

# Strengthening Data Science Capacity and the Ecosystem: Enabling Data-Centered Public Health Interventions

## Request for Proposals

Applications due no later than January 13, 2022, 11:30 a.m. U.S. Pacific Time

### Background

The scale and scope of the COVID-19 pandemic have reiterated the need for access to and analysis of timely accurate health data to support the implementation of control measures and to guide health interventions and evidence-based practices. The supporting infrastructures and capacities through which researchers access and analyze these data, however, vary greatly. For several reasons, researchers in the Global South often are skeptical of broad calls for data sharing when they are left out of the analysis and results interpretation. Additionally, deploying “one-size-fits-all” analysis tools and governance processes across different regions from where they were created may produce poor analysis results and could even lead to misleading interpretations and conclusions when scientists analyze data without insight from the context in which it was collected.

As such, data science infrastructure and capacity needs to be proximally focused in the areas where the data emerge, to produce solutions that enable strong evidence-based practices to guide health interventions.

### The Challenge

Collecting, sharing, analyzing and interpreting health-related data are essential to the planning, implementation, and evaluation of public health practice as well as enabling advanced knowledge for health emergency preparedness. Many data resources are not reusable if they have not been sufficiently documented and contextualized, or appropriately collected and licensed for use, particularly in cases where secondary analyses are planned that were not originally anticipated when the data were originally generated.

The desired outcome of this request for proposals is development of tools and processes that are “fit-for-purpose” to facilitate collaboration among researchers in low- and middle-income countries. This effort is focused on filling data science gaps and challenges whose solutions will address global health problems while fostering multi-disciplinary collaborations among researchers. We are specifically interested in funding proposals led by early and mid-career researchers. We expect each primary applicant to have a senior scientific advisor who would commit to provide scientific and technical guidance and mentorship to the team and facilitate connections with international networks as needed.

Specifically, the **objectives** of the challenge will be to:

- Develop and improve foundational tools, standards and protocols that enable data-centered health research, interventions, and health surveillance
- Improve access to curated and linked datasets for research and training purposes

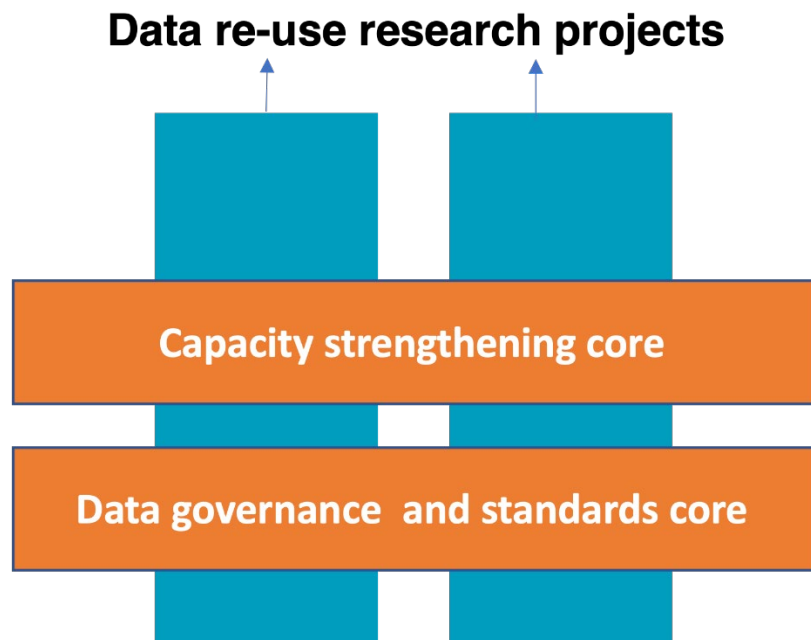
- Enable South-South data-centered collaborations and build on and strengthen existing initiatives and ecosystems
- Enhance Data Science infrastructure, analysis capabilities and local leadership to address public health problems
- Foster innovation with health data in the interest of their respective local communities
- Bridge the gap between data-centered research and policy through collaborations and design of appropriate data science communication tools and channels
- Increase health data reliability and enable reuse by developing tools and processes to curate, standardize and aggregate data
- Improve data collection, sharing, governance, regulatory compliance, and analysis processes to enable data-centered public health research

### What We Are Looking For

We will be developing a coordinated program in order to support these objectives. This will include cross-cutting cores which will provide specific functions to the program as well as individual research projects focusing on one of more specific disease areas.

Applications are invited from researchers working in healthcare, academia, industry, or research foundations, and will be welcomed from researchers based in any of the following regions: Africa, India, Bangladesh, Pakistan, Nepal, and Brazil. **Each proposal should be submitted by one primary applicant, but the awards require the participation of at least two research institutions from the regions listed above and not both in the same country.**

### Program structure



## **Data governance and standards core**

We seek to fund one data governance and standards core. This core would:

- Support the research projects funded through this initiative including the capacity strengthening core
- Leverage the funded research projects as use cases for development and refinement of data sharing processes and governance
- Seek to strengthen global communities and multi-disciplinary collaboration from researchers in low- to middle-income countries (LMICs) by developing processes to enable South-South collaborative analysis
- Develop adapted standards or Standard Operating Procedures that address data sharing issues including anonymization and perturbation protocols
- Develop re-usable tools, protocols, and standards to collect, clean, harmonize, and commonly structure individual data and meta-data related to one or many disease areas to increase health data reliability and enable reuse. New tools and process development should leverage and build on existing work
- Develop data informed processes to overcome barriers to the collection of complete and accurate information; and challenges related to data and information governance and data reuse

## **Capacity strengthening core**

We seek to fund one capacity strengthening core to:

- Support the research projects and the data infrastructure and governing core funded through this initiative
- Bridge the gap between data scientists, modelers, regulators and policy makers, through trainings and best practices development
- Develop re-usable training materials and handbooks
- Lead and coordinate collaborative analysis efforts through different formats (workshops, hackathons, jamborees, etc.)
- Lead capacity-building efforts targeting existing gaps in data science trainings in LMICs and building on existing initiatives (reusing available resources when available to avoid duplication of efforts)
- Develop data-informed processes to overcome barriers to the collection of complete and accurate information, and challenges related to data and information governance and data reuse

## **Data re-use research projects**

We seek to fund two data re-use projects where research will focus on one of more specific disease areas:

- Build data-driven public health platforms, tools and dashboards that guide policy making and public health interventions

- Identify, use, curate and link clearly defined datasets (administrative, clinical, environmental, etc.) to address policy gaps or deliver insights and conceptual and/or technological leaps on Global Health research priorities including:
  - o Maternal, Newborn and Child Health
  - o Infectious diseases
  - o COVID-19 and pandemic preparedness
  
- Proposals could focus on a specific disease area but need to demonstrate commitment to providing access to all research products including tools and processes developed

### **Funding level**

Successful proposals will receive an award of up to USD \$250,000 with an 18- to 24-month grant duration. While proposals require the collaborative participation of at least two investigators at institutions in different countries based in any of the regions listed above, proposals should be submitted by one primary applicant/institution. While restricted to one application per institution as the primary applicant, researchers may participate as collaborators in multiple collaborative applications.

Selected proposals will become part of a coordinated program aiming at Strengthening Data Science Capacity and the Ecosystem to Enable Data-Centered Public Health Interventions. We seek to build a network of researchers working together in close collaboration with the existing [Data Science Grand Challenges network](#) and its partners to build on lessons learned in different regions and further develop processes, tools and governance that maximize the impact of data-focused research studies, as well as to identify and address the global data-related challenges.

### **Successful proposals to be considered should:**

- Be driven by a shared commitment to open science, data sharing, and building collaboration and analysis infrastructure to enable discoveries that will benefit people everywhere
- Seek to strengthen global communities and multi-disciplinary collaboration from researchers in LMICs, and involve patients and the wider public
- Involve substantial collaboration between at least two research institutions in different countries encompassing Africa, India, Bangladesh, Pakistan, Nepal, and Brazil. The suggested collaboration should be key to advancing the project goals and yield insights that are unlikely in the absence of the co-produced approaches. In addition, however, applications could include collaborations with institutions in other geographical areas
- Have the potential to have impact on addressing public health issues within the proposed budget and timeframe of 18 -24 months
- Data re-use proposals should use well-defined datasets that directly address the core research questions and have a clear governance plan and demonstrate how the data could benefit the broader community at a later stage.
- Demonstrate a commitment to impact policies and the general public

**We will not consider funding for:**

- Studies that aim to collect and generate new data through this funding
- Proposals led by institutions not based in any of these regions: Africa, India, Bangladesh, Pakistan, Nepal, and Brazil
- Proposals that will involve a single research institution or solely involve collaborators from multiple institutions from the same country
- Proposals that do not demonstrate a clear commitment to open science and making their findings, processes and/or tools developed accessible and reusable
- Proposals that do not demonstrate how they would leverage existing work and networks
- Proposals that do not demonstrate how they can benefit from and support other projects funded through this coordinated program
- Proposals that are not accomplishable within the grant term