Request for Proposals: Climate and Energy - 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 philanthropic 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 The Rockefeller Foundation and GIZ on behalf of the German Federal Ministry for Economic Cooperation and Development.

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.
- Are 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 RFP - 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. energy systems, climate or emissions modeling) 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 that can support efforts to develop open and accessible 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 demonstrates interdisciplinary collaboration between qualified experts in language, research, machine learning, and data management. Partnerships between well-resourced actors with strong research capacity and actors serving marginalized populations and low-resourced languages are encouraged.

- **Transformational Impact –** Datasets should improve machine learning and have the potential to enable social benefit for underserved populations. Examples include, but are not limited to: a) labeling, cleaning, collecting validation data for, or pooling existing datasets to unlock additional value or ensure greater accuracy in the existing dataset; b) creating a new, high-value labeled dataset for a low-resourced language; c) making an existing dataset more representative and inclusive of race, gender, ethnicity, ability, etc.

- **Equity –** There is a compelling theory of change demonstrating how the dataset will be applied to help vulnerable and underserved communities.

- **Participatory Approach –** Datasets should center the needs of affected communities and work with partners to identify community benefit. Specifically, projects should engage community members in data governance decisions, (e.g., what data is curated and how it is used). If the dataset has a geographical scope, the team is predominantly located in the respective area and/or sustains close ties to local actors to ensure sustained maintenance and usage of the dataset by the local community.

- **Ethics –** The project 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 considered 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. While the EOI is not expected to include a detailed budget, we do expect EOIs to include a broad estimate of the overall budget expected to complete the project.

- **Accessibility -** The dataset will be made widely accessible under open-source licensing pursuant to Lacuna Fund's IP Policy. If this is not possible, a compelling case is made for more restrictive licensing in order to protect privacy or prevent harm, along with a mechanism for providing access under the proposed licensing.

- **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
Applicant webinar	17 May 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 Energy 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

This section articulates the purpose of the RFP and dataset needs that the Technical Advisory Panel has identified. Applicants should feel free to propose training and evaluation datasets with a compelling use case and rationale outside of the articulated needs.

Purpose

The purpose of this Request for Proposals is to support efforts to develop open and accessible datasets for machine learning applications that will enable equitable climate and energy outcomes in Africa, South and Southeast Asia, and Latin America. Climate change disproportionately affects people in low- and middle-income contexts who have done the least to contribute to it. Machine learning holds 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, the effective use of machine learning is hampered by a lack of ground truth data that can be transformational in its application and is created in local contexts. Machine learning could help map energy infrastructure needs, enable effective deployment of renewable energy technologies globally, and increase energy access and resilience. Unfortunately, many parts of the world that would benefit the most from such technologies do not currently have the data to power them.

Infrastructure, for example, is one of the most essential variables to energy planning as it heavily influences the choice of least-cost technological solutions that can bring electricity faster, cheaper, and cleaner to a particular region. Understanding demand is important for optimizing power systems, transitioning to clean energy, and expanding electricity access. Major uses for energy, such as agriculture, can provide insight into innovative energy sources and demand estimation. In addition to energy demand and use data, other variables such as climatology, emissions, grid infrastructure and reliability, and buildings data are essential to understand gaps in energy access and planning for a changing climate. In order for AI to make a broad and lasting contribution to clean, affordable, and reliable energy for all, more open and freely accessible AI training data and models are needed to develop locally adapted solutions.

Need

Lacuna Fund seeks proposals from qualified organizations to develop open and accessible training and evaluation datasets for machine learning applications for climate and energy in Africa, South and Southeast Asia, and Latin America. Datasets can cover single countries, single or several regions or the entire globe, depending on the proposed use-cases.

Proposals may include, but are not limited to:

- Collecting and/or annotating new data;
- Annotating or releasing existing data;
- Augmentation of existing datasets in all areas to decrease bias (such as geographic bias or other types of bias or discrimination) or increase the usability of data and technology related to climate and energy in low- and middle-income contexts.

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

The TAP sees a need for datasets that enable better energy planning, access, deployment, optimization, reliability tracking, and transition, as well as applications of energy in other sectors. **Datasets may include, but are not limited to, the following:**

- Energy demand data, at the building, community, or grid level. Energy demand data could be obtained by aggregating demand estimation data from distribution companies or household survey data or satellite data scaled with an ML model to achieve demand estimation at scale. Teams may also consider developing proxies for demand forecasting. For example, agriculture is a major energy use, and data from the sector could be useful to estimate energy demand.
- Energy supply data, including power generation and power systems planning data.
- Grid infrastructure location data, including obtaining existing data or developing proxy information.
- Electricity reliability, hazards, exposure and vulnerability, as well as fault detection data.

- Household-level survey data, including socioeconomic information for consumer demand estimation. Survey data could provide financial information to support the green energy transition and expanded energy access.
- Climatological data, emissions data (including from point sources or fugitive emissions), resource mapping, and other variables related to energy planning.
- Buildings data for energy efficiency, demand management, building decarbonization, and other climate friendly applications.
- Other energy relevant data, for example related to transport, manufacturing, cement, steel, storage, or links to water-use, land-use, air pollution, and other social or economic systems.

Across all projects, the TAP encourages teams to consider:

- Leveraging existing resources, including collection methods or technologies as well as existing resources in other contexts;
- Strong collaboration and contribution from system operators, government agencies, and/or the infrastructure/engineering field on project teams;
- Methodologies that will ensure datasets center the needs of affected communities, work with partners to identify community benefit, and engage community members in data governance decisions;
- Plans to make data widely accessible for use in community-centered projects; and
- Privacy considerations for both grid-level and household-level information.

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.

In the future, the applicant will be asked to provide information about:

- The affiliated institution(s) ethical review processes;
- The team's ability to gain national approvals.
  - Which jurisdictions in which research will take place require national approvals for any research?
  - Does the research proposed meet the definition of research that requires approval in each jurisdiction? If not, please provide explicit justification.
  - Please discuss your plan to seek approvals for the project if selected.

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, Proposed Solution and Dataset and existing efforts** - Describe the problem or gap in training or evaluation data and the proposed solution (dataset labeling, creation, aggregation, augmentation, or maintenance). *Please be clear about how your project addresses a gap and complements existing work.* Please summarize the dataset you intend to create, augment, or maintain, and the specific machine learning needs the dataset would fill in the context of applications in climate change adaptation, mitigation, and/or planning. List existing efforts you are aware of and if you plan to engage them in the project.

- **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 list potential current and future use cases for the proposed datasets. Ensure that use-cases are not generic but show how you imagine

better decisions can be made by specific organizations or actors with the new data being available and how it addresses the problem you identified above. Use-cases can be several levels removed, for example how machine learning results based on your data can enable new use-cases. 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. Explain permissions in place or steps you will take to secure national or other required approvals. 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.
  - 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 societal impact (particularly beyond just creating new models or research applications). If applicable, describe how the products could motivate multiple and durable paths of research, investment, planning or commercial application. For example:
  - How AI training data and models support predicting the microclimate instead of consuming significant computational resources using traditional physics-based models.
  - Another use case would be ML algorithms guide the building's dynamic operation to achieve the higher energy efficiency.

- **Data Management and Licensing** – Please describe:
  - Plans for licensing to maximize responsible downstream use, in particular communities who may not be able to afford commercially licensed data. Per Lacuna Fund's IP Policy, the dataset and any related IP, such as collection methods, models, 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.
  - 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, and implications for long-term maintenance and expansion of the dataset.
  - 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 and that the dataset generation will be described in either a detailed guide or a publication.
- **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 within 18 months from execution of agreements. 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 $2.5 million USD. We are anticipating proposed budgets in the range of $10k – 200k for small to medium-sized projects and up to $400k for large, complex projects.

We anticipate being able to fund 2-3 large projects and a larger number of smaller projects. 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 12%.**

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 Matchmaking Platform

Lacuna Fund has partnered with [Climate Change AI](#) to create an online platform to facilitate matchmaking for applicants to the upcoming Lacuna Fund calls for proposals in Climate & Energy. In addition to applicant matchmaking, the platform will facilitate mentorship and technical assistance from experienced researchers for early career data scientists. The platform will facilitate matching and connections between interested potential applicants with other entities that have complementary skills and expertise. Lacuna Fund hopes such a community-building platform will enable robust, collaborative proposals for this call and in the future. [Register here](#) for the partnership and mentorship platform.

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.