Request for Proposals: Climate and Forests - 2023

Lacuna Fund: Our Voice on Data

7 March 2023

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.

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 Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH (GIZ).

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. **For this specific call, preference will be given to non-profit entities and research institutions.** 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 and forests, etc.) to allow high quality AI/ML analytics to be performed by multiple entities.

** The Climate & Forests call is made possible with generous support from GIZ's FAIR Forward Programme on behalf of the German Federal Ministry of Economic Cooperation and Development (BMZ). Teams selected for funding will be required to follow specific record keeping, reporting, and procurement terms. For example, for subcontractors or consultants, teams will be required to obtain at least three offers for comparison for awards between EUR 1,000 and 200,000 and conduct a public tender for contracts over EUR 200,000. Detailed guidance will be provided to finalists.*

Selection Process and Evaluation Criteria

Lacuna Fund seeks proposals to unlock, create, aggregate, and/or improve training and evaluation datasets for machine learning applications that will enable equitable climate and forest 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, which are based on the principles that guide Lacuna Fund operations:

- **Quality** -- The organization or team proposing the project includes qualified experts in a) the domain of interest; b) machine learning; and c) data management. Collaborations with government agencies and community groups are encouraged. The team provides clear use case(s) for the dataset. The proponent situates the proposed dataset within existing resources (or lack of resources) in the domain and proposes to use effective data collection and labeling techniques and tools to speed the collection, cleaning, and sharing of data.
- **Transformational Impact** – The project makes existing dataset(s) more representative, inclusive, and/or sustainable or creates a new, high-value labeled dataset for an underserved population or problem. A proposal may be considered transformational if it has the potential to address a particularly important or timely problem with equity in ML/AI

or have significant reach in terms of number of underserved people or geographies affected.

- **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. There is a compelling theory of change demonstrating how the dataset will create greater access to the benefits of ML/AI for vulnerable and underserved communities.
- **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. In-country partners are involved in strategic elements of the project. The proposal describes how the team will engage affected stakeholders, seek informed consent for data collection and use, and share project outputs and benefits with data providers and/or the community.
- **Ethics** -- The project has a plan 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. The project's proposed goals and methodology are unbiased and ethical.
- **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), a robust governance model for the open dataset, and possible use cases for the dataset. The plan may include who will update and manage the dataset, potential funding sources, proposed engagement strategies for impacted populations and data users, and measures to keep data open and accessible.
- **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. The subgrantee will prioritize releasing the intellectual property under a permissive open-source licensing structure such as Apache 2.0 for any code or other inventions, or CC-BY 4.0 International for any other intellectual property (e.g., creative works that are not code, or patentable).

Timeline

Request for Proposals opens	7 March 2023
Question deadline Please submit questions to secretariat@lacunafund.org	11 April 2023
Answers posted	21 April 2023
Full proposals due	1 June 2023

Question and Answer Period: All questions related to the RFP should be submitted to secretariat@lacunafund.org with "Climate and Forests RFP 2023 Question" in the subject line. Questions submitted by 11 April 2023 will be de-identified and answered publicly by 21 April 2023 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 that will help communities better understand and address the relationship between climate change and forests 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 (ML) holds great promise to advance efforts across the forestry domain to understand, mitigate, and adapt to climate change.

Forests support livelihoods of diverse communities globally, host the world's largest share of terrestrial biodiversity, and provide important ecosystem services such as carbon sequestration. The vast majority of the world's forests are not located in legally established protected areas, leaving them vulnerable to deforestation due to agriculture, mining, logging, and urban development, contributing to global greenhouse gas emissions. These stressors on the world's forests negatively impact human health, exacerbate food insecurity, undermine the rights of Indigenous Peoples and local communities, and contribute to habitat and wildlife loss.

ML could help unlock the potential for forests to serve as natural climate solutions. However, particularly in low- and middle-income contexts globally, the effective use of ML is hampered by a lack of ground truth data. Filling in data gaps and increasing accessibility to ML technologies can help policymakers, researchers, and local communities take advantage of ML tools to better understand future trends and inform action to mitigate and adapt to the impacts of a changing climate on forest ecosystems.

Need

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

This Request for Proposals (RFP) is funded by GIZ on behalf of the German Federal Ministry for Economic Cooperation and Development and as such, primary applicants must originate from or be partnered with organizations that are in nations on the BMZ Country list. Of particular interest are proposals originating from FAIR Forward partner countries (Ghana, Rwanda, Uganda, Kenya, South Africa, India, Indonesia), and preference is for non-profit entities or research institutions.

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 forests in low- and middle-income contexts;
- Improving prior datasets based on learnings to date;
- Linking and harmonizing existing datasets (such as across regions, climate events, and time);
- Maintaining/updating an existing high value dataset from one of the regions mentioned above;
- 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 see a need for training and evaluation datasets that will help us better understand the impacts of climate change on forests and interventions that could mitigate these impacts, demonstrating how to unlock forests as natural climate solutions. 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:**

- **Land cover / land use change datasets with spatial signatures that improve understanding of the relationship between land cover/land use change and climate change** – for example, datasets that delineate and/or link: agriculture (field/farm) and forest boundaries; tree species distribution, classification, and diversity; tree crops (such as oil palm and rubber); tree cover gain and loss across the tropics (from natural occurrences and anthropogenic deforestation activities); forest degradation and reduction in ecosystem services; and energy use of forests as fuel.
- **Afforestation and reforestation datasets that inform strategic restoration efforts to mitigate climate change** – for example, by linking weather, water, and soil data to identify which trees will grow best amidst changing precipitation patterns and desertification without degrading the soil.
- Datasets that **show the value of non-extractive forest ecosystem services and products that could help Indigenous Peoples and local communities adapt to climate change and/or receive compensation for activities that preserve forest ecosystems and mitigate climate change** – for example, datasets on clean water; forest products (such as acai or Brazil nuts); local market information.
- **Contextualizing Indigenous knowledge of biodiversity into labeled training data to highlight how climate change is impacting biodiversity in forest ecosystems** – for example, cross-referencing local Indigenous languages with satellite earth data on biodiversity, forests, and climate to make scientific predictions; data showing linkages between biodiversity loss, habitat loss, and loss of Indigenous knowledge.
- Datasets that **correlate supply chain interventions with forest loss and climate change** – for example—to: improve accuracy of deforestation estimates, increase traceability of supply chains (including through asset-level datasets), verify sustainability certifications; data disaggregated by specific commodity or driver of deforestation.

- Datasets that **link environmental/nature crimes including illegal logging, illegal mining, and conflict timber to climate change, including information to detect and prevent these crimes** – for example, encroachment on forest areas and/or Indigenous territories; acoustic data capturing use of equipment in forest areas or forest buffer zones; digitizing forest boundaries to detect illegal logging (such as by employing/overlaying existing datasets on illegal activities in an area); drone detection of illegal mining (causing deforestation and coastal erosion) and/or of wildlife poaching hotspots and routes.
- Datasets that **can be used to show the impact of deforestation on wildfires (or vice versa) and climate change** – for example, wildfire detection, prediction, and classification; risk mapping (such as risk reduction, response, and recovery) and rapid damage assessment.
- **Carbon sequestration datasets utilizing biomass estimation or species dominance** – for example, determined through field sampling, environmental DNA (eDNA), and/or LiDAR to reduce uncertainty around carbon estimation or to verify claims around carbon to combat greenwashing and to report reference emission levels; datasets containing evidence regarding the extent to which blue carbon capture happens in mangroves.
- **Invasive species datasets that link with climate change data** – for example, tracking the impact on biodiversity in forests, carbon capture in forests, ecosystem services, altered landscapes, and coastal areas.
- Datasets **showing how climate change affects animal species movements** – for example, changing density and migration patterns of various species resulting from climate change; monitoring hotspots of human-wildlife conflict and wildlife movements in forest areas.
- Datasets **linking changes in forest health to climate change** – for example, linking species health and vegetation indices with hydrology/water quality, meteorological patterns, and soil moisture; datasets in coastal areas with mangrove reforestation or regions with lake erosion.

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
- Budget & Timeline

Applicant 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.
- Information about the affiliated institution(s) ethical review processes;
- Information about 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 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.
  - Quality control measures, such as the quality all data samples must meet for the dataset. Please include how the team plans to address outliers that may affect the quality of the dataset.
  - A plan to assess and mitigate error and bias (e.g., gender bias or other biases).
  - How you plan to leverage existing resources, including collection methods or technologies, linking pre-existing datasets across the domain, as well as existing resources in other contexts.

  - 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.
  - 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 use in research or commercial contexts. Note any practical constraints you may face (e.g., internet penetration).
- **Data Management and Licensing** – Please describe:
  - 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.
  - 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.
  - 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.*
  - Plans to include a metadata file and datasheet as documentation for your dataset, according to Lacuna Fund's Dataset Hosting and Documentation Guidance. For earth observation data, metadata should be listed in the STAC standard. Please consider how your documentation will support FAIR data principles (Findability, Accessibility, Interoperability, and Reuse).
  - The hosting platform you intend to use. Hosting platforms must assign a digital object identifier (DOI) to the dataset, quantify downloads of the dataset, and collect contact information for dataset downloads. Please see more guidance for suggested hosting platforms in the Dataset Hosting and Documentation Guidance.
- **Risks, Including Ethics and Privacy** - Identify issues and potential risks, including but not limited to potential privacy and ethical concerns, and describe steps you will take to mitigate them. Specifically:
  - A reflective statement on the communities you come from and identities you hold, and how those may impact your work.
  - State how you will ensure informed consent if appropriate. (This should include notification of potential future use cases for data.)
  - Describe how you will ensure equity in project labor, including but not limited to fair compensation for labeling.

  - Describe how gender diversity and other demographic considerations are incorporated in the project team, collection of training data, and model development, in order for datasets to accurately represent impacts on different communities/groups.
  - 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. Please incorporate privacy considerations at both the individual and community level.
  - Discuss potential adverse impacts in the production and use of the dataset and steps to mitigate them, including potential human rights risks and the potential for high power consumption in AI technology leading to increased carbon dioxide emissions.
- **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) and how a potential use case could sustain the project.

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 30 April 2025. For planning purposes, you can expect that agreements will be completed and work may begin by January 2024.

Budget

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

The total pool available is approximately $2 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;
- data storage
- computing power
- model development.

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%.**

In-kind cloud storage and computing power may be available from Lacuna Fund partners. If you would like to utilize this resource, please include it in your budget. Current in-kind support may be available from Google Cloud and Amazon Web Services (AWS). Selected teams will receive instructions for how to apply when they receive their award.

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 adaptation, mitigation, and planning in the forests domain. We look forward to reviewing your submission!

NOTE: Mentorship Opportunity

Lacuna Fund is pleased to partner with Deep Learning Indaba to offer mentorship opportunities for applicants. The Deep Learning Indaba is an organization whose mission is to Strengthen Machine Learning and Artificial Intelligence in Africa and works towards the goal of Africans being not only observers and receivers of the ongoing advances in AI, but active shapers and owners of these technological advances.

On a one-time basis for this Lacuna Fund call, the Deep Learning Indaba will offer applicants from around the world a special opportunity to join the Deep Learning Indaba community and match those who are interested with a mentor who will review your draft proposal and discuss avenues for strengthening it. Interested parties may apply for a session with a mentor by visiting the Indaba Mentorship Programme - Deep Learning Indaba and selecting to participate **"*As a mentee*".** When prompted, select for assistance **"*Writing a research proposal*".** In the field **"*Please tell us more about your draft proposal*",** please type **LACUNA FUND,** followed by a detailed description of your proposed dataset. Providing a detailed description will support with matching applicants to mentors that are aligned in your interest area. Interested applicants are encouraged to apply for a mentorship session

after you have a detailed outline or draft proposal and *at least 4 weeks before* the proposal due date. We will do our best to identify a mentor for everyone who requests one, but we cannot guarantee that mentors will be available for all. Mentors will be assigned on a first come, first-served basis. All applicants are expected to read and abide by the mentorship programme's *code of ethics and conduct*.