

# Create a Dataset record in LSHTM Data Compass

## Purpose

This document outlines the process for creating a Dataset record and uploading files to LSHTM Data Compass (<https://datacompass.lshtm.ac.uk/>).

## What is a Dataset record?

A Dataset is a type of metadata record that has been published in LSHTM Data Compass for the purpose of describing a research output produced by, or with the involvement of, an LSHTM researcher.

Examples of research outputs that are eligible for inclusion including survey data, interview transcripts, executable software, code and scripts, search strategies, survey questionnaires, interview guides, and other resource that support the research process.

A Dataset record may be created to describe research outputs hosted in LSHTM Data Compass or in an external repository, such as Figshare. The latter is used to ensure LSHTM maintains a list of research outputs produced with its involvement and acts as a source for the LSHTM Current Research Information System (CRIS) and staff profiles.

If you have produced several research outputs in your research study, it is advisable to create a Dataset record for each, so that it can be assigned a unique DOI. Research outputs held in LSHTM Data Compass can be linked to indicate their relationship to each other and the broader research study.

## Sign in to the repository

To sign-in to the data repository, click the “Login” button, located on the sidebar or the top-right corner of the screen.



A login screen should be displayed, where you are prompted to enter your LSHTM username and password. This uses the same authentication system as other LSHTM services. Your username may begin with ‘lsh’, ‘eph’, ‘itd’, ‘php’ or similar – it is not your email address.

If you have successfully logged into the repository, a home icon should appear on the top right of the screen and a “Workarea” option will appear in the navigation menu.



Once you have successfully logged in to the repository, you will be taken to a ‘workarea’ page that lists metadata records that you have created and/or own.

## Understanding the Work Area

The Workarea is your personal homepage in the repository. Here, you can view any Collection or Dataset records you've previously edited, as well as create new ones.

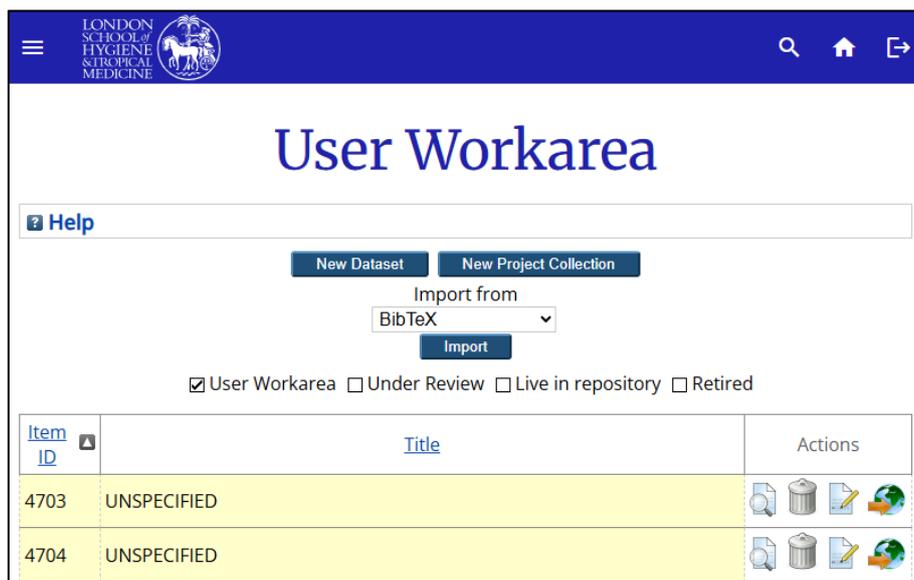


Figure 1: User Work Area screen

By default, the Workarea lists only draft items that you can edit. You can also view submitted item by selecting the appropriate checkboxes:

1. User Workarea: Draft items that are in your work area for editing.
2. Under Review: Items submitted for review that have yet to be published in the repository. These items cannot be edited, but you can recall them to your work area to make updates.
3. Live in Repository: Approved items that are available in the repository. These items cannot be edited while they are live, but it is possible to create a new version of the record.
4. Retired: Items that have been withdrawn from the repository.

Labels and colour coding are used to differentiate between the different item types

The Actions section indicates actions you may perform upon an item in the repository:



The View button displays a preview of how the item will appear in the public repository.



The Delete button removes an item from your work area. Deleted items cannot be recovered. This is only displayed for items editable in your work area; it will not appear for items under review or public.



The edit button will take you to the editing page for an existing record.



Submit item for review.

## Create a Dataset record

To create a Dataset record, click the “New Dataset” button at the top of the screen.

The Dataset workflow consists of four pages:

1. Resource description: Describe the research output that has or will be made available – its title, content, how was it acquired, and where.
2. How was it obtained: Outline the people and projects that were involved in its acquisition - creators, contributors, funders, and associated resources.
3. File upload: Upload and describe files associated with the record. File upload is optional.
4. Licence: Review and accept the licence agreement for submission to the repository.

You can navigate between steps using the page name or the “Next” button. Use: “Save and return” to save progress and go back to the Workarea; or “Cancel” to exit without saving.

★ Mandatory fields are labelled with an asterisk symbol.

### 1. Describe the content

The first page collects core information on the research output to be made available.

Name	Description	Status
Title	The name of the research output. This should indicate its content ( data, code) and differ from other research outputs, such as publication titles.	Mandatory
Alternative title	A second name used to identify the data, if needed	Optional
Description	A brief description of the research study, the content of the research output, and other relevant information	Mandatory
Resource type	A controlled list indicating the type of resource– data, code, etc.	Mandatory
Resource description	A text description of the resource type, e.g. qualitative data	Optional
Keywords	Subject or scientific terms that will make the resource easier to find	Mandatory
Barnard subject terms	Medical subject terms relevant to the research output. Type the first few letters of a word to look-up the relevant term. Multiple terms may be added to each line. The Barnard Classification scheme is available at <a href="https://doi.org/10.56920/lshtmp-1">https://doi.org/10.56920/lshtmp-1</a>	
Data collection period	The time period when information contained in the resource was created, captured, or generated.	Optional
Date information	The date that the research output was submitted to the LSHTM repository or a third party platform. This is used for citation purpose.	Mandatory
Capture method	A multi-choice field used to specify the method of acquisition or creation.	Mandatory
Description of data capture	Describe the methodology used to collect or obtain information. Relevant details may include the population being study, sample size, procedures applied, and location.	Optional
Geographical area covered	The broad region where data collection took place. A maximum of 4 locations may be added using the Google Maps drawing tool.	Optional
Language(s) of written material	A multi-choice field used to specify one or more languages in which content is written.	Optional
Additional Information	Additional information on the resource that is not provided elsewhere, e.g. data availability statement.	Optional

Table 1: Dataset record workflow – first page

## 2. Describe how it was created

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

Name	Description	Status
Creator(s)	One or more people associated with the creation of the research output who should be included in the citation. Provide their ORCID if known.	Mandatory
Associated Role(s)	The name and ORCID of people who had an indirect role in the research. These names will not appear in the citation. See below for role types.	Optional
Contact email for Corresponding authors	Email address for one or more people who are able to answer questions on the research output and make decisions on access requests.	Mandatory
LSHTM Research Online look-up	A look-up used to link the item to any papers in LSHTM Research Online that have used it. Type the paper title to autocomplete.	Optional
Related resources	Web accessible resources associated with the research output. For example, a project website, journal publication, alternative host, etc. If the record is describing a resource held elsewhere, the website type should be set as 'Online Data Resource'.	Optional
Funder information	<p>If the research output was produced as part of a funded research study or trial, please provide relevant details, such as the title, funder name, and funder grant number.</p> <p>The funder URI is a unique machine-readable ID used to link the research output to the funder. If the funder is not found, their website can be listed.</p>	Optional
Participating institution(s)	One or more organisations that have made a contributed to the item. E.g. the researcher's host institution, an organization that has have made a financial contribution to its development.	Mandatory
LSHTM Faculty / department	A multi-select list used to indicate the LSHTM faculty/unit and department/group in which each creator and contributor are based.	Optional
Research Centre	One or more LSHTM Research Centre's associated with the item.	Optional
Research Group	Other research groups with which Creators are affiliated that are not indicated in the Research Centre list.	Optional
Copyright holder(s)	Institutions and/or individuals that hold rights over the item.	Mandatory
Publisher	The name of the organization or platform that has made the item available. E.g. Figshare.	Mandatory
Place of publication	Location of the organization that has made the item available (if relevant)	Optional
Embargo date for metadata record	<p>State any embargo period that apply before the metadata can be published.</p> <p>Used only if metadata and DOI are covered by publication blackout that prevents further disclosure.</p>	Optional
Identification number	The unique ID for this item, e.g. a DOI. This is a system-defined field.	Optional

### 3. File upload

The third page is used to upload and describe files associated with the data record. If you are creating a metadata only record, this page can be bypassed by pressing the NEXT button.

Files can be uploaded using one of three methods:

- Choose file: One or more files can be selected on the depositor’s computer.
- Dropper tool: One or more files can be dropped onto the web page to upload them.
- From URL: Imports a file from a public source, such as an external repository.

After upload, please open the files to verify they have been added correctly.

Files added to the record can be edited using one of four options.



A simple file format Converter



Displays the file MIME type and checksum.



Alters the order in which files are listed on the record page



Delete file

### File description

Name	Description	Status	Examples
File description	A brief description of the file's content. This may repeat details found elsewhere.	Mandatory	Dataset containing variables on 128 participants, covering X, Y and Z.
File purpose	The controlled list specifies the purpose of each file. Files without an assigned type will not appear on the record page.	Mandatory	
Visibility	This controlled list specifies the access level for the file. See below	Mandatory	
Embargo date	An optional file embargo can be set – this specifies a date when a file will be made openly available.	Optional	
Embargo reason	Explain the reason an embargo is needed.	Optional	
Licence	Licence conditions that specify how the file may be used.	Mandatory	CCBY - recommended for open data  Data Sharing Agreement – a custom licence for restricted data
Content type	The controlled list specifies the content of the file. See below for controlled list.	Mandatory	
Retention period	The time period the file should be retained, before it should be reviewed.	Mandatory	Indefinite

Table 2: File-level metadata

### 4. Licence agreement

The final page contains a licence agreement that depositors must agree to as part of submission. Once the licence agreement has been read, press ‘Submit Item for review’ button to notify the repository administrator that it is ready to check.

## Controlled lists

### 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 3: Data collection mode list

### File purpose

File purpose	Description
Data	The primary research output to be made available, e.g. qualitative/quantitative data
Data visualization / representation	A visual representation of the data, such as a bar chart in a PDF.
Documentation / metadata	Supporting information that enables the data to be understood.
Study Instrument	A questionnaire, interview guide, search strategy, or similar resource
Licence form	Documentation that specifies access and use conditions. E.g. Data Sharing Agreement.

Table 4: Purpose of file

## Access level

Access level	Description
Anyone	File is available to all, without registration
LSHTM and other registered users	File is accessible to users with an LSHTM login. Others must complete a request form.
Request access for all	Users must complete a request form, which is sent to custodians (corresponding authors) for review

Table 5: A controlled list used to specify the security level to be applied to the file

Content type	Description
Unspecified	Default value. An error will display for items submitted without changing this value
Compressed Archive	Files stored in ZIP or other compressed format
Audio	A sound file, such as WAV or MP3
Dataset	Qualitative or quantitative data
Video	A video recording such as an MPG
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 Powerpoint presentation or similar
Software	Source code or compiled software
Textual content	Content stored as unstructured text
Still Images	Photographs or other images
Other	Everything else

Table 6: A controlled list used to specify the content contained within the file