Immigrant Activity-Space Segregation: Spatio-Temporal Patterns of Assimilation and Separation in Barcelona

Proposal 657956-IASS (MSCA-IF)

DATA MANAGEMENT PLAN

This Data Management Plan (DMP) provides an analysis of the main elements of the data management policy that will be used for each data set generated by the IASS project. This is not a fixed document, but will instead evolve during the lifespan of the project

The IASS will generate two data sets: (1) the spatio-temporal data of approximately 400 volunteers collected using mobile phone geolocation (IASS-GEOTRACKS), and (2) the interview notes and survey responses collected from a subset of these volunteers (IASS-INT/SURV). In both cases, the data sets will be made open access after ensuring that no participant may be personally identified in them.

Spatio-Temporal Mobile Phone Data (IASS-GEOTRACKS)

Data set reference and name

Reference: IASS-GEOTRACKS
Name: IASS Spatio-Temporal Mobile Phone Data

Data set description

This data set will consist of time-stamped geolocations showing the activity spaces of each of approximately 400 volunteers. The data will be collected via the Space Mapper mobile phone application, which anonymizes geolocations by placing each precise location into a predefined grid of 4 hectare cells and recording only the cell identifier rather than the exact location within the cell. This information will be linked to data on each participant's ethnicity, country of birth, and immigration history, but no personally-identifying data will be included.

The time-stamped geolocation table will consist of approximately 400,000 records, with fields for time, anonymization cell ID, and anonymized user identifier. The participant table will consist of approximately 400 records, with fields for user identifier, ethnicity, country of birth, year of arrival in Spain, year of arrival in Barcelona.

This data underpins the quantitative analysis of activity-space segregation, which will form part of various scientific publications. It could be useful to other researchers interested in segregation, activity spaces, or human mobility more generally.

Standards and metadata

The metadata will consist of:

1. The project name (IASS), description, and implementation dates;
2. The name and contact information of the researcher (John Palmer);
3. The name and contact information of the host institution (UPF);
4. The funding source (MSCA-IF);
5. A complete description of the sampling method by which participants were selected;
6. A description of the mobile phone application through which the spatio-temporal data was collected, and a link to the application's source code on Github (https://github.com/JohnPalmer/SpaceMapper).
7. An explanation of the parts of the data that have been removed from the open access version to ensure anonymity of participants.

Data sharing

This data set will be made open access by placing it on the Zenodo platform (https://www.zenodo.org/) after ensuring that no participant may be personally identified in it. Each table will be uploaded as a comma separated values file, making it broadly accessible with different analysis tools. The data set will be placed in the public domain, with a Creative Commons "No Rights Reserved" (CC0) license.

Since one goal of the data collection methodology is to ensure that no collected data is personally identifiable, almost the entire data set will be shared as is. However, the subsequent review to ensure anonymity may dictate that certain information be removed. For instance, exact dates may be removed, leaving only time-of-day, and day-of-week, and some participant characteristics may be removed to ensure that re-identification of users through links with other data sets is not possible. In all cases, the only information not made public will be that necessary for ensuring the privacy of participants.

Archiving and preservation (including storage and backup)

All original data will be encrypted and preserved on a USB flash drive as well as on a CD-ROM for at least 10 years from the end of project implementation. All open access data will be stored in the CERN data center through the Zenodo platform.

The total volume of the data set is not expected to exceed 8 gigabytes. The cost of the USB flash drive and CD-ROM storage is negligible (less than 20 euros), and the Zenodo storage is free.

Interview Notes and Survey Responses

Data set reference and name

Reference: IASS-INT/SURV
Name: IASS Interview Notes and Survey Responses

Data set description

This data set will consist of notes from oral interviews and written responses to surveys from approximately 40 volunteers (a subsample of the 400 participants in the IASS-GEOTRACKS data set). The data will be collected by the researcher personally meeting with selected participants. It will be initially recorded using a digital audio recording device (for the interviews) or by hand (for the surveys) in the language used by each participant, and later translated to English and transcribed to digital text to make it fully searchable. The information will be linked to the anonymous participant identifiers user in the IASS-GEOTRACKS data set.

This data underpins the qualitative analysis of activity-space segregation, which will form part of various scientific publications. It could be useful to other researchers interested in segregation, activity spaces, or human mobility more generally.

Standards and metadata

The metadata will consist of:

1. The project name (IASS), description, and implementation dates;
2. The name and contact information of the principle investigator (John Palmer);
3. The name and contact information of the host institution (UPF);
4. The funding source (MSCA-IF);
5. A complete description of the sampling method by which participants were selected and subsampled;
6. A description of the interview technique.
7. A copy of the survey form.
8. An explanation of the parts of the data that have been removed from the open access version to ensure anonymity of participants.

Data sharing

This digital text portion of the data set will be made open access by placing it on the Zenodo platform (https://www.zenodo.org/) after ensuring that no person may be personally identified in it. The anonymization process will include removing all names and other identifying information from the text. The data will be uploaded as a set of simple text files, making it broadly accessible with different analysis tools. The data set will be placed in the public domain, with a Creative Commons “No Rights Reserved” (CC0) license.

The audio recordings will not be shared and all names and identifying information will be removed from the text files before sharing them in order to ensure the privacy

of participants. In addition, the links to the IASS-GEOTRACKS data set will not be shared for the same reason.

Archiving and preservation (including storage and backup)

All original data will be encrypted and preserved on a USB flash drive as well as on a CD-ROM for at least 10 years from the end of project implementation. All open access data will be stored in the CERN data center through the Zenodo platform.

The total volume of the data set is not expected to exceed 1 gigabyte. The cost of the USB flash drive and CD-ROM storage is negligible (less than 20 euros), and the Zenodo storage is free.