Data Management Plan for Research Students

Project title	Exploring the role of human insight in infecctious disease forecasting
Author name	**Nikos Bosse**
Supervisor	**Sebastian Funk**
Contact email	**Nikos.bosse@lshtm.ac.uk**
Date of last edit	**11.07.2021**

Guidance on writing a Data Management Plan can be found at https://lshtm.sharepoint.com/Research/Research-data-management/ and http://servicedesk.lshtm.ac.uk

Advice and feedback can be obtained from: researchdatamanagement@lshtm.ac.uk

DESCRIBE YOUR RESEARCH

1. What digital resources – data, code, collection tools, etc. - will you collect/obtain and use?

Relevant details to mention: topics covered, type (e.g. survey), source (collected by self or others), format (e.g. STATA) and amount (e.g. 10 interviews). Draw attention to human or other data that require additional protection.

1. Collection of forecasts for COVID-19 through a self-developed shiny app

2.

3.

4.

2. What hardware and software will be used in your research?

List any hardware and software to be used, their intended purpose (e.g. collection, analysis), and (if relevant) the number needed. E.g. 20 Samsung 10" tablets, LSHTM's Open Data Kit software, STATA and MS Access for analysis.

R shiny (data collection), R (data analysis), Google Sheets (data storgage)

3. What data-related activities will be performed during the research?

List key data-related activities that you and/or others will perform during the research. For instance, trial draft survey in month 6, collect data in month 8-10, clean and anonymise data in month 11, analyse data in month 12-18.

Task	**Description**
Collect predictions	Voluntary participants make predictions of COVID-19 through a shiny app.
Aggregate predictions	Predictions will be aggregated
Submit forecasts to Forecast Hubs	Aggregated forecasts will be submitted to the German and Polish as well as the European Forecast Hub

4. What quality checks will you perform to ensure resources are fit for purpose?

Outline any quality checks to be performed before, during and after the above activities, e.g. to ensure data are captured correctly, remain accurate and complete, or ensure you avoid recognised problems. The UK Data Services offers guidance at http://ukdataservice.ac.uk/manage-data/format/quality.aspx.

Manual checks for plausibility and visual inspection.

5. How will you address ethical & legal issues within your research?

- What permissions are needed? E.g. to collect data in country, analyse data for specific purpose, share data
- From whom must approval be obtained? E.g. study participant, ethics committees, data provider
- How will permissions be provided? E.g. ask participants to sign a consent form, sign a Data Transfer Agreement

Informed consent of participants is given prior to collecting any data (by clicking ‘agree’).

Ethics approval from LSHTM is obtained.

6. What documentation will be created to ensure resources can be understood?

What aspects of the research will be documented and how? E.g. processes could be documented in Standard Operating Procedures, workflows applied described in a lab book, a codebook written to describe variables, etc.

Detailed explanation of the project is provided for the participants before any forecasts are submitted.

All code is public and documented.

STORAGE AND SECURITY

7. Where will resources be stored at key stages of your research?

Identify where resources will be held during capture, processing, analysis and other stages, and who will have access to them. Consult https://lshtm.sharepoint.com/Services/IT-Services/ServiceDesk/LSHTM-data-storage-options.pdf

Intermediate data and participant information will be stored on Google Drive. Pseudonymised forecasts are made publicly available on github.

8. What labelling conventions will you apply to manage your resources?

Briefly describe any naming conventions or classification systems you will apply to resources. E.g.

- Filenames: key characteristics you will record to group files, e.g. FG1_transcript_2018-10-01
- Variable: conventions to be used for question IDs, completed responses & missing variables
- Versions: how will you identify changes to resources over time (e.g. v1.1, v1.2)

Forecasts are stored in dated files. Version control (git) ensures changes can be tracked over time.

9. How will you keep data safe and secure? (choose one or more)

Only anonymised data will be used - personal, sensitive, or otherwise confidential data is not needed for the research		Store personal details in a separate secure location & link it via an identifier	X	Delete personal & confidential details at earliest opportunity (specify when below)	X
Use digital storage that require a username/password or other security feature	X	Physical security (such as locked cabinet or room)		Protect portable devices using security features, e.g. biometric	
Encrypt storage devices		Encrypt during transfer	X	Avoid cloud services located outside EU	
Take 'Information Security Awareness training'	X	Ensure backups are also held securely			
Notes:	Any user data will be deleted as soon as the project is finished				
Identify additional steps you will take to avoid, reduce, or eliminate risks that may affect your resources.					

ARCHIVING & SHARING

10. What resources should be kept as evidence of your research?

Research often has value beyond the lifespan of the project that produced it. For this reason, many researchers are required to keep data for a set time period, typically 10 years following completion, to comply with funding or journal publication requirement. List the resources in Q1 that will be kept and for how long. If some resources can't be retained for some reason (e.g. it contains personal data), state the reason that this is not permitted.

Pseudonymised forecasts will be kept after the project has finished.

11. Where will these resources be hosted?

Identify where each resource will be hosted following research completion. E.g.

- Files intended for sharing may be hosted in the LSHTM data repository (http://datacompass.lshtm.ac.uk) or a 3rd party repository, such as UK Data Service, ArrayExpress, Zenodo, etc.
- Internal and confidential files can be held on the LSHTM Secure Server
- My supervisor will look after them

Github.com

12. When will the resources be made available? (choose one or more)

During the research life		At the same time as findings are published in an academic journal		A set time after research end, e.g. 12 months. Specify below	
Resources already available (provide details below)	X	On completion of my thesis		Other (provide details below)	
Further information / Other					
https://github.com/epiforecasts/covid.german.forecasts https://github.com/epiforecasts/europe-covid-forecasts					

13. How will you make other researchers aware that the resources exist?

Publish a metadata record describing the resources in a repository or other catalogue		Obtain a Digital Object Identifier (DOI) or other permanent ID	X
Cite resources in future research papers, e.g. in the data access statement or		Cite resources in project reports	

reference list			
Publish a description for the project website		Write and publish a Data Paper	
Add resources to a list of your academic outputs			
Other measures / Further details			

14. What steps will you take to ensure resources are easy to analyse and use in future research? (choose one or more)

Prepare a codebook or other documentation that provides an accurate description of content		Store resources in open file formats such as CSV, Rich Text, etc. See https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats	X
Write a user guide that provides a high-level overview of research		Apply a standard licence that allows a broad range of uses (e.g. Creative Commons, Open Data Commons)	X
Designate a corresponding author / data custodian who will handle data-related questions		Use domain-specific standards that make it easy to import and analyse data	X
Other / Further information			

15. If resources can be made available, but not openly, what conditions on access/use must be met?

E.g. data can be used for specific types of research only. Leave blank if not applicable.

Requirement:	**To be addressed by:**

RESOURCING

16. What are the primary data management challenges in your research?

E.g. uncertainty on data management practice, data security, data-related costs, staff resources, etc.

17. How can LSHTM & others help you to better manage your data?