Health Informatics Centre
University of Dundee

IS 802159 IS 802159

Data Security

PURPOSE

This SOP describes HIC Services' measures to provide security, confidentiality and privacy to data received, managed, and supplied by HIC Services in the following scope of;

- An overview of HIC Services data security
- Access to HIC secure rooms, NHS network and data
- Receiving new data
- Providing a project dataset to data users
- Project Level Anonymisation
- The HIC Services Safe Haven

PROCEDURE

Policy

For overall Policy see Legal and Governance Policy.

Overview

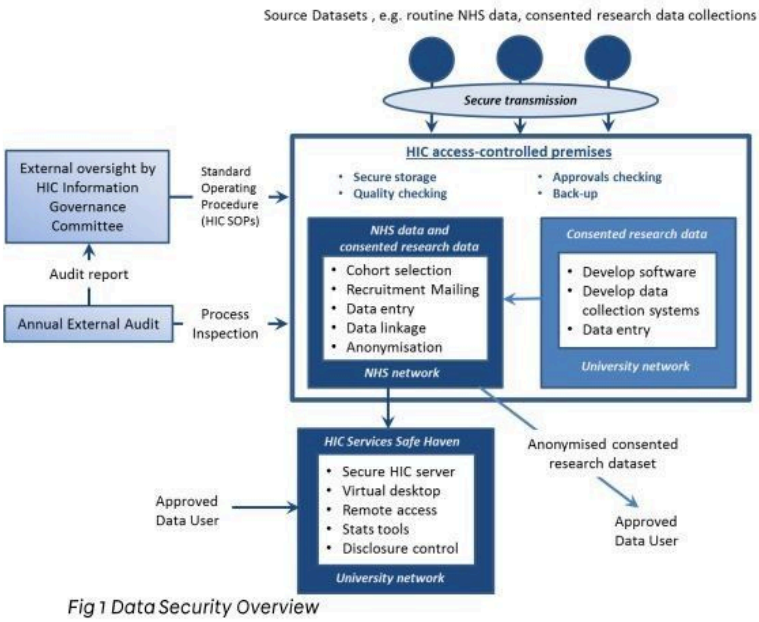


Fig 1 Data Security Overview

1. Identifiable data will be received by HIC Data Analysts via secure, encrypted transmission and will be stored, quality assured and processed within HIC's secure computing environment.

2. The HIC physical premises have access-controls to the rooms containing any physical data. These are always secured when no HIC staff are present.
3. The access-controlled server rooms, accessible only to IT Administrators, contain multiple separate networks:
    a. NHS
    b. University
    c. Cloud
4. Strict login access controls are in place across all HIC networks, including the NHS network and Safe Haven, in line with University of Dundee password best practice guide.
5. Records of all projects, approvals and data releases are kept on the HIC Services Project Management (PM) System.
6. HIC Services are audited at least once annually by external auditors as part of HIC Services ISO27001 Certification.
7. HIC Services procedures are reviewed and approved by the HIC Information Governance Committee. The Committee receives a copy of the annual external audit and the actions HIC Services have taken in response to issues raised by the auditors.
8. New HIC staff, are added to the document HIC Services SOP Appendix C - Contacts and Roles.
9. New staff will not be given access to confidential HIC data or live HIC external production systems until their Disclosure Scotland security check has been completed and a certificate received and unless their job role requires such access. For new HIC Technical Staff not already resident in the UK, a local police record check will be required in place of a Disclosure Scotland check.

Access to HIC Physical Premises Steps

1. HIC staff will have controlled access to secure rooms, as required.
2. All HIC staff will read and sign a copy of the HIC Staff Confidentiality Agreement.
3. Other University staff and visitors will not be given access to HIC's secure physical premises unless they have signed-in and are accompanied by someone that has secure area access.
4. HIC operate a hot desking system where staff are appointed a mobile device / laptop that is their responsibility to keep safe and/ secure in line with the Remote Working section of this SOP.

Access to HIC-Networks Steps

1. Access to specific datasets within the HIC NHS and University networks will be decided and actioned by the relevant HIC Services Process Manager or their designated deputy.
2. Data access will only be provided to HIC Technical Staff on request to specific named datasets.
3. HIC Technical Staff must request access to a named dataset via the Process Manager, or their designated deputy.
4. Each dataset access request must be endorsed by the HIC Technical Staff member's supervisor.
5. When access is no longer required, the HIC Technical Staff member, or their supervisor, will request removal from the dataset's access list.
6. Every 6 months, the network and dataset access lists will be audited to reconfirm the Technical Staff's access. The outcomes of this audit will be stored in the PM System.

Physical Premises Access and Dataset Access Removal Steps

1. When HIC Technical Staff no longer need access to the HIC IT network, or a specific named dataset, their supervisor is responsible for making a request to the HIC System Administrator, copying in the Process Manager, to have their access rights removed.
2. When a HIC Technical Staff member leaves HIC, all physical premises, HIC IT network and dataset access will be removed. This will be triggered by relevant Process Manager, who will inform the HIC System Administrator of the changed access requirement.
3. When HIC Technical Staff no longer need remote access to NHS Tayside environment, their supervisor, or deputy is responsible for making a request to NHS Tayside IT to have this access revoked.

Remote Working Steps

1. HIC policy is to allow employees to work remotely when it has been determined that this will allow work to be performed effectively, securely, and productively, and provided that access to HIC data and networks is:
    - In keeping with the University "Hybrid Working Policy".

    - Via approved University or NHS Tayside VPN access.
    - Data remains on those networks unless otherwise authorised.
2. Only University "Permission to Connect" authorised, encrypted devices are used, as defined in University of Dundee Network Connection Policy.
3. No hard copy data falling into the University classification of confidential (See University Information Classification Scheme) should be taken off site unless approved in advance. Guidance can be provided by the University Archive, Records Management and Museum Services (ARMMS).
4. When working remotely care must be taken to prevent others from being able to view potentially sensitive information.

Clear Screen and Clear Desk Steps

1. To avoid inappropriate access to computing devices and to restrict unauthorised access to information on display screens:
    - Users must log off or lock their computers when unattended.
    - A password protected screensaver, or automated lock screen should be set to activate after a maximum of 15 minutes of inactivity on all devices capable of such.
2. When working remotely and with confidential HIC systems or data, staff will not work in an area that is overlooked, or that allows unauthorised persons to view the systems or data.
3. All information held in a physical format marked as, or falling into the University category of confidential, must be appropriately secured when staff are absent from their workplace and at the end of each working day, to reduce its potential exposure to unauthorised access.

Portable and Personal Devices Steps

1. Laptops, tablets, and other mobile devices connecting to the University or NHS network or holding University or holding potentially confidential organisational information must be protected by a password or pin code in line with the University's password best practice guide.
2. Use of personal devices must be used in adherence with the University's Acceptable use policy, including receiving University "Permission to Connect" and encryption before connecting to the University network. For example, devices must be encrypted.
3. When working in public areas such as restaurants, on trains or aircraft, care must be taken to prevent others from being able to view potentially sensitive information.
4. Any offline files must be copied back to the correct storage location at the next opportunity where applicable.
5. Data on portable and personal devices must be deleted as soon as is no longer required.
6. In the event of loss or theft of a mobile device the user must inform a manager or Team Lead who will then contact the HIC Governance Manager who will follow SOP03 handling significant events.

Releasing a Project Dataset Steps

1. Release of a Project Dataset is undertaken only by a HIC Data Analyst or by a documented and approved delegated process.
2. The Project Dataset will only be released to an Approved Data User by the HIC Data Analyst when all the requirements for such release are met.
3. The release only provides data for the cohort, detailed in the Data Requirement specification which is written and agreed by the Approved Researcher to meet the aims and methods specified in the Project Description. Both are stored on the PM System.
4. An Approved Data User must be linked to the project, identified on the PM System.
5. The Data Governance section of the PM System must have been completed by the HIC Governance Manager, Data Linkage Manager, or other appointed deputy, confirming that all approvals are in place and stored on the PM System.
6. Details of all data releases will be recorded by the HIC Data Analyst on the PM system unless otherwise approved by the Data Controller. The project Data Requirements Specification will be stored on the PM System, and details of data released from NHS datasets will be stored within the NHS Project folder. Project level anonymisation will be carried out on the Project Dataset. This pseudonymisation ensures that each Project Dataset has a unique series of pseudonymised codes. This step is not required if the Approved Project has approvals to receive patient identifiable data.
7. All Approved Project Datasets will be released onto a secure, HIC Services-managed server within the HIC Services Safe Haven environment. Approved Data Users registered on the PM System will be provided a login to remotely access and carry out analyses on their dataset. Data will not be provided directly to Approved Data Users except where:

- Project datasets are of consented participants and the project approvals do not require use of the HIC Services Safe Haven.
- Data is being exported to other Safe Havens.
- Data is being sent to a data supplier or controller with appropriate approvals already in place.
- If the PM system is not available, any required governance checks will be recorded on paper, and appended to PM system as soon as practicable, once access is restored.

Project-Level Pseudonymisation

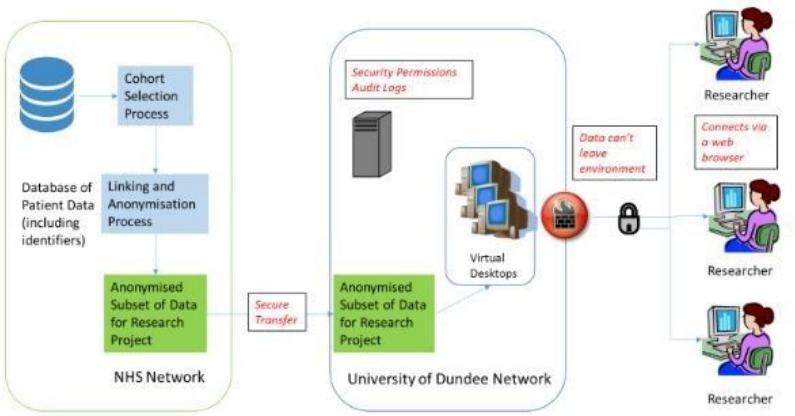
1. HIC's pseudonymisation process utilises various methods, using the method deemed most sensible based on the format and its use. Primarily, these anonymisations will be by:
2. Mapping – Also known as pseudonymisation. A random mapping is generated to replace the identifiable data item. The mapping is stored securely by HIC Services to allow reidentification if required.
3. Dilution – The identifiable data item is diluted to a point where it covers a much larger range of people, for example, reducing postcode (average of 15 households) to postcode district (average of 20,000 people).
4. Removal – removing the data item completely by not offering it for extraction or release.
5. Obfuscation – hiding (and thus removing) the identifiable data in some way whilst retaining the main aspects of the data. For example, this could be free-text data where names have been replaced with a placeholder, or image data where pixels containing identifiers are over-written and replaced by a coloured box.
6. Every project dataset will be uniquely pseudonymised before being made available to an Approved Data User unless identifiable data is required and is specifically approved by the Data Controller.
7. The Pro-CHI is generated by the HIC Data Analyst to uniquely pseudonymise a typical NHS dataset CHI number. This pseudonymisation is carried out by generating a mapping between a CHI number and a randomly generated Pro-CHI. The pseudonymisation cannot be reversed without the mapping table. This table, and therefore the capability to both pseudonymise and reidentify is located on a secure server located on the NHS network and is accessible only by the HIC Data Analyst and the System Administrator.
8. The pseudonymisation process also removes all names and addresses from the dataset. As standard it also:
    - Uniquely pseudonymises the CHI into a project-specific pseudo-CHI (Pro-CHI)
    - Dilutes the date of birth – standard pseudonymisation is 3 months whereby the day will become '01' and the month will become the middle month of the appropriate quarter year e.g. 24/01/2005 will be pseudonymised as 01/02/2005).
    - Dilutes the postcode-to-postcode district / outward postcode only by removing the last 3 digits, e.g. DD1 9SY becomes DD1.
    - The GP code, the GMC number (General Medical Council registration number), the GP Practice code and the Pharmacy code are all replaced by pseudonymous versions.
    - If data is being provided via HIC Services that has already been effectively pseudonymised elsewhere, this procedure may not need to be repeated by HIC Services.
    - Where there is no CHI a unique pseudonymised identifier will be allocated to each individual.
    - Aggregated data provided for study feasibility will not show values <5.
9. Over-writing identifiable portions of images with new pixels, usually of a single colour, thus removing the identifiers completely from the image.
10. Metadata in DICOM (Digital Imaging and Communications in Medicine) or equivalent image files is anonymised using software with any fields which may contain Identifiable data being processed according to the rules set here, for example free text is removed, patient CHI is replaced by Pro-CHI. Free text fields, which may contain Identifiable data, are not provided as part of a Project Dataset unless specific Caldicott Guardian approval is obtained.
11. These steps will help ensure that identification of individual patients is not possible, while retaining the ability to link patient data across multiple data sets for a particular project.
12. Any request to the HIC Data Analyst for more detail about any of the Anonymised Data items listed above will be treated as a request for identifiable data, which will require study specific Caldicott Guardian approval. The Pro-CHI is project-specific ensuring that data provided to any one project remains within the bounds of linkage as approved for that project.
13. The Pro-CHI allows traceability of any HIC Services Project Dataset to the project the data was released for, and the subsequent Approved Data Users for that dataset.

Reversing the Pseudonymisation Process Steps

1. There are occasions when it is necessary to reverse the pseudonymisation process and go back to the original source of the data. For example:
2. Over the course of a study additional data is sometimes required to help achieve the outcomes of the study, potentially from a wider data source. Individuals would need to be identified to be able to request this additional data.
3. To validate findings, e.g. from information in the patient paper file.
4. To identify individuals who may, for their own benefit, need further tests or treatment (patient safety). This action would only be initiated by the opinion of a qualified clinician collaborating with the study.
5. When reversal becomes necessary, permission must be sought from the appropriate Data Controller or relevant delegated authority such as Caldicott Guardian approval.
6. This permission must specify which individuals can have access to any identifiable data (e.g. a patient's GP, or the Approved Data User who will be viewing patient files).
7. Once this approval has been obtained, it will be recorded in the PM System.
8. Details of the request from the Approved Data User and release of any identifiable or additional pseudonymised data will also be added to the PM System.

HIC Services Safe Haven Overview

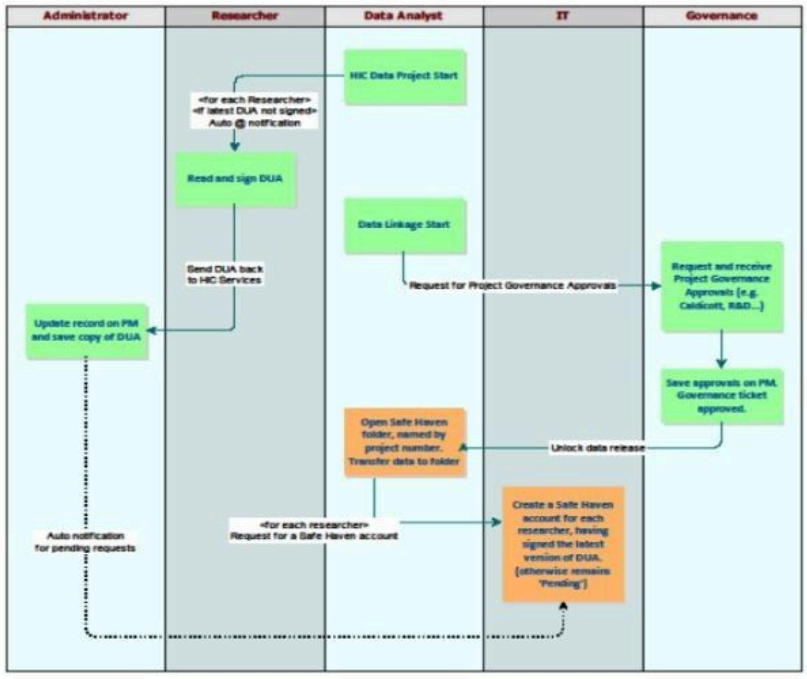
The infrastructure model is illustrated below:



1. The HIC Services Safe Haven utilises a secure remote-access environment. In this model data are no longer released externally to Approved Data Users for analysis on their own computers but placed on a HIC managed server, within a restricted, secure IT environment, where the Approved Data User is given remote access to analyse it. Access to the HIC Services Safe Haven by Approved Data Users is restricted and controlled by access control lists and security groups.
2. Commonly used tools for data analysis are provided for use within this environment and Approved Data Users can securely access this environment from anywhere. The Approved Data User's access is restricted. Users are not able to print, copy and paste out of the environment, or access the internet. Approved Data Users are not permitted to copy individual-level data outside the environment via any means including, for example but not limited to, photographic, recording, screen grabbing and note taking.
3. User-specific files (e.g. look-up tables, stats scripts) can be imported to a user's personal folder via a HIC System Administrator after review by a HIC Data Analyst or by following a documented delegated process approved by the HIC Data Analyst. Users can make copies available to project collaborators via the shared project folder if required.
4. To aid researchers in reuse of cleaning techniques applied to data in other projects, files can be transferred from other projects with prior agreement. This work must be carried out by a HIC Data Analyst and entails:
    - The pseudonymisation process being reversed, then reapplied for the new project to prevent cross project linkage.
    - The file being restricted to cohort members.
    - The file being restricted to datasets and fields already available for the project, and as documented in the Data Requirement specification.
5. The System Administrator will provide details about current available software and versions on request.

6. New software can be added on request, providing it is licenced appropriately, and poses no security risk to the environment. HIC reserve the right to refuse to add tools to the environment, or to remove them when they are seen to affect the security of the environment.

Safe Haven Account Creation Steps



1. All account creation requests must be approved by a HIC Data Analyst and a project's Principal Investigator. Accounts will only be granted to allow an Approved Data User to access their active Approved Project. All project approvals must be in place and all Approved Data Users will need to have read and signed the latest HIC Services Data User Declaration and completed appropriate training, before being given access to the Safe Haven.

Output Disclosure Control Steps – Unconsented NHS Data Studies

1. Individual-level data are not permitted to be removed from the HIC Services Safe Haven, only analysis outputs and user created documents or code e.g. reports, summaries, aggregates, graphs etc. may be removed.
2. To enable a file to be removed from the Safe Haven environment, the Approved Data User will follow the appropriate process documented in the HIC Trusted Research Environment knowledge base.
3. HIC reserve the right to withhold any files prior to output to allow for completion of a detailed risk assessment.
4. HIC expect requests for Artificial Intelligence or Machine Learning (AI/ML) models to be removed from the Safe Haven. These models and some other file types are often binary files, or similar, where it's not possible to directly inspect the contents for the presence individual-level data.
5. If HIC are unable to ascertain whether individual-level data is included in requested output files, HIC will assess the request in the frame of risk – risk of the files including individual-level data, and risk to disclosure of personal data and to the rights and freedoms of the individuals who the data is about. To assess risk, HIC will consider criteria such as:
    - HIC AI/ML Triage Form
    - Model Attack Report
    - Data Minimisation
    - Model Risk Assessment questionnaire
    - The ethics and transparency of the output requestor
    - Contractual Agreement (where an external 3rd party is involved).
    - DPIA
6. HIC are seeking satisfaction that outputting the model presents minimal risk. Analysis and outcomes will be recorded and stored in PM System. Criteria will be judged together, lack of one does not mean HIC will withhold the requested output. The final decision will be

documented and stored in PM System.
7. All output files will be reviewed by a HIC Data Analyst or documented and approved delegated process, once verified as safe to output, files will be made available to the Approved Data User.
8. HIC will only release files where they are confident that the risk of them containing or revealing individual level data is minimal.

Output Disclosure Control Steps – Consented Bio-Resource Data Studies

1. To enable a file to be removed from the Safe Haven environment, the Approved Data User will follow the appropriate process documented in the Safe Haven User Guide. Individual-level data are permitted to be removed from the HIC Services Safe Haven, but must obey the following rules:
    - No dates will be included in the release of derived data.
    - Study ID
    - Sex (M/F)
    - Age (not DOB)
    - Diagnostic or event status (aggregated from multiple hospital entries and multiple ICD10 codes and other sources)
    - Aggregated Biochemical values. (e.g. mean untreated Cholesterol)
    - Drug response. (e.g. model derived beta or odds ratio; or absolute or percentage change in biochemical parameter)
    - Drug adherence (% prescription encashment)
    - Duration of treatment (time-not calendar period)
2. The output files will be reviewed by a HIC Data Analyst, System Administrator or delegated documented process approved by the HIC Data Analyst. Once verified as meeting the above rules the file will be emailed to the Approved Data User.

Importing and hosting Research Datasets from external sources Steps

1. Research data not supplied by HIC, may be hosted in the HIC Services Safe Haven for analysis by Approved Data Users. This will require an approval from the appropriate Data Controller for the data.
2. The Data Controller can request the removal of the data.
3. Output disclosure control will be assumed to meet the same needs as Unconsented NHS data studies unless approved and agreed otherwise by the Data Controller.
4. Synthetic or open data must also have approval from the Data Controller or be evidenced as acceptable to use in its documentation or licencing agreement.

Transfer of Data to and from HIC Steps

1. Datasets will be managed and processed by HIC Data Analysts or appointed deputy, or by delegated documented process approved by the HIC Data Analyst.
2. Datasets will be transferred by the HIC Data Analyst through an agreed process.
3. HIC Services will require written approval from Data Controllers (e.g. Caldicott Guardian for NHS data) prior to releasing data from these datasets.
4. Identifiable data will be received by approved methods of transfer documented in this SOP.
5. Datasets should not be transferred via portable media (e.g. CD/DVD, memory stick or portable storage)
6. Large scale data including, but not limited to, imaging and genomics datasets may be transferred on encrypted storage in cases where the network infrastructure is not capable of transferring the required volume of data, e.g. limited bandwidth availability where data cannot be transferred in an acceptable amount of time without disruption to NHS clinical and business network traffic. In the case of NHS identifiable data, these must be NHS approved items.
7. Datasets should not be transferred via unsecured non-NHS email. If non-NHS email must be used, the dataset itself must be secured using an approved method.
8. When HIC Services receives data via an unapproved method, the data will be transferred to a secure server on the HIC IT network, and any portable media used archived, cleared, or securely destroyed if practicable. The sender will be told not to use portable media or unsecured email for any subsequent transfers of data.
9. On receipt of the dataset, the HIC Data Analyst will load the data onto secure servers. The servers reside on the HIC IT network and are only accessible to the HIC Data Analyst and the System Administrator.

10. If data is to be routinely loaded, HIC will use RDMP to do so where possible. RDMP creates an Event Log which can be accessed within RDMP itself, which records:
    - The number of INSERT and UPDATE operations performed on the live database.
    - The username, start and end time, unstructured progress messages of the data load.
    - Errors.
    - Source filenames.
    - Linked to archived source file, which is stored after the successful data load.
11. HIC Business Support Team staff may have restricted access to add CHI numbers to datasets, as required, using the appropriate software program. Access is only granted to those staff members trained to carry out this task.

HIC Approved Methods of Transfer Steps

1. data transferred to and from HIC will utilise one of the following methods:
2. Internal NHS email
3. NHS Secure File Transfer system.
4. Transfer via secure web method (https).
5. Secure File Transfer Protocol (e.g. SFTP).
6. Alternative transfer methods. If the above methods are not available other transfer methods such as unsecured email or FTP may be considered. Any file must be secured using one of the following approved secure methods:
    - Advanced Encryption Standard compression (e.g. AES-256 Zip).
    - Asymmetric encryption (e.g. PGP Public Private Key cryptography).
7. Identifiable, unconsented NHS data will always be transferred within the NHS environment only.

APPLICABLE REFERENCES

- Data Access Approvals
- Significant Events
- Archiving a Project Dataset
- Staff Confidentiality Agreement
- Data User Declaration
- Legal and Governance Policy
- For Definitions see ISMS Glossary
- Appendix C – Roles and Contact Details
- University of Dundee Acceptable Use Policy
- University of Dundee Remote Access Policy

DOCUMENT CONTROLS

Process Manager	**Point of Contact**		
Chris Hall	hicbusiness-support@dundee.ac.uk		
Version Number Major is defined by if it is a revision of content/procedure and approval is required. It's represented as: **3**.0, **4**.0, **5**.0. Minor is defined as not requiring approval and	**Effective Date**	**Edited By**	**Edition Details**

includes small changes to spelling, grammar and formatting. It's represented by 3.**1**,3.**2**, 3.**3**.			
1.0	01/09/13	Duncan Heather (HIC Governance Manager)	This is a new SOP which expands on a subset of policies and supersedes Version 6 of the HIC Standard Operating Procedure, Management of HIC Datasets. Effective from 1st August 2011.
1.1	15/05/14	Duncan Heather (HIC Governance Manager)	Edit governance diagram to add "HIC Governance Committee"
2.0	01/08/14	Duncan Heather (HIC Governance Manager)	Add new section: Access to HIC secure rooms, NHS network and data
3.0	01/01/15	Duncan Heather (HIC Governance Manager)	General review and update
4.0	01/08/15	Duncan Heather (HIC Governance Manager)	Added sections for Teleworking, Clear screens and desks and Personal & Portable devices
5.0	01/01/16	Duncan Heather (HIC Governance Manager)	Extended types of secure data transfer methods permitted and added more detail about Safe Haven restrictions for users. Revised safe haven diagram at 5.8 to reflect move away from Citrix
6.0	01/04/16	Duncan Heather (HIC Governance Manager)	Add details re: restricting release of free text (5.7.1.4), anonymisation of image files (5.7.1.9), transferring image files (5.9.1.4 & 5.9.1.5) and updating name of Governance Committee to Information Governance Committee

7.0	01/12/17	Duncan Heather (HIC Governance Manager)	5.1.1.4 changed from UoD Policy to UoD best practice guide. 5.1.1.6 updated to reflect ISO 27001 Certification requirement. 5.2.1.2 removed unnecessary verbiage. 5.2.2.2 updated to Non-Disclosure agreement 5.2.3.6 updated to reflect appropriate roles
8.0	08/11/18	Duncan Heather (HIC Governance Manager)	Added section 5.8 reversing the anonymisation process- moved from HIC SOP 03 and 5.6.1.8 to reflect a backup process if PM system is down when releasing data
9.0	17/03/20	Duncan Heather (HIC Governance Manager)	Updated staff access section 5.2. Added new 5.9.4 section about releasing individual-level derived data from Safe Haven for consented genetic studies. Amended section 5.3 Teleworking
	14/05/20	Rachael Torano (Business Support)	Updated with new and approved Header Changes made to cover page to include Op's review date as agreed at the HIC Information Governance Meeting As agreed by the HIC Governance Committee signatures have been removed from cover page and Document Review and Revision History table has been updated to include review dates and notes
10.0	15/01/21	Duncan Heather (HIC Governance Manager)	Added to security check requirement for non-UK resident new HIC staff in 5.2.1.2
	11/11/21	Jenn Johnson (HIC Business Support)	Updated header and footer. This revision does not need to be approved by Committee's as previously agreed and does not constitute a version change.

	17/01/22	Duncan Heather	Reviewed, no changes made
11	14/12/23	Symone Sheane Chris Hall	General content review and update. Updated into new template.
11.1	19/02/23	Bruce Miller	Moved SOP from SharePoint Site to Confluence

Last Review Date	**Triggers** • Annual Review • Internal Audit (conducted by TASC annually) has identified opportunities for improvement (OFI) or nonconformities (NC) • External Audit has identified OFI or NC • Significant Event has highlighted OFI • Client Complaint has highlighted OFI • HIC All Staff (individuals of) have identified OFI
December 4th, 2023,	HIC All Staff (individuals of) have identified OFI

Approved by the HIC Ops Group Date	**Approved by the HIC Exec Date**
December 4th 2023	N/A