Create a Data Collection in LSHTM Data Compass
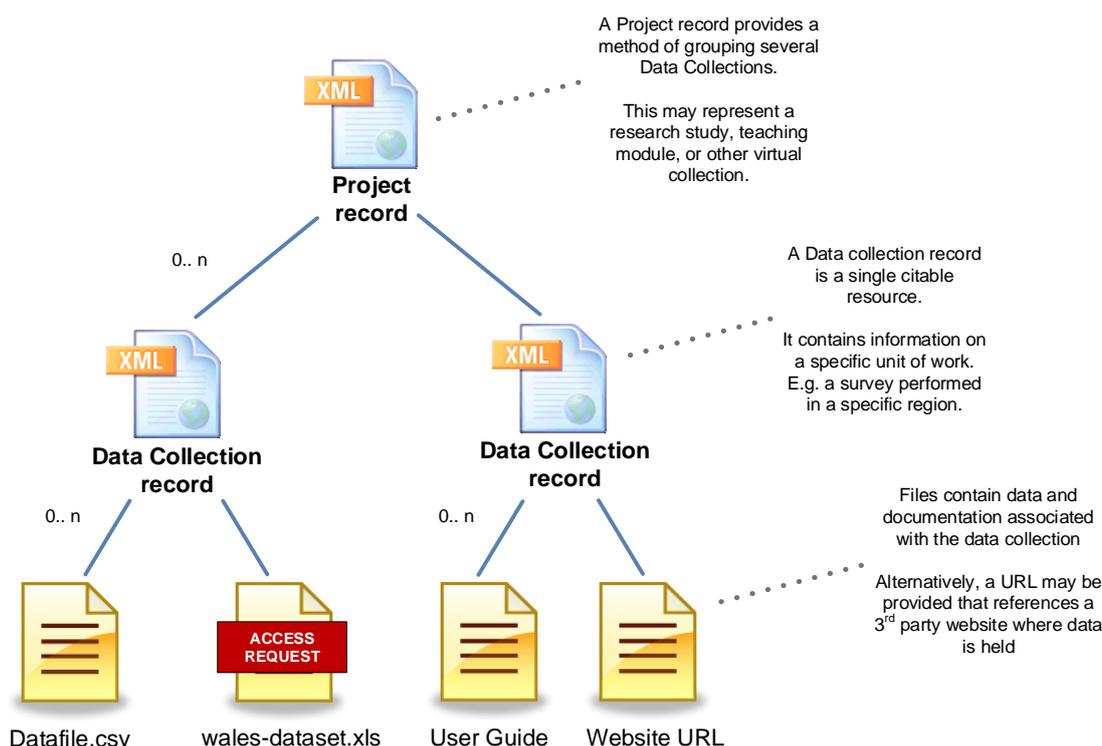
About this Guide

LSHTM Data Compass is a curated digital repository of research data and other reusable resources produced by LSHTM researchers and their collaborators.

This User Guide explains how LSHTM staff and students can create a Data Collection record and upload their research data.

Organise resources held in LSHTM Data Compass

LSHTM Data Compass uses a simple data model comprised of three levels: Project, Data Collection and File. Data Collection is the key object in the repository – this is a metadata record that describes a distinct digital output intended for reuse, such as a dataset, search strategy, or other research outputs.

Studies that have published two or more Data Collection records may wish to create a Project record to link them. Consult the 'Creating a Project record in LSHTM Data Compass' tutorial for details.

Getting started

1. Visit https://datacompass.lshtm.ac.uk/cgi/users/login and login using your LSHTM username & password.
2. Press the 'New Data Collection' button in the user work area
3. Complete the Data Collection workflow using the information below and upload relevant files
4. Submit the record for review using the 'Deposit Item Now' button.

Completing the Data Collection form

Field	Status[1]	Description	Examples
Tell us about your resource			
Title	M	The name of your Data Collection. This should differ from your paper's title to avoid confusion	IDEAS project - Outbreak Risk dataset Data for: "Paper title"
Alternative Title	O	A second title used to identify the dataset, e.g. in an accompanying paper, the title in a non-English language	
Description	M	Briefly describe the data content (e.g. topics covered, variables contained) and high-level details on the research context in which it was obtained	Dataset 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, containing variables on their species, capture date/time, day of censoring, and whether it was alive or dead on collection.
Keywords	M	Subject or scientific terms or phrases relevant to your research domain that describe the data content. Use widely adopted vocabulary where feasible. E.g. International Disease Classification	Planetary health Anthropocene Climate change Biodiversity Ecosystem Change Land use change
Data Collection Period	O	The start and end date of data capture	1 September 2013 – 30 October 2014
Date Information	M	A key data in the data's lifecycle. E.g. completion or publication date	1 October 2018
Data collection mode	O	A multi-choice field use to specify the collection methods. See Appendix 1 for examples.	"Compilation/ Synthesis"
Description of data capture	O	May be used to briefly describe the methodology applied to create/capture data. Relevant details may include the sample size, procedures applied, and location	Full enrolment lists for each school were collected by the study team, compiled as master enrolment lists and used to populate class registers. These were distributed to teachers in five-week periods and were then completed daily for each child.
Geographic area covered	O	One or more geographic regions relevant to the data, e.g. villages in which data collection took place. A broad region should be identified; the region *must not* identify the exact location of study participants.	

---

[1] Status: M = Mandatory, O = Optional

Tell us how it was created			
Creator(s)	M	Name, email address and ORCID (if possible) of one or more people <u>directly responsible</u> for data collection. This will be used to generate the author list for citation. The list may include LSHTM & non-LSHTM staff. Those involved in an indirect capacity (e.g. advisors) should be listed in the Associated Role field.	Jane Falconer orcid: 0000-0002-7329-0577
Associated Roles	O	LSHTM & non-LSHTM researchers who had an indirect role in the research. See Appendix 2 for a list of role types.	Sarah Whitmee, Andy Haines, Lancet Commission on planetary health
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
Tell us about your project and faculty			
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 researchers belong, and other organisations that made a significant contribution to the work	London School of Hygiene & Tropical Medicine, London, United Kingdom University College London, London, United Kingdom
LSHTM Faculty/Department	O	One or more LSHTM faculty & departments in which the LSHTM researchers are based.	EPH : IDE
Research Centre	O	One or more LSHTM Research Centre's associated with the research	Cancer Survival Group
Research Group	O	Other LSHTM research groups to which the researcher(s) belongs	

Copyright Holder	O	Institutions or individuals that own rights over the data E.g. a funder, institutions in which creators are based	London School of Hygiene & Tropical Medicine, London, United Kingdom Elsevier, 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	Used only if the metadata publication & DOI registration must be delayed as part of a publication blackout, e.g. to await Nature publication	
Describe each file			
File description	M	A short description of the file's content. This may duplicate the title or other detail	Dataset containing variables on 128 participants, covering X, Y and Z.
File purpose	M	A controlled list used to classify the file. See Appendix 3 for list.	Study instrument
Visibility	M	Controlled list used to specify the security level to be applied. See Appendix 4 for list.	Visible to all
Licence	M	Licence conditions associated with the file that specify how it may be used.	Creative Commons Attribution - recommended for open data Data Sharing Agreement – a custom licence recommended for restricted data
Content type	M	A controlled list used to describe the file content. See Appendix 5 for controlled list.	Dataset Audio
Retention period	O	The time period that the file should be held and made available. Anonymised data will often be kept for an indefinite period	Indefinite

Appendix 1: 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 on a particular topic that has been 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, by reducing it to its main points. E.g. an interview summary, a condensed report
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

Appendix 2: Associated Roles

A controlled list that allows the depositor to specify the type of contribution made to the research:

- Contact Person
- Data Collector
- Data Manager
- Distributor
- Editor
- Hosting Institution
- Producer
- Project Leader
- Project Member
- Registration Agency
- Registration Authority
- Related Person
- Researcher
- Research Group
- Rights Holder
- Sponsor
- Supervisor
- Work Package Leader
- Other

Other contributor types may be suggested by emailing researchdatamanagement@lshtm.ac.uk.

Appendix 3: 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
Documentation / metadata	Supporting information that enables the data to be understood.
Study Instrument:	A questionnaire, interview topic guide, test, search strategy, or other resources used to capture or measure data
Licence form:	Documentation that specify conditions for access and use for the data, e.g. a Data Sharing Agreement or Data Transfer Agreement.

Appendix 4: File Visibility

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

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. Registration of external users is not supported at present, but may be implemented at a later date.
Request access for all	All users must complete and submit a request form, which is sent to the corresponding author.

Appendix 5: Content Type

A controlled list used to specify the content contained within the file. This information is important for 3rd party services that use metadata published by the LSHTM data repository.

Type	Description
Unspecified	The default value. An error message will be displayed if the record 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