Terms of Reference

Analysis of automating data collection elements for GAM 2022-2025

- using machine learning for NLP and text analytics for GAM

1. Background and objective

In May 2016, the United Nations General Assembly High-Level Meeting on Ending AIDS mandated UN Joint programme on HIV and AIDS (UNAIDS) to collect HIV and AIDS data and annually report back on progress. The reporting covers all the 195 UN member states, using the existing about 75 key indicators. While the declaration of commitments from 2016 expired in 2020, a new declaration is presented in June 2021 High-Level Meeting on HIV. The new more specific and numerous targets for 2025 have been set, and for their monitoring UNAIDS needs to set up an efficient monitoring framework, minimizing (manual) reporting burden, and maximizing analytical value.

Currently, the Global AIDS Monitoring (GAM) data collection is done through an online reporting platform, and the data gets published in the annual Global AIDS Report and on AIDSinfo data visualization platform. Much of this work includes manual processing: data entry (country rapporteurs), data management and organization, validation and presentation/visualization (strategic information department, SID).

While the new UNAIDS strategy and targets imply a far more detailed and broader data framework, it is anticipated that the burden on collecting, validating and analyzing the data is likely to increase both for country rapporteurs and UNAIDS. It is necessary to explore new ways of automating the data collation from publicly accessible sources and platforms, reducing the reporting burden on countries. Furthermore, with the increase in data, iterative yet fast cross-analysis and learning are required to identify issues such as inequities in services and affected communities, as well the role of societal enablers affecting peoples access to the service they need.

Many countries publish more frequent and more granular (sub-national) data or other observations on their formal and civil society web-sites, but do not necessarily share such with UNAIDS as it means transforming the data to another format and entering it manually into the GAM or other platforms.

It would be more efficient to harvest the data from the publicly accessible web-sites (HIS, surveys, civil society reports, research web-sites etc.), which would remove the manual labour and reduce chances for error in data entry. The analysis needs increase exponentially, and for that machine learning methods can serve as a starting point to identify the opportunities and to explore areas where such data and analysis can provide to UNAIDS and its monitoring framework(s) the best value.

2. Scope of work

The contractor will deliver an analysis of Global AIDS Monitoring framework and opportunities to collate and analyse data with machine learning solution for NLP and text analytics, and produce a proof of concept with prioritized data sets.

3. Main activities

A. Document the identified data sources used for GAM framework (2 days)
B. Organize consultation sessions with SID, select the NLP and text analytics applicable sources, and directly sourced (via API) data sets (3 day)
C. Analyze, test and document the options for each data source (18 days)

D. Present and discuss the technical options and findings with SID (1 day)
E. Revise and finetune testing, and document final options/recommendations (3 days)
F. Produce final recommendations and exchange with SID on findings (3 days). The final report needs to include:

    i. mapping of alternative methods;
    ii. recommending one methodology, or a combination;
    iii. Assessment of which indicators (or elements) are available from public websites.
    iv. Prepare analysis on a combination of using a direct pull from public websites, and a data entry platform.
    v. Recommendations to implement a direct pull approach for the data which is publicly accessible and purposeful for extraction.

4. Timing

The work will commence on signature of the contract. All activities and reports must be completed at latest by the end of June 2021.

5. Qualifications and Professional Experience Required:

- Data scientist / Senior data scientist: Advanced university degree (Master's degree or equivalent) in data science, mathematics, statistics, engineering or any related field.
- Ability to develop and deliver data science solutions to better enable the planning, decision making, and implementation of organization programmes.
- Knowledge to design products that provide situational awareness, insight, and predictive assessment to guide programme management and execution.
- Experience in identifying significant issues and opportunities to implement innovative approaches and evidence-based reasoning to complex organizational issues.
- Interpersonal skills and ability to work in a team, forging relationships and producing results in multi-cultural and multi-disciplinary settings within a tight deadline.
- Command of English required.

6. Level of effort

The working days are estimated to be allocated as per above outlined activities. Working time estimate: 30 days over a period of 3 months (April – June 2021).

7. Payment schedule

1st payment (25%)	on return of countersigned contract;
2nd payment (60%)	upon receipt and acceptance of a concept paper report in May 2021;
Balance (15%)	upon receipt and acceptance of final report in June 2021.