Request for Proposals: Climate and Health - 2022

Lacuna Fund: Our Voice on Data

28 April 2022

Table of Contents

1 – Introduction

Overview and Purpose of Lacuna Fund

Lacuna Fund supports the creation, expansion, and maintenance of equitably labeled datasets that enable the robust application of machine learning (ML) tools of high social value in low- and middle-income contexts globally.

The Fund aims to:

- Disburse funds to institutions to create, expand, and/or maintain datasets that fill gaps and reduce bias in labeled data used for the training and/or evaluation of machine learning models.
- Make it possible for underserved populations to take advantage of advances offered by AI.
- Deepen understanding by the machine learning and philanthropy communities of how to fund development and maintenance of equitably labeled datasets most effectively and efficiently.

Principles of Lacuna Fund

The following principles guide the operations of the Lacuna Fund.

- **Accessibility** – The Fund is committed to ensuring that labeled datasets created through its funding are accessible to and benefit underserved communities in service of the goals outlined above. Datasets and related intellectual property will utilize appropriate open data licensing to maximize responsible downstream use. (See the Fund's IP Policy for additional details.)
- **Equity –** The Fund aims to make AI more equitable by supporting the creation, expansion, and maintenance of datasets that are created by and responsive to the needs of those with underrepresented identities globally. These datasets should not create or reinforce bias, (e.g., they should be gender inclusive and representative of people of color globally) nor should they support the systems or technologies that create harm.
- **Ethics –** The Fund will fund data collection in a manner consistent with ethical labor standards and requires sub-grantees to outline steps they will take to protect privacy and prevent harm in the collection, licensing, and use of datasets created with grant funds.
- **Participatory Approach –** The Fund strives to meet the needs of affected stakeholders by encouraging the leadership or strong engagement of local experts, beneficiaries, and end-users in the governance of the Fund and in supported projects. The Fund will consider participation in a manner consistent with our principles on equity and ethics.
- **Quality** – Data generated by Lacuna-funded efforts should be of high quality, enabling beneficial applications in research, communities, and industry.
- **Transformational Impact** – The Fund aims to unlock the advances offered by AI for poor and underserved communities by funding datasets that address fundamental gaps in AI.

Philosophy of Grantmaking

Lacuna Fund values a collaborative and locally-driven approach to data creation, expansion, and maintenance. We recognize that the continued usefulness and maintenance of open data derives from a community invested in that data.

Lacuna Fund hopes to fund datasets that contribute to multiple applications of high social value, whether through research, commercial innovation, or improved public sector services. **While Section 3: Purpose and Need sets out needs identified by the Technical Advisory Panel (TAP), Lacuna Fund welcomes novel ideas within the domain area that have a clearly articulated benefit aligned with the selection criteria listed below.**

This call for proposals is supported by Wellcome Trust and Google.org.

2 – Overview

Organizational Eligibility

Lacuna Fund aims to make its funding accessible to as many organizations as possible in the AI for social good space and cultivate capacity and emerging organizations in the field.

To be eligible for funding, organizations must:

- Be either a non-profit entity, research institution, for-profit social enterprise, or a team of such organizations. Individuals must apply through an institutional sponsor. Partnerships are strongly encouraged, but only the lead applicant will receive funds.
- Have a mission supporting societal good, broadly defined.
- Be headquartered in or have a substantial partnership in the country or region where data will be collected.
- Have all necessary national or other approvals to conduct the proposed research. The approval process may be conducted in parallel with the grant application, if necessary. Approval costs, if any, are the responsibility of the applicant.
- Have the technical capacity – or the ability to build this capacity through a partnership described in the proposal - to conduct dataset labeling, creation, aggregation, expansion, and/or maintenance, including the ability to apply best practice and established standards in the specific domain (e.g. climate changes, health impacts/outcomes) to allow high quality AI/ML analytics to be performed by multiple entities.

Selection Process and Evaluation Criteria

Lacuna Fund seeks proposals to unlock, create, aggregate, and/or improve labeled datasets for machine learning applications that will enable equitable climate and energy outcomes in Africa, South and Southeast Asia, and Latin America. Lacuna Fund and its partners will perform an initial screen of the proposal for organizational eligibility and feasibility. Following the initial screen, a Technical Advisory

Panel of domain experts, data users, and stakeholders will evaluate the proposals based on the selection criteria outlined below. Technical Advisory Panel members may not submit a proposal in response to an RFP for which they are a reviewer (see Lacuna Fund's Conflict of Interest Policy).

The Technical Advisory Panel for this call will review the submissions and select a set of proposals to be funded. The selections will be based on the degree to which the full proposals meet the following criteria:

- **Quality** -- The organization or team proposing the project includes qualified experts in: a) the domain of interest; b) machine learning; and c) data management.

- **Transformational Impact** – The project either: a) labels or collects validation data for existing labeled datasets or upcoming surveys and in so doing, unlocks additional value in the existing dataset; b) creates a new, high-value labeled dataset for an underserved population or problem related to the Sustainable Development Goals (SDGs); c) makes an existing dataset more representative and inclusive of low- and middle-income contexts; d) aggregates existing datasets; or, e) makes a widely used and equitable dataset more sustainable.

- **Equity** – The team states the equity issue they are proposing to address and describes how the dataset will fill gaps and make the dataset more representative and equitable. NLP addition: The proposed dataset represents, or lead to applications representing, populations currently underserved by language technologies.

- **Participatory Approach** - The team is headquartered in the geographic area where the data will be collected and/or has a substantial partnership with an institution headquartered in the region to ensure sustained maintenance and usage of the dataset by the local community. The proposal describes how the team will engage affected stakeholders, seek informed consent for data collection and use, and share project outputs with data providers and/or the community.

- **Ethics** -- The project has a plan to address and is able to pass an ethical screen (e.g., an institutional review board) that probes: a) privacy concerns, b) potential for downstream misuse c) possible discrimination vectors (e.g. gender), and d) fair and equitable working conditions, if paid labelers are involved in the project.

- **Efficiency** – The proponent has taken into account existing datasets and proposes to use effective data collection and labeling techniques and tools to speed the collection, cleaning, and sharing of data.

- **Feasibility** – The project is feasible in relation to the budget and scope of work proposed.

- **Accessibility** - The dataset will be made widely accessible under open source licensing, or if this is not possible, a compelling case is made for more restrictive licensing in order to protect privacy or prevent harm.

- **Sustainability** – The project has a plan to ensure sustainability and future maintenance of the dataset e.g. by a dedicated community or a pool of interested parties (for-profit and/or not-for-profit) and a robust governance model for the open dataset.

Timeline

Request for proposals opens	28 April 2022
Question and Answer deadline Please submit questions to secretariat@lacunafund.org	20 May 2022
Answers posted	3 June 2022
Full proposals due	17 July 2022

Question and Answer Period: All questions related to the RFP should be submitted to secretariat@lacunafund.org with "Climate and Health RFP 2022 Question" in the subject line. Questions submitted by 20 May 2022 will be de-identified and answered publicly by 3 June 2022 on the Lacuna Fund website in a document posted on the "Apply" page and shared with all the applicants via email.

3 – Purpose and Need

Purpose

The purpose of this call for proposals is to support efforts to develop open and accessible datasets for machine learning applications related to understanding and mitigating climate harms on health in Africa, South and Southeast Asia, and Latin America.

From variations in spread, prevalence, and location of infectious diseases to health problems caused by the increasing frequency of natural hazards, changes in climate are negatively impacting human and community health in several ways. These changes disproportionately affect people in low- and middle-income contexts who have done the least to contribute to them. Machine learning and other data science tools hold great promise to advance efforts across a variety of sectors to understand, mitigate, and adapt to climate change. However, particularly in low- and middle-income contexts globally, a lack of ground truth data hampers the effective use of data science tools to generate information for decision making. Filling local data gaps can help policymakers, researchers, and local communities take advantage of machine learning tools to better understand future trends and take action to mitigate the harms of a changing climate on human health. Specifically, machine learning models driven by local data could inform infrastructure planning, strengthen healthcare and food security, help make resource allocation decisions, and identify interventions to mitigate and prepare for climate harms on health.

Need

Lacuna Fund seeks proposals from qualified, multidisciplinary teams to develop open and accessible training and evaluation datasets for machine learning applications for climate and health in Africa, South and Southeast Asia, and Latin America.

Proposals may include, but are not limited to:

- Collecting and/or annotating new data;
- Annotating or releasing existing data;
- Augmenting existing datasets from diverse sources to fill gaps in local ground truth data, decrease bias (such as geographic bias, gender gaps or other types of bias or discrimination), or increase the usability of data and technology related to climate and health in low- and middle-income contexts;
- Linking and harmonizing existing datasets (such as across regions, diseases, climate events, and time).

We see a need for training and evaluation datasets that will help us better understand the impacts of climate change on human and community health and interventions that could mitigate these impacts. We seek datasets identified by local experts designed to address locally identified needs, so the following are illustrative examples only. **Datasets may include, but are not limited to the following:**

- **Linking digitized vital registration datasets to climate data**, including creating and digitizing comprehensive vital registration systems information in communities where these are not present
- **Disease vector, outbreak, and climate datasets,** including collecting and linking data on climate change and disease outbreaks to better administer preventive care or immediate aid
- **Non-communicable diseases and climate change data,** such as linking the impact of climate factors with non-communicable health outcomes (e.g., health impacts of natural hazards caused by climate change such as heat waves, wildfires, particulate matter and smoke, droughts, storms, rising sea levels, and warming oceans)
- **Extreme weather data** for early warning systems (e.g., predicting drought, disease, heat)
- **Aggregating Water, Sanitation, and Hygiene (WASH) system** data with climate datasets, including datasets showing vulnerability to extreme weather events and associated health effects/interventions
- Datasets showing impacts of climate change on health **in both rural and urban areas and informal settlements**
- Datasets illuminating the **effects of climate change on mental health** as well as the effectiveness of nature-based and other interventions on mental health
- **Health co-benefits of climate action,** linking data on health benefits to climate mitigation measures
- Datasets that provide **gender disaggregated data** and allow a deeper understanding of the ways in which climate change affects women and men differently.

- **Datasets quantifying the indirect impacts of climate change on health** such as its effects on food security and understanding the human and environmental factors related to food production, climate, and weather changes

While the focus of Lacuna Fund is primarily on dataset creation, annotation, augmentation, and maintenance, proposals may include the development of a baseline model(s) to ensure the quality of the funded dataset and/or to facilitate the use of the dataset for socially beneficial applications.

We encourage teams to consider:

- Leveraging existing resources, including collection methods or technologies, linking pre-existing datasets across climate and health, as well as existing resources in other contexts;
- Strong collaboration and contribution from data scientists, health professionals, government agencies, and/or the meteorological field on project teams;
- Reflecting on gender and other factors associated with inequalities, in order for datasets to accurately represent how climate change impacts different communities/groups;
- Working across areas (e.g. across diseases, regions, climate events, time) where possible and relevant;
- Privacy considerations at both the individual and community health level.

4 – Proposal Information

Note: The Lacuna Fund website includes various resources, such as relevant references on data quality, documentation, and format, to help applicants prepare a competitive application.

Proposal submissions will only be accepted through the SurveyMonkey Apply application portal available at www.lacunafund.org/apply. A description of application questions is available below for information only. The following sections are required:

- Applicant Information (accessible in the portal)
- Proposal Narrative
- Timeline, Budget, and Budget Narrative
- Due Diligence Self-Reporting Form

Applicants' Information

This section will prompt the applicant to provide:

- A 200-250 word proposal abstract;
- Details about the institution(s) and/or team applying;
- Where the work will take place;
- CVs for key team members.

The applicant will also be asked to provide information about:

- The affiliated institution(s) ethical review processes;
- The team's ability to gain national approvals.

Proposal Narrative

Please limit your proposal narrative to 10 pages not including references, with 2.5 cm margins and a minimum of 11-point font. Appendices or proposal narrative material beyond 10 pages will not be reviewed.

This section will prompt the applicant to upload a cohesive narrative, in PDF or Word form that addresses the following:

- **Qualifications** - Describe the organization(s) or partnership(s) applying, how they satisfy the eligibility criteria articulated above, and your unique qualifications to undertake the proposed work.

- **Problem Identification and Proposed Solution and Dataset** - Describe the problem or gap in training or evaluation data and the proposed solution. Summarize the dataset(s) you intend to create, augment, aggregate, or maintain. *Please include how your project addresses a gap and complements existing work.*

- **Specifications and Deliverables for Proposed Data and Documentation** – Include the following:
  - Quantity of data that will be included in the dataset.
  - Types and format of data and/or labels, as well as sample frame and size or a plan to ensure representation, if applicable.
  - Metrics to be used to assess desired outcomes of data creation (e.g., fairness metrics in the dataset, QA/QC against a benchmark, etc.)

- **Intended Beneficiaries and Use Cases** - Describe previous consultation and/or proposed collaboration with intended beneficiaries and outline potential current and future use cases for the proposed datasets. State how the proposed quality, collection methods, and other details make the data suitable for use in that particular operational context.

- **Methodology** – Provide a brief overview of the proposed steps (and key assumptions) for developing and implementing the project. Please include:
  - Proposed data collection and labeling techniques and information on interoperability. Please include consideration of existing or common infrastructure and the latest techniques and tools to speed the collection, cleaning, and sharing of data.
  - A plan to assess and mitigate error and bias (e.g., gender bias or other biases).
  - Plans for data storage, ownership, and discoverability, in line with the Lacuna principles. For example, the proposed data format should be well documented and include information required for the end user to use the data.
  - Permissions in place or steps you will take to secure national or other required approvals. Consider which jurisdictions require approvals and whether the proposed

research meets the definition of research in that jurisdiction. If you determine that local, national, or regional approvals are *not* required, please explain why not.

- o Any anticipated challenges or uncertainties in data collection and proposed countermeasures.

- **Transformational Impact** - Explain how the proposed labeling or dataset will contribute to achieving the desired impact. If applicable, describe how the products could motivate multiple and durable paths of research or commercial application. For example, datasets that could help better understand the relationship between climate change, heat stress, and morbidity and mortality could inform early warning systems and interventions in multiple geographies globally. Note any practical constraints you may face (e.g., internet penetration).

- **Data Management and Licensing** – Please describe:

  - o Any anticipated issues related to copyright for source data and collaboration with the copyright holder. Please also address any anticipated issues for copyright and licensing of secondary data.

  - o Plans for licensing to maximize responsible downstream use. Per Lacuna Fund's IP Policy, the dataset and any related IP, such as collection methods, datasheets, how to load or read datasets, or other information to ensure usability should be made available under an open source, by-attribution license (CC-BY 4.0 or similar). If more restrictive licensing is proposed, provide a rationale for this. The budget may include resources for licensing.

  - o If you intend to use an existing dataset for your project, please indicate that your team has received the necessary permissions from the dataset's owner that the dataset can be released in accordance with Lacuna Fund's IP Policy, or provide justification for another licensing structure. *Letters of support from existing data holders are optional but encouraged.*

  - o Ensure your metadata is well-described.

- **Risks, Including Ethics and Privacy** - Identify potential risks, including but not limited to potential privacy and ethical concerns, and describe steps you will take to mitigate them. Specifically:

  - o State how you will ensure informed consent if appropriate. (This should include notification of potential future use cases for data.)

  - o Describe how you will ensure equity in project labor, including but not limited to fair compensation for labeling and gender parity for field agents.

  - o Present a plan for anonymization of personally identifiable information (PII) and compliance with privacy laws if applicable. If a national legal framework is not available, the proposal should outline or refer to best practice.

  - o Discuss potential adverse impacts in the production and use of the dataset and steps to mitigate them.

- **Sustainability Plan** – Describe how the labeled dataset will be maintained and/or expanded beyond the initial funding (e.g., through resultant ML applications, by a dedicated community, or a pool of interested parties with a robust governance model for the open dataset).

- **COVID-19 Considerations** – Please state if you will need to adapt your work due to restrictions related to COVID-19. If so, describe the steps you will take to adapt, and in the case of field work, ensure the safety of project staff.

Project Timeline

This section will prompt the applicant to submit a table with a timeline for the completion of major activities and deliverables. The timeline may include, but is not limited to, staff training, data collection, labeling, quality assurance, validation/cleaning, and data publication. Deliverables may include, but are not limited to, portions of the dataset to demonstrate proof of concept, the full dataset, and accompanying documentation or collection methods to be open sourced.

All timelines should include a date by which data will be publicly available with all documentation.

Note: Proposed projects must be completed, datasets published, and final reports submitted by June 2024. For planning purposes, you can expect that agreements will be completed and work may begin by December 2022.

Budget and Budget Narrative

Provide a detailed activity-based objective budget for the completion of the steps included in the deliverables table submitted through the SM Apply portal. This should be formatted in the Lacuna Fund budget template, available in the applicant portal

The total pool available is approximately $1 million USD. We would like to fund at least one project in each of the target regions (Africa, Latin America, South and Southeast Asia) and anticipate supporting 4-6 smaller projects with budgets between $50-100k and 2-3 larger, more complex projects with budgets ranging from $200-300k. The Technical Advisory Panel will assess **the feasibility and suitability of the budget** as well as the **linkage between the budget and grant narrative** as part of the selection criteria. Budgets may include, but are not limited to, costs for:

- capacity building related to data collection and quality assurance/quality control;
- data collection;
- data labeling;
- QA/QC or verification;
- post-processing of data;
- data publication;
- licensing;
- open access publication of results;
- time to prepare a data statement for the dataset;
- crowd-sourcing efforts, such as label-a-thons.

Funds may not be used for the direct payment of any customs, import, or other duties or taxes levied with respect to importation of goods or equipment into any country or jurisdiction. **Indirect rates are**

strictly limited to 20% of direct research costs for institutions based in a low- or middle income country or 15% for institutions based anywhere else.

See the instructions sheet in the budget template for further information on budget guidelines, including information on allowable staff costs.

Thank you for your interest in Lacuna Fund and your efforts to support greater equity and accessibility for machine learning applications to support climate change adaption, mitigation, and planning. We look forward to reviewing your submission!

NOTE: Online Partnership Platform

Lacuna Fund has partnered with Climate Change AI to create an online platform to facilitate formation of teams for applicants to the upcoming Lacuna Fund calls for proposals in Climate & Health. In addition to applicant matchmaking, the platform will facilitate mentorship and technical assistance from experienced researchers for early career data scientists. Specifically, the platform will facilitate partnerships between interested potential applicants with other entities that have complementary skills and expertise. We believe such a community-building platform will enable robust, collaborative proposals for this call and in the future. More information about the platform can be found here.

Annex A: Due Diligence Self-Reporting Form

Submission of this form and all other materials should be completed through the Lacuna Fund application portal, accessible at https://lacunafund.org/apply.

Meridian Institute must comply with requirements of contributors to the fund as well as legal requirements related to being an organization providing funding based in the United States.

To the extent permitted by law, all information supplied will be held in confidence and not disclosed to any third parties without prior notice and approval.

1. Please state your form of legal incorporation and the relevant law/regulation. Please also provide the month and year when your organization started operating as this entity.

2. Does your organization have a Conflict of Interest Policy and are there any instances of Conflict of Interest Meridian Institute should be aware of?

3. Does any Public Official or government entity have any financial, management or controlling interest in your company/organization? If so, provide details and level of interest below.

4. Please list the full names of all Principals for your company/organization. (Note: "Principal" means the executive officers, partners, owners, directors, trustees or others who exercise control over your company/organization).

5. Does your company/organization employ any Public Officials? If yes, please list the name(s) and position(s) of the Public Official(s) below or on a separate sheet of paper.

6. Does any Principal of your organization have a close relative who is a Public Official (close relative means a mother, father, sister, brother, wife or child)? If yes, please list the name, relationship, title, responsibilities, and government department or agency of the Public Official(s) below or on a separate sheet of paper.

7. Has your company/organization, or any subsidiary or affiliate of your company/organization, or Principal of your company/organization ever been convicted in any felony matter or been the subject of a criminal investigation, indictment or similar proceeding? If yes, describe below:

8. During the prior three years, has your company/organization, or any subsidiary or affiliate of your company/organization, been the subject of past or pending litigation, or government investigation? If yes, describe below.

9. Is there additional information about your company/organization or any employee of your company/organization that would assist Meridian Institute in performing its anticorruption due diligence obligations? If so, describe below.

10. The undersigned, being duly authorized to respond to this questionnaire, and new certification as to the matters set forth below, hereby certifies as follows:

    o To the best of my knowledge, all information set forth in this anticorruption due diligence form is truthful, correct and complete;

    o I have read the information concerning the UK Bribery Act 2010 (http://www.justice.gov.uk/downloads/legislation/bribery-act-2010-guidance.pdf), Australian Criminal Code (http://www.oecd.org/daf/anti-bribery/anti-

briberyconvention/2027148.pdf), and the U. S. Foreign Corrupt Practices Act
http://www.sec.gov/investor/alerts/fcpa.pdf), at the noted websites and I am familiar
with the requirements of these anti-corruption statutes;

o   I have read and fully understand the Meridian Institute Code of Conduct and, as a
condition of doing business with Meridian Institute, I agree to abide by such Code of
Conduct;

o   Neither I nor my company have ever paid, approved or otherwise provided anything of
value, directly or indirectly, to a Public Official for any improper, corrupt or illegal
purpose, nor will we;

o   Neither I nor my company have ever created a false invoice or otherwise manipulated
documentation to disguise making a payment to a Public Official for any purpose, nor
will we;

o   My/our accounting practices will accurately and completely describe any payments
made, including any payments made to a Public Official.

**NOTE: "Public Official" means any person, whether elected or appointed who holds an executive,
legislative, administrative or judicial office or position in any public entity, including any international
agency. In addition, "Public Official" includes any person who performs public functions in any branch
of the national, state, local or municipal government of any country or territory or who exercises a
public function, by employment or under contract, for any public entity, agency or enterprise of such
country or territory, including state owned or controlled enterprises. The definition of "Public Official"
also includes any official of a political party or any candidate for political office.**