GEOGRAPHIC DISPLACEMENT PROCEDURES AND DISSEMINATION OF GEOCODED DATA IN THE MULTIPLE INDICATOR CLUSTER SURVEYS (MICS) PROGRAMME

MICS METHODOLOGICAL PAPERS

Paper No. 12, 2024

Data and Analytics Section
Division of Data, Analytics, Planning and Monitoring

United Nations Children's Fund
Data and Analytics Section, Division of Data, Analytics, Planning and Monitoring
3 United Nations Plaza, New York, NY 10017, USA
Tel: +1 (212) 326-7000
Email: mics@unicef.org

Suggested citation: Gashi, N., Burdziej, J., and Beshanski-Pedersen B., (2024). *Geographic Displacement Procedures and Dissemination of Geocoded Data in the Multiple Indicator Cluster Surveys (MICS) Programme*. MICS Methodological Papers, No. 12, Data and Analytics Section, Division of Data, Analytics, Planning and Monitoring, UNICEF, New York.

Geographic Displacement Procedures and Dissemination of Geocoded Data in the Multiple Indicator Cluster Surveys (MICS) Programme

MICS METHODOLOGICAL PAPERS

Paper No. 12, 2024

Data and Analytics Section
Division of Data, Analytics, Planning and Monitoring

About MICS

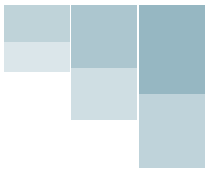
The Multiple Indicator Cluster Surveys, MICS, is one of the largest global sources of statistically sound and internationally comparable data on children and women. MICS data are gathered during face-to-face interviews in representative samples of households. The surveys are typically carried out by government organizations, with technical support from UNICEF.

Since the mid-1990s, MICS has supported more than 120 countries to produce data on a range of indicators in areas such as health, education, child protection and HIV/AIDS. MICS data can be disaggregated by numerous geographic, social and demographic characteristics.

As of 2024, six rounds of surveys have been conducted: MICS1 (1993-1998), MICS2 (1999-2003), MICS3 (2005–2010), MICS4 (2009–2013), MICS5 (2012-2017), and MICS6 (2017-2023). The seventh round of MICS (MICS7) is currently taking place in 2023–2026. Survey results, tools, reports, micro-data, and information on the MICS programme are available at <*mics.unicef.org*>.

About the MICS Methodological Papers

MICS Methodological Papers are intended to facilitate exchange of knowledge and to stimulate discussion on the methodological issues related to the collection, analysis, and dissemination of MICS data; in particular, thepapers document the background methodological work undertaken for the development of new MICS indicators, modules, and analyses. The findings, interpretation and conclusions do not necessarily reflect the policies orviews of UNICEF.



Acknowledgements

This report has been developed by Geographic Information System (GIS) and household survey experts in the Data and Analytics Section, Division of Data, Analytics, Planning, and Monitoring, UNICEF. The views expressed in this report are those of the authors and do not necessarily reflect the views of UNICEF. The authors would like to express their appreciation for the Demographic and Health Surveys (DHS) program, which developed the geographical displacement methodology utilised by the Multiple Indicator Cluster Surveys (MICS) Programme.

1

Introduction

The Global MICS Programme of the United Nations Children's Fund (UNICEF) has been recommending the capture of geospatial data in MICS surveys since its third round in 2005 (1). At the beginning, the goal was to ensure the availability of a single coordinate corresponding to the geographical location of each survey cluster. In 2005, for most surveys, this required the implementation of GPS[1] data collection within the main MICS fieldwork operation itself. However, since then, census mapping in the 2010 and 2020 rounds of Population and Housing Censuses have integrated digitised enumeration areas. This generally makes geospatial data collection less necessary, but it is still occasionally used for monitoring fieldwork, or when the required geocodes are not available from census mapping.

Since 2015, the MICS Programme has maintained a package of survey tools to support GPS data collection, including a variety of training and field manuals, a data collection form, and a data entry platform.[2] The Programme has also tested innovations in GPS data collection, such as automatic data capture and extended data capture.[3] However, with the technological shift of household surveys from paper-based, through early PDA-based[4] experiments, to now full-fledged tablet-based surveying (CAPI[5]) and the increase in the availability of digitised sampling frames, the package of tools has become less relevant and is maintained primarily for occasional use.

Regardless, outside encouraging GPS data collection or more recently recommending that a link to the geospatial data of the sampling frame is ensured, the MICS Programme never engaged in the editing or compilation of the data itself. Nowadays, the demand for geospatial data is ever-increasing, as is demand for harmonisation for the purpose of global monitoring and research. As with each element of the MICS

[1] The Global Positioning System (GPS) is a satellite-based navigation system that accurately determines the location of a radio receiver on or above the Earth's surface.
[2] https://mics.unicef.org/tools
[3] See for example: http://groundtruth.in/category/swaziland/
[4] Personal Digital Assistant (PDA) is a handheld device that combines computing, communication, and organizational features.
[5] Computer-Assisted Personal Interviewing (CAPI) is a research method in which interviewers use electronic devices, such as computers or tablets, to administer and record responses to survey questions during face-to-face interviews.

Programme, the value of providing technical support and hosting a central, harmonised spatial data repository has become more apparent than ever.

In response, the MICS Programme has moved towards an objective of making geospatial data available for as many surveys as possible, including past surveys: An initiative to support countries in collecting, editing, compiling, and disseminating geospatial data, as part of MICS surveys. The methodology and tools used for geographical displacement have been tested and have demonstrated effectiveness in ensuring both confidentiality and data integrity across multiple surveys and diverse locations (4).

Geocoded data offer numerous advantages. For instance, all characteristics of children, adolescents, women, and men found in the entire MICS datasets can be linked to the spatial locations of their corresponding clusters. By integrating MICS geocoded data with information from population, environment, or climate factors, MICS data can be aggregated to additional units of analysis, such as climatic zones, Degree of Urbanisation (2) classifications, or health facility catchment areas, thereby extending geographic studies beyond national or subnational boundaries.

When information about a respondent is revealed by a data user from a publicly released dataset, it is termed 'disclosure' (3). To ensure both the availability of georeferenced data and the protection of respondents from disclosure, the MICS Programme supports National Statistical Offices (NSO) to anonymise the cluster locations before making them publicly available. This paper presents the technical approach for anonymising cluster location data and making it accessible through the MICS website. This approach is grounded in the Programme's close collaboration with NSOs involved in MICS surveys, facilitated through UNICEF Country Offices.[6] This effort aims to gradually expand the availability of geocoded MICS datasets. By fostering partnerships with NSOs, the Programme endeavours to enhance access to spatially referenced MICS data, thereby enabling a broader range of analyses and insights.

This paper provides a comprehensive explanation about the methodology of displacement of cluster geocodes, the use of MICS Geocode Plugin tool, and the policy of GPS data release of MICS. Each chapter details the necessary inputs required for the tool and the resulting outputs. Specifically, the paper covers the methodology for determining cluster centroids and performing geographical displacement. Additionally, it discusses how these processes ensure both data usability and respondent confidentiality. Through step-by-step instructions and practical examples, users will learn to effectively apply the tool in MICS surveys or other household surveys.

[6] MICS surveys are implemented by national implementing agencies, in technical collaboration with UNICEF. The data collected is the property of the individual governments. National teams implementing MICS surveys rely on global guidance, protocols, and tools produced by the Global MICS Programme. Due to the absence of such global guidance on the collection and dissemination of geospatial data, past MICS surveys that collected geospatial data did not make plans to disseminate it. As a result, the MICS Programme does not have immediate access to raw geospatial data for the majority of those surveys. This limitation has constrained retrospective work, but the MICS Programme is working closely with NSOs to make past data publicly available. Regardless, the primary focus has been on adopting protocols for upcoming surveys.

2

Methodological Overview

Background

The overall process for the geographical displacement of clusters in a MICS survey is derived from the methodology developed by the DHS program (4), with modifications detailed in this paper. It is important to highlight that the DHS methodology undergoes continuous refinement, as outlined in the Guidance Note on Spatial Anonymisation (5), prepared jointly by DHS Program and the Development Data Group of the World Bank for the Inter-Secretariat Working Group on Household Surveys. The guidance underscores ongoing efforts to improve methodologies in this domain. Furthermore, it's worth noting that the MICS Programme will also adapt and refine its methodology in the future as needed to ensure the accuracy and effectiveness of survey operations, underpinned by the overall protection of respondents.

To ensure both the availability of georeferenced data and the protection of respondent privacy, the methodology used for MICS surveys consists of two layers of protection:

1. Determination of MICS Survey Cluster Centroid:
   Initially, household locations within the same cluster or within the boundary of the cluster are consolidated into a single coordinate point.
2. Geographical Displacement of MICS Survey Cluster Centroid:
   Subsequently, for added confidentiality, the coordinate undergoes displacement through the cluster centroid displacement process.

After determining the cluster centroid, the coordinate is randomly displaced both in angle and distance. The displacement angle is selected randomly between 0 and 360 degrees. The displacement distance depends on the cluster's area type: urban clusters may be displaced up to 2 kilometres, while rural clusters may be displaced up to 5 kilometres. Additionally, 1 percent of randomly selected rural clusters may be displaced up to 10 kilometres. Clusters are displaced within the same administrative unit, one level lower than the sample stratification of the survey.

The geographical displacement of MICS survey cluster centroids undergoes review and validation through a collaboration between the NSO and the Global MICS Team to ensure accuracy and acceptability.
To accommodate local geographical and administrative conditions in different regions around the world, it is essential to fine-tune the application of the methodology. While this methodological paper outlines the overall approach for each of these phases, it does not address the establishment of minimum requirements for public dissemination, which is an important consideration. Some datasets may present limitations that require context-specific adjustments beyond the described methodology. For instance,

clusters located in island states may be surrounded by water, making geographical displacement very difficult. Similarly, if clusters are sampled based on population characteristics that do not exist within the displacement buffer, such as ethnic minority groupings or camp settings, the data will need to be aggregated. This means either further masking the clusters within a larger group or simply excluding the displaced coordinates, leaving the data associated only with the entire administrative area (e.g., a district, municipality, or commune). Other less obvious issues may also arise, necessitating the establishment of minimum criteria and testing of data against these criteria, eventually leading to the aggregation of such data. Any deviation from the standard methodology described in this paper will be detailed in files accompanying the data.

Tools

The tool used to carry out the methodology for the determination of cluster centroids, geographical displacement, and extraction of geospatial covariates is the MICS Geocode Plugin[7] (Figure 1), which is specifically designed for use with QGIS 3.x.[8] The use of QGIS was chosen because it is a free and open-source geospatial software. By using a desktop GIS software, users can easily view and manipulate both the input and output data, as well as compare it with other relevant layers, such as administrative boundaries, in real-time. While the Plugin is designed for MICS, it can also be utilised for other household surveys.

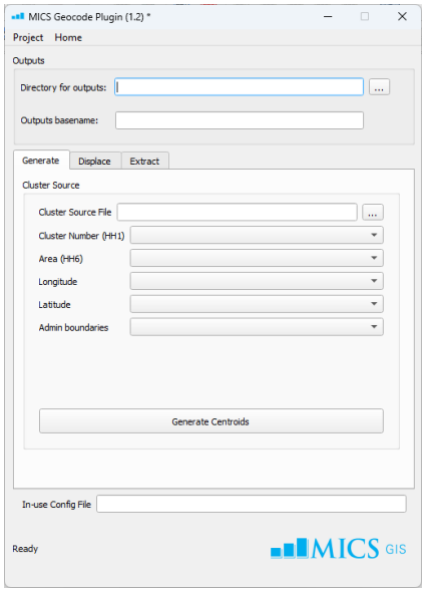


Figure 1. MICS Geocode Plugin

[7] The MICS Geocode Plugin is accessible for download on the Tools page of the MICS website. Manuals detailing installation and usage instructions for the Plugin can also be found in the same location. Similar to all MICS tools, the MICS Geocode Plugin and its accompanying manuals bear MICS branding and terminology but are applicable to any household survey.

[8] QGIS is a professional, user friendly open-source GIS application that runs on Windows, macOS, Linux, Android, and iOS. It supports viewing, editing, printing, and analysis of geospatial data in a range of data formats. https://qgis.org/en/site/

3

Cluster Centroids

This chapter marks the first phase of the geographic displacement procedure of cluster geocodes. In this stage, the centroids of survey clusters are determined, which serve as the geographic points for subsequent displacement procedures. Establishing these centroids correctly is essential for ensuring the effectiveness of the entire geographic displacement process.

The method of determining centroids depends on the available data: Either cluster locations are specified as coordinates with latitude and longitude in a spreadsheet or similar (Scenario 1), or locations are delineated as points, multi-points, or polygons in a shapefile format (Scenario 2).

In Scenario 1, survey clusters can be defined by one or more points, represented by pairs of geographical coordinates. The method used to determine the final centre of each cluster depends on the number of points comprising it. Ideally, the centre is calculated as a pole of inaccessibility, which represents the farthest location to reach within a given area.[9,10] This calculation occurs when the convex hull (outer boundary) of all input points forming a cluster results in a valid polygon (third and fourth examples in Figure 2). However, if the pole of inaccessibility cannot be determined (e.g., due to the presence of only one or two points in the cluster), the centre of the cluster is determined as the centroid of all input points (first and second examples in Figure 2).

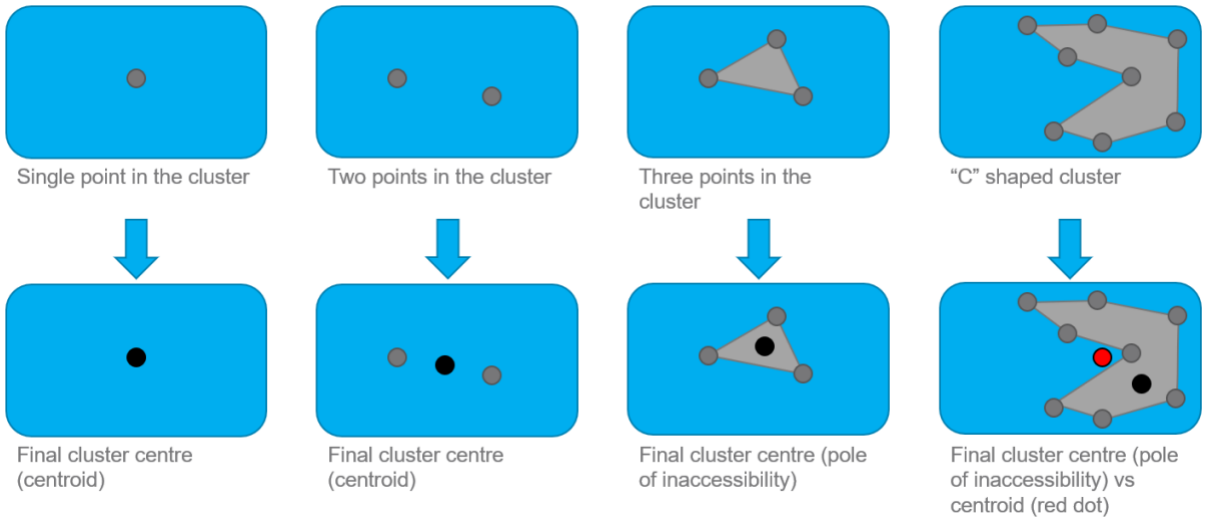


Figure 1. Cluster centroids and pole of inaccessibility from different number of cluster points

[9] https://docs.qgis.org/3.22/en/docs/user_manual/processing_algs/qgis/vectorgeometry.html#pole-of-inaccessibility

[10] https://www.geographyrealm.com/pole-inaccessibility/

In Scenario 2, when clusters are defined as points, the cluster centre coincides with the point location. In cases where a cluster is defined as a multi-point shapefile and consists of only two points, the centre of the cluster will be calculated as the centroid of those two input points. However, for clusters defined as multi-points or polygons, the cluster centres are determined using the pole of inaccessibility method.

Inputs

For Scenario 1, the following inputs are required to determine the centroids of the clusters:

- A single CSV file in a comma-separated format, encoded in UTF-8.
- A single column with a cluster number value.
- Latitude and longitude values in the WGS84[11] coordinate system, expressed as decimal values in separate columns.
- A single column with a cluster area type value, expressed as a string with only "R" or "U" values to indicate rural or urban clusters, respectively.
- All points within a cluster must have the same cluster area type.
- A single column with subnational administrative unit names or identification numbers (IDs) (e.g., Pcode) to which the cluster belongs.
- The file must be saved on a local disk.

For Scenario 2, the following inputs are required to determine the centroids of the clusters:

- A (point, multi-point, or polygon) shapefile in the WGS84 coordinate system.
- The shapefile must be topologically correct (e.g., if it's a polygon, it shouldn't have any overlaps).
- A single column with a cluster number value.
- A single column with a cluster area type value, expressed as a string with only "R" and "U" values for rural and urban clusters respectively.
- All points within a cluster must have the same cluster area type.
- A single column with subnational administrative unit names or IDs to which the cluster belongs.
- The file must be saved on a local disk.

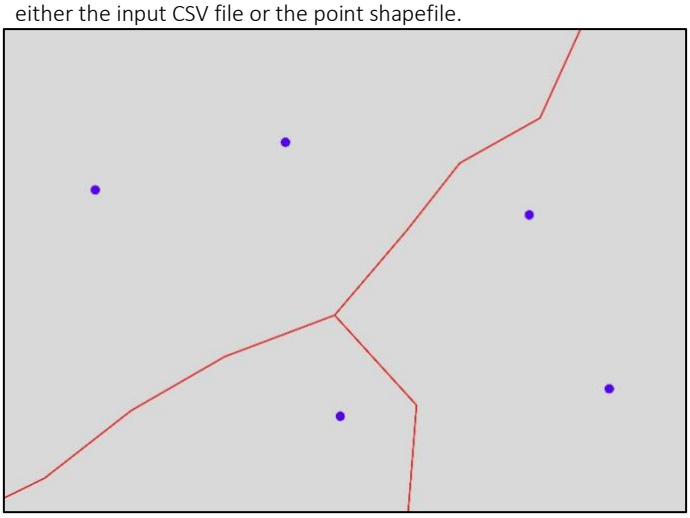
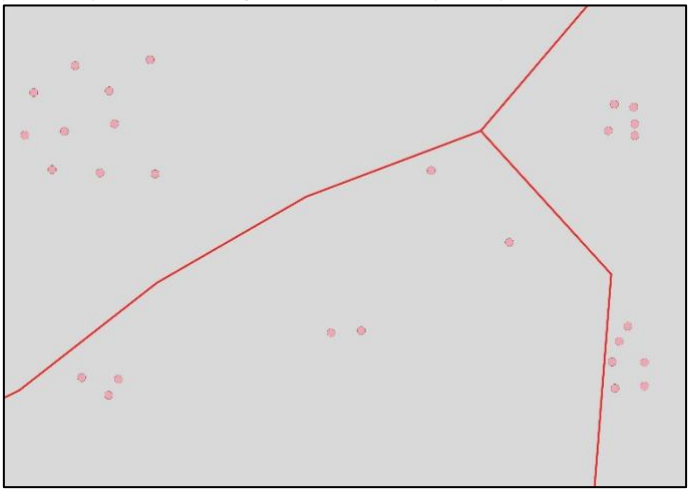
Outputs

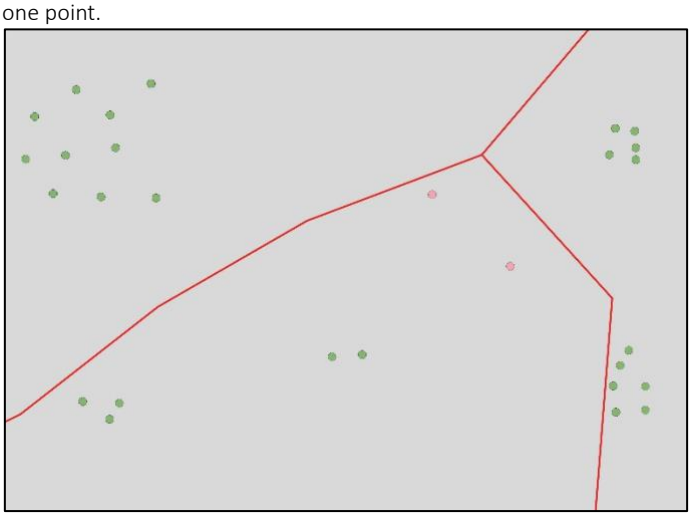
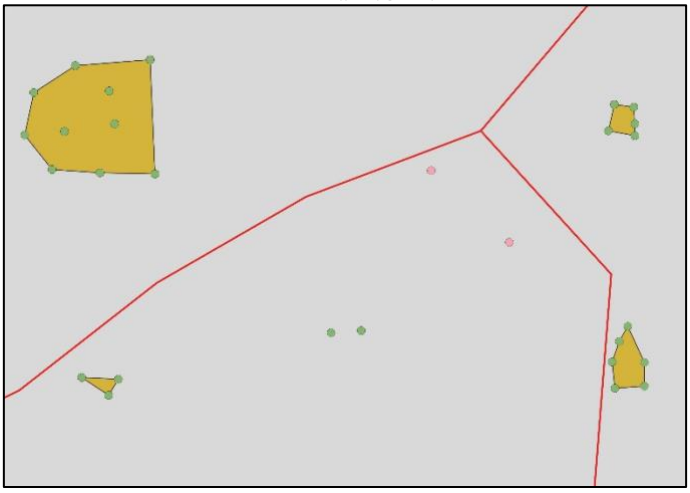
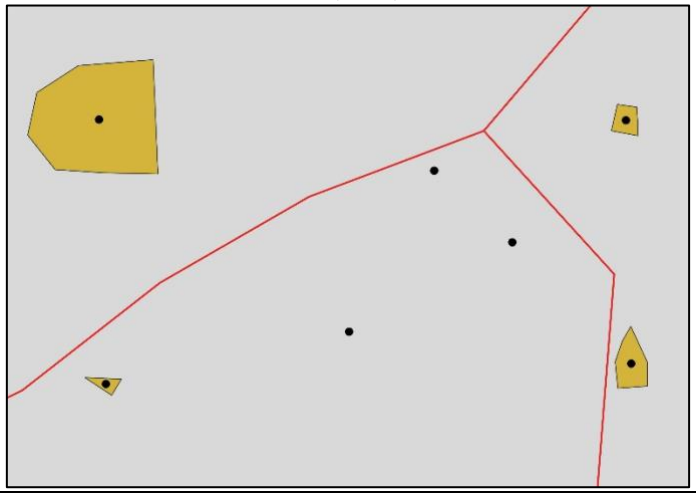
The MICS Geocode Plugin will generate multiple outputs in the process of determining the cluster centroids. Some of them will be added to QGIS as scratch (temporary) and others as permanent vector layers. The number of outputs will depend on the input type. Table 1 illustrates three types of outputs

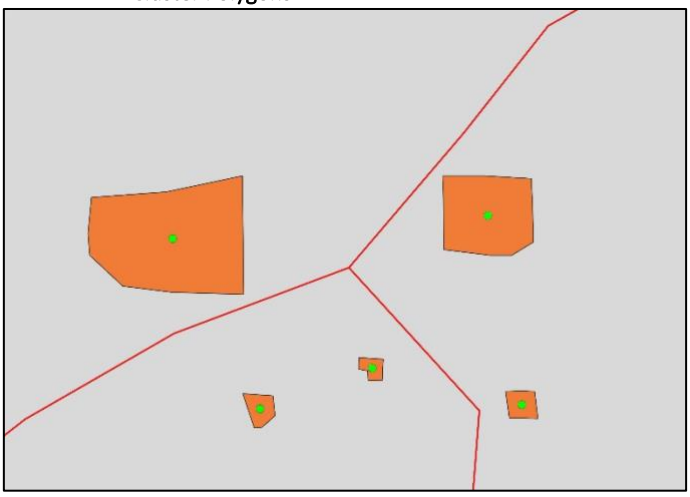
[11] The World Geodetic System 1984 (WGS 84) is a global standard used in cartography, geodesy, and satellite navigation. It serves as the reference system for GPS, enabling accurate positioning by aligning satellite-based measurements with the Earth's surface.

corresponding to the following inputs (either in CSV file or Shapefile format): point, multi-point or polygon. In each output, the red borders delineate the administrative unit boundaries.

Table1: Type of Outputs from Phase 1: Determining cluster centroids

Output	Description
Input: Point (CSV file or Shapefile)	
Cluster Centroids (blue dots)	One point per cluster. Raw GPS coordinate for each cluster, which has been imported from either the input CSV file or the point shapefile. 
Input: Multi-point (CSV file or Shapefile)	
Cluster Points (pink dots)	Multi point per cluster. Raw GPS cluster coordinates imported from the input CSV file or multi-point shapefile. There might be one or more points per cluster. 

Cluster Multi-Points (green dots)	Cluster Multi-Points refer to **Cluster Points** that have been grouped together based on their cluster number, resulting in multi-point geometry. This applies only to clusters with more than one point. 
Cluster Convex Hulls (yellow polygons)	Convex hulls are outer boundaries (polygons) of **Cluster Multi-Points**. 
Cluster Centroids (black dots)	Cluster centroids are calculated either as poles of inaccessibility or centroids derived from **Cluster Convex Hulls**. For clusters with only one point, the centroid is the same as the **Cluster Point**. 

Input: Polygon Shapefile	
Cluster Polygons (orange polygons)	Cluster polygons imported from a polygon shapefile. 
Cluster Centroids (green dots)	Cluster centroids are determined as poles of inaccessibility or centroids of cluster polygons derived from **Cluster Polygons**. 

4

Cluster Displacement

In this phase, the determined centroids of survey clusters undergo geographical displacement to ensure respondent confidentiality. The centroids are displaced randomly using the following rules:

- Displacement is conducted using a random angle and random distance approach.
- All clusters are displaced in an angle of 0 to 360 degrees.
- Urban cluster centroids are displaced up to 2 kilometres.
- Rural cluster centroids are displaced up to 5 kilometres, with 1% of the clusters randomly selected for displacement up to 10 kilometres.
- All clusters remain within the original subnational area, typically corresponding to one level lower than the sample stratification of the survey (usually admin 2[12]). If a cluster falls into a different administrative unit after displacement, the displacement process is repeated until it falls within the correct administrative unit. The maximum number of iterations allowed for this process is 20. If even after 20 times, the cluster is displaced in the wrong administrative unit, then a manual verification by the GIS specialist takes place.

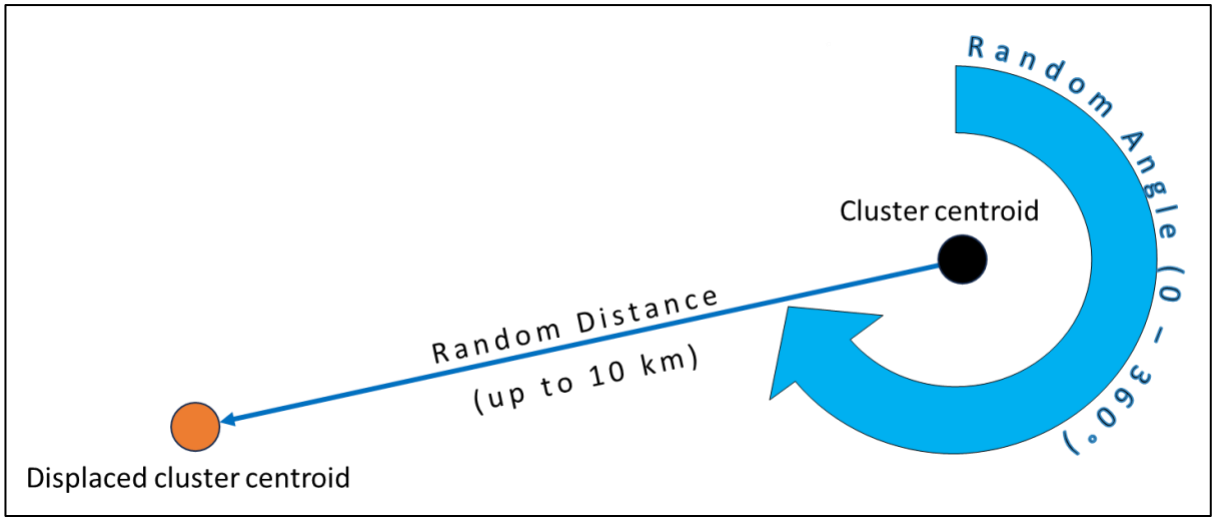


Figure 2. Process of cluster centroids displacement

[12] The second-level administrative division (Admin 2) refers to the subordinate administrative units within a country. These divisions are below the first-level administrative divisions (Admin 1). Examples include Local Government Areas (LGAs) within Nigeria or districts in Lao People's Democratic Republic.

When carrying out the displacement or building of buffers[13], the geometries are transformed from WGS84 (EPSG:[14] 4326) into Web Mercator (EPSG: 3857). This is done in order to make use of the Cartesian coordinate system. Web Mercator is chosen due to being based on the same datum (WGS84) and being a commonly used projected coordinate system in web mapping with a global coverage. However, it's worth noting the effect this may have on distance calculations at varying latitudes.

Inputs

Input 1: A point shapefile of cluster centroids:

- This shapefile is an output from the first phase - **Cluster Centroids**.
- Must be in the WGS84 coordinate system.
- A single column with a cluster number value.
- A single column with a cluster area type value, expressed as a string with only "R" and "U" values for rural and urban clusters respectively.
- A single column with subnational administrative unit names or IDs to which the cluster belongs.
- The file must be saved on a local disk.

Input 2: A polygon shapefile of administrative unit boundaries:

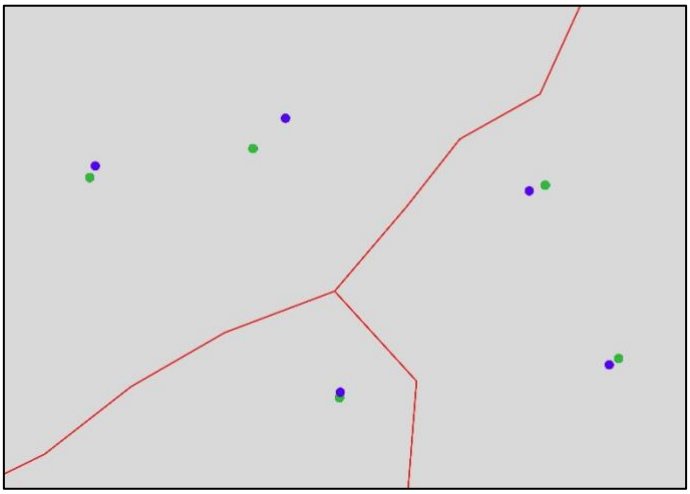
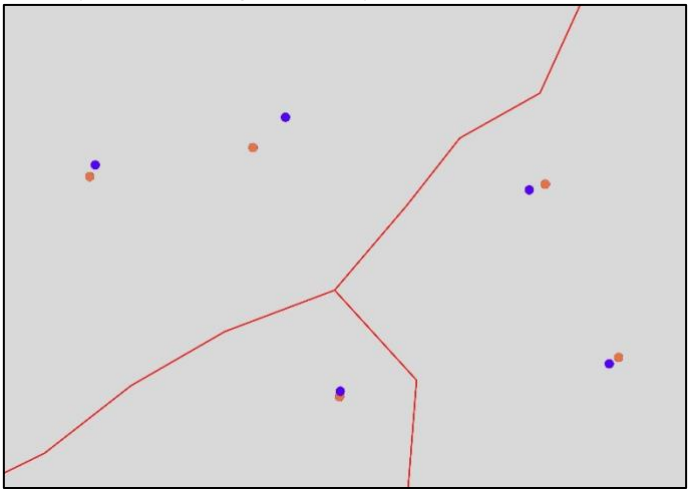
- Must be in WGS84 coordinate system.
- Must be topologically correct (e.g., no overlaps).
- Must be one subnational level lower than the sample stratification of the survey (usually admin 2).
- Must have a column with unique names or IDs (e.g., Pcode) that should perfectly match the subnational administrative unit names or IDs from the shapefile of cluster centroids.

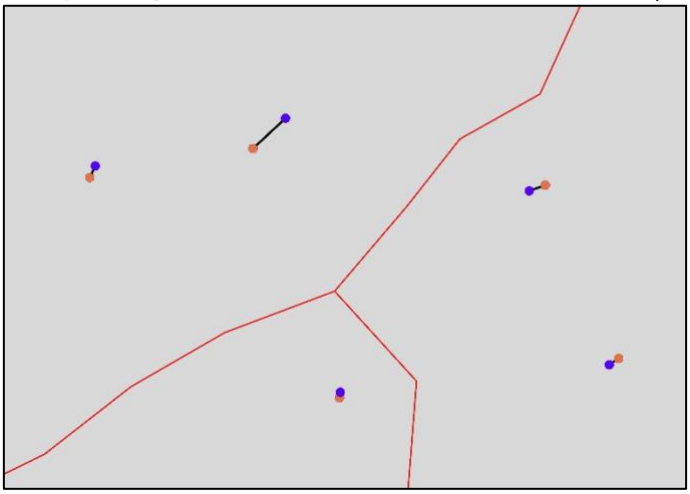
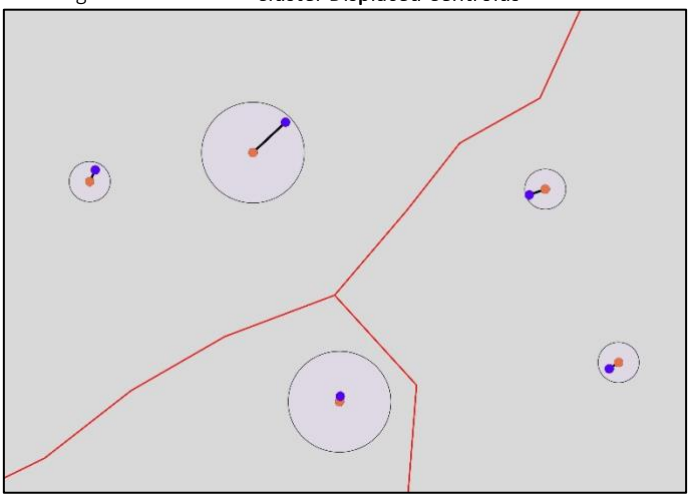
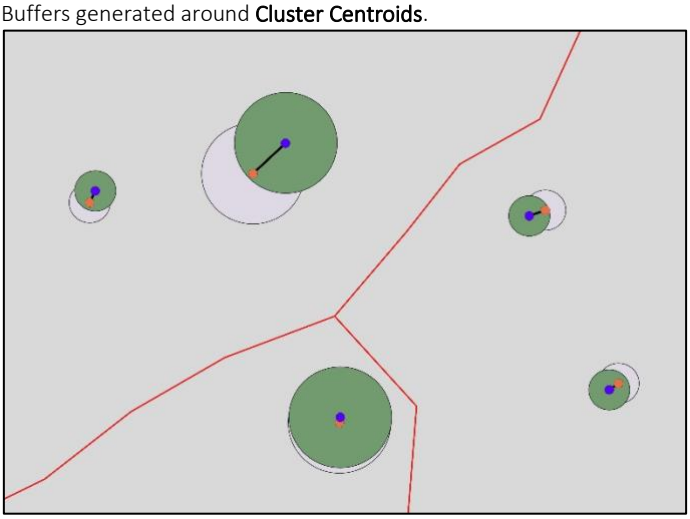
Outputs

The MICS Geocode Plugin will produce multiple outputs, some of which will be added to QGIS as scratch (temporary) vector layers, while others will be permanent layers. The number of outputs will be the same regardless of the input type, whether it is a CSV file or a Shapefile. The output figures listed in Table 2 below were generated using a CSV file input with one point per cluster. In each output, red borders delineate the administrative unit boundaries, and blue dots represent the cluster centroids generated from a CSV file input with one point per cluster.

[13] In spatial analysis, buffer refers to a zone created at a specified distance around each point feature in a geographic dataset. This buffer zone, typically represented as a circular area, extends outward from the point to define a region of interest.

[14] An EPSG code is a unique identifier used to represent coordinate systems and other geodetic properties like datums, spheroids, and units. The acronym EPSG stands for European Petroleum Survey Group, which maintains a geodetic parameter database with standard codes.

Table 2: Outputs from Phase 2: Displacing cluster centroids	
Output	Description
Cluster Unanonymised Displaced Centroids (green dots)	**Cluster Centroids** displaced by a random distance and direction within the same administrative area. 
Cluster Anonymised Displaced Centroids (orange dots)	Similar to **Cluster Unanonymised Displaced Centroids**, but with an anonymised attribute table. This output from the Plugin can be exported as a CSV file as well. 

Cluster Displacement Links (black lines)	Line layer linking each **Cluster Centroid** with the relevant **Cluster Displaced Centroids**. 
Cluster Anonymised Buffers (grey buffers)	Buffers generated around **Cluster Displaced Centroids**. 
Cluster Original Buffers (green buffers)	Buffers generated around **Cluster Centroids**. 

The output data for public use includes the file named 'Cluster Anonymised Displaced Centroids,' containing the displaced coordinates of each cluster. Additionally, buffer layers are provided to allow for the extraction of geospatial covariates from global databases containing data on population, climate, environment and other topics that can be linked to MICS survey cluster locations. Additional details regarding geospatial covariates data extraction will be presented in a separate methodological paper.

5

Pre-Release Verification of GPS Data

Before releasing the GPS dataset, the cluster coordinates undergo additional verification to ensure:

1. Each cluster's GPS coordinates are checked to confirm they fall within the correct administrative unit, which is one level lower than the sample stratification.
2. The names of the subnational administrative units in the GPS dataset, Survey Boundaries, and MICS survey datasets are matched.

During the first step of verification, GPS coordinates are checked to ensure that even after displacement, they remain within the correct administrative unit. The survey boundaries used are typically provided or confirmed by the NSOs and do not imply any endorsement by the United Nations.

Accepted boundaries may not always perfectly align with the survey's sampling frame, but they serve as a reference for displacing cluster locations. Consequently, the following method is used in cases where a cluster's GPS coordinate is located in a neighbouring administrative unit:

- If close to the boundary of the correct administrative unit, manual displacement is performed to place the cluster in the proper administrative unit.
- For rural clusters, manual displacement is conducted if their GPS coordinate is less than 10 kilometres from the border of the correct administrative unit.
- Urban clusters are manually displaced if their GPS coordinate falls within 2 kilometres of the border of the correct administrative unit.

6

Public Dissemination

The MICS Programme began releasing GIS datasets in 2022. Georeferenced household surveys offer many benefits, allowing researchers to study the impact of respondent locations on the well-being of children, adolescents, women, and men. Additionally, the geocoded MICS data enables users to disaggregate MICS indicators by various units, such as population, environment, or climate topics, providing a more comprehensive understanding of the data.

However, the importance of making geographic data available to researchers must be balanced with the need to protect respondent's privacy. Therefore, as described above, the MICS Programme utilises a methodology for anonymising cluster geocodes. Moreover, the Programme has established a specific policy for releasing geographic data. Initially, the data are not freely available; users are required to submit a research scope and agree to the conditions of use before access is granted.

On behalf of the NSOs, the MICS Programme will review the research scope and grant access to the data only if researchers agree to:

- Use date only for the research described.

- Treat all microdata confidentially, making no effort to identify any household or individual respondent.

- Not redistribute data, either directly, within any tool/dashboard, or by other means.

- Inform and reference the NSO and UNICEF of any shared or published research.

7

Bibliography

(1) Multiple Indicator Cluster Survey. 2024. https://mics.unicef.org.

(2) European Commission, Eurostat, Applying the degree of urbanisation – A methodological manual to define cities, towns and rural areas for international comparisons – 2021 edition. 2021. *Publications Office of the European Union*. https://data.europa.eu/doi/10.2785/706535.

(3) Hundepool, Anco, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Eric Schulte Nordholt, Keith Spicer and Peter-Paul de Wolf. 2012. Statistical Disclosure Control. *John Wiley & Sons, Ltd.*

(4) Clara R. Burgert, Josh Colston, Thea Roy, and Blake Zachary. 2013. Geographic Displacement Procedure and Georeferenced Data Release Policy for the Demographic and Health Surveys. Calverton, Maryland, USA: *ICF International: DHS Spatial Analysis Reports No. 7.*

(5) Inter-Secretariat Working Group on Household Surveys. 2021. Spatial Anonymization, Guidance Note prepared for the Inter-Secretariat Working Group on Household Surveys.

For information on the report, please contact:

United Nations Children's Fund
Data and Analytics Section
Division of Data, Analytics, Planning and Monitoring
3 United Nations Plaza, New York, NY 10017, USA Tel: +1
(212)-326-7000
Email: mics@unicef.org