NBDC Security Guidelines for Human Data (for Data Users) ver. 7.0

Feb 1, 2024

Ver. 7.0

Introduction

The Database Center for Life Science (DBCLS) / the Joint Support-Center for Data Science Research (DS) of the Research Organization of Information and Systems (ROIS) operates the NBDC Human Database in accordance with the NBDC Guidelines for Human Data Sharing (hereinafter, the Data Sharing Guidelines). This “NBDC Security Guidelines for Human Data (for Data Users)” (hereinafter, the User Security Guidelines) provides the minimum set of requirements that should be fulfilled in order to safely utilize the registered-access data and controlled-access data defined in the Data Sharing Guidelines for the purpose of research activities, while protecting data confidentiality.

The controlled-access data may contain data that could be used to identify individuals in combination with other information. Therefore, measures must be implemented as required for the security level (standard-level (Type I) or high-level (Type II)) designated by a data submitter for each data set.

Because the information technology (IT) environments surrounding data users are diverse and ever-changing, merely complying with the User Security Guidelines may not be sufficient for data security. Data users are responsible for understanding their IT environments to be used for saving and calculating controlled-access data well, and taking additional security measures as deemed necessary, e.g., by referring to the security rules defined by the administrator of each IT environment as well as other guidelines.

The User Security Guidelines will be updated appropriately in response to IT developments.

1. Definitions

1. Controlled-access data, Data
    - The “controlled-access data” as defined in the Data Sharing Guidelines.
2. Data accessible to registered users（Registered-access data）
    - The “registered-access data” as defined in the Data Sharing Guidelines.
3. Principal investigator (PI)
    - The “PI” as defined in the Data Sharing Guidelines.
4. Data user

- Data users who have been approved by the Human Data Review Board for the use of the controlled-access data, persons who have been entrusted by a data user to perform a work under the supervision of the data user, and data users of the registered-access data who have completed the registration for the use of the registered-access data.

5. Data server (see Figure 1)
    - A computer for data users to store and calculate controlled-access data, owned by the data user or the organization to which data users belong, or the “available server outside of affiliated organization (hereinafter, “off-premise-server”)” as defined in the Data Sharing Guidelines. In the IT environment including the data server, it is necessary to satisfy the following (1) to (4) as preconditions (except in the case of using only the off-premise-server).
        - (1) Devices with high mobility, such as notebook PCs that are at high risk of loss or theft are not used.
        - (2) The equipment of the data server and the storage device / medium storing the data are managed by the organization that owns them.
        - (3) When installing the data server in a LAN, the LAN must be owned by the data users’ affiliated organization. In addition, on the LAN on which the data server is installed (hereinafter, data server-installed LAN), a firewall that restricts communication between the external network and the data server-installed LAN must be installed by the network administrator in the affiliated organization, and access from/to the outside is kept necessary minimal (Example: IP address and port of source and destination are limited) to maintain high security.
        - (4) In the data server-installed LAN, if there is a computer used by a person other than the data users, the communication with other computers is appropriately managed by the firewall function.

6. Data access terminal (see Figure 1)
    - A device that is for data users to access data in the data server and that does not permanently save the data locally. When transmitting data between

the data access terminal and the data server, via a communication path outside the data server-installed LAN, it is necessary that all communication paths are encrypted with sufficient strength or that the data themselves are encrypted before being transmitted.

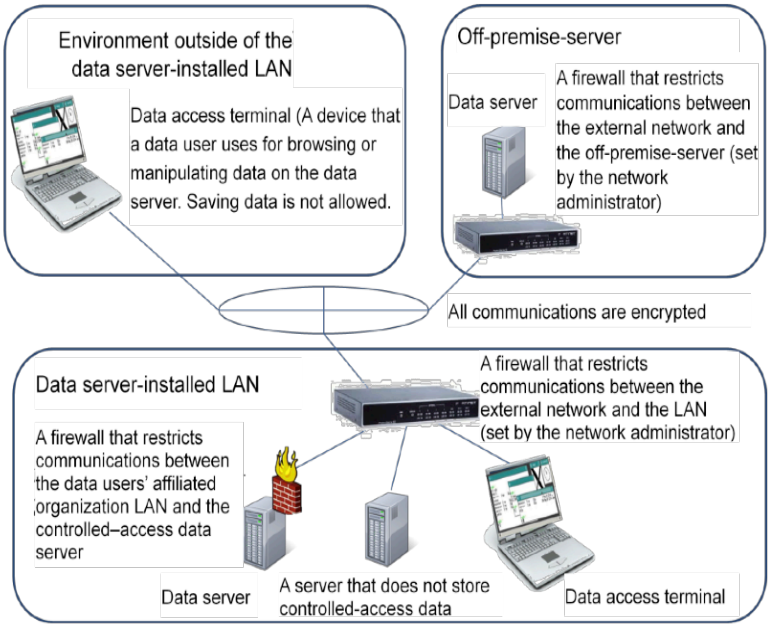

Figure 1 Data server-installed LAN, off-premise-server, data server and data access terminal

2. Measures to Be Taken under Standard-Level (Type I) Security

2.1 Basic Rules for Data Use

Data users must use the controlled-access data based on the following basic rules.

1. Data users must store the controlled-access data in the data server and, in principle, must not move outside of the data server.
2. In cases where it is unavoidable to temporarily move the controlled-access data outside of the data server but within the data server-installed LAN, data users must delete the data outside of the server promptly after use in a way that does not allow restoration.
3. Data users must not duplicate data except for the following cases. In any case, however, duplicated data must be deleted promptly after use in a way that does not allow restoration.
    - Creation of a backup copy of the data
    - Temporary duplication for data transfer
    - Temporary duplication performed by software

4. Access to the controlled-access data is granted exclusively to the data users and must be conducted solely from data servers or data access terminals.
5. Because IT environments surrounding data users are diverse and ever changing, data security is not necessarily guaranteed only by complying with these guidelines. Therefore, data users must understand the IT environment used for data storage and data calculation well, and take additional security measures as deemed necessary based on the security rules specified by the administrator of each IT environment and other guidelines[1][2].

2.2 What the Principal Investigator Must Do

Data use in general

1. The PI should ensure that all data users fully understand and comply with the User Security Guidelines (For Data Users).
2. The PI should confirm that the data users have received an education on information security implemented by their affiliated organization or the like.
3. The PI should keep a record of information regarding data users and the data server (including information on the data storage place in the file system) in an electronic file or the like accessible to only data users, and update the record every time a change occurs. The record must be managed so that the update history can be reviewed.
4. The PI should accept an audit conducted by the Human Data Review Board or a third party commissioned by the DBCLS with regard to the state of implementation of security measures.
5. The PI should submit "Checklist for the NBDC Security Guidelines for Human Data" to the Human Data Review Board Office at the time of application for data use and, in principle, every year thereafter.
6. In case of a security incident such as data breach, the PI must follow the procedure described in “Responsibilities of Data Users” in the Data Sharing Guidelines and take measures such as notification to the DBCLS.

Data server

When an “off-premise-server” is used, the PI must clarify the responsibility sharing with the “off-premise-server“ by means of the server usage rules etc.

1. The PI should prepare for a data server (including a virtual server) and file system that are dedicated to the study as described in the Application Form for Data Use. When there is no choice but to use a server shared with other persons who are not data users, the access to the folders containing the controlled-access data should be limited only to the data users.

2. If a computer used by a person other than data users exists within the data server-installed LAN, the PI should at least enable the firewall functions provided by the operating system (OS) (e.g., iptables in Linux) or equivalent to the function, and restrict communication from/to the inside of the data server-installed LAN properly.
3. The PI should not allow sharing of a user ID or a password for the data server, even among data users. In addition, the PI must set a sufficiently strong password that cannot be guessed by others. (It must be at least 8 characters long. It is desirable to combine numbers, upper case letters, lower case letters and symbols. Do not use those that are easy to guess such as name, phone number, birthday, and the like.)
4. The PI should apply the latest security patches insofar as possible for all software installed on the data server.
5. The PI should not install unnecessary software, particularly, file sharing software (also called file exchange or P2P software; e.g., Winny, BitTorrent).
6. The PI should install antivirus software, and perform virus scan at once whenever moving a file from the outside of the data server. The PI must keep the antivirus software and virus definition file up to date.
7. The PI should not start unnecessary processes as many as possible when the OS boots up etc.
8. Desirably, as security monitoring, the PI should periodically acquire and analyze various logs of the data server.
9. When discarding a device that saved the controlled-access data, the PI should initialize the data storage or physically destroyed in a way that cannot be restored.
10. In case of a security incident such as data breach, the PI should disconnect the relevant data server or data access terminal immediately from the data server-installed LAN.

2.3 What the Data User Must Do

Items 1. through 7. are to be complied with by data users of the registered-access data and data users of the controlled-access data, and items 8. through 13. are to be complied with only by data users of the controlled-access data.

1. The data user should receive an education on information security implemented by the affiliated organization and the like and comply with the security rules set by the affiliated organization.
2. The data user should not allow sharing of a user ID or a password for the data server, even among data users. In addition, the user must set a sufficiently strong password that cannot

be guessed by others. (It must be at least 8 characters long. It is desirable to combine numbers, upper case letters, lower case letters, and symbols. Do not use those that are easy to guess such as name, phone number, birthday, and the like.)

3. The data user should not access data from a terminal application on a device that can be used by many and unspecified persons (e.g., a PC in an Internet cafe).
4. The data user should apply the latest security patches to the data access terminal whenever possible.
5. When leaving a data access terminal, the data user should log out from the data server or lock the terminal. In addition, the terminal should be configured so that the screen is locked after a certain period of inactivity (around 15 minutes).
6. The data user should disable a cache function, if any, which automatically saves data on the data access terminal.
7. When it is inevitable to print data out (including the browse screen capture), the data user should strictly manage the printout to protect the confidentiality of the data, and shred the printout after use.
8. When logging in to the data server from a data access terminal via computational network outside the data server-installed LAN, the data user should encrypt all the communication paths using a sufficiently strong encryption method every time data are transmitted between the data access terminal and the data server or encrypt the data themselves before transmitting them. It is desirable to perform similar encryption when logging in to the data server from the inside of the data server-installed LAN.
9. The data user should not copy or save data displayed on a data access terminal screen to the local disk. It is desirable to use a data access terminal application that does not permit copying and saving of data displayed on the terminal screen to the local disk.
10. When obtaining backup data, the data user should ensure that one of the following requirements is met:
    - The backup is saved on the data server.
    - When the backup is saved in a mobile device (e.g., a tape, USB memory, CD-ROM, notebook PC), the data is encrypted and deleted after use in a way that does not allow restoration. A record of information regarding the mobile device should be kept, e.g., in an electronic file accessible to only the data user, to minimize the risk of theft or loss and to enable early detection of an incident of data theft or data loss.

11. When it is inevitable to use a mobile device for temporary data transfer, the data user should handle the data in the same way as backup data.
12. After finishing data use, the data user should delete all data including all backups from all the devices in a way that does not allow restoration of the data. If the data cannot be deleted by the above method, such as for paper or a mobile device, the data should be physically destroyed in a way that cannot be restored by cutting or the like. It is preferable to delete any temporary file that was generated during computations as soon as it becomes unnecessary.
13. In case of a security incident such as data breach, the data user should immediately disconnect the relevant data server or data access terminal from the data server-installed LAN and report the incident to the PI. In case where an “off-premise-server” is used, the data user should immediately take actions according to the off-premise-server usage measures etc.

3. Measures to Be Taken under High-Level (Type II) Security (except for using only an “off-premise-server”)

In addition to the measures listed in the previous section, “2. Measures to Be Taken under Standard-Level (Type I) Security,” the following measures must be taken with regard to the data server.

1. The PI should place the data server in a server room that meets all of the following requirements.
    - (1) Access to the room is limited, using multi-factor authentication with at least two of the following three authentication methods.
        - Biometric authentication (e.g., vein, fingerprint, iris, and face recognition).
        - Property-based authentication (e.g., IC card, one-time password, and USB token).
        - Knowledge-based authentication (e.g., password).
    - (2) Record of access to the room is automatically obtained and made available for later audit.
    - (3) The server room must be dedicated to the purpose as described in the Application Form for Data Use. If a dedicated server room cannot be set up, the data server must be stored in a locked dedicated server rack.

4. Contact Information for Inquiries about the User Security Guidelines

The Data Sharing Subcommittee Office

humandbs@dbcls.jp

References

[1] NCBI. NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy. (Online) March 9, 2015. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?document_name=dbgap_2b_security_procedures.pdf

[2] Ministry of Health, Labor, and Welfare. *Iryojoho shisutemu no anzenkanri ni kansuru gaidorain* (Guidelines for Security Management for Medical Information Systems) [in Japanese]. Version 5, May 2017. http://www.mhlw.go.jp/file/05-Shingikai-12601000-Seisakutoukatsukan-Sanjikanshitsu_Shakaihoshoutantou/0000166260.pdf