to maintain the security of data in transit, it is essential to keep in mind the encryption standards while sharing files in any form or medium. Advanced Encryption Standard (AES) is most commonly used by governments and security organizations as well as everyday businesses for classified communications.

Step 6: Protecting data against loss

Besides securing against threats of misuse, it is equally important for data security to prevent the loss of data. Data and processing information must be backed up securely and regularly, with timely maintenance of passwords. These may be device-level, institutional-level, or cloud-based backups, depending on resource availability and sensitivity of data. An encrypted hard drive may also be used to maintain backups, especially in areas with low connectivity. The access to this storage must be regulated and closely monitored using a digital log system for the personnel intending to use data from the storage. The back-up storage must be updated regularly based on the frequency of update of the data. General practice is to backup every 24 hours, with incremental backup every 3 hours for real-time data, and differential backups for sporadic data flows. The data backup must be protected both physically in case of instances of fire or flood, and also protected from intentional unauthorized attacks by using a strong firewall system.

Reference

- Sub-section on "Data Flow" - J-PAL Research Resources[7]
- Data Classification Table - Harvard University[8]
- Section 5.2 Methods: Balancing Privacy and Data Usability[9]
- Section 2.3.3 Encryption: Chapter 2 Physically Protecting Sensitive Data[10]

- Physically protecting sensitive data - IDEA Handbook
- Data Classification - Administrative Examples by Harvard University and corresponding requirements for security
- J-PAL Guide on data security procedures for researchers
- J-PAL Guide on using administrative data for randomized evaluations
- Resources from MIT's Information Systems & Technology Department:

a. Secure computing
b. Encryption (including software recommendations) and whole-disk encryption
c. Removing sensitive data