**Statistics Without Borders**

**Statement of Work Form**

**SECTION I. BASIC INFORMATION**

**SWB Project Number**: Project 171

**Client Name**: HelpAge International

**Project Name**: HelpAge Data Disaggregation

**SWB Project Contact**: Keri Wheatley

**Email**: wheatley.keri@gmail.com

**Backup SWB Project Contact**: Alexander Newton

**Email**: Alexander.Newton@ons.gov.uk

**SECTION II. PROJECT REQUIREMENTS**

**About the Client Organization.**

From website ([www.helpage.org](http://www.helpage.org)): “Our Global Network is a diverse group of 158 like-minded organisations in 86 countries, which supports millions of older people to live safe, dignified and healthy lives.”

**Objectives: please state the objectives of the project briefly and concisely, including expected outputs, purpose, and any other expectations.**

**Purpose:**

HelpAge International wants to challenge established norms for statistical reporting on older persons by proving that data disaggregation to a lower, more granular level is possible and statistically robust. Nongranular statistics reinforces oversimplified picture of inequalities, and the inadequate data itself becomes a barrier to the inclusion of at-risk and marginalized groups in policy and program responses.

**Project Objective:**

In this project, volunteers will perform a data disaggregation on 9 large household surveys (low-level data has been acquired) across 3 countries to determine the lowest level of disaggregation based on minimum reporting dimensions. Outputs for this project include the final indicators for each survey and commentary on the statistical methodology.

*Specific research questions:*

In relation to the specific survey type and different levels of analysis (individual and household):

i. What is the statistically robust and most granular level of disaggregation of indicators in intersection of sex, age and disability (i.e. minimum level of disaggregation)?

· What should the age bands (i.e. 5 year band, 10 year band, or 20 year band) and the upper age cohort (e.g. 80+, 85+, etc.) be?

· How significantly different are results across broad levels of disaggregation (e.g. both sexes aged 60+ with or without disability) and more granular disaggregation (e.g. women aged 60-69 with or without disability)?

ii. What is the statistically robust and most granular level of disaggregation of indicators in intersection of sex, age, disability (i.e. minimum level of disaggregation), and additional characteristics of geographic location and ethnicity (i.e. secondary level of disaggregation)?

iii. Based on the findings of the analysis and with respect to each survey type, what general recommendations or considerations can be made on data disaggregation for non-project countries with similar surveys?

**Project Detail:**

Each survey will be processed in the following way:

1. Understand the sampling design of the survey (sample size, variance, robustness, etc.)
2. Aggregate individual data points and re-summarize according to a target indicator
   1. Each survey will have two target indicators (an indicator, like spending, may have multiple associated metrics like spending on food, spending on housing, and spending on transportation). Please, see Section III for the list of survey-specific indicators.
3. Produce the minimum level of summarization for two different levels of reporting:
   1. Individual level: Binned age, gender, and disability (ex. women ages 60-64 with disability). See section III for information on minimum level of disaggregation (Note: some surveys do not collect data on disability)
   2. Household level (see Definitions section)
4. Include a secondary reporting dimension to the disaggregation (see Definitions section)
5. Only add one secondary reporting dimension for each analysis
6. Secondary reporting dimensions have been derived in most cases
7. Provide summarization of results in report format (PDF) and include statistical methodologies used
8. Some surveys may not have data to support a secondary reporting dimension

**Outputs:**

1. Specify the final indicators found with each survey and supporting information, such as margin of error, sampling size, etc.
2. Provide the client with the actual code (with comments) and stats.

**Additional Considerations:**

* + 1. If possible, determine a common statistical approach for each survey type: DHS, LFS, HEIS, and Census. (Note: harmonisation of data and indicators either across surveys or across countries is not required)
    2. During the project, consider any patterns of disaggregation that may appear between surveys or across countries in relation to the same survey type. As a nice to have, client would like a summary of any common disaggregation levels found.
    3. As a nice to have, provide commentary or recommendations for any future data producer, such as the Census Bureau, on methodology, appearance of biases, understanding limitations etc.

**Background: please include the general background of the project.**

The vision of the Agenda 2030 is the world without poverty, inequality, injustice, and environmental degradation. Central to this vision is the promise of ‘leaving no one behind’, realized through the inclusion of ‘all, at all ages’ and ‘reaching the furthest behind first’.

However, monitoring the SDG progress in relation to at-risk and marginalized groups is challenging as these groups are practically invisible in statistics. Partially, this is due to inadequate data disaggregation, for example focus on siloed and binary categories like ‘woman/man’, ‘disabled/able’, or ‘young/old’, etc. This fails to consider individual’s multiple and intersecting identities, and how these identities interplay with norms and structures to create a complex web of disadvantages.

Nongranular statistics reinforces oversimplified picture of inequalities, and the inadequate data itself becomes a barrier to the inclusion of at-risk and marginalized groups in policy and program responses.

Initiatives like [IAEG-SDG working group on data disaggregation](https://unstats.un.org/sdgs/iaeg-sdgs/disaggregation/) and the [Titchfield Group on ageing and age-disaggregated data](https://unstats.un.org/unsd/methodology/citygroups/Titchfield.cshtml) are working to address these issues. The following project aims to support these initiatives as well as to influence dissemination of data on marginalized groups as part of the census 2020 round.

**Scope: please list what is in scope for the project and what is out of scope.**

In Scope

* We will analyze the surveys listed in Data section

Out of Scope

* We will not be combining any surveys results
* We will not look into additional countries and surveys outside of SOW
* We will not be adding any new data sources
* We will not enforce a common methodology across surveys
* We will not enforce a common level of disaggregation across surveys

**Additional Requirements: please list any specific skills required. such as language, software,** **specific methodologies, etc.**

**Language:** English

**Software/Tools:** Tools for data formats: STATA, SAS, SPSS, ASCII, RAR

Feature engineering tools, such as Python

Some statistical analysis required, but software-agnostic

**Methodology:** Methods to disaggregate survey data and predict population densities for finer scale

**Other:** Report-writing ability, good scientific and technical communication

Check all applicable, if any:

☐ This project requires that volunteer travel to the project site.

☐ The client agrees to fund or reimburse travel expenses. Please describe specifics and any limits: None

**SECTION III. PROJECT PARAMETERS**

**Definitions: please list exact, repeatable definitions including the population of interest, the target event, and any other necessary definitions, as applicable.**

**Lowest desired (minimum) reliability for analysis**

* SWB will need to make recommendations for measure of accuracy required for reports

**Survey-specific indicators to be analyzed**

|  |  |  |
| --- | --- | --- |
| **Country** | **Type** | **Indicator** |
| Tanzania | Census | **i. Household level:** Proportion of households using improved drinking water (variable [TZ2012A\_WATSUPIMP](https://international.ipums.org/international-action/variables/TZ2012A_0140#codes_section))  **ii. Individual level:** Employment status of older people during the week before the census (variable [TZ2012A\_EMPSTAT](https://international.ipums.org/international-action/variables/TZ2012A_0435#codes_section)) |
| DHS | **i. Household level:** Proportion of population using safely managed drinking water services ([Technical note](https://unstats.un.org/sdgs/metadata/files/Metadata-06-01-01.pdf) for indicator)  **ii. Individual level:** Educational attainment of older people by the highest level of school attended (as per country specific levels identified in the questionnaire) |
| LFS | **i. Household level:** Proportion of the population with large household expenditure on health (survey section 6 Non-food expenditure, Q3 (code 2022-2024), Q4 (the survey indicator might not fully match the SDG methodology described in the [Technical note](https://unstats.un.org/sdgs/metadata/files/Metadata-03-08-02.pdf). However the note might still be useful)  **ii. Individual level:** Average hourly earnings of employees by sex, age, occupation and person with disabilities ([Technical note](https://unstats.un.org/sdgs/metadata/files/Metadata-08-05-01.pdf) for indicator) (occupation classification NCO-2004 available [here](http://microdata.gov.in/nada43/index.php/catalog/127/variable/V330)) |
| HIES | **i. Household level:** Proportion of the population with large household expenditure on health ([Technical note](https://unstats.un.org/sdgs/metadata/files/Metadata-03-08-02.pdf) for the indicator)  **ii. Individual level:** Average hourly earnings of employees by sex, age, occupation and person with disabilities ([Technical note](https://unstats.un.org/sdgs/metadata/files/Metadata-08-05-01.pdf) for indicator) (occupation variable in the survey [hh\_b11](https://microdata.worldbank.org/index.php/catalog/3455/data-dictionary/F189)) |
| India | DHS | **i. Household level:** Proportion of population using safely managed drinking water services ([Technical note](https://unstats.un.org/sdgs/metadata/files/Metadata-06-01-01.pdf) for the indicator)  **ii. Individual level:** Educational attainment of older people by the highest level of school attended (as per country specific levels identified in the questionnaire) |
| LFS | **i. Household level:** n/a  **ii. Individual level:** Average hourly earnings of employees by sex, age, occupation and person with disabilities ([Technical note](https://unstats.un.org/sdgs/metadata/files/Metadata-08-05-01.pdf) for indicator) (occupation variable in the survey [s42q9a](https://microdata.worldbank.org/index.php/catalog/3549/data-dictionary/F23)) |
| HIES | **i. Household level:** Household expenditure on health (absolute value as total household expenditure is not available). (Relevant ([Technical note](https://unstats.un.org/sdgs/metadata/files/Metadata-03-08-02.pdf))  **ii. Individual level:** Older person’s state of economic independence  (Variable [b10\_q5](http://microdata.gov.in/nada43/index.php/catalog/135/datafile/F9/V351)) |
| Ukraine | Census | **i. Household level:** Proportion of population by type of premises (variable [UA2001A\_TYPEHH](https://international.ipums.org/international-action/variables/UA2001A_0027#codes_section))  **ii. Individual level:** Individual’s sources of livelihood, by type (variable [UA2001A\_INCSRC](https://international.ipums.org/international-action/variables/UA2001A_0413#codes_section)) |
| DHS | **i. Household level:** Proportion of population using safely managed drinking water services ([Technical note](https://unstats.un.org/sdgs/metadata/files/Metadata-06-01-01.pdf) for the indicator)  **ii. Individual level:** Educational attainment of older people by the highest level of school attended (as per country specific levels identified in the questionnaire) |

**Minimum Level of Disaggregation**

Type 1: Individual Level

|  |  |  |
| --- | --- | --- |
| **Dimension** | **Variable Type** | **Description** |
| Age Cohort | Ordinal | Three binning methods:   1. Older age in 5-year cohorts   (Under 60, 60-64, 65-69, 70-74, 75-79, 80-84 etc.   1. Older age in 10-year cohorts   (Under 60, 60-69, 70-79, 80-89, etc.)   1. Broad age cohorts: 60+, 70+, 80+   Current standard is to use 10-year cohorts. HelpAge would like to challenge the idea and see if we can get more granular statistics. Final binning capabilities will depend on the minimum level of confidence required by client. |
| Sex | Categorical | Refers to male, female and ‘both’ categories, as given by survey. |
| Disability | Categorical | Not all surveys identified within the project collected data on disability. Those that do, vary in their approaches to measuring disability, ranging from a ‘Yes/No’ question to a set of questions like the Washington Group Short Set on Functioning (WG-SS). For the purpose of this exercise any definition of disability is accepted.  If a data tool uses [Washington Group Short Set on Functioning](https://www.washingtongroup-disability.com/question-sets/wg-short-set-on-functioning-wg-ss/) to measure disability please, use relevant technical guidance to analyze disability data (available [here](https://www.washingtongroup-disability.com/analysis/wg-short-set-on-functioning-wg-ss-syntax/)):   * the suggested threshold for defining disability is: any 1 domain/question is coded ‘A LOT OF DIFFICULTY’ or ‘CANNOT DO AT ALL’ |

Type 2: Household Level

|  |  |  |
| --- | --- | --- |
| **Dimension** | **Variable Type** | **Description** |
| Household Level | Categorical | These are the specified variables that will need to be derived from the survey. If not possible with a survey, skip this analysis.   1. SNGL: One-person household 2. PRTN: Two-person household (spouses/partners only) 3. ADLT: Household with an older person(s) and persons aged 20 and over 4. MINR: Household with an older person(s) and person aged 19 or younger 5. MLTG: Household with an older person(s) and other generations 6. OTHR: Any other type of living arrangement (e.g. institutional setting, informal settlement, homeless, etc.) |

**Secondary Level of Disaggregation**

*To be combined with the minimum level of disaggregation;* *Add only 1 secondary level of disaggregation to reporting dimensions.*

*Example: Use reporting dimensions female, age 60-64, with disability, in rural area*

|  |  |  |
| --- | --- | --- |
| **Dimension** | **Variable Type** | **Description** |
| Urban/Rural  (first priority) | Categorical | Predetermined data for: DHS, Ukr. census, Ind. LFS, Ind. HIES, Tnz HIES; do not derive  Variable is not predetermined for: Tanzania LFS (if not possible to derive urban/rural disaggregation, please, disaggregation by predefined administrative regions) |
| Ethnicity (secondary priority) | Categorical | Predetermined data for some surveys; do not derive |

**Data: please list the sources, data types, volume, and general description of the contents for each data set.**

|  |  |  |
| --- | --- | --- |
| **Country** | **Survey** | **Format** |
| Tanzania | Demographic and Health Survey (DHS) | STATA, SAS, SPSS, ASCII |
| Tanzania | Labor Force Survey (LFS) | CSV, SPSS, STATA |
| Tanzania | Household Integrated Economic Survey (HIES) | CSV, SPSS, STATA |
| Tanzania | Census | DAT |
| India | Demographic and Health Survey (DHS) | STATA, ASCII |
| India | Labor Force Survey (LFS) | Text, RAR |
| India | Household Integrated Economic Survey (HIES) | SAS, SPSS, ASCII |
| Ukraine | Demographic and Health Survey (DHS) | STATA, SAS, SPSS, ASCII |
| Ukraine | Census | DAT |

*Definition of target indicator - The target indicator is the set of metrics about each reporting dimension. Average, maximum, minimum spending in (housing, transportation, medical) for women aged 60-70 with disability living in rural area.*

**Analysis Periods: please list the dates of the data collected. If the project involves predictive modeling, please specify the prediction windows.**

N/A

**Assumptions and Limitations: please list any major assumptions and limitations (statistical and otherwise) that impact the successful completion of the project.**

Statistical analysis in some circumstances will change based on subsequent analysis needs for the data. Currently, the HelpAge has not specified any subsequent analysis they would like to perform with project results.

**Dependencies: please list all project dependencies.**

1. Timely response by HelpAge to questions and requests for information and feedback by SWB, and conversely from SWB, is required for the timeliness and success of this project.

**Project Risks: please list all known project risks not covered above.**

None known at this time.

**Analysis Methodology and Approach: please list the project’s general methodology or approach (e.g. regression, DOE, etc.).**

Small area estimation, data disaggregation

**Software & Tools: Please list any specific SWs/tools required or preferred (e.g. SAS, R, SPSS, Tableau Public, etc.).**

STATA, SAS, SPSS, ASCII, RAR, Python, R

**Follow-Up Analysis and Tasks: please state any specific analyses and tasks in the scope of the project. If none, please state that no follow-up tasks are in scope.**

No follow-up tasks are in scope after the review and recommendations are delivered and finalized (after sign-off by HelpAge International).

**Deliverables: please list specific deliverables, including the format of the output.**

The following will be delivered to HelpAge International:

1. The actual code (with comments) and stats, including a Readme file with access instructions. Format in coding language used for project.
2. A written report outlining the steps undertaken to disaggregate various data source and any considerations and recommendations from project. PDF format.
3. A virtual meeting between the review group and HelpAge International to present the report and discuss its findings and recommendations.

**SECTION IV. PROJECT LOGISTICS**

**Roles and Responsibilities: please list the SWB point of contact and the client point of contact for the project. Please list any other roles and responsibilities for the project.**

**SWB point of contact:** Keri Wheatley

**Client point of contact:** Alex Mihnovits

**Other roles and responsibilities:** Backup PCM, Alexander Newton; Statistical Consultant, Richard Griffin; DQA, Nathaniel Martin

**Timeline and workload:**

**Expected project duration (in weeks):** 7 weeks

**Expected team size (#volunteers):** 10 volunteers excluding PCMs

**Expected workload (hours per week):** 6-8 hours/week, including 1 live call/week

**Hard Project End Date, if any (see dependencies and assumptions):** March 7, 2021

**Please include any details or specifics for the project plan. The more specific the project plan, the better.**

Project team will work in five parallel workstreams. Each workstream will consist of one Statistical Expert and one Data Engineer. The first phase of the project will involve the assignment and analysis of 5 surveys (priority determined by HelpAge). The second phase of the project will involve taking the learnings from the first phase analyses and applying the same methodology to the remaining surveys.

**Expected frequency of progress review:** Weekly

**SECTION V. MISCELLANEOUS**

**Other Comments: please include any additional comments**

N/A

**SECTION VI. ENGAGEMENT APPROACH AND CLIENT RESPONSIBILITIES**

Statistics Without Borders (SWB) provides clients access to volunteers who have been deemed qualified to participate in a project as requested, and provides management of its volunteers under its generally established standards. However, SWB expressly disclaims any and all warranties, whether express or implied, of the project outputs. Neither the ASA nor the SWB will sponsor projects directly, sanction, warrant, certify, or otherwise guarantee the work of its volunteers, project deliverables, or their products or derivatives. All services and deliverables are provided as is by the volunteers, and the client is responsible for reviewing the deliverables to ensure their accuracy and completeness and for the results obtained from its use of the deliverables. The volunteers have acknowledged to and are expected to perform project work to the best of his/her ability in a professional and ethical manner.

Furthermore, Client agrees to hold harmless, defend and indemnify Statistics Without Borders and the American Statistical Association and their respective members, directors, executive committees, officers, employees, agents, advisors, and representatives, and their successors and assigns (each, an “Indemnitee”), from and against any claim, demand, cause of action, complaint, suit, proceeding, judgment, award, loss, damage, or other liability, including reasonable attorneys’ fees, imposed upon or incurred by such Indemnitee by reason of or resulting from or in connection with this Statement of Work including, but not limited to, any work or other actions (or omissions) that are undertaken or occur with respect to this Statement of Work.

In addition,

* Client agrees to the Client Code of Conduct (<https://higherlogicdownload.s3.amazonaws.com/AMSTAT/ea9d01bf-dbfe-46f6-a681-fd7890cf1357/UploadedImages/About_SWB/SWB_COC_clients.pdf>).
* Client will notify SWB prior publishing and attributing the results or derivatives of this project to SWB.
* Client will not alter or otherwise misrepresent any results, findings, and/or recommendations delivered to the client by the SWB project team, publicly or privately.
* Client agrees that a lack of timely response to SWB request for input may result in a termination of the project at the discretion of SWB.
* The client agrees that post-delivery review and feedback will be limited to a maximum of three (3) iterations, and any further iterations will be considered a new project.

If any rework is required beyond the control of SWB, SWB reserves the right to close the project and initiate a separate project.

**SECTION VII. AGREEMENTS**

Check all applicable, if any:

☐ The client consents that the project results will be made public following the completion of the project. This includes public presentation and publication.

☐ The data contains sensitive information. In order to protect the privacy of the data subjects, each volunteer with the access to the data will be required to sign the Data Security and Privacy Agreement. This may include future volunteers yet to be identified but who need to access the data after the project completion to fulfill tasks related to SWB’s operations; in this case, SWB will make a good faith effort to contact Client and will keep a copy of the agreement signed by the volunteer in case the contact cannot be established.

☐ The client will be sharing information proprietary and confidential to the client, other than or in addition to the data. In order to protect the client secrets, each volunteer will be required to sign the Confidentiality Agreement. This may include future volunteers yet to be identified but who need to access the data after the project completion to fulfill tasks related to SWB’s operations; in this case, SWB will make a good faith effort to contact Client and will keep a copy of the agreement signed by the volunteer in case the contact cannot be established.