Data management for MSc summer projects

Introduction

As a researcher, it is important that you demonstrate your ability to effectively manage research data and keep it secure in accordance with current good practice. This document outlines your responsibilities and provides references where further information and support can be obtained.

1. Data security requirements

Research at LSHTM that involves the processing of personal data on individuals must comply with the General Data Protection Regulation (GDPR) 2016/679 and UK Data Protection Act 2018, which set out provisions and requirements for how it is handled.

LSHTM students should read and ensure they comply with LSHTM's Information Security policy and its supporting documents, available at https://www.lshtm.ac.uk/aboutus/organisation/information-management-and-security.

2. Using personal devices in research

It is recognised that it is often convenient for LSHTM students to use their own personal devices for work purposes. LSHTM permits such use, on condition that requirements and guidelines outlined in the following documents are followed:

- LSHTM Mobile and Remote Working Policy
  https://www.lshtm.ac.uk/sites/default/files/2018-05/LSHTM-mobile-and-remote-working-policy-Apr2018.pdf
- LSHTM Bring Your Own Device (BYOD) policy
  https://www.lshtm.ac.uk/sites/default/files/bring-your-own-device-policy.pdf

3. Data sources

All research involving humans, their tissue and/or their data must have ethical approval (and any other required approvals) in place before the project can begin, this includes accessing or being provided with datasets that are not fully in the public domain. When you submit your CARE form (required for all MSc projects) the Research Governance and Integrity Office will assess the project and inform you whether ethical review is required.

MSc students preparing for their summer project may obtain data from several sources:

3.1. Supervisor-provided data

Your research supervisor or other LSHTM staff may be willing to provide a sample dataset to analyse. The LSHTM repository at https://datacompass.lshtm.ac.uk/ provides a list of datasets generated by LSHTM staff.

3.2. Collecting primary data

Students collecting primary data should obtain participant consent for their research in accordance with LSHTM-SOP-005: Informed Consent for Research[1].

Web-based surveys:

LSHTM currently hosts several open source survey tools, including Open Data Kit (ODK), which may be used by students performing semi-structured and structured surveys in their research. Survey forms are accessible via up-to-date web browsers and mobile devices. On completion of the survey, data is securely transferred to a 256-bit encrypted ODK server hosted at LSHTM in London. Further information is available at http://opendatakit.lshtm.ac.uk/

[1] https://lshtm.sharepoint.com/Research/Research-Governance/Pages/standard-operating-procedures-(sops).aspx

Interviews and focus group discussion:
Students should follow good practice when performing interviews. The Oral History Society has published advice on remote oral history interviewing during the Covid-19 pandemic https://www.ohs.org.uk/advice/covid-19/.

Personal data must be managed in compliance with GDPR. Servers must be located within the EU and provide functionality such as encryption, user authentication and other security features. The use of cloud-based tools such as MS Teams is permitted.

Some students use audio transcription/translation services to process interview recordings. A list of suppliers can be found at https://lshtm.sharepoint.com/Services/Procurement/Pages/Suppliers/Translation.aspx.

3.3. Finding public data

Research data is often made available openly or via an application process. Several specialised search engines may be used to locate research data:

- DataCite Search: https://search.datacite.org/
- Elsevier Data Search: https://datasearch.elsevier.com/
- Google Dataset Search: https://datasetsearch.research.google.com/
- DataMed: https://datamed.org/

Other examples can be found by searching for 'Find research data' on ServiceDesk.

4. Data classification

Students are encouraged to classify and label their research files based upon its confidentiality and risk level, using the levels described in the LSHTM Data Classification and Handling Policy[2]. Specific care should be taken when handling Confidential and Highly Confidential Data.

Type	Description	Examples
Public	Information that is accurate and can be made available without risk to confidentiality	Information for use in public reports and the web.
Internal	Anonymised data which is restricted to LSHTM members & specific collaborators. Disclosure may result in some inconvenience or annoyance, but this is minor and recoverable.	Project document, anonymised data that can't be re-identified, training materials, non-sensitive meeting minutes.
Confidential	This applies to personal and sensitive personal data. Data is restricted to specific individuals and can't be made available further without significant impact.	Interview notes, datasets with sensitive personal data
Highly Confidential	Limited to specific individuals working in a very restricted manner. Risk of significant legal liability, severe distress/danger to individuals, and/or organisational reputation impact.	Health data relating to identifiable individuals; Bank details.

5. Data storage and security

Consider where research files will be held and who will have access to them at each stage of your research:

5.1. Using electronic devices in research

Electronic devices that contain research data must be encrypted to protect personal and confidential from unauthorised access. This requirement covers desktop computers, laptops, tablets, mobile phones, USB sticks, external hard drives, and other storage media.

[2] https://www.lshtm.ac.uk/sites/default/files/data-classification-and-handling-policy.pdf

Many operating systems contain built-in encryption software, but this is often disabled by default and must be setup by the user. Tutorials are available for Windows 10 BitLocker[3], Apple MacOS FileVault[4], iOS devices (iPhone, iPad)[5], and Android devices[6]. Alternatively, you can use a free encryption tools such as VeraCrypt[7] to encrypt specific locations on your device.

Encryption tutorials and frequently asked questions are available on LSHTM ServiceDesk:
https://lshtm.topdesk.net/tas/public/ssp/content/detail/service?unid=18ed4f8c03854d5e99aa0b83edc61829

6.2. LSHTM storage systems

LSHTM provides several storage systems for use by MSc students. These are listed in the LSHTM Storage Options guide at https://www.lshtm.ac.uk/files/LSHTM-data-storage-options.pdf. Care should be taken to choose storage that is appropriate to the classification level being handled.

6.1. Accessing data via Horizon

Horizon Remote Desktop provides a convenient method to access network files and applications via a web browser. However, due to the limited number of LSHTM licences available, Horizon should be used only when the task cannot be achieved using local tools. Research applications available in Horizon, include Stata, EndNote, R, SAS, SPSS and ArcGIS. Horizon can be accessed via Windows, Mac, Linux, iOS and Android devices via http://horizon.lshtm.ac.uk/. Tutorials are available on http://servicedesk.lshtm.ac.uk/

6.1. Using cloud services

Caution must be taken when using free cloud services. These services are often provided with little or no guarantee of service, which may result in files being deleted, corrupted, or unavailable when needed. Factors to consider when choosing data storage include:

- Where is data stored? Is the location compliant with ethical and legal requirements?
- What measures are in place to prevent loss, unauthorised access and change? Access controls, backups, etc.
- Are systems monitored and do audit logs exist to identify potential breaches?
- If you're using an app or storage/processing facility provided by a 3rd party, what does the data processing agreements allow them to do? Some service provider agreements permit them to use data for any purpose.
- Can you delete data from the service and have it completely removed? How long will copies remain?

These services *MUST NOT* be used to store personal data, due to the difficulty in determining the exact location of a server (which may result in a data protection breach) and the possibility it will be accessed by unauthorised users.

Contact IT Services via http://servicedesk.lshtm.ac.uk/ for advice on the use of cloud services.

6. Anonymisation

Anonymisation is a process performed to remove personal identifiable information contained within a dataset. A person's identity can be determined using many types of information:

- Direct identifiers such as name, address, GPS location, email, username, photograph, voice recording
- Indirect identifiers used in combination to recognise a person. E.g. age, occupation, employment location.

A Data Protection Impact Assessment (DPIA) should be performed to decide the type of information needed for research and that which requires processing. This should take into account: the likelihood of re-identification using

[3] https://support.microsoft.com/en-gb/help/4028713/windows-10-turn-on-device-encryption
[4] https://support.apple.com/en-gb/guide/mac-help/mh11785/mac
[5] https://support.apple.com/en-gb/guide/security/sece3bee0835/1/web/1
[6] https://source.android.com/security/encryption
[7] https://www.veracrypt.fr/en/Downloads.html

the information provided; the anonymisation techniques available that may be applied; and the feasibility of addressing the research question after anonymization has been performed. Students may contact the LSHTM Data Protection Officer (DPO@lshtm.ac.uk) for advice.

Common techniques applied to reduce the risk of participant identification include:

- Removing or replacing direct identifiers, such as names, addresses, etc.
- Re-sorting and replacing unique IDs that can be used to link the dataset, e.g. NHS number
- Reduce the accuracy of values recorded, e.g. place age into 5 year categories
- Limiting/removing upper and lower ranges that may be identifiable, e.g. patients aged 100+ years

When performing anonymisation action, it is important to maintain a record of the processing performed.

Consult LSHTM-SOP-036: Confidentiality and Anonymisation of Research Data for further information https://lshtm.sharepoint.com/Research/Research-Governance/Pages/standard-operating-procedures-(sops).aspx

Further guidance:

- UK Data Service: https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation.aspx
- Data sharing for demographic research: Managing deductive disclosure risk https://www.icpsr.umich.edu/web/pages/DSDR/disclosure.html
- LSHTM Service Desk: How do I anonymised and redact qualitative data? https://lshtm.topdesk.net/tas/public/ssp/content/detail/knowledgeitem?unid=937b22c8a97c4c30b9278279b20650f8
- Data Management for interview and focus group resources in health https://doi.org/10.17037/PUBS.04646631

7. Data retention and destruction

Research data that contains personal and sensitive information must be disposed of securely after the end of the project by a date agreed with the supervisor. This requirement covers all copies of the data, irrespective of whether it is in digital or physical form. The student should work with their supervisor to ensure disposal arrangements are appropriate and perform action necessary to remove the data in most circumstances.

A record and the reasons for destruction of essential documents should be documented and signed by a person with appropriate authority. The record may be kept as an MS Excel spreadsheet that lists the documents destroyed.

If a data provider requests a Data Destruction Certificate or similar document that confirms data has been fully and unrecoverably deleted, contact IT Services via https://servicedesk.lshtm.ac.uk/ for advice.

LSHTM procedures for data destruction are outlined in LSHTM-SOP-043, available at https://lshtm.sharepoint.com/Research/Research-Governance/Pages/standard-operating-procedures-(sops).aspx

Support services for MSc students

LSHTM provides a range of IT services for student use. Please visit:

- IT Services for students: https://lshtm.sharepoint.com/Services/ITServices/Pages/students.aspx
- Service related information and FAQs: http://servicedesk.lshtm.ac.uk/

Version control

Version	Date	Authors
1.0	11 September 2020	Gareth Knight, Peter Wright, Phil Rogers