**NBER's data security policies and procedures for hosting confidential data.**

**Revised: November, 2018**

1) **Computing Environment:**

NBER's computing facilities are based on centralized UNIX and Linux servers with network attached storage servers. All of the computer systems with hard drives and other "online" data storage facilities containing restricted data will be kept in our centralized locked server rooms. Physical access to the locked server room is provided via electronic keypad. Entries to the server room are logged on NBER Office Manager's computer, who also manages access to the building premises. Access card/keys to the server rooms are programmed and provided only to NBER IT staff. Should there be a need for an outside IT Consultant, entry into the room will be in the presence of authorized NBER personnel. Except when authorized NBER personnel are physically present in these server rooms, they will be kept locked at all times. Further, the main entrance to the NBER offices is on the 3rd floor of the building, where a receptionist is present. All other entrances to the NBER premises have electronic locks and are monitored with closed-circuit televisions.

Remote login access to authorized users is provided via secure shell, and all of the processing, computation, and analytical work using sensitive data is to take place on these servers and the storage devices. Statistical software such as SAS and Stata, and computational packages such as R, Python, Octave etc. are loaded on these compute servers. So all the analyses including the data can remain within the confines of these servers located in the locked rooms, even though the users may be located at terminals outside of the room. Users pledge not to take out any output containing any identifiable data. All physical (removable) disks and tapes containing restricted data at the NBER are kept in locked cabinets in these locked server rooms. On-site backups are also kept in these locked server rooms. Offsite backups are encrypted and will be taken to a secured locker at Cambridge Trust Company Bank. Only the NBER corporate officers or personnel from IT Staff appointed by the NBER and who will register with the bank may have access to this locker. NBER may also use a commercial secure offsite backup facility(such as CompuVault, or Iron Mountain), that will at the very least include 27/4 security, closed circuit television monitoring, access restricted storage, and tracked media transportation service.

Several additional measures have been put in place on the LAN. Servers for computational use are separated out from systems for general activities such as webserver, mailserver etc. Any programs that transmit information in clear (unencrypted) text, such as "telnet" and "ftp" are disabled on these systems. Research computers are booted "diskless" from a central storage server. They are deliberately booted in read-only mode so even under a breach of an individual research server the central core is not compromised. User account passwords are vetted using "pam_passwdqc" utility, and strong passwords are enforced. A vulnerability scanner (Openvas) runs once a week on the entire address space on the LAN. A honeypot computer has been setup to bait and block intrusion attempts.

Many additional security procedures are applied to communications between the in house NBER computer systems and computers outside the walls of the NBER offices. At our gateway we have an application firewall and a Cisco router that provide additional protection. The application firewall is placed just inside the router in our network topology. We subscribe to realtime virus checking from the firewall vendor. All web and email traffic are subject to this virus checking therefore the office LAN is protected from virus and malware at the gateway. Access controls (acls) are used extensively on the Cisco router. By default, no ports except the ssh port are open to the outside world. Specific ports based on functionality of a server are open for that server. For example, smtp & (s)pop/(s)imap ports are open only for the mailserver(s), while http/https ports are open only to the webserver(s). As mentioned before, research computing servers and storage appliances are not part of such roles and therefore they can be accessed only via the ssh protocol.

**VPN ENVIRONMENT**: Further, the servers hosting confidential data covered by this agreement are placed inside a private LAN isolated from the general office LAN behind a NATting firewall that requires an encrypted VPN connection. These servers inside the private LAN have an IP address in RFC1918 space (a private address space not forwarded to the outside world). Since connection to the private LAN requires an established VPN connection, all transactions to and from the private LAN are tunneled via the VPN in encrypted form. Additionally, each VPN connection is authenticated using a two factor authentication system. The first authentication is based on UNIX username/password (strong passwords are required vetted via pam_passwdqc) via a radius server. This is followed by a phone call/SMS to a registered user phone number that the user acknowledges by answering and pressing # on their phone or the user can provide OTPW (one time password) from a registered hardware token. The IT staff at NBER will register each phone number for each user authorized to use these systems. This mechanism will act as a protection against all password compromises including a remote client that could be virus/malware infected, keystroke logging security breaches, shoulder surfing, phishing, etc. The VPN is configured to lock an account after three consecutive failed log-in attempts. Certain ports are open between the VPN and the non-VPN environment. These are primarily to assist in system administration and monitoring (for example NAGIOS and nrpe, or logging to central syslog service).

Unix user groups/ACLs will control access to directories within these secure servers. Unix groups will be made based on DUA/Project and researchers are appropriately associated with their respective groups. The data management team has its own group. Each DUA/Project will be assigned specified directory space where researchers on the DUA/Project must keep all their files and conduct their analyses.

2) **Transmission of Electronic Confidential Data:**

NBER will receive confidential data from the source agency in encrypted form. The encryption key shall be sent separately from the encrypted data. Confidential data will not be shared with any other institution or facility unless proper approval has been obtained from the data provider.

**3) Storage of Confidential Data:**

All "online" confidential data will be stored within NBER's centralized computing environment as described in item 1.

Original data received from the data provider after being uploaded to these secure servers will be kept in locked cabinets/safe within these locked server rooms or a locked safe. Only authorized personnel will have access to the cabinets.

Onsite backup of confidential data will also be stored on separate servers within the locked server rooms. Only systems administrators will have access to these backup servers.

Weekly, offsite backup of confidential data will be performed on removable disks on backup servers located within the locked server rooms. Data on these removable disks will be encrypted using FIPS 140-2 compliant AES-256 technology. These disks containing encrypted data will then be taken by NBER's IT Director, on a weekly basis, to a bank locker at Cambridge Trust Company, Harvard Square, 1336 Massachusetts Avenue, Cambridge, MA 02138. Only the NBER corporate officers or personnel from IT Staff appointed by the NBER and who will register with the bank may have access to this locker. Each secure locker has the following features. The locker requires two keys, one is in the possession of the NBER, and the other is with the bank official. Both keys are required to open the locker. To access the locker, the authorized NBER personnel would need to produce proper identification and sign in. The vault attendent would accompany this NBER personnel and will use the bank key to open their part of the locker. The NBER personnel will open the locker with their key. The bank premises are secured and are equipped with security cameras monitoring everything. There is a security guard present at the entrance. The lockers themselves are inside a locked area within the bank. Lockers are not shared. The lockers are accessible only during bank operating hours. Robbing a bank is a federal crime. NBER may use a commercial secure offsite backup facility(such as CompuVault, or Iron Mountain), that will at the very least include 27/4 security, closed circuit television monitoring, access restricted storage, and tracked media transportation service.

**4) Authorization Procedure:**

Each project will apply and receive appropriate approvals or Data Use Agreements. A copy of the agreement will be submitted to the Data Custodian/Data Manager and NBER IT. The project will also undergo NBER IRB approval.

Prior to receiving authorization to use restricted data at the NBER, potential new researchers or staff will take a Human Subjects Protection training course. The researcher will then be briefed on the NBER's computing facilities and procedures on maintaining data security in effect at the NBER, and the penalties for failing to comply with these procedures. After completing this training session the researcher/staff will sign a pledge of data confidentiality with NBER.

5) **Access to Data:**

After receiving appropriate authorization, a user will be assigned a login userid. They will be required to create a password that will be vetted by "pam_passwdqc" to ensure a strong password. The user will then be assigned to an appropriate user group(s) based the project DUA(s). Login access to the secure servers will then be enabled for the user. Procedures to login will follow the guidelines described in item 1 (Computing Environment). Unix user groups along with ACL/permission settings will control access to these directories.

All work using the confidential data will be performed on the secure servers. Each project/DUA will be allotted specific directory folders where all the work for the DUA must be conducted. In some cases datasets are large and multiple projects obtain authorization to use the same data. In order to have efficient use of resources of potentially enormous datasets, datasets common to multiple projects will be kept in a common pool of directories with read-only access. However, access to individual files within these common pool of directories will be granted for each project based on data authorized under the project's DUA. The default setting is a denied access for all users but for the data management team, and ACLs will be set to grant access on a per DUA and per dataset basis.

6) **Printouts:**

Users will be instructed to not make any printouts of restricted/identifiable data. Only analytical results and aggregates not containing any identifiable data can be printed.

7) **Data Destruction and Disposal:**

The Principal Investigator of a DUA will inform the Data Custodian/NBER IT staff of closure of a project. At the end of a project (as informed by the Researcher) the restricted data will be deleted from the servers using a program like "shred" or equivalent. Shred will over-write all deleted sectors of the data so the data is not recoverable. Any retired data disks will be first "zeroed out". Any CD/DVD ROMs will be destroyed and magnetic tapes will be degaussed with an onsite degausser. If required by the data provider, the PI/Data Custodian will then complete the Certification of Data Destruction as needed and submit to the source agency.

8) **Termination of access:**

The PI will inform the Data Custodian/NBER IT staff of any termination of participation of a researcher in their project DUA. The NBER IT staff will then remove the user from the unix group membership for those projects, and if necessary remove from login access to the servers that are dedicated for these restricted data analyses or the private LAN itself. Correspondingly, the IT Director will handle access changes for any other member of the NBER IT staff, and the President of NBER will initiate and supervise access changes for the IT Director, should such a need arise. As a secondary step, annually, the Data Custodian/NBER IT staff will present the PI with a list of authorized users and seek an update.

**9) Privacy Breach**

The PI has the main responsibility for informing source of any suspected incidents wherein the security and privacy of the secure data may have been compromised. The IT Administrator(s) will inform the PI if they identify any incidence of breach on the computing environment.