Terms of Reference

Microdata Curation Consultant

Statistics Unit, Field Information and Coordination Support Section (FICSS), UNHCR Copenhagen

UNHCR, the UN Refugee Agency, is seeking a Consultant to deliver on its open data vision of publicly available survey and disaggregated data and statistics through the implementation of the UNHCR microdata library, including data identification, cleaning, curation, anonymization, benefit/risk assessment of sharing data, and completion of appropriate metadata according to agreed standards.

UNHCR is a global organization dedicated to saving lives, protecting rights and building a better future for refugees, forcibly displaced communities and stateless people. Every year, millions of men, women and children are forced to flee their homes to escape conflict and persecution. We are in over 125 countries, using our expertise to protect and care for millions.

Title: Microdata Curation Consultant
Duty Station: Copenhagen, Denmark (possibility home-based)
Duration: 1st January 2019 through 30th June 2019
Contract Type: International Consultant
Closing date: 9th December 2018

Background and Organizational context

UNHCR's **Field Information and Coordination Support Section (FICSS)** is the focal point in UNHCR for the collection, analysis and dissemination of operational data and statistics, and for technical support, including training, on related methodologies and activities. The effective management and analysis of operational data is a prerequisite for successful interventions that have an impact on the well-being of UNHCR's beneficiaries. It is the goal of FICSS to support all UNHCR offices, at Headquarters and in the Field, to better manage operational data.

FICSS routinely supports a number of household survey data level collections each year, conducted directly by UNHCR or implemented by partners. Traditionally this data has not been retained in formats and locations that make it available for future use, either in country, by UNHCR globally, independent researchers, or by partners such as the soon-to-be-established UNHCR-WB data center.

Consistently with the drive towards open data, UNHCR is, therefore, in the process of creating an internal data library (RIDL) and the Microdata library, an online open repository which will allow public and/or licensed access to anonymized data collected in these surveys, and users to perform directly their own analysis.

Microdata are data that hold information collected on individual respondents, such as people or households. Such data usually come from surveys or censuses and, to a lesser extent, administrative registers. The data may be available for all or a subset of variables covered by the data collection exercise. Making anonymized microdata publicly available ensures that maximum use is made of the data, improves data quality and credibility and may prevent duplication of surveys and thereby reduce the costs for humanitarian and development organizations that are in need of data to inform their programming.

In most cases, data cleaning, validation and missing data SOPs or guidance need to be established and applied to datasets where appropriate. Proper and secure microdata dissemination requires the establishment of policies and procedures that formally define the conditions for accessing microdata, and to apply anonymization methods to data before release. Datasets must therefore be properly anonymized before they can be shared. At the moment, at UNHCR there are no existing SOPs or guidance for the anonymization of publicly shared microdata except for a draft guidance for refugee registration data, and the capacity to do this at the field level is limited.

Duties and responsibilities

Under the overall supervision of the Senior Statistician, the Consultant will be responsible to assist with the data cleaning, missing data management and anonymization of household surveys and begin populating microdata library with datasets and corresponding metadata. It will more specifically perform the following activities:

- Liaise with prioritized country operations to identify and obtain historical datasets suitable for sharing.
- Work with country offices to have data stored within RIDL according to data security and data protection protocols.
- Work with field offices to ensure relevant metadata is provided along with the data.
- Develop instructions and establish procedures to evaluate the quality of datasets and assess suitability for publication, removing and correcting errors if necessary, and work with field locations as relevant.
- Apply statistical disclosure control (SDC) techniques used to anonymize all datasets ensuring that no person or household is identifiable from the results of an analysis of survey data (at an appropriate level of risk), or in the release of microdata, utilizing the R-based sdcMicro package. This will be done in consultation with the data producer that will guide in the determination of key and sensitive variables. Such anonymization will be performed both at the household level and individual level, and applying different methods for public use and licensed use files. Perform statistical analysis on pre vs post anonymization utility variables to verify anonymized data maintains validity and should be recommended for sharing.
- Based on the work performed on available datasets, assist with the development of SOPs to remove personally identifiable data and prevent re-identification of anonymized data, including guidance for data owners on the public release of the anonymized version of the dataset, striking an acceptable balance between identification risk and information loss.
- Generate the appropriate metadata according to the Data Documentation Initiative (DDI) metadata standard and any other documentation needed to catalogue, record and communicate to users all the salient characteristics of the micro-datasets shared on the MicroData Library.

Monitoring and Progress Controls

- Delivery Periodicity: first payment in February 2019 upon delivery of interim report containing draft data management and anonymization SOPs and details on substantial progress towards population of RIDL and microdata library with suitable datasets. Second payment upon delivery of second report detailing progress towards achievement of final goals. Final payment, upon achievement by June 2019 of the following goals:
  - RIDL to be populated at a minimum with all raw data used to populate Microdata library.

- o Microdata library to be populated with at least 10 clean, anonymized datasets by end of consultancy. Each dataset will ideally be from a different country. Should that not be feasible, more than one dataset from the same country may be acceptable as long as it refers to a separate survey or dataset relative to a different refugee population and a different time period.
- o Drawing from the experience gained in populating the data library, establishment of SOPs (including a template and samples for guidance to field staff) regarding data management workflow for data from production to inclusion in RIDL, cleaning and anonymization, to public release.

Essential minimum qualifications and professional experience required

- Advanced University Degree (Master's or PhD equivalent) in Statistics, Economics, or related field.
- At least 5 years of relevant work experience,
- Previous demonstrated experience manipulating large household datasets;
- Previous demonstrated experience with social research methods, including data and collection, anonymization, analysis.
- Proficiency in R (preferred) or any other equivalent statistical package;
- Good self-management skills to work both independently and as part of a team, including remotely.
- Excellent communication skills, including the ability to liaise efficiently with operations and communicate and disseminate information effectively to a broad range of stakeholders.
- Work experience in international environment and in refugee/migration substantive issues is an asset.
- Excellent writing skills.
- Excellent knowledge of English and working knowledge of another UN language.

Location and conditions

The successful candidate will be based with the UNHCR team in Copenhagen, Denmark. The possibility of remote work for candidates located outside the duty station will be considered.

The salary is in line with the UNHCR 'B' Level Consultancy fee.

To apply

Interested applicants should submit their letter of motivation, Personal History Form (P11), and CV to dencocmuhr@unhcr.org indicating *"Microdata Curation Consultant"* in the subject of the email.

P11 forms are available at www.unhcr.org/recruit/p11new.doc

Closing date for receipt of applications: Sunday 9th December 2018

Female candidates are encouraged to apply