Plan Overview

A Data Management Plan created using DMPonline

Title: Beprepared task 2.1

Creator: Peter Lugtig

Principal Investigator: Peter Lugtig

Data Manager: Peter Lugtig

Project Administrator: Peter Lugtig

Affiliation: Utrecht University

Funder: ZonMw (Nederlands)

Template: Data management ZonMw-template 2019

ORCID iD: 0000-0001-8548-6828

Project abstract:

Pandemic Preparedness – task plan for WP 2.1

Summary of entire project

In response to the COVID-19 pandemic, governments introduced a range of prevention measures to reduce transmission in order to avoid overburdening healthcare and preventable deaths. However, prevention measures such as minimizing social contacts may also negatively affect people's well-being (social, mental, physical, financial) and liberties; and may in particular affect people in disadvantaged positions. This, in turn, may influence whether people support COVID-19 policies and adhere to prevention measures. Although these challenges fall in the realm of social and behavioural sciences, these disciplines were not well-prepared for a major infectious disease outbreak.

The current research proposal seeks to improve pandemic preparedness in the Netherlands, through advancing knowledge about the social and behavioural aspects of pandemics and translating those insights into a toolkit for pandemic preparedness. This in turn should lead to faster and more targeted policy advice for a future pandemic. The main aims are to advance:

1. Understanding of how individual and contextual factors influence prevention decisions and behaviours during a pandemic (work package 1),
2. Methodologies for collecting behavioural data and translating that into mathematical an statistical models for predicting effects of behavioural interventions on pandemic outcomes (work package 2)
3. Understanding processes and characteristics of pandemic resilience at the individual, community and organisation level, and how to support that (work package 3), and
4. Methods for more effectively advising policy makers on behavioural and social science insights, and integrating those in multi-disciplinary policy advice (work package 4).

This consortium application entails four work packages (WP) comprised of 13 research questions, that can be addressed within a 12-month period (duration of this Phase 1 funding for pandemic preparedness). We will be applying a range of qualitative (e.g., interviews, focus groups, co-creation methods) and quantitative (e.g., experiments, surveys, agent-based modelling, simulation studies) methods, using both existing data and new data. In each WP an expert in socially disadvantaged groups will ensure this perspective is fundamentally integrated in the research process.

Results from this consortium are both relevant for science and pandemic preparedness at the level of public health institutions, funding agencies and government. For science, we will share novel insights and methods through delivering scientific papers, conference presentations and webinars. For public health, we will translate the results with the Dutch National Institute for Public Health and the Environment (RIVM) into a pandemic preparedness toolkit and create the basis for an interdisciplinary research network for a future pandemic infrastructure (under development). For research funders, this Phase 1 funding research will uncover key questions that need to be addressed in future research prior to and during a future pandemic.

Context of Workpackage 2 – Data monitoring, modelling and prediction

One of the early challenges during Covid-19 was monitoring relevant social and behavioural indicators, and effects of prevention measures. Today there is still limited methodological evidence on which data sources are most suitable for monitoring, how to include behavioural data into epidemiological models of pandemic spread, and estimate the impact of behavioural and social interventions on transmission or hospitalisation.[5-8] The aim of this WP is to identify the most suitable social and behavioural data sources, develop methods for linking social and behavioural effects of behavioural interventions to viral transmission models at a micro level, and identify the potential impacts of behavioural and social interventions on the course of a pandemic at a macro level. Overall, these three tasks help (future) social and behavioural researchers to more effectively assist and advice policy makers in the future.

Specific task plan for Task 2 .1 – Data monitoring

Research question 2.1: What methods for collecting behavioural data during a pandemic (social activity data, adherence to prevention measures) perform best (e.g., relevance, accuracy, timeliness, etc) and for what purpose (e.g., monitoring, modelling, rapid policy advice)?

Methods: In a first stage**,** data sources that have been used (inter)nationally during the Covid pandemic (2020-2022) to monitor social activity and adherence behaviours are reviewed.

The inventory of data sources will be based on a literature review and a survey among experts (recruitment via snowballing) who have worked with social and behavioural data during the Covid pandemic. The literature search will be conducted via academic search tools, such as google scholar, but will also focus on reviewing grey literature, documentation that accompanies dashboards designed for policymakers. Should the data description and documentation suffice to review the quality of the data sources, this documentation will be used to in a second stage review the data source along a set of quality dimensions (see below). If the documentation does not suffice, experts that have worked with the data will be consulted. Experts will be identified via contact persons or authors of documentation that does exist. Should no author be available, experts in the pandemic preparedness consortium will be used as a first step to identify experts on the sources identified. Should – after snowballing – little documentation be available, we will try to use the available information.

Should no information be available, and no experts found, then the datasource will be discarded from the review process. No micro-data will be analysed. Analyses will be based on meta-data of datasets, data documentation and reports, (grey) literature, and expert consultations. The result of the inventory is a spreadsheet with datasources in rows, and metadata, locations of documentation/experts in columns.

In a second stage, the identified data sources will be reviewed. The review will focus on data sources from countries with good data infrastructures, notably the European Union, United States, Canada, Republic of Korea, Taiwan Australia, New Zealand and Japan. Should data-sources have been used in- and across multiple countries, they will only be included once in the review. It is possible we will identify and review data-sources from other countries listed above.

The review itself will be based on a review of every datasource using the data quality framework by Biemer and Lyberg (2003), along the following 6 dimensions: relevance, accuracy, credibility, timeliness, accessibility, and interpretability of the data. On every dimension, we will rate the datasource on a 5-point scale, ranging from 1) very bad 2) bad 3) neither good nor bad 4) good and 5) very good. In short, the quality dimensions mean the following:

1. - Relevance: is the datasource relevant for the statistic that the researcher is interested in. As an example, mobility data on car traffic are probably a good source for measuring the effects of ‘work from home’ policy measures, but a bad source of evaluating the effects of advice to keep social distance.
2. - Accuracy: this evaluates the validity and reliability of data. Does the intended measurement of the datasource (google mobility data measures car traffic intensity) actually measure this, and does it do so consistently?
3. - Credibility: how credible are the data that are being produced. Do they come from a reputable source, have they been produced consistently over time, and is documentation available that for example describes the data production process in detail?
4. - Timeliness: how up-to-date are data sources. Are mobility data updates daily, weekly, monthly, and how long is the delay between the production and release of the data?
5. - Accessibility: are data accessible for use by government/policy makers, researchers and/or the wider public?
6. - Interpretability: is it easy to process the data to such a degree that they are easy to interpret by the person using the data?

As a result of stage 2, strengths and shortcomings of each data source are summarized, along with potential barriers for future use in the Netherlands. Should barriers exist that can be removed, possible ways to methodologically mitigate these will be described. This intermediate deliverable will consist of a word document. Stage 2 will start when the inventory of stage 1 is about ¾ complete (estimate: M5).

In stage 3, a 2-round Delphi study (online voting then face-to-face meeting) with 10-15 experts (methodology, modelling, behavioural and social, public health) will be conducted to assess consensus on the rating of data sources and their application, and formulate recommendations for monitoring during future outbreaks. Future research should investigate how these data sources can be accessed and maintained in practice in the context of the existing Dutch infrastructure for pandemic preparedness. Stage 3 will start when Stage 2 is completed.

Intermediate Deliverables:

Stage 1: an inventory of data sources (M8)

Stage 2: a review of data sources (M11)

Stage 3: results from the Delphi study (M14)

Final deliverables (all due Month 15)

1. A scientific article reviewing the social and behavioural data sources, based on the narrative review from stage 2.

2. For the toolkit: methodological recommendations for monitoring of behavioural data during future infectious disease outbreaks based on Delphi study from stage 3, and methodological barriers/solutions from stage 2.

ID: 142669

Start date: 01-10-2023

End date: 28-02-2025

Last modified: 24-01-2024

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

Beprepared task 2.1

1. General features of the project and data collection

1.1 Project leader contact details

Peter Lugtig
dept. of Methods & Statistics
p.lugtig@uu.nl
+31 612697816 (private)

1.2 I have composed my DMP with the assistance of a data stewardship (or management) expert. List his or her name, function, organisation/department, phone number and email address.

- The expert is connected to my department or institution (please explain his/hr expertise related to data stewardship)

Pascal Pas
https://www.uu.nl/staff/PPas

1.3 In collecting data for my project, I will do the following:

- Use existing data (please specify)
- Generate new data

We will not use microdata (individual data), but review data sources that are/were used in Covid monitoring. For that we will use only publicly available metadata. We will curate a new dataset of exisiting sources.

1.4 In my research, I will use:

- A combination of quantitative and qualitative data

The curated dataset will be reviewed. The review will consist of a qualitative assessment of the quality of each dataset, as well as a graded response (quantitative) assessment.

1.5 I will be reusing or combining existing data, and I have the owner's permission for that.

- No permission is required, since the data are openly accessible

All metadata we will use are openly accessible. These will be descriptions of datasets in academic papers, working papers, documentation, and websites.

1.6 In collecting new data, I will be collaborating with other parties.

- No

We will review the data ourselves at Utrecht University. No other partner is involved.

1.7 I am a member of a consortium of 2 or more partners. Clear arrangements have been made regarding data management and intellectual property. (also consider the possible effect of changes within the consortium on issues of data management and intellectual property)

- Yes, clear arrangements have been made regarding data management and intellectual property through a consortium agreement

The consortium agreement has been in place, and is currently awaiting signing by the members.

1.8 I can give an estimate of the size of the data collection; specifically, the number of participants or subjects (“n=”) in the collection and its size in GB/TB

- Not yet (please explain)

No participants. We expect to review the metadata of about 150 datasources.

1.9 The following end products I will make available for further research and verification (please elaborate briefly)

- Documentation of the research process, including documentation of all participants
- (Several versions of) processed data

We will publish the curated list of data sources, along with our data quality assessment as a processed dataset

We will document and summarize the quality in a paper.

1.10 During the project, I will have access to sufficient storage capacity and sites, and a backup of my data will be available. (please elaborate briefly)

- Yes, I will make use of my institution's standard facilities for storage and backup of my data

2. Legislation (including privacy)

2.1 I will be doing research involving human subjects, and I am aware of and compliant with laws and regulations concerning privacy sensitive data.

- No, I will not be doing research involving human subjects; proceed to section 3 (Making data findable)

2.2 I will be doing research involving human subjects, and I have (a form of) informed consent from the participants for collecting their data.

Question not answered.

2.3 I will be doing research involving human subjects, and I will protect my data against misuse.

Question not answered.

2.4 I will stick to the privacy regulations of my organisation

- Yes

3. Making data findable

3.1 The data collection of my project will be findable for subsequent research. E.g., on a catalogue, a web portal, or through the search enginge of the repository (note: this is key item 3, which you should report to ZonMw at the end of your project).

- Yes, it can be found through the search engine of the archive or repository in which it is stored (please specify)

The projects findings will all be published on OSF. We will also publish a copy on Github.

3.2 I will use a metadata scheme for the description of my data collection (note: this is key item 7, which you should report to ZonMw at the end of your project).

- Yes, I will use a metadata scheme specific for my field of research (please specify)

See project plan. We will use the framework of Biemer and Lyberg (2003) to describe the metadata along quality framework. Next to this, we will describe the source of the data, name of the data, doi of the source, and time of data access.

3.3 I will be using a persistent identifier as a permanent link to my data collection (note: this is key item 1, which you should report to ZonMw at the end of your project).

- Yes, I will be using the DOI code

Just a doi code.

4. Making data accessible

4.1 Once the project has ended, my data will be accessible for further research and verification.

- Yes, immediately

We will use the license by OSF (CCBY 4.0)

4.2 Once the project has ended, my data collection will be publicly accessible, without any restrictions (open access).

- Yes, proceed to section 5 (Making data interoperable)

4.3 I have a set of terms of use available to me, which I will use to define the requirements of access to my data collection once the project has ended (please provide a link or persistent identifier; also note that this is a key item 4, which you should report to ZonMw at the conclusion of your project).

Question not answered.

4.4 In the terms of use restricting access to my data, I have included at least the following:

Question not answered.

5. Making data interoperable

5.1 I will select a data format, which will allow other researchers and their computers (machine actionable) to read my data collection (note: this is key item 5, which you should report to ZonMw at the end of your project).

- Yes (please specify)

We will publish the data in a .csv format

5.2 I will select a terminology for recording my data (e.g., code, classification, ontology) that allows my dataset to be linked or integrated with other datasets (note: this is key item 6, which you should report to ZonMw at the conclusion of your project).

- Yes, metadata standard (please specify)

Yes. dois of sources. Quality framework of Biemer and Lyberg

5.3 I will be doing research involving human subjects, and I have taken into account the reuse of data and the potential combination with other data sets when taking privacy protection measurements.

- No (please explain)

No human subjects. The project just uses metadata

6. Making data reusable

6.1 I will ensure that the data and their documentation will be of sufficient quality to allow other researchers to interpret and reuse them (in a replication package).

- I will perform quality checks on the data to ensure that they are complete, correct and consistent (please explain)
- I will document the research process (please explain)

We will write a paper about the dataset as output for the project
We will ensure that the curated dataset is complete. Missing data may appear for dois (e.g. when documentation is a website).

6.2 I have a number of selection criteria, which will allow me to determine which part of the data should be preserved once the project has ended. (see also question 1.9 and 6.1)

- Yes

All data are preserved. Note that we will not create a copy of all data that are reviewed. So only the relevant metadata are stored, not the original documents/websites.

6.3 Once the project has ended and the data have been selected, I can make an estimate of the size of the data collection (in GB/TB) to be preserved for long-term storage or archival.

- Yes (please specify)

The dataset is foreseen to be about 300 kB (1 .csv file containing about 150 rows, and 20 columns)

6.4 I will select an archive or repository for (certified) long-term archiving of my data collection once the project has ended. (note: this is a key item, which you should report to ZonMw at the conclusion of your project)

- Yes, and this archive has a different form of certification (please specify the archive and certification)

We will publish on OSF and Github

6.5 Once the project has ended, I will ensure that all data, software codes and research materials, published or unpublished, are managed and securely stored. Please specify the period of storage.

- Yes, in accordance with VNSU guidelines (please specify the number of years)

Utrecht University storage policy stipulates that data are stored for at least 10 years at UU servers.

6.6 Data management costs during the project and preparations for archival can be included in the project budget. These costs are:

- Amount ..... (please elaborate)

0 (time to curate the data not included)

6.7 The costs of archiving the data set once the project has ended are covered.

- Yes (please elaborate)

This is part of OSF, UU and Github poilicies. Data storage there is free of charge