

Wellcome Trust Sanger Institute
HUMAN GENETICS DATA SECURITY POLICY
February 2011

Although almost all human genetics research at the Wellcome Trust Sanger Institute (WTSI) uses anonymised data without directly attached personal identifiers such as names and full addresses, much of it is obtained from collaborators under an obligation to preserve the confidentiality of clinical interactions, and due care with respect to privacy protection must be demonstrable.

The Institute's policy is to ensure that the level of security applied to the access, movement, use and storage of human genetics data sets is commensurate with the risks associated with such processes.

The following institute-wide guidelines aim to ensure consistent data security procedures for this class of data and to enable WTSI researchers and their collaborators to use this data effectively in support of the Institute's mission.

Data Security Guidelines

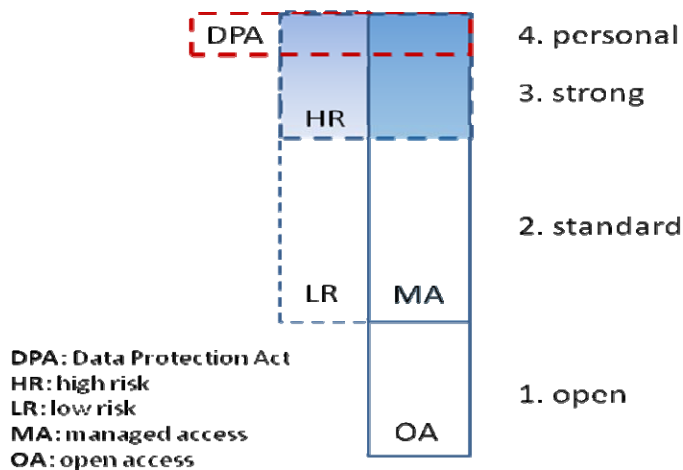
I. General procedures

1. **Data category** - All projects involving human genetics data must be assigned to one of the four data security categories at inception (please see section II. Data security categories). If WTSI researchers receive data from an external collaborator they should confirm the security category with the external investigator prior to transfer of data if possible, else immediately on receipt of data (by sending them a copy of these guidelines and stating the level at which data will be handled). In addition to the specifications of the data security category to which data have been assigned, access agreements may specify additional conditions which must be met.
2. **Responsibility** - For data at levels 2, 3 and 4, data sets must be assigned to a WTSI faculty member (the 'responsible group leader') who will take responsibility for implementation of the WTSI human genetics data security policy, notably for implementing access controls (e.g., UNIX groups) with the support of the IT teams and for compliance with the terms of relevant access agreements.
3. **Register** – A central register of data in categories 3 and 4, as well as data in categories 1 and 2 if required by an access agreement, should be maintained. A brief description of the data set, its category (*Open, Standard, Strong* or *Personal*), name of the responsible group leader and any access agreements associated with the data should be filed centrally. It is the responsibility of the WTSI group leader to ensure data sets are registered. The register will be reviewed annually for group leaders to confirm whether projects are still active,

whether security levels are still appropriate if projects are active, and whether data management is currently consistent with security levels designated. Group leaders may also be asked to provide information on data sets in categories 1 and 2.

4. **Training** - For data in security categories 2, 3 and 4, the responsible group leader must ensure that users are aware of access permissions and usage guidelines, including any conditions of access agreements that govern data use. It should be made clear to WTSI employees and others with access to WTSI computer systems that using human genetics data in security categories above *Open* for which the individual does not have permission is forbidden and may lead to the instigation of disciplinary procedures.
5. **Portable devices** – Remember to store portable devices securely. Never store data in security categories 3 and 4 on USB keys or other portable hard drives, or data in category 4 on a laptop. Data in category 2 must be encrypted on USB keys or other portable hard drives. Data in category 3 must be encrypted when stored on laptops.

II. Data security categories – specific procedures



- **Level 1: Open** (data sets that are published openly without any security restrictions)
No data security restrictions on how the data are handled.
- **Level 2: Standard** (genomic data with relatively limited demographic/phenotypic information)
We expect that most projects will fall into this category. The following data security procedures apply:
 - a) **Do not re-identify** No attempt must ever be made to link genetic data sets to names or other directly identifying information such as full addresses.
 - b) **File access** Data can be held in unencrypted files on the WTSI compute system, with Unix group read/write access for one or more appropriate groups (normally pre-existing) but not Unix world read/write access. Laptops holding this data should have password protected logins and

screenlocks (set to lock after 5 min of inactivity). If the data is held on a USB key or other portable hard drive, it must be encrypted. Though it may be convenient to list everyone in the Unix group(s) used, this is not required – the policy can be met by making clear who may and may not use the files. Data must not be shared beyond the faculty group (or set of listed names if relevant) without explicit approval of the responsible group leader. In some cases, access agreements may require that the list of people who will have access be registered. In this case, it would be the responsibility of the WTSI group leader to ensure that all those who need access have been named, and that nobody who is not named accesses the data.

c) **Transfers**

- **Email** Any data sent by email should be sent to individuals not mailing lists, after confirming with the responsible group leader that the sender is authorised to send the data and the recipient to receive it.
- **Web/ftp** Data may be placed on secure web or ftp sites so long as they require password login and transfers are encrypted, not on open web, ftp sites.

If data are to be sent back and forth between the Institute and collaborating institutions, collaborators should be asked to follow a similar protocol when sending data to us.

- d) **Potentially identifying information** Data at this level should not contain participant names or full addresses. Birth dates/age will usually be approximated. Postcodes should preferably include the first component only. Other potential identifiers may be regarded as inappropriate.
- e) **Check data upon receipt** Data which are received from collaborators containing potentially identifying information that would qualify for a higher level of security should either be modified to remove it, or treated at a higher level of security instead.
- f) **Termination** Once projects are completed, data should be removed from laptops and either also removed from the WTSI Unix system, or archived in an encrypted archive with the permission of the responsible group leader. There is no specific time limit by which this must be done: some data sets and projects will continue to be used in ongoing research, but that should be in the context of an active project.

- **Level 3: Strong** (genomic data with more extensive demographic/phenotypic information, data from small/vulnerable populations)

Stronger security is required when more sensitive anonymised information is stored, or when the conditions of consent or data access explicitly require it. The same policies as for *Standard* security are required, with the following additions/changes. In some cases, additional safeguards may be required, dependent on the project.

- a) **File access** Data can still be kept on the main Sanger Unix file systems, but should only be accessible to named users. Files should either have only user Unix read/write access, not group or world access, or project-specific Unix groups should be used for group access that contain only those names authorised to access the data. User IDs within groups should

be reviewed at 6 monthly intervals by the responsible group leader. Data should not be held on USB keys or other portable hard drives. Data kept on laptops should be encrypted when not in active use, either in individual encrypted files or in encrypted directories/partitions. Users may be asked to sign an agreement addressing their responsibilities with respect to access to such data.

- b) **Transfers (Email)** Data received by email for which the email directories are on a laptop should be detached from the email and encrypted. When email folders are within the Sanger firewall it is acceptable to leave unencrypted data in them. Data can be sent to collaborators by email if it is encrypted, and any encryption keys/passwords should be sent by a different route.
- c) **Potentially identifying information** Data at this level should not contain participant names or full addresses. However more anonymised phenotypic information may be appropriate than at the *Standard* level.

- **Level 4: Personal**

In some cases we may have data that are “personal data” as defined by the Data Protection Act, such as names and full addresses, or are determined by group leaders to require a higher level of security. In addition to the requirements of the *Strong* security level, the following apply.

- a) **File access** The entire data set should not be available on the standard Unix file systems, nor on laptops, and should not be sent by email in any form. It should be available only on a separate file system with limited controlled access. Any remote access to this file system should require a higher level of authentication than simple username/password, such as a card-pad system. Data subsets for analysis that require the main Sanger compute resources to be processed should be extracted from the complete data set, and handled at either the *Standard* or *Strong* level as designated appropriate on a case-by-case basis by the responsible group leader. Such subsets should never include names, full addresses, or other directly identifying information.
- b) **Transfers** If data are transferred to other institutions, data should be encrypted during the transfer and recipients must provide a level of security similar to that outlined in these guidelines.
- c) **Identifying information** Any degree of personal data can in principle be kept at this level.
- d) **Legal** The Data Protection Act may apply and advice should be taken as to how to conform to its requirements.