Plan Overview

A Data Management Plan created using DMPonline

Title: PhD Van Rooyen 2020

Creator: Cairan Van Rooyen

Principal Investigator: Cairan Van Rooyen, Dr Cliff Elwell

Data Manager: Cairan Van Rooyen

Affiliation: University College London

Funder: Engineering and Physical Sciences Research Council (EPSRC)

Template: EPSRC Data Management Plan

ORCID iD: 0000-0003-1058-1091

Project abstract:
 The aim of this research is to investigate the relationship between occupant ventilation
 practices and IEQ factors, including: overheating, indoor air quality and noise and relate this
 to health and well-being thresholds and guidance.

ID: 50932

Last modified: 29-01-2020

Grant number / URL: EP/H009612/1

Copyright information:
 The above plan creator(s) have agreed that others may use as much of the text of this plan
 as they would like in their own plans, and customise it as necessary. You do not need to credit
 the creator(s) as the source of the language used, but using any of the plan's text does not
 imply that the creator(s) endorse, or have any relationship to, your project or proposal

PhD Van Rooyen 2020

Data Collection

What data will you collect or create?

Physical data about participants dwelling.
The following data will be collected about the physical characteristics of each participant's dwelling.

1. Measurements of all rooms, window and door positions and sizes, to create a detailed floorplan.
2. Details of the windows: how far the windows open, the proportion of windows that open, the position of the opening window, how the windows open/hung, if windows can be left open securely.
3. Details of the trickle vents: sizes, model number (if present), if they are open or closed during the survey, check operation and if there is a hole all the way to outside.
4. Survey building fabric to estimate the constructions and build-ups.
5. Details of mechanical ventilation: locations and model numbers of equipment, types of grilles, sizes of grilles, how open are grilles (eg: air valves), are cooker hoods ducted to outside or recirculating, locations of external grilles, how are systems controlled, is there fan run on.
6. Measure door sizes and the amount that they are undercut from the floor finish.
7. Record any doors that are self-closing.
8. Record locations and details of sources of indoor and outdoor: Air contaminants/pollution, heat sources indoors and noise/sound.

Physical data about participants indoor and outdoor environment.
The following data will be collected about the physical environment of each participant's dwelling.

1. Kitchen, living and outdoors: Temperature, humidity, PM1, PM2.5, PM10, CO, CO_2, NO and TVOC.
2. Bedrooms: Temperature, humidity and CO_2.
3. Bathrooms: Temperature, humidity and light.
4. All windows, doors and ventilation systems: times when they are opened/closed.
5. Living rooms: VOC's for lab analysis.
6. Living rooms and outdoors: noise levels.

Social data about participants and their behaviour.

1. General household data.
2. Comfort and heating behaviours.
3. Ventilation behaviours.
4. Ventilation and health relationships.
5. Indoor environmental quality satisfaction.

How will the data be collected or created?

Physical data about participants dwelling.
This data will be collected by the researcher during a detailed survey of the individual dwellings using a measuring tape, pen, paper and a camera.
The researcher will develop the raw site survey information into a detailed floor plan of each dwelling, with all the collected information recorded on the drawing. The drawing will be created electronically using AutoCAD (.dwg) and stored in PDF format. Both the .dwg and .pdf will be stored on UCL N drive.

Physical data about participants indoor and outdoor environment.

1. Kitchen, living and outdoors: Temperature, humidity, PM1, PM2.5, PM10, CO, CO_2, NO and TVOC. This data will be automatically collected every 5 minutes by specialist Eltek AQ110 instruments.
2. Bedrooms: Temperature, humidity and CO_2. This data will be automatically collected every 5 minutes by specialist Onset MX1102 instruments.
3. Bathrooms: Temperature, humidity and light. This data will be automatically collected every 5 minutes by specialist Onset U12-012 instruments.
4. All windows, doors and ventilation systems: times when they are opened/closed. This data will be automatically collected every time a window/door/ventilation system is opened/operated by specialist Onset UX90-001M instruments.
5. Living rooms: VOC's for lab analysis. VOC's will be collected by sorbent tubes installed in the living rooms, which will be sent off to an accredited laboratory for gas spectrometry analysis to identify the compounds.
6. Living rooms and outdoors: noise levels. This data will be automatically collected every 5 minutes by custom made Raspberry Pi based noise instruments.

All of the above data is collected and stored as time-series data for every 5 minutes. The raw data is stored in CSV format.

Social data about participants and their behaviour.

This data will be collected by the researcher from the participants using a pseudonymised questionnaire. One, possibly two semi-structured walkthroughs are proposed during the winter and potentially summer seasons.

The walkthrough(s) will be recorded using an audio recording device. The audio (or AV) files from the walkthrough will be transcribed into a Microsoft Word document.

Formats and sharing

All data formats are widely used and can be opened by a range of popular software packages - this will facilitate sharing and long term access to the data by others.

File and folder structure

The structure below shows the proposed folder and file structure for the collected data.

```
Project data folder
|
+-- Participant A, B, C, D etc (folder)
| |
| |
| +-- Metadata
| |
| +-- metadata.xls (file) metadata on the dwelling, which equipment has been installed and where, orientation
| |
| |
| +-- Survey data (folder)
| |
| +-- site_notes.jpg (file): scan of site notes.
| +-- floorplan.dwg (file): AutoCAD file of floorplan.
| +-- floorplan.pdf (file): PDF of floorplan.
| |
| |
| +-- Technical data (folder)
| |
| +-- eltek_aq110.csv (file): time-series of all AQ110's.
| +-- mx1102.csv (file): time-series of all MX1102's.
| +-- u12-012.csv (file): time-series of all U12-012's.
| +-- ux90-001m.csv (file): all events and times.
| +-- noise_data.csv (file): time-series of noise data.
| |
| |
| +-- Social data (folder)
| |
| +-- initial_interview.doc (file): transcript of the initial interview.
| +-- summer_walkthrough.doc (file): transcript of the summer walkthrough.
| +-- winter_walkthrough.doc (file): transcript of the winter walkthrough.
| |
| |
| +-- Photos
| |
| +-- photo1.jpg (file): photos
| |
| |
| +-- Checklists (folder)
| |
| +-- site_prep.pdf (file): site preparation checklist
| +-- download.pdf (file): data download checklist
|
|
| +-- Cross calibration (folder)
| |
| +-- cross_calibration.csv (file): raw data from 2 weeks for cross-calibration.
| +-- offsets.xls (file): calculated offsets for each sensor.
|
|
|
```

```
| +-- MIDAS data (folder)
| |
| +-- weather.csv (file): raw weather data from MIDAS
|
|
| +-- Equipment Log
| |
| +-- equipment_log.xls (file): file with all equipment references on loan and where they are installed - dwelling code and room
```

Versioning

Only the raw data will be stored. The researcher will use a Python script and code to process the data - the processed data will never be permanently stored. Python holds the processed data in RAM and this is cleared every time Python is closed.

Versioning will be used for the Python code that the researcher develops. The researcher will use Git and GitHub to document all changes to the code. GitHub will be used to document changes and as a change control log.

This will record all changes to the code, the reasons for the changes and allow the researcher to revert to previous versions of the code if required.

Documentation and Metadata

What documentation and metadata will accompany the data?

Documentation

A comprehensive overarching document will be produced to accompany the data. This document will contain the following general information:

1. Title of the project and the data set.
2. Contact details of the researcher.
3. Contact details of the principal investigator.
4. Dates the data was collected.
5. Information on conditions of access or data confidentiality.

The following project-specific information will be detailed in the documentation:

1. A description of the PhD project - aims, objectives and research questions.
2. A description of the layout of the data - describing the structure of folders and files and the relationship between files.
3. A description of how the data was collected: sampling, data collection process, instruments used, hardware and software used, scale and resolution, temporal and geographic coverage and secondary data sources used.
4. Details of data validation, checking, proofing, cleaning and quality assurance procedures carried out.
5. Details of any changes made to data over time since their original creation and identification of different versions of data files.

The following data-level information will be included in the documentation (Eynden et al. 2011):

1. Names, labels and descriptions for variables, records and their values.
2. Definitions of specialist terminology or acronyms used.
3. Codes of, and reasons for, missing values.
4. Derived data created after collection, with code, algorithm or command file.
5. Data listing of annotations for cases, individuals or items.

Metadata

The metadata should be published within 12 months of the creation of the research data - in accordance with EPSRC recommendations.

The metadata will be created using the DataCite metadata schema. The metadata will include the following minimum information:

1. A digital object identifier (DOI).
2. Creator/researchers details.
3. Title.
4. Publisher.
5. Year and dates.
6. Details of the type of resource.

Ethics and Legal Compliance

How will you manage any ethical issues?

The ethical aspects of this research will be managed by the researcher and the principal investigator in full compliance with UCL ethics requirements.

This research requires internal departmental ethics approval, as it involves human participants and/or their data. This includes social and technical data from participants and their homes.

This research is not high risk and therefore does not require approval from an external (outside of BSEER) ethics committee.

This research is classified as low risk.

The research involves non-vulnerable participants covering non-sensitive topics, where:

1. where participation will not induce undue stress or anxiety, and
2. it will not possible to identify (directly or indirectly) individuals from the data or publications because all personal information will be separated from the research data, and
3. no personal data will be shared outside of the research team.

An ethics application has been submitted to the BSEER ethics committee in accordance with its requirements.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

I am not sure how copyright and IPR relates to this project...?

Storage and Backup

How will the data be stored and backed up during the research?

Pseudo anonymised technical data from the monitoring campaign, technical data from the surveys, social data from questionnaires and interviews will all be securely stored on the UCL N drive (100GB of centrally managed storage). No personal or sensitive data will be held on UCL N drive. Backups are made every night of all files created or modified since the preceding night's backup.

Sensitive and personal data, such as personal and contact details of participants will be stored on the UCL Data Safe Haven. This service provides a technical solution for storing, handling and analysing identifiable data. It has been certified to the ISO27001 information security standard and conforms to the NHS Information Governance Toolkit.

How will you manage access and security?

Only the principal investigator (Dr Cliff Elwell) and the main researcher (Cairan Van Rooyen), will have access to all data stored on the UCL N drive and UCL Data Safe Haven.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Pseudo anonymised technical data from the monitoring campaign, technical data from the surveys and social data obtained from questionnaires and interviews are all of long-term value and will be retained, shared and preserved.

In line with the expectations of the EPSRC, this data will *'be securely preserved for a minimum of 10 years from the date that any researcher 'privileged access' period expires or, if others have accessed the data, from last date on which access to the data was requested by a third party.'*

What is the long-term preservation plan for the dataset?

The data discussed above will be uploaded to the UK data service website, where it will be made freely available to others, including researchers outside of the EEA.
This data will be fully anonymised and stored on UCL's long-term Research Data Repository.

Data Sharing

How will you share the data?

Data will be archived and shared in line with EPSRC's expectations, where:
"*Publicly funded research data should generally be made as widely and freely available as possible in a timely and responsible manner*";
Within 12 months of the data being created, appropriately structured metadata, describing the research data will be published.
Following any publications, and on the successful completion of the PhD, all data generated from this research will be uploaded to the UK data service website, where it will be made freely available to others, including those outside of the EEA.
https://www.ukdataservice.ac.uk/deposit-data
Data will be uploaded with documentation and relevant meta-data to allow future use of the data.

Are any restrictions on data sharing required?

Yes. Data will only be shared following any publications and on the successful completion of the PhD, including any viva corrections.

Responsibilities and Resources

Who will be responsible for data management?

Both the principal investigator (Dr Cliff Elwell) and the main researcher (Cairan Van Rooyen), will be responsible for data management.

What resources will you require to deliver your plan?

Access to UCL N drive - free for UCL students. Access granted.
Access to UCL Data Safe Haven - free for UCL students. Application in progress.
Access to the UK data service website - free. Application at a later date.
The PhD is funded by EPSRC and Public Health England. There will be no further requirement for funding to deliver this data management plan.