

# UNDP Innovation Facility | 2018 Call for Proposals

**Q1: Title of the Initiative:** Big Data for Development – Scaling Out, Up and Deep for Development Impact in Sudan

**Q2: Contact Person for the Initiative (UNDP E-Mail address):** anisha.thapa@undp.org

**Q3: Is this initiative part of an existing project/programme?**

Nation-wide scaling up of the big data for MPI experimentation previously funded by UNDP Innovation Facility 2017

**Q4: This initiative will take place in:** A Country Office

## SECTION | THE COUNTRY OFFICE INNOVATION INITIATIVE

**Q5: Please select a country or territory where the initiative will take place**

Sudan (the)

## SECTION | THE INNOVATION INITIATIVE

**Q10: What is the current stage of the initiative?**

Transition to Scale: the initiative has demonstrated small scale successes; it has developed a model and attracts partners to help fill gaps in their capacity to scale.

**Q11: Requested amount in US\$ (including GMS at 8%):** US (\$) 95,040

**Q12: Will further/other funding be available?** Yes

**Q13: If cost-sharing is available for your initiative, please share:**

- How much (in US\$)? 58,000
- From whom? UNDP Sudan, Telecom and Postal Regulatory Authority

**Q14: What is the expected impact?**

- More targeted, in reaching the identified group
- Changes/enhances programming (at any stage) through newly gained insights

**Q15: SUSTAINABLE DEVELOPMENT GOALS | To which SDG is the innovative initiative contributing the most?**

Select up to 3.

- Goal 1: No Poverty
- Goal 5: Gender Equality
- Goal 17: Partnerships for the Goals

**Q16: INNOVATION METHODOLOGIES: What innovation method will you be using?**

- Big Data (including new and emerging data)

**Q17: Please choose the funding window you are applying to:** Development Impact | Scaling

# UNDP Innovation Facility | 2018 Call for Proposals

## FORM B | Transitioning to Scale Stage or Scaling Stage

### **Q28: What problem does your innovation respond to? Please share evidence and/or data to support the problem statement.**

In Sudan, timely and frequent development planning and progress monitoring suffers due to political sensitivity and lack of data. Conflict and instability in many areas of Sudan make on-the-ground data collection difficult. Low domestic funding for statistics adds to the data shortage as traditional data collection methods are expensive. Until 2017, Sudan's development programme design and targeting the most vulnerable (for e.g. the poorest) was based on figures from National Baseline Household Survey 2009 [1]. Just in 2018, updated poverty level figures from the 2014 Household Budget and Poverty Survey were published. This challenge will continue to persist. The last census was in 2009, and the next census is still put on hold, and most probably will take place after the presidential election in 2020.

References: [1] Sudan Interim Poverty Reduction Strategy Paper, 2013  
<https://www.imf.org/external/pubs/ft/scr/2013/cr13318.pdf>

### **Q29: Describe your current solution to address the problem?**

In a quest for innovative and efficient data generation, which can inform development policy directions, UNDP Sudan embarked on testing variety of big data sources such as household electricity consumption, night time lights (in 2015) and mobile phone use data (in 2017) to calculate proxy of poverty levels in Sudan at scales deeper than state levels. The 2015 test results showed weak predictive capacity of the selected big data source which led to the choice of another in 2017 – mobile phone use. With support again from UNDP Innovation Facility, UNDP Sudan in collaboration with the Telecom Postal Regulatory Authority (TPRA), Central Bureau of Statistics (CBS), Zain Telecom (biggest telecom operator) and Berlin University piloted the testing of the ability of mobile phone use (Call Detail Records – CDR) to predict proxy poverty levels by using MPI composite index for the state of Al Gezira in Sudan. The results are highly promising [2] with strong correlation and high explanatory power (Adj R sq. > 0.7) between multi-dimensional poverty (MPI) index of Al Gezira localities [3] – control variable (measured by 2017 CBS Household MPI Survey) and the covariates from CDRs at locality level – test variable (analyzed by Berlin University and Zain Telecom) with ability to predict poverty proxies at antennae level which can be corresponded to sub administrative unit level from locality.

However, there are limitations to this proof of concept that become drivers to our scale up proposal as below:

#### i. Coverage:

Al Gezira is a relatively stable state out of 18 states in Sudan. Sudan's complexity lies in its size and the variety of challenges it faces in its regions – Darfur in the west, East Sudan and Southern region [Map1] trickling down to its states. Hence this proof of concept although highly representative and effective at state and lower levels, the approach should be refined enough to predict poverty proxies for other regions. Hence scaling up horizontally to wider space covering more vulnerable people is imperative for a more representative data at national level.

#### ii. Telecom Partners:

Given Zain Telecom's dominant coverage in Al Gezira, only Zain provided their data and processing services pro bono as a partnership for this initiative under their agreement with TPRA on BDI4SD [4]. Other major telecom operators - MTN and Sudatel were not included in this initiative. However their network coverage in other parts of the country is more. Hence, accuracy of proxy poverty can be improved more if other telecoms' CDR datasets are included in the analysis.

# UNDP Innovation Facility | 2018 Call for Proposals

## iii. Data disaggregation:

Disaggregation of the proxy values space wise was well demonstrated in this pilot. However, given the priorities for advancing gender equality and youth under UNDP CPD, the pilot analysis did not explore gender or youth disaggregated poverty proxies. Hence scaling deeper disaggregation in terms of other dimension – sex, and age is equally critical to address the priorities and make meaningful contribution to SDG monitoring.

## References:

[2] [Technical Report \(May 2018\)](#) – Monitoring multidimensional poverty using mobile phone metadata: A Feasibility Study from Al Gezira, Sudan (Initiative Blog overdue because of delay in finalization of Non-Disclosure Agreement)

[3] [Gezira State Pilot Multidimensional Poverty Survey \(CBS, 2017\)](#)

[\[Map1\]](#)

[4] Big Data Initiative for Sustainable Development Agreement (TPRA- MNOs, 2017)

## **Q30: What is your vision of scale?**

This proof of concept established a high-correlation between the big data from cellphone usage and antenna location and MPI composite index at locality level. The big data will give us a fairly good overview of spatial disparity – inequality across different localities based on MPI composite index. This would be ground-breaking to measure spatial inequalities at locality level, contributing to well-informed debates for public expenditure planning, donor coordination and local governance. It will be also contributing to the achievement of SDG 10 on addressing inequalities. It is expected that the primary users of the big data would be the government, donor communities and civil society, well beyond the United Nations. Furthermore, for nation-wide scale up, UNDP will ensure gender disaggregated data be available, so that it will feed into SDG 1 and SDG 5 indicators. At the same time, in advancing big data for MPI, UNDP will make clear of limitations of big data to maintain integrity of our development policy advisory support. The genesis of MPI is to present a degree of “deprivation” along with MPI indicators ranging from health, education, child mortality, access to electricity/IT technology, to water. The big data does not inform at the level of sectoral needs as the score is at aggregated level MPI, perhaps except the indicator on access to electricity/IT technology. Eventually, better monitoring of impact indicators for two priority goals – SDG 1 and SDG 5 leads to better understanding of needs allowing better, agile, inclusive programmatic response and ultimately equitable reduction in poverty.

## **Q31: What is the evidence of your innovation’s initial success and scalability potential? Share data that demonstrates the initial success of your solution; including links to studies, articles or relevant information.**

The success of this pilot’s initial success can be highlighted as below:

i. Multi-stakeholder Partnerships established in a politically sensitive setting: First and foremost, conducting a successful pilot using telecom sector big data in a post conflict setting like Sudan where the leadership faces restrictions imposed by the ICC resulting in its national establishments and stakeholders to be extremely sensitive to any sort of data collection. The partnership secured with TPRA, CBS and eventually the involvement of private sector telecom operator – Zain Telecom is a big success indicator ensuring sustainability of the initiative. With the MoU between UNDP Sudan and TPRA through this first initiative, the scope of involvement of two other leading telcos – MTN and Sudatel on a pro bono basis reducing efforts on negotiation and administration. As intended this initiative has created an appetite amongst key stakeholders to foster ecosystem for exploration of innovative data towards measurements of SDG status and progress.

# UNDP Innovation Facility | 2018 Call for Proposals

ii. Highly promising results on predictive capacity of telecom sector big data [6]: The pilot proved that the CDR can successfully predict proxy poverty values [Fig. 1] in Sudanese context going further down to telecom antennae level resulting proxy poverty values at higher geographical resolution [Fig. 2], therefore opening the scope for precisely targeted evidence-based development interventions, adjustments and progress tracking within Al Gezira state alone with full potential to replicate in other states, at a regional and national level.

iii. Low cost and efficient solution: For lower income country like Sudan and mainly for relatively under prioritized portfolio such as statistics by national budget, this solution offers high promise for institutional take up as a complementary low-cost alternative to traditional household level surveys and censuses. Metadata is already being collected by the mobile phone operators on an ongoing basis. Consequently, using this data for statistical purposes does not bear the costs of in-field data collection. The high frequency of the data collection further offers the possibility of ongoing future observation/monitoring and evaluation.

## References

[6] ] [Technical Report \(May 2018\)](#) – Monitoring multidimensional poverty using mobile phone metadata: A Feasibility Study from Al Gezira, Sudan (Initiative Blog overdue because of delay in finalization of Non-Disclosure Agreement)

[Fig. 1] [MPI Proxies and MPI Survey values at locality level](#) [Fig. 2] [MPI Proxies at antennae level](#)

**Q32: Describe the primary scaling pathway(s) and the enablers for moving your project forward through these. What are the most effective scaling pathways (channels, institutions, partnerships, and approaches) that will enable you to increase the reach of your innovation?**

As highlighted under limitations described in Section 2, the scaling pathways will be mainly along three dimensions.

### Scaling Out:

It has been informally agreed with the government partners to expand the geographic sample coverage of our proof of concept. As a pilot, the method will be scaled out at regional level including at least three more states which will include vulnerable populations in protracted crisis such as IDPs, refugees and returnees. This coverage is at least 25% of the total Sudanese population.

### Scaling Up:

With TPRA and CBS already leading this initiative from government side, more national institutions from the academia – University of Khartoum and the national SDGs monitoring entity – National Population Council will be brought on board. From private sector side, through the BDI4SD agreement with TPRA, more network operators – MTN and Sudatel have already shown strong interest to come on board. Forwarding the efforts due from our 2017 initiative, in collaboration with Sudan CBS, we will initiate linkage of the proposed scale up to UN ECOSOC Statistical Commission’s Big Data for Official Statistics WG [7] and International Telecommunication Union (ITU) “Big Data for Big Impact” [8] as a pilot case on social impact of mobile technologies through the use of insights provided by telecom sector big data to address SDG challenges in Sudan.

### Scaling Deep:

The scaling deep dimension of the proposal would be to explore the disaggregation capacity of the proof of concept to produce proxy poverty data by gender & perhaps age group mainly the youth (15 -29) and as well as the measurement of MPI at lower administrative unit level from locality. Encouraged by UNDP joining the

# UNDP Innovation Facility | 2018 Call for Proposals

consortium of Partnership for AI [9] recently, Sudan CO wants to harness its resources to not only generate more intelligent processing model and hence better automation, but also comprehensively address the DO NO HARM principle in the newly formulated data privacy by-law under Sudan's Telecom and Post Act 2017, seeking advisory from the consortium partners as well as UN Stats Big Data Working Group.

Other scale up considerations include:

IT Infrastructure: Standard IT infrastructure (both hardware and database) will not be adequate for the current initiative to expand to wider regions and deeper data analysis. With the increase in volume (more states), variety (data formats from more telecom operators) and velocity (time interval), the data processing will become more complex and challenging sooner or later; requiring bigger memory, high-end servers, CPU cost, data warehouse and networking infrastructure [10].

References:

[7] <https://unstats.un.org/bigdata/taskteams/mobilephone/>

[8] <http://www.itu.int/en/wtisd/2017/Documents/WTISD17-Call-for-Action-signed-en.pdf>

[9] <http://www.undp.org/content/undp/en/home/news-centre/news/2018/undp-joins-tech-giants-in-partnership-on-ai.html>

[10] <https://sp.ts.fujitsu.com/dmsp/publications/public/wp-bigdata-solution-approaches.pdf>

**Q33: What is your proposed strategy to implement the scaling pathways that will expand your solutions' reach? What do you plan to do over the next 9 months in to bring the solution to a larger scale and set the stage for its continued scaling beyond this period?**

The strategy to implement the scaling pathway will be as follows:

**PHASE 1: HOUSEHOLD SURVEY (October 2018 – February 2019)**

Having learnt lesson from 2017 proof of concept, the sample frame design for Multidimensional Poverty Index (MPI) measurement through Household Survey will be at level lower than locality – administrative unit. The HH survey will be conducted in collaboration with CBS and its result will serve as control data for the next activity.

**PHASE 2: Strengthening the “Big Data Mathematical Model” (November 2018 – March 2019)** This activity will be an iteration exercise to improve the previously formulated mathematical model to test its proxy poverty predictive capacity with more variety of datasets from three different mobile operators – Zain, MTN and Sudatel and on higher volume of data covering a region (approx. 25% of Sudanese population). The target region will be decided in consensus with TPRA and CBS in coming weeks. The locality level MPIs will again be the test data while the admin unit level calculations will be used for the evaluation of predictive capacity of the mathematical model.

**PHASE 3: BIG DATA for Development ARCHITECTURE and SOP DESIGN (April – July 2019)**

A design on national big data information flow, infrastructure, responsible parties, regulation, Standard Operating Procedures (SOP) and core team to harness telecom sector big data for SDGs will be produced in collaboration with the scale up partners proposed in Section 5.

**PHASE 4: Strengthen the relevant government entities, including the National Population Council (the entity assigned as the SDG focal point) with enhanced capacity of Big Data literacy and usage (July – September 2019)** This phase will have two main activities:

# UNDP Innovation Facility | 2018 Call for Proposals

- i. Capacity development and infrastructure setup: Building on the training received from previous initiative, and based on the output of Phase 3, relevant stakeholders will receive further training on SOP of improved big data mathematical model. Required IT infrastructure will be set up in an appropriate entity to ensure national ownership.
- ii. Knowledge event: In parallel, a knowledge and promotion event will be organized to share the scale up results as well as secure longer term and more commitment from new stakeholders with an intention of approval by the highest level within the government. Supporting the Scaling Up pathway described in Section 5, activities and outputs here will target UN level efforts on exploration of big data for measuring SDGs with an aim to inspire states with similar context.

## **Q34 How will you ensure sustainability and manage potential risks? Can the innovation be operationally, financially, socially and politically viable when serving millions of people? If so, how?**

The sustainability of the scale up is addressed by the Phase 4 output which will have participation of multiple influential stakeholders – Government entities directly reporting to Council of Ministers, state academic institutions and leading companies from private sector. In addition, the national SDGs monitoring mandated entity on board along with TPRA who has a high priority stake in this initiative guarantees to an extent allocation of federal government resources.

### Risks & Mitigation Measures:

1. Data privacy concerns while handling CDR from telecom operators.  
Similar data sharing model with data providers (Zain, MTN and Sudatel) as previous initiative will be established with clear Non-Disclosure Agreements signed. The previous protocol integrated three measures as below:
  - in house production of selected statistics/
  - remote access/ or moving of algorithms to data owners' server.This will meanwhile be addressed by support to comprehensively address the data privacy by-law under the Telecom Act. TPRA has already requested this support from UNDP and is a key deliverable of the scale up proposal.
2. Political interference and market risk.  
As this is a TPRA led initiative with all telecom operators on board through BDI4SD agreement, and CBS being a co-partner - this risk is low.
3. High cost of IT infrastructure.  
Commitment from national partners on this. UNDP can provide technical and procurement support on selected items.
4. Delay in HH survey and analysis due to natural disasters such as floods (rainy season is until October).  
Prioritized survey exercise in states which have less risk of being affected by floods will be conducted first.

## **Q35 How will you measure impact and monitor the progress of scaling?**

The project team initiative from 2017 – involving UNDP, TPRA, and CBS will continue for the scale up with additional team members.

A National Project Consultant will be hired who will be responsible for overall project implementation, convening of partners old and new, including monthly or fortnightly progress reporting on the ongoing &

# UNDP Innovation Facility | 2018 Call for Proposals

planned activities; and preparing communication products including working out loud through-out the implementation period.

A Big Data Analyst team (preferably from University of Berlin with advisory from the AI consortium) will be responsible for support of Phase 2 and Phase 3 activities in close collaboration with telecom partners and TPRA.

The above two will form the core project team and will conduct periodic meetings in addition to real time monitoring involving govt partner focal points. They will also report to concerned managements within CBS, TPRA, and UNDP.

## [Scaling Activity and Progress Monitoring Plan](#)

### **Q36 Introduce the team!**

1. Anisha Thapa, UNDP Big Data Pilot Lead
2. Mohamed Abdalla, TPRA Big Data Pilot Manager
3. Dr. Karamalla, CBS - Director General
4. Till Koebe and Fabian Bruschen, Berlin University Big Data Scientists
5. Abdalatif Hassan, UNDP Sudan Economist (Advisor)
6. Zain, MTN and Sudatel, Telecom Operators (CDR Data Providers and Preprocessing Support)
7. National Population Council (SDGs Coordinator and Focal Point for Sudan)

To be filled:

7. Big Data Scale Up Project Consultant (National)
8. Big Data and AI Advisor (From Consortium of Partnering for AI)

### **Q37: Propose a budget**

#### [Estimated Budget Breakdown:](#)