

DATA MANAGEMENT PLAN

Project Reference: PTDC/GES-OUT/1662/2020

Project Acronym: HUG

Project title: The health impacts of inner-city gentrification, displacement and housing insecurity: a quasi-experimental multi-cohort study



HUG Health, —
Urban Displacement
— & Gentrification

History of changes

Version	Authors	Date	Changes/Content
V0	Ana Isabel Ribeiro, José Pedro Silva, Sofia Ribeiro	2022-05-11	draft
V1	Ana Isabel Ribeiro, José Pedro Silva, Sofia Ribeiro	2022-06-03	First version

Table of Contents

- History of changes.....2
- List of tables4
- Abbreviations and acronyms4
- 1. Introduction5
 - 1.1. Project overview.....5
 - 1.2. Data management plan structure6
- 2. Data Summary6
- 3. FAIR data.....9
 - 3.1. Data availability and open access 10
 - 3.2. Documentation and metadata 11
 - Data documentation..... 11
 - File formats and software..... 12
 - File naming convention 12
 - Folder structure..... 13
 - Metadata..... 14
- 4. Data security 15
 - Data access 15
 - Data handling and sharing 15
 - Data storage and backup 16
- 5. Data protection 16
 - 5.1. Data protection principles and legal framework 16
 - 5.2. Data protection measures 17
 - Informed consent 17
 - Anonymisation 17
 - Data Processing Policies and Agreements 18
- 6. Ethical aspects 18
- 7. DMP update 19
- 8. References..... 19

List of tables

Table 1. Data Summary.....	7
Table 2. Research outputs' availability	10
Table 3. Folder structure	13
Table 4. Metadata elements used in Zenodo	14

Abbreviations and acronyms

DMP – Data Management Plan

G21 – Generation XXI Cohort

GDPR – General Data Protection Regulation (REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC)

GIS – Geographical Information System

ISPUP – Instituto de Saúde Pública da Universidade do Porto

PI – Principal Investigator

UP – University of Porto

1. Introduction

1.1. Project overview

Cities and neighbourhoods are under constant change. Since the end of the 2008 economic recession, many European cities are facing an unprecedented housing crisis characterized by increasing housing prices and a lack of public investment in housing provision, which is catalysing neighbourhood change processes such as gentrification, residential insecurity and displacement [1]. Gentrification is the process by which higher income people occupy disadvantaged neighbourhoods. This process poses several challenges related with housing availability and affordability, leading to the displacement of the poorest residents and to profound physical and sociocultural transformation of the neighbourhoods [2]. Portugal is amongst the countries most severely affected by the housing crisis. In just four years, between 2015 and 2019, housing prices rose by 40% [3] with staggering growth rates in the two main cities, Porto and Lisbon. With the current COVID-19 pandemic, a reduction in family income and an increase in the financial effort on housing-related expenses are expected [4]. Further, the lockdown precipitated by COVID-19 led to a greater exposure to the residential environment, which can be critical for the 15% of the population living in precarious housing [4]. Thus, the availability and affordability of decent housing and the provision of healthful living environments has become an important societal and public health concern.

Our health is shaped by our genetics and behaviours, but also by the neighbourhoods into which we are born and live [5]. So, gentrification and displacement might have both positive and negative health consequences among residents. On the one hand, the revitalization of gentrified neighbourhoods is associated with improvements in environmental quality and access to community resources, which may benefit populations' health [6]. On the other hand, changes in neighbourhood composition and forced displacement may be a source of stress and residential insecurity [2]. Despite the relevance and contemporaneity of this topic, there is a dearth of research on the health effects of these urban shifts and existent literature suffers from a number of methodological weaknesses that limit their validity and generalizability [2].

HUG is an innovative and multidisciplinary project that integrates knowledge from geography, sociology, epidemiology, urban planning and environmental sciences to investigate the health effects of gentrification, displacement and housing insecurity using for that geospatial data from two population-based cohorts from Porto Metropolitan Area (PMA) covering different life stages (childhood, middle and later adulthood). In this project, we will measure gentrification, housing insecurity and displacement at neighbourhood-level over the last two decades using mixed data sources from the census, commercial and administrative databases and novel statistical methods. Because participants' locations are geocoded, these neighbourhood-level measures will be integrated with the information about the physical and mental health of the two ongoing cohort studies.

Based on these longitudinal datasets, and using the current neighbourhood change processes as a natural experiment, we will fit mixed-effects regression models to compare the same individual's health before and after neighbourhood change [7]. Because neighbourhood change processes may have a

stronger health impact on particular population groups, we will also investigate the moderation effects of socioeconomic position, tenure status, age, gender and marital status. Additionally, using structural equation modelling, we will explore to what extent the observed associations derive from improvements or deterioration in neighbourhood physical quality (e.g. green space, pollution), which will be assessed through Geographical Information Systems and a virtual audit. This strong quantitative approach will be seasoned with a community-based participatory qualitative study, which will convey depth to our analysis and empower citizens. For that, we will conduct semi-structured interviews and a Photovoice study in a sample of cohort participants to capture their experiences and opinions about the health, social and community consequences of current neighbourhood change processes [8].

The primary expectations of the HUG project are to impact on the scientific community through the production of impactful and high-quality publications, to empower the community and to spread the resulting knowledge through a strict dissemination plan in order to contribute to sustain policies that guarantee healthy and inclusive urban environments.

1.2. Data management plan structure

Given the large amount of data collected and generated by the HUG project (listed on [Table 1](#)) a data management plan (DMP) is needed in order to specify what data is being processed and to manage how that data will be handled, stored and shared.

The DMP is structured as follows:

- The [Data Summary](#) section outlines the data expected to be collected and generated throughout the project. It also justifies that collection by specifying the project objectives and attributes data management responsibilities.
- The [Fair Data](#) section outlines the guidelines and measures for promoting the FAIR principles (findability, availability, interoperability and re-usability) in the project such as making the data available in an open repository, assuring data documentation, using metadata and following standards and conventions. Data sharing restrictions are also mentioned.
- The [Data Security](#) section provides an overview of data security measures to be adopted in the project.
- The [Data Protection](#) section provides the legal framework and data protection measures to manage the processing of sensitive and personal data.
- The [Ethical aspects](#) section addresses any ethical issue involving the project's data processing activities.

At the end of this document is also defined the DMP update frequency.

2. Data Summary

HUG is an innovative and multidisciplinary project that integrates knowledge from geography, sociology, epidemiology, urban planning and environmental sciences to investigate the health effects of gentrification, displacement and housing insecurity. The project has the following specific objectives:

- Measure neighbourhood socioeconomic change and gentrification since 2000 until 2020;
- Establish the residential histories of the cohort participants and estimate associations between residence in gentrified/non-gentrified neighbourhoods and displacement, and mental and physical health;
- Measure housing insecurity and investigate its space-time evolution, namely in the context of the COVID-19 pandemic, and how it affects participants' health;
- Evaluate the impact of neighbourhood change on the magnitude of socioeconomic inequalities in health, since neighbourhood change processes may increase social segregation;
- Quantify the degree of physical change occurring in participants' neighbourhoods using GIS and virtual audits and, thereby, explore the potential mechanisms beyond the health consequences of gentrification and displacement;
- Investigate if the observed effects are similar across gender, age, socioeconomic position and tenure status, as certain groups may be more vulnerable to housing conditions neighbourhood's change;
- Understand and collectively discuss the community's experiences and perceptions about the social and health effects of the actual and future neighbourhood change.

To achieve these project objectives, the following data will be collected:

Table 1. Data Summary

Type of data	Source	Description/observations
Contact, demographic and socioeconomic data	EPIPorto and Generation XXI cohorts	This includes addresses for georeferencing, i.e., assigning a geographical coordinate to each participant. Moreover, demographic and socioeconomic variables at individual level such as gender, age, income, education, occupation among other will be used.
Mental health information		Data regarding quality of life, depression and cognitive function,

		behavioural, socialization, and attention problems and biomarkers of stress.
Physical health information		Data such as medical diagnoses, hospitalizations, mortality, pain symptoms, functional limitations, lung function, body mass index and other cardio-metabolic risk factors.
Housing data		Overcrowding, poor housing conditions, residence in social housing and in other substandard housing types, housing overburden, COVID-19 housing payment assistance, difficulties in having their ends meet and, frequent house moves.
Population and housing census	INE (Instituto Nacional de Estatística – Statistics Portugal)	Socioeconomic variables from the 2001 and 2011 census. Data from the 2021 census is not yet available.
Socioeconomic deprivation indexes	European Deprivation Index 2001 and 2011 (available on figshare)	These indicators calculated by the project PI will be used to measure neighbourhood socioeconomic change/gentrification.
Property costs data	Geographic Information System of the Municipal Property Tax	Location coefficient: measure of the property costs used for tax purposes to measure neighbourhood socioeconomic change/gentrification.
Tourist accommodations	Registo Nacional de Turismo (RNT)	Number of tourist accommodations in a specific location to measure touristification/gentrification.
Omnidirectional imagery	Google Street View	To assess the presence and quality of environmental features, such as green space, food outlets and community services, as well as micro-scale features of the built environment that impact how people feel about being in specific places
Satellite imagery	Landsat 5 and 8 (NASA and the U.S. Geological Survey (USGS))	To estimate green space exposure, using the NDVI (Normalised Difference

		Vegetation Index) and to assess temperature (thermal band data).
Climate and meteorological data	Portuguese Institute for Sea and Atmosphere (IPMA)	Data from meteorological observations to calculate and validate climate extremes estimates
Air pollution levels	QUALAR (Agência Portuguesa do Ambiente)	Hourly air pollution levels from the National Monitoring Network
Land cover	CORINE Land Cover (European Environment Agency) and COS (Direção Geral do Território) databases	Ancillary data to produce air pollution estimates.
Traffic information	HERE Technologies	Ancillary data to produce air pollution estimates.
Noise levels	City Councils	
Participant's experiences and opinions (audio recordings and photographs)	Semi-structured interviews conducted in the HUG project	The interviews and the photovoice study will generate photos, field notes and transcripts from audio recordings that detail the participant's experiences and opinions about the health, social and community consequences of gentrification and displacement.
	Photovoice study conducted in the HUG project	

At this time, it's not possible to estimate the size of the data to be collected and generated.

With these data the research team will be able to build a geodatabase containing all participants' locations, as well as neighbourhood-level and gentrification-related exposures to improve versatility and usability, optimize performance and storage and to facilitate spatial analysis. The data analysis will also generate webmaps and technical reports about environmental neighbourhood-level exposures.

These data will be useful to other researchers working the in geography, sociology, epidemiology, urban planning and environmental sciences.

Responsibility for data management will be shared by the PI and co-PI of the project and ISPUP's IT department. In the case of qualitative data, the researcher conducting the qualitative study (Photovoice and interviews) will also be responsible for managing the generated data.

3. FAIR data

Data management should follow the 'FAIR' guidelines, which means that data should be findable, accessible, interoperable and re-usable [9]. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation solution.

HUG is focused in sharing and opening as much data collected within the project as possible, within the limits imposed by confidentiality obligations and data protection requirements.

3.1. Data availability and open access

The projects research outputs will be made openly available by default except those containing any personal and sensitive data (e.g. locations and health data from the cohorts' participants). **Table 2** lists the various outcomes of the project, their availability and place of publication.

All data susceptible of open access will be made available through **ISPUP's community** on Zenodo making it the preferred deposit location for dissemination and long-term preservation.

Table 2. Research outputs' availability

Research outputs	Availability	Location
Ecologic indicators	Open-access	ISPUP's community on Zenodo Project's website
Webmaps	Open-access	Project's website
Photographs	Open-access if authorised by the participants	ISPUP's community on Zenodo Project's website
Interviews transcripts and field notes	Not publicly available	--
Technical reports	Open-access	ISPUP's community on Zenodo Project's website
DMP	Open-access	ISPUP's community on Zenodo Project's website
Statistical analysis codes	Open-access	ISPUP's community on Zenodo
Research protocols and ethics approvals	Open-access	ISPUP's community on Zenodo
European Deprivation Index (Portuguese version)	Open-access	ISPUP's community on Zenodo ISPUP's website figshare
Preprints	Open-access	ISPUP's community on Zenodo
Articles	Open-access	Peer-reviewed open-access journals

Ecological indicators will be made available as soon as they are validated by a scientific publication.

Regarding the photographs produced by participants during the photovoice study, an authorisation for making them openly available will be requested from its authors. These photos will not be exhibited/published online before the physical exhibition (estimate date: December 2022).

Text files resulting from the interviews and photovoice will not be made available, as they will likely contain very detailed information that can hardly be fully anonymised and remain useful. Therefore, in order to protect the research participants, they will not be shared.

The Portuguese version of the European Deprivation Index (regarding the years 2001 and 2011) previously made available on figshare by the project PI will also be made available on Zenodo.

The results will be published in peer-reviewed open-access journals and the pre-prints (before peer review) will be published in Zenodo to ensure all scientific outputs are accessible to all the interested parties (including lay people, non-academic professionals and decision makers).

There will be no access restrictions for further non-commercial uses of the data made available by the project in open repositories. Both the photos and the ecological indicators will be licensed through a Creative Commons licence: Attribution - Non-Commercial 4.0 International (CC BY-NC 4.0). Since this license is not foreseen in Zenodo's metadata options, the license will be indicated in the description field or in an attached file.

Pre-existing data used by the project will be available for use in accordance to the policies of their custodians. Qualitative textual data will not be reusable in order to protect the privacy of participants, as mentioned above.

3.2. Documentation and metadata

Data documentation

Data documentation is an extremely important process carried out throughout the project life cycle that allows the capture of the work being done in a way that enables others (or even the research team) to understand, interpret, assess, and repeat that work. This documentation can include information about the purpose of data collection and data collection procedures, the structure and organization of the data files, context metadata of data collection, data confidentiality, access, and use conditions, among other information [10].

To ensure future interpretation of the tasks carried out, the research will use the following documentation methods:

- Inclusion of text files on the folders next to the research results, e.g., a text file will be created on the photographs folder stating the code of author, date of delivery, in which group session it was discussed, discussion results;

- Inclusion of descriptive data within text files, e.g., for an interview transcript: at the top of the transcript it'll be included the code, date, location and duration of the interview and other annotations deemed relevant;
- Inclusion of codebook and data dictionaries within worksheet files;
- Inclusion of information about the data made publicly available and its interpretation in the description field of the repository used and/or in an attached *readme.txt* file.

File formats and software

Data will be stored in popular file formats, which can be opened by multiple programs in different operating systems: .shp and .geoTIFF for maps, .csv for ecological indicators, .jpeg for photos, .wav for sound files and .doc or .txt for text files.

To open .shp and .geoTIFF files GIS software (such as ArcGIS or QGIS) will be necessary. For .csv and .doc files any office suite can be used (such as Microsoft Office, OpenOffice, etc.). For .txt it can also be used Notepad or other word processor software. For .jpeg and .wav files any multimedia software can be used.

File naming convention

File names should be clear and meaningful in order to facilitate data access and understanding of its contents. The file name should convey metadata elements such as:

- Description of the content
- Name of the creator (initials are preferred)
- Version and/or date of creation/last modification

Given the variety of the data collected and generated, this DMP will not establish a file naming policy for the whole project. However, these guidelines should be followed:

- Special characters should be avoided (can cause readability issues within certain operative systems);
- Underscore or hyphen should be used over spaces (can cause readability issues within certain operative systems);
- Preference is given to minuscular letters instead of capital letters;
- The files should have short and relevant names that convey the necessary context and descriptive information;
- File naming should be consistent throughout the project and among the research team;
- The dates should follow the YYYY-MM-DD format in order to maintain the chronological order of the files and facilitating sorting and browsing. This also allows for the compliance with the international standard for representation of date and time – [ISO 8601](#).

Folder structure

To facilitate the use and recuperation of information a predetermined folder structure for the project's documents will be created.

Table 3 presents a base folder structure that can be adapted depending on the research continued needs. Although this structure can be changed throughout the project, some rules should be followed:

1. The folder names should be clear and concise to better convey their contents;
2. The folder structure should not exceed four levels of subfolders;
3. If multiple folders have the same type of subfolders, these subfolders should have a prefix that indicates its preceding folder (see folders 'Discussion-sessions' and 'Interviews' on Table 3);
4. The use of spaces should be avoided (using hyphen or underscore instead).

Table 3. Folder structure

Root folder (located on the internal server): HUG-project			
Level 1 folder	Level 2 folder	Level 3 folder	Level 4 folder
Cohort-data	G21		
	EPIPorto		
Ecological-data	Gentrification		
	Physical-environment		
	European-Deprivation-Index		
Qualitative-data	Photographs		
	Discussion-sessions	DS-audio-files	
		DS-transcripts	
		DS-field-notes	
	Interviews	I-audio-files	
I-transcripts			
Scientific papers	Paper 1	P1-Statistical analysis	
		P1-Main-text	
		P1-References	
	Paper 2	P2-Statistical-analysis	
		P2-Main-text	
		P2-References	
Reports	Technical-Reports		

	Scientific-Reports-FCT		
Dissemination	Website -materials		
	Social-Media		
	Press-releases		
	Urban-Health-Conversation-Series		
	Conference-presentations		

Metadata

Regarding metadata, some information may already be integrated in the software used or repository where the data will be made available, namely in the metadata provided by Zenodo. The Zenodo repository stores all metadata in a JSON-format according to a defined JSON schema [11]. The metadata can also be exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema [11] which facilitates data interoperability due to their popularity.

The following metadata elements from Zenodo will be used:

Table 4. Metadata elements used in Zenodo

Metadata element	Content
Type	Dataset, Photograph, Preprint, Data Management Plan, Report
DOI	Digital Object Identifier generated by Zenodo or DOI already assigned
Publication date	Format: YYYY-MM-DD
Title	The title of the dataset/publication
Creator	The person(s) responsible for creating the dataset/publication
Description	Information about the data made available and its interpretation or abstract
Version	The version of the dataset
Language	ENG or PT
Keywords	Topics related to the dataset
Access Rights	Access right for record. Default: Open Access
License	Creative Commons Attribution - Non-Commercial 4.0 International
Funding	FCT - PTDC/GES-OUT/1662/2020
Related identifiers	Related research outputs.

Any other research specific metadata will be made available in a *readme.txt* file attached to its respective data or publication.

GIS's software (such as ArcGIS and QGIS) also allows the extraction of metadata to a XML file. These files store the following metadata from the maps: title, tags, summary, description, credits, use limitations, extent and scale range. These files will be made available along the information processed by the software.

4. Data security

This section provides the security measures and provisions to be applied in the HUG project to assure data security.

Data access

1. The computers where the files with personal data are decrypted and accessed should have access control (user-name/password) at user level (authorized researchers);
2. Due to the sensitive nature of the cohort's data, namely, individual-level clinical, behavioural and sociodemographic data, these data will only be accessible upon request to the cohort coordinators;
3. Participant's neighbourhood-level exposures will be deposited in the cohort central databases and these data will only be accessible upon request to the cohort coordinators;
4. Files should be compressed into archives (zip, rar, 7z, etc) with encryption of the archive with a password.
5. Regarding passwords:
 - a. The same password must not be reused in different files.
 - b. The passwords used for encrypting the files should be known only to the researchers directly responsible for the confidentiality of the data.
 - c. Passwords should be long, minimum 20 alphanumeric characters, but they should be easy to memorize and verbally transmitted, in order to avoid writing them down.
 - d. Passwords should be stored in password registration applications, with local registration (on the computer and not in the cloud), such as Keepass, where the local registration is protected by a password/password known only to the respective researcher.

Data handling and sharing

6. All software used to process, generate and transfer data should be updated and located in ISPUP or UP servers/computers;

7. For sharing data within the research team (and eventually with a subcontracted third party), the platform [FileSender](#) (hosted by UP) will be used to allow the safe transfer of research data and sensitive information;
8. Audio recordings of interviews, photovoice analysis and discussion sessions will be created, transcribed and then deleted by the end of the project.

No international data transfers are envisaged in the project.

Data storage and backup

9. All data processed or generated in the project will be safely stored in ISPUP's internal server on a folder only accessible to the research team;
10. Since there's no VPN solution available, these data will also be stored on the researcher's laptops for off-site work. When this occurs all data access and sharing measures should be strictly followed and the information should be deleted after the research is done.
11. The data available for further uses will be stored in a certified repository for long term preservation and curation, namely in the ISPUP community on Zenodo;
12. External portable storage devices, such as external hard drives and USB drives and cloud storage solutions such as Google Drive are not recommended for long-term storage and are strongly ill-advised for the storage of sensitive information;
13. To prevent data loss, backups will be made regularly and stored in two different media (on ISPUP's internal server and on the PI's computer).

5. Data protection

5.1. Data protection principles and legal framework

All personal data processing within the project must comply with applicable law. The following applicable legal framework will be considered:

- REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) – GDPR.
- Portugal: Law n.º 58/2019, of 8 August (LPD). Data Protection Authority: Comissão Nacional de Protecção de Dados – CNPD

Giving this framework, the following principles apply:

1. Only personal data collected according to law may be processed;

2. All personal data processing must comply with the GDPR principles, namely, fairness, lawfulness and transparency; purpose limitation; data minimisation; data quality; security, integrity and confidentiality (article 5 of the GDPR) [12].
3. The basis for processing personal data within the project is the informed consent of data subjects, as established by articles 6 and 9 of the GDPR. This informed consent shall be designed and collected in accordance with GDPR criteria, besides complying with international ethical requirements commonly accepted (e.g. Helsinki Declaration);
4. Datasets and other data generated from the project may not be used for any other purpose than the project scientific goals.

5.2. Data protection measures

To ensure compliance with the legal framework mentioned above, the following provisions will be put in place:

Informed consent

Due to the use of cohort data, there's two major instances of informed consent:

- Generalized informed consent given at each cohort's periodical assessments to use their personal data to research purposes (out of the HUG project scope)
- Project specific informed consent given when contacting and interviewing for the HUG project (within the HUG project scope)

When requesting the cohort's data to its coordinators, the research team will ensure that the informed consent was renewed on the last cohort's assessment.

Within the HUG project informed consent will be requested for (1) the discussion sessions, (2) the interviews, (3) the photovoice study, (4) the exhibition of a selection of photographs taken by the participants, and (5) the publishing of the photographs on Zenodo and of a selection of photographs in a book. Consent for (4) and (5) is regarding the publishing of the photograph's author name (to also assure intellectual property rights) – if the participants wish to remain anonymous, they will be free to do so.

The signed informed consents will be stored separately in a locked cabinet.

The templates for informed consent will be attached to the DMP.

Anonymisation

The datasets made publicly available will only contain ecological indicators and any personal data will be removed to ensure total anonymity. Anonymisation of datasets will be made manually or by the [Amnesia software](#).

The recordings of the photovoice sessions and interviews will be anonymised as fully as possible after transcription. These recordings will be encrypted/password protected and retained until the end of the project.

Data Processing Policies and Agreements

A micro-website will be created for the HUG project in order to promote its research outcomes. A privacy policy and a cookie policy will be necessary to ensure the website's compliance. This task will be made in coordination with the Data Protection Group and the website's developer subcontractor.

The interview transcriptions will be carried out by a third party. If carried out by a subcontracted company, a Data Protection Agreement will be made establishing the conditions of the processing of data by the company. If a research fellow is hired for the specific task, a Non-Discloser Agreement will be signed to ensure confidentiality of any personal data revealed on the recordings.

The audio recording will be sent to this third party by UP's FileSender with password protection and the files will be named with the ID given to the participants (pseudonymization). This way only the research team can identify to each participant the recording and transcription belong to.

6. Ethical aspects

This project will involve humans from two well-established cohorts – G21 and EPIPorto. G21 and EPIPorto comply with national legal/ethical frameworks. The studies were approved by the University of Porto Medical School/ Hospital S. João Ethics Committee and a signed informed consent was required for all participants. All the phases of these studies comply with the Ethical Principles for Medical Research Involving Human Subjects expressed in the Declaration of Helsinki. No invasive tests will be performed.

The collection of qualitative data will be approved by Ethics Committee the Institute of Public Health of the University of Porto. The participants in the study will receive information about the purposes of data collection and how it will be managed, and written informed consent will be requested from them.

The photovoice study will include a discussion with the participants about the ethical implications of using photography for research purposes and the political aspects of this methodology.

To conduct the present project additional ethical approval (addendum), and approval by ISPUP's Data Protection Office (DPO) will be obtained.

Concerning the protection of personal data, both G21 and EPIPorto projects involve the collection of identifiable personal data and participants' locations. G21 and EPIPorto projects were submitted and approved by the national data protection authority, CNPD (Comissão Nacional de Proteção de Dados).

Due to the sensitive nature of the cohort's data, namely, individual-level clinical and behavioural data, it will only be accessible upon request to the cohort coordinators and any personal and sensitive data will not be made available on the repository.

Ethics approvals and informed consent documents will be attached to the DMP.

7. DMP update

Updated versions of the DMP will be provided on a biannually fashion or whenever is justified.

8. References

- [1] A. Pittini, G. Koessler, J. Dijol, E. Lakatos, and L. Ghekiere, "The state of housing in the EU," Brussels, 2017. [Online]. Available: <https://www.housingeurope.eu/resource-1000/the-state-of-housing-in-the-eu-2017>
- [2] A. S. Schnake-Mahl, J. L. Jahn, S. V. Subramanian, M. C. Waters, and M. Arcaya, "Gentrification, Neighborhood Change, and Population Health: a Systematic Review," *Journal of Urban Health*, vol. 97, no. 1, pp. 1–25, Feb. 2020, doi: 10.1007/s11524-019-00400-1.
- [3] INE, "Índice de preços da habitação (Base - 2015) por Localização geográfica e Categoria do alojamento familiar," 2015. https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&contecto=pi&indOcorrCo d=0009201&selTab=tab0&xlang=pt
- [4] Housing Europe, "#StayAtHome and Europe's housing crisis," Mar. 20, 2020. [Online]. Available: https://www.housingeurope.eu/resource-1376/stayathome-and-europe-s-housing-crisis?mc_cid=a7cf809c48&mc_eid=63616f2b07#.XnoQ2ckbtSw.facebook
- [5] A. I. Ribeiro, "Public health: why study neighborhoods?," *Porto Biomedical Journal*, vol. 3, no. 1, p. e16, Aug. 2018, doi: 10.1016/J.PBJ.0000000000000016.
- [6] M. Steinmetz-Wood, R. Wasfi, G. Parker, L. Bornstein, J. Caron, and Y. Kestens, "Is gentrification all bad? Positive association between gentrification and individual's perceived neighborhood collective efficacy in Montreal, Canada," *International Journal of Health Geographics*, vol. 16, no. 1, p. 24, Jul. 2017, doi: 10.1186/S12942-017-0096-6.
- [7] X. Basagaña *et al.*, "Analysis of multicentre epidemiological studies: contrasting fixed or random effects modelling and meta-analysis," *International Journal of Epidemiology*, vol. 47, no. 4, pp. 1343–1354, Aug. 2018, doi: 10.1093/IJE/DYY117.
- [8] C. C. Wang, "Photovoice: a participatory action research strategy applied to women's health," *J Womens Health*, vol. 8, no. 2, pp. 185–192, 1999, doi: 10.1089/JWH.1999.8.185.
- [9] European Commission's Directorate-General for Research & Innovation, "Guidelines on FAIR Data Management in Horizon 2020," Jul. 2016. [Online]. Available: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- [10] University of Illinois, "Introduction to Data Management for Undergraduate Students," 2022. <https://guides.library.illinois.edu/introdata/documentation> (accessed May 16, 2022).

- [11] Zenodo, “General Policies.” <https://about.zenodo.org/policies/> (accessed May 11, 2022).

- [12] European Parliament and Council of the European Union, *Regulation (EU) 2016/679 of the European Parliament and Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Prot*, vol. L, no. 119. 2016. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679&qid=1605975122588>