

Plan Overview

A Data Management Plan created using DMPTool-Stage

DMP ID: <https://doi.org/10.48321/D1FG63>

Title: Strengthening Resilience Across Scales: Moving Cities beyond COVID-19

Creator: Andrea Lampis - **ORCID:** [0000-0002-1561-5409](https://orcid.org/0000-0002-1561-5409)

Affiliation: Universidade de São Paulo (www5.usp.br)

Principal Investigator: David Simon, Michael A. Cohen, Diego Silva Ardila, Célio Bermann, Manya Van Ryneveld, Margarita Gutman

Funder: São Paulo Research Foundation (fapesp.br)

Template: Digital Curation Centre (português)

Last modified: 08-07-2023

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 customize 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

Strengthening Resilience Across Scales: Moving Cities beyond COVID-19

At the city scale, each research team will undertake baseline studies and participant observation, working with the respective local governments, health authorities, and other stakeholders. At the household scale, each “house interview” will have two components: 1) a questionnaire to one or more residents covering no more than 10 key questions/issues, phrased differently according to the particularities of each locale; and 2), a schematic cross-section diagram capturing spatial information of the different users and use types within the residential unit or property.

The high detail expected from the interviews at the household level makes it necessary to guarantee a certain degree of rapport between the interviewer and interviewees. Hence, the sampling will be purposive rather than random, with a quota of 30 interviews per neighborhood. To capture a fair range of cases within selected neighborhoods, the criteria for purposive sampling in each site will consider the spread of the sample, racial and cultural diversity of the population, and density of commercial activity. With these criteria, sample saturation will likely take place via snowballing, or chain referral from 4/5 initial cases to reach the agreed quota.

Detailed research encompassing all neighborhoods across the five cities is clearly impracticable. Hence we will select two broadly representative low/lower-middle class neighborhoods to avoid the extremes of wealth and poverty. The two areas will have had different experiences of the pandemic - one severe and the other relatively moderate – thus enabling study of the underlying structural and proximate causes and testing of hypotheses about the implications for building community resilience and moving beyond COVID-19. Noting substantial differences in urban scale, data availability and level of disaggregation on relevant variables among the five cities, we will use practicable approaches in each city to select the most appropriate pair of neighborhoods in:

Slough: Inspection of municipal ward-level time-series epidemiological data on morbidity and mortality, cross-referenced with socio-economic data and local knowledge by the researchers and Slough Borough Council partners.

New York City: Building on 2021 studies by The New School of the Astoria neighborhood in Queens and the State of New York on the Bronx, the team will use the framework developed by the Taskforce on Racial Inclusion and Equity (TRIE) of the City Government which identified the 33 neighborhoods most vulnerable to COVID-19. Zip-Code data within these neighborhoods will be analyzed to select local communities and to organize household interviews.

Bogotá: The selection strategy will be based on 2020 and 2021 studies conducted by Secretaria Distrital de Salud, Instituto Nacional de Salud and Data-Lama Research Group focused on the urban scale. For neighborhood observation, granular data at the block level will be available from the same sources. The team will use neighborhood demographic, anonymized epidemiologic, health infrastructure and mobility data to support selection criteria and the field work (household in-depth interviews, neighborhood data collection, and community workshops).

São Paulo: The team will build its selection strategy on existing studies cross-cutting epidemiological and socioeconomic evidence. The LabCidade, the Observatory of Metro São Paulo and the network of the Rede de Pesquisa Solidaria (Solidary Research Network) have built a street-level detailed Covid-19 impact database (morbidity and mortality). Data will be cross-analysed with more qualitative insights related to areas where communities have self-organized. Our selection process will use in-depth knowledge of energy-related issues such as access to energy, energy services, and connectivity to refine the selection process.

Cape Town: Using provincial government data from its Covid-19 Hotspot Strategy and in conversation with the

City of Cape Town's ward-level data, the team will select two sub-districts. An approach of spanning multiple "spaces" with relevance to all four scales will be implemented, including digital spaces, digitally produced resources, as well as a selection of key physical spaces in these areas.

As for the data from the questionnaires, they will be stored in Excel (.xls), and for the interviews, the formats will be .mov/.mp4, and finally the digital map data will be worked with Arcgis software in .shp format. Dados em Excel: tamanho aproximado 200 MB.

Interview data: Approximate size 10 GB.

Digital photos: approximate size 20 GB.

Digital Map Data: Approximate size 20 GB.

Formats can and can be opened in any audio and video format.

This data management plan complies with Royal Holloway's Research Data Management Policy and addresses the capture, management, integrity, confidentiality, preservation, sharing and publication of research data.

The project will build on and analyse existing publicly available quantitative data sets and will also generate significant amounts of qualitative data through workshops, focus groups, interviews and site visits. The collection, storage and sharing of the data will comply with the respective research ethics procedures of Royal Holloway, the New School, University of São Paulo, University of Rosario and the Human Sciences Research Council.

All these research ethics policies emphasise informed consent and the anonymity of data. In addition, the research will also need research approval from the respective local authorities (these research approval processes also include approval of research ethics). We will therefore seek informed participant consent in all instances and participants will be informed where there will be multiple future uses of the data.

The data that will be generated will include: anonymised transcripts of focus groups and selected interviews (not all interviews will be able to be recorded and transcribed); photographs (with accompanying permission forms where identifiable people are depicted in the photographs); selected field notes (from research visits); an inventory/catalogue of research locations; and maps. During the project's life, any hard copies will be stored at the offices of the respective university/research institute departments in lockable filing cabinets, while electronic records will be held in password and firewall protected computers and other devices. Appropriately aggregated and anonymised data appropriate for wider audiences in the public domain will be lodged on the project website.

Key variables in shaping risk and vulnerability of study area populations include, class, gender and other socially constructed factors. These need to be factored into all aspects of urban policy and decision making, including safety and accessibility. Addressing gender relations, inequalities and intersectionality with other characteristics such as class and race will receive due attention throughout the project, including in relation to project partners' outputs and policy documents. This has particular importance in Cape Town's post-apartheid context and the highly polarised inequalities in São Paulo. Full attention is paid to social inequality through principles of equality, diversity and inclusion in the research design as explained in the Narrative.

The research team is gender balanced and diverse agewise, and its activities will be gender sensitive. An intersectional approach will be used and monitored throughout. In accordance with institutional and professional ethical codes of research conduct, all participants will be fully appraised of the nature and purpose of the research, with participation on the basis of informed consent. Personal data will be aggregated and anonymized, with participants given the option of anonymity.

Institutional ethical approval procedures will be followed, with appropriate risk assessments conducted for all field-based activities. We will not work directly with children or other particularly vulnerable groups. Each team will submit the application to the respective institution's IRB/ethics review committee. For example, Royal Holloway has a standard ethical application document that determines – on the basis of answers supplied – the appropriate level of approval and hence detailed procedure required. The other institutions have equivalent procedures.

Throughout the research, ensuring informed consent will be key, e.g. requiring all interviewees and focus group participants to sign letters of consent where this itself does not arouse suspicion or create risk to the participants. The team and relevant municipal officials will be formally well-educated professional adults, so no children or vulnerable people will be involved as researchers. The team will ensure that ethical principles focusing on rights, independence, well-being and safety of all participants – who might include vulnerable people – are upheld throughout.

The team and participants will be fully informed of the purpose, methods and intended outcomes of the workshops and what their participation entails. We acknowledge and agree to implement the UKRIs and RHUL's zero tolerance approach to harm and abuse, including bullying and harassment within the team and by team members. The organisations within the team for this application have robust policies and codes of practice in this regard, and we all take adherence to them seriously. RHUL complies with the 2012 Concordat and its research policies can be found at:

<https://intranet.royalholloway.ac.uk/staff/research/research-and-innovation/research-enterprise/research-integrity/research-culture-new.aspx>.

<https://www.researchnewschool.com/hrpp>

https://www.ime.usp.br/wp-content/uploads/legislacao/codigo_de_etica_da_usp.pdf

<https://www.urosario.edu.co/Documentos/Investigacion/Soportea-la-investigacion/Descripcion-de-las-Consideraciones-Eticas-25-de-ab.pdf>

<http://www.hsrc.ac.za/en/about/research-ethics>

Once the project is completed, original records and other data sets will be anonymised and stored securely in the respective institutions' secure storage facilities and will be destroyed after an agreed period.

Each partner in the project shall own the arising intellectual property generated by its employees, students and/or agents under the project (where the intellectual property has been jointly generated, the intellectual property will be jointly owned). Each partner will be granted an irrevocable, non-transferable, royalty-free right to use all arising intellectual property generated during the life of the project for academic and research purposes.

Metadata for the preservation of information contain information necessary for archiving and preserving a particular resource.

In the case of information under management by the project, the file name and the URL relative to its location in the Drive backup will be the main types of metadata used.

The project plans to build a data dictionary with the information collected.

The capture of all data will be carried out through mobile devices, such as smart cell phones and tablets, however their processing will be done with software such as Arcgis for digital maps, N-Vivo for qualitative interviews and Excel for questionnaires. The metadata pattern used is described below:

1. **Title:** Dataset name.
2. **Description:** A brief explanation of the data.
3. **Source catalogue:** Page (URL) of the agency where the dataset is published.
4. **Responsible agency:** Name and acronym of the agency or entity responsible for publishing the dataset.
5. **Categories in the VCGE:** Taking Brazil as an example, the Electronic Government Controlled Vocabulary is a hierarchical list of government affairs that uses common terms and is geared towards society. To browse and choose the categories, access the VCGE at <https://vocab.e.gov.br/2011/03/vcge>. Similar protocols shall be adopted according to the regulations and guidelines of the countries and the institutions participating in the call under the umbrella of the present proposal.
6. **Features:** A dataset can consist of more than one data file. The basic criterion for separating multiple features in more than one dataset is the finding that they diverge across multiple metadata.

1. **Identifier:** Persistent URL that points to the resource on the web.
2. **Title:** Name of the resource.
3. **Format:** Resource format. E.g.: XML, JSON, CSV, etc.
4. **Description:** Brief description of the content of the resource. **Tags:** List of keywords related to the dataset, and that are useful in sorting and searching it.

1. **Authorship:** Institution or person responsible for producing the resource.
2. **Documentation:** Document URL that exposes details about the dataset.
3. **Geographic coverage:** Location or geographic region to which the data refer. E.g.: Recife.
4. **Temporal coverage:** Date or period to which the data refer. E.g.: 03/2012.
5. **Geographic Granularity:** Geographical accuracy of geographic coverage. E.g.: municipal.
6. **Temporal Granularity:** Temporal accuracy of temporal coverage. E.g.: month.
7. **Update frequency:** Time frequency with which the dataset is updated.
8. **References:** Relationships to other datasets.
9. **Methodology:** Data creation process.

17. **Vocabulary/ontology:** Documents structured with dataset-specific metadata.

The project will carry out several backup copies of the files, as failures or inconsistencies can be replicated in the backups and the multiplication of copies of the original (at least 3) reduces this risk. The protocol for performing the backups will involve an automated procedure (e.g. iCloud for machines using the iOS operating system, or similar systems for machines using the Android system). Finally, a backup copy will be performed on an external high-capacity hard drive that will be acquired by the project.

Data managers will be responsible for backups and eventual data recovery, which, in addition, will have the assistance of USP's information and technology team.

The backups will be carried out weekly according to the maximum-security protocols, monitoring and review, however the linkages of the machines with the storage clouds guarantee the daily backups. Information security policies are generally expressed as codes of conduct to which users of computer systems must fully adapt. A data security policy adherence format signed by each project participant will be generated.

Drives will not be shared but kept under the control of the data managers and the project administrator.

All material collected in the field as specified in the 'data collection' section will be shared.

The personal data of the people interviewed will be encrypted after the publication of the results to guarantee the possibility of carrying out a panel-type survey, that is, the process that includes interviewing exactly the same initial sample again.

The data will be transferred to an external HD under the possession of the project administrator to guarantee the impossibility of digital leaks.

Data related to people's identities, workplace and other sensible issues will be protected and accessible to principal and associated investigators.

The Lead PI, Co-PI and the Data Managers selected amongst the AI

Hardware and software usually available in the participating institutions and team coordination and leadership.

Planned Research Outputs

Data paper - "Hypotheses testing"

Testing of hypotheses related to how employment, location in relation to places of employment, nature of various formal and informal social protection systems, assets, domestic skills and levels of digital connectivity affect the incidence and impact of COVID- 19 in the neighborhood.

Data paper - "Systematization of pre-epidemic situations "

Systematization of pre-epidemic situations in each neighborhood in terms of these and other variables to build a coherent baseline profile.

Data paper - "Assessment number of Covid-19 cases and deaths"

Assessment of how the number of cases and deaths of COVID-19 affects the economic and social well-being in the neighborhood and household.

Data paper - "Peer reviewed papers "

The principal research outputs are expected to be at least 2 academic articles in appropriate international peer-reviewed urban studies/planning, geography and development journals and a team-written book, along with blog posts, other social media interventions and press articles.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Hypotheses testing	Data paper	2023-01-31	Open	None specified	1 GB	Creative Commons Attribution 4.0 International	None specified	No	No
Systematization of pre-epidemic situations	Data paper	2023-01-31	Open	None specified	5 GB	Creative Commons Attribution 4.0 International	None specified	No	No
Assessment number of Covid-19 cases and deaths	Data paper	2023-01-31	Open	None specified	5 GB	None specified	None specified	No	No
Peer reviewed papers	Data paper	2023-01-31	Open	None specified	1 GB	Creative Commons Attribution 4.0 International	None specified	No	No