

# How to submit data to LSHTM Data Compass

## Overview

This document outlines the process for submitting research data and other reusable resources to the LSHTM Data Compass, the institutional research data repository.

### 1. Login to the data repository



Figure 1: Login screen

1. Visit <https://datacompass.lshtm.ac.uk/> and press 'Deposit Items' or 'Login' to go to the login screen
2. Enter your LSHTM username and password (i.e. the same used to access your email) and press the Login button. If successful, you should be presented with a 'User Workarea' screen.

The User Work area outlines

[1] The metadata records that you are able to create

[2] Config options for the type of item listed in your work area

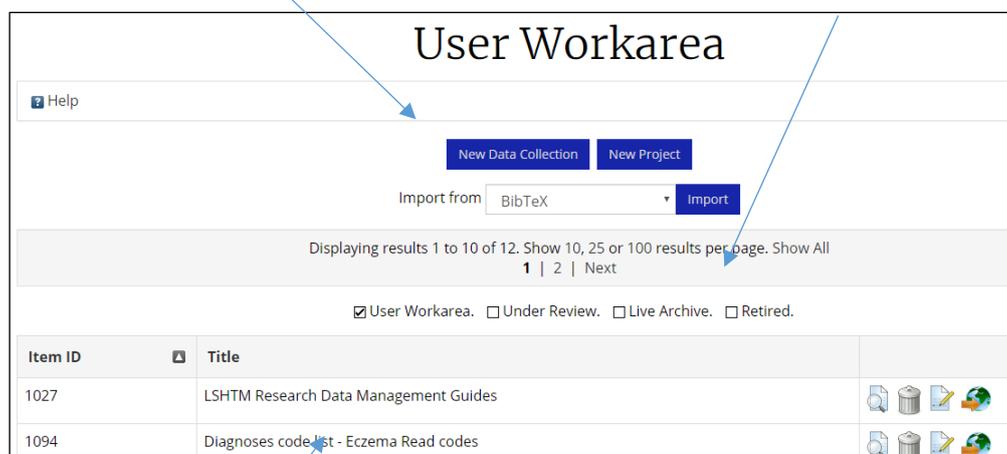


Figure 2: User work area

[3] A list of data records assigned to the user

[4] Actions that can be performed – view, delete, edit, submit

The following actions may be performed:

#### 1. New metadata record

LSHTM users are able to create two types of metadata record:

- a. *Data collection*: A record used to describe a research output (dataset, code, search strategy) and supporting material necessary to understand and use it. If in doubt, this is the one that should be used.
- b. *Project*: A virtual group used to link several data collections held in LSHTM Data Compass and assign a single DOI to cite them. See <https://doi.org/10.17037/DATA.00001266> for an example.

In most cases, you will need to create a Data Collection record. Press the '[New Data Collection](#)' button to display the metadata entry form,

## 2. Record list configuration

By default, the work area will list only work-in-progress records that have not been submitted. However, you can view submitted records by selecting the relevant checkbox:

- Under review: Records that have submitted for review, but have not been made public
- Live archive: Records that are publicly available in the data repository
- Retired: Records that were publicly available, but have been subsequently removed, e.g. they describe a dataset held on a 3<sup>rd</sup> party server that is no longer available.

## 3. Metadata records list

A list of metadata records for which the user is the designated owner.

## 4. Metadata records list

A set of icons that can be used to perform various actions

Icon	Description
	View record. An edit button can be clicked to modify specific fields.
	Delete record and associated files
	Edit record
	Submit record for review

## 2.1. Describe your data's content

The first page collects information on the research output to be made available. Complete the form and press the 'Next' button to move to the next page. Mandatory fields (M) are labelled with a star/asterisk symbol

The 'Save and return' button will save the item to your work area for future editing.

Field	Status <sup>1</sup>	Description	Examples
Title	M	The name of your dataset. This should differ from your paper's title to avoid confusion	IDEAS project - Outbreak Risk dataset Diagnoses code list - Eczema Read codes Data for: "Paper title"
Alternative Title	O	A second name can be used to identify the dataset.	E.g. The title in a non-English language
Description	M	Write a brief description of the data content (e.g. variables contained) and how it was collected.	Data produced as part of a study to explore the incidental effect of ivermectin MDA on 5-day mortality amongst Anopheles mosquitoes. Each row refers to a mosquito within the study, indicating their species, capture date/time, day of censoring, and whether it was alive or dead on collection.
Keywords	M	Subject/scientific terms that make the resource easier to find. Examples can be found at <a href="https://datacompass.lshtm.ac.uk/view/keywords/">https://datacompass.lshtm.ac.uk/view/keywords/</a>	Hookworm Public-private partnership
Data type	M	Indicates the type of resource described. E.g. quantitative data, qualitative data, software. The free-text field can be used for specific details <sup>2</sup> .	Dataset – qualitative Software – R script Other – data collection tools
Data capture Period	O	Start and end date of data collection	1 September 2013 – 30 October 2014
Date Information	M	A key date in the lifecycle. E.g. completion date, deposit date	1 October 2019
Data capture mode	O	A multi-choice field use to specify the collection methods. See Appendix 1 for examples.	"Compilation/ Synthesis"
Description of data capture	O	A free text field that may be used to describe the methodology applied to create/capture data. Relevant details may include sample size, procedures applied, location	Full enrolment lists for each school were collected by the study team, compiled as master enrolment lists. These were distributed to teachers in five-week periods and completed daily for each child.
Geographic area covered	O	The broad region where data collection took place. Up to 4 locations may be added using the Google Maps drawing tool	
Language(s) of written material	O	A multi-choice field used to specify one or more languages in which content is written.	English
Additional Information	O	A free text field for other relevant details.	

Table 1: Data collection form - 1st screen

<sup>1</sup> M = Mandatory, O=Optional

<sup>2</sup> Data type is used to populate the categories at <https://datacompass.lshtm.ac.uk/view/resourcetype/>.

## Data Collection Mode

A multiple-choice field used to specify data collection methods. The LSHTM data repository uses the following controlled list provided by the Data Documentation Initiative Controlled Vocabularies Group.

Term	Description
Aggregation	Statistical information that has been processed (e.g. averaged, totalled) to the point where individuals cannot be identified.
Compilation/Synthesis	The combination of several, often heterogeneous sources to create new content
Content Coding	The conversion of qualitative information into quantitative
Experiment	The manipulation of one or more variables included in a hypothesis
Experiment: Field Intervention	An experiment performed in a natural, uncontrolled setting in which a researcher manipulates 1 or more variables and measure its impact.
Experiment: Laboratory	An experiment performed in a controlled physical setting, in which a researcher manipulates 1 or more variables and measures its impact
Experiment: Web-based	An experiment that implements an artificial situation via the web
Focus Group	A group discussion organised for research purposes
Interview	An interviewer addresses questions to one or more participants.
Measurements and tests	The application of specialised instruments, techniques or standards to assess specific properties of a person/item/process/phenomena. E.g. education achievements, physical properties, personality traits
Observation	Data collected through observation of a person, event, or other phenomena, without attempt to intervene or manipulate variables
Observation: Field	Observation conducted in a natural, non-controlled environment,
Observation: Lab	Observation conducted in a controlled, artificially created setting
Recording	Registering by mechanical or electronic means. E.g. audio, video
Questionnaire	A Self-administered questionnaire in which the respondent provides a response, without live interviewer participation
Writings and Diaries	Participants keep an audio/video/text diary or write a story
Simulation	Modelling or imitative representation of real-world processes, events, or systems.
Summary	The presentation of information in a condensed form
Transcription	Recordings or live observations are converted into a written form
Unknown	For use when the data collection mode are unknown
Other	For use when data collection mode is known, but not found on the list

Table 2: Data collection mode terms

### 2.2. Describe how it was created

The second page collects information on the study and people who created the item.

Field	Status <sup>3</sup>	Description	Examples
Creator(s)	M	The family Name, forename, email and ORCID (if possible) of LSHTM & non-LSHTM people associated with the data. This is used to generate the citation.	Jane Falconer orcid: 0000-0002-7329-0577
Associated Roles	O	LSHTM & non-LSHTM researchers who had an indirect role in the research. See below for role types.	<ul style="list-style-type: none"> <li>• Contact Person</li> <li>• Data Collector</li> <li>• Data Manager</li> <li>• Distributor</li> <li>• Editor</li> <li>• Hosting Institution</li> <li>• Producer</li> <li>• Registration Authority</li> <li>• Related Person</li> <li>• Researcher</li> <li>• Research Group</li> <li>• Rights Holder</li> <li>• Sponsor</li> <li>• Supervisor</li> </ul>

<sup>3</sup> Status: M = Mandatory, O = Optional

			<ul style="list-style-type: none"> <li>• Project Leader</li> <li>• Project Member</li> <li>• Registration Agency</li> </ul>	<ul style="list-style-type: none"> <li>• WP Leader</li> <li>• Other</li> </ul>
Email of LSHTM Data Contact	M	Email address of the corresponding author. This is particularly important for controlled access data	jane.falconer@lshtm.ac.uk	
Associated publication in LSHTM Research Online	O	A look-up field used to link the data to any published papers in LSHTM Research Online that has used it	Safeguarding human health in the Anthropocene epoch: report of The Rockefeller Foundation-Lancet Commission on planetary health.	
Related resources	O	Web resources associated with the data, e.g. project website, journal publication.	https://doi.org/12345	
Project Title	O	The title of the study in which the data was captured/used	School-based Programme of Malaria Diagnosis and Treatment in Southern Malawi	
Grant number	O	The research funder's grant number		
Funder/sponsor	O	One or more organisations that financially supported the work. This field will autocomplete for common funders.	Medical Research Council Wellcome Trust	
Participating Institutions	O	One or more institutions to which the researcher(s) belong and other organisations that made a significant contribution.	London School of Hygiene & Tropical Medicine, London, United Kingdom	
LSHTM Faculty/Department	O	One or more LSHTM faculty & departments in which LSHTM researchers are based.	EPH : IDE	
Research Centre	O	One or more LSHTM Research Centre's associated with the research <sup>4</sup>	Cancer Survival Group	
Research Group	O	Other LSHTM research groups to which researcher(s) belongs		
Copyright Holder	O	Institutions or individuals that own rights over the data E.g. the funder, institutions in which creators are based	London School of Hygiene & Tropical Medicine, London, United Kingdom	
Publication details	M	The name and location of the organisation/system that is making the data available. This will be LSHTM in most cases, unless the record refers to a dataset hosted elsewhere.	London School of Hygiene & Tropical Medicine, United Kingdom  Figshare	
Embargo date for metadata record	O	Use only if the metadata publication & DOI registration must be delayed as part of a publication blackout.		

<sup>4</sup> This is used to generate the list at [https://datacompass.lshtm.ac.uk/view/research\\_centre\\_feed/](https://datacompass.lshtm.ac.uk/view/research_centre_feed/).

Identification Number	O	Ignore - used only when recording data held elsewhere.	
-----------------------	---	--	--

Table 3: Data collection form - screen 2

### 2.3. Upload files

The third page of the workflow allows you to upload and describe files for inclusion in the data repository. Files can be uploaded using the 'Choose file' button or dropped onto the area shown below.

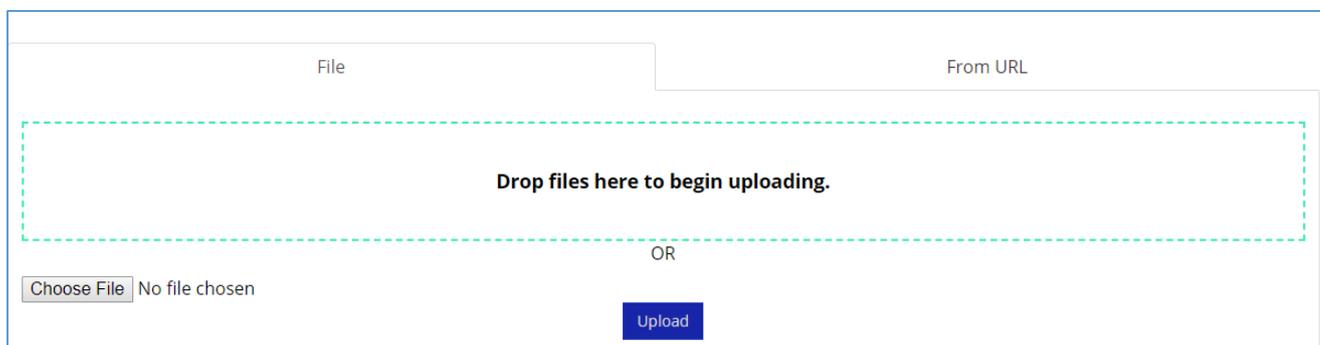


Figure 3: Upload files menu

Each data collection in LSHTM Data Compass is different. However, they often contain four resource types:

- The research output: A de-identified dataset, code list, search strategy, software code, etc.
- Supporting documentation: A codebook that contains information necessary to understand the output. E.g. a table that lists the name and description of each variable in a dataset, permitted values and their meaning.
- Ethics forms: Participant information sheet & Informed Consent forms that outline ethical permissions for data capture and sharing
- Processing scripts: Code associated with the data, e.g. R code, Stata Do files

The following metadata should be assigned to each file.

Field	Status <sup>5</sup>	Description	Examples
File description	M	A short description of the file's content. This may duplicate detail elsewhere	Dataset containing variables on 128 participants, covering X, Y and Z.
File purpose	M	A controlled list used to classify the file. See below for list.	Study instrument
Visibility	M	A controlled list used to specify the security level to be applied. See below	Visible to all
Licence	M	Licence conditions associated with the file that specify how it may be used.	CCBY - recommended for open data Data Sharing Agreement – a custom licence for restricted data
Content type	M	A controlled list used to describe the file content. See below for controlled list.	Dataset Audio
Retention period	O	The time period the file should be held.	Indefinite

Table 4: File-level metadata

#### Purpose of File

File purpose	Description
Data	Qualitative or quantitative data held in a machine processable form, e.g. CSV
Data visualization / representation	A visual representation of the data, e.g. a bar chart held in PDF

<sup>5</sup> Status: M = Mandatory, O = Optional

Documentation / metadata	Supporting information that enables the data to be understood.
Study Instrument	A questionnaire, interview guide, test, search strategy, or similar resource
Licence form	Documentation that specify conditions for access and use for the data, e.g. a Data Sharing Agreement or Data Transfer Agreement.

Table 5: Purpose of file

#### File Visibility

Access level	Description
Anyone	File can be accessed by all users, without registration
LSHTM and other registered users	File can be accessed by users with an LSHTM login. Other users must complete and submit a request form.
Request access for all	Users must complete & submit a request form, which is sent to a corresponding author.

Table 6: A controlled list used to specify the security level to be applied to the file.

#### Content Type

Type	Description
Unspecified	Default value. An error will be displayed if item is submitted without changing this value.
Compressed Archive	Files stored in a ZIP archive or other compressed format
Audio	Sound files, such as WAV, AIFF and others
Dataset	Qualitative or quantitative data
Video	Video recordings, such as AVI, MPG, MOV and others
Interactive Resource	A tool that includes some form of user interactive, e.g. a learning resource
2D/3D model	A 2D or 3D model created using a CAD package or similar tool.
Script	A script file created to process the data in some way, e.g. Python, STATA DO file.
Slideshow	A presentation such as those created in PointPoint or Prezi
Software	Source code or compiled version of a software tool
Textual content	Content stored as unstructured text
Still Images	Photographs or other images
Other	Everything else

Table 7: A controlled list used to specify the content contained within the file

#### 2.4. Deposit item

Finally, review the licence agreement and press 'DEPOSIT ITEM NOW' button to submit the item for review.