

Lacuna Fund: Our Voice on Data

Request for Proposals: Labeled Datasets for Agriculture in Sub-Saharan Africa

13 July 2020

Table of Contents

1 – INTRODUCTION	2
PURPOSE AND GOALS OF THE FUND.....	2
PRINCIPLES OF THE FUND	2
PHILOSOPHY OF GRANTMAKING	3
2- OVERVIEW.....	3
ORGANIZATIONAL ELIGIBILITY	3
SELECTION PROCESS AND EVALUATION CRITERIA	3
TIMELINE	4
3 - PURPOSE AND NEED	5
PURPOSE	5
NEED	5
4 – PROPOSAL INFORMATION	6
YOUR INFORMATION	7
PROPOSAL NARRATIVE	7
TIMELINE	9
BUDGET AND BUDGET NARRATIVE	9
ANNEX A: DUE DILIGENCE SELF-REPORTING FORM	11

1 – Introduction

Purpose and Goals of the Fund

Lacuna Fund supports the creation, expansion, and maintenance of equitably labeled datasets that enable the robust application of machine learning 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 machine learning.
- 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 most effectively and efficiently fund development and maintenance of equitably labeled datasets.

Principles of the 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 creating datasets that are representative of the Global South and people of color globally and their needs. These datasets should not create or reinforce bias.
- **Ethics** – The Fund will fund data collection in a manner consistent with ethical labor standards and requires subgrantees 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 involving local beneficiaries and end users in the governance of the project as well as in data creation.
- **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.**

Lacuna Fund is supported by The Rockefeller Foundation, Google.org, and Canada's International Development Research Centre.

2- Overview

Organizational Eligibility

The 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 welcome but only the lead applicant will receive funds.
- Have a mission supporting societal good, broadly defined.
- Be an organization headquartered in Africa or have a substantial partnership with organization(s) headquartered in Africa.
- Have all necessary national or other approvals to conduct proposed research. The approval process may be conducted in parallel with grant application, if necessary. Approval costs, if any, are the responsibility of the applicant.
- Have the technical capacity to conduct dataset labeling, creation, expansion, and/or maintenance, including the ability to apply best practice and established standards in the specific domain (e.g. agriculture) to allow high quality AI/ML analytics to be performed by multiple entities.

Selection Process and Evaluation Criteria

The Lacuna Fund and partners will perform an initial screen for organizational eligibility and completion of all proposal requirements. Following the initial screen, a Technical Advisory Panel of domain experts, data users, and stakeholders will evaluate 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 assess proposals for funding based on the degree to which they 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; or d) makes a widely used and equitable dataset more sustainable.
- **Equity** – There is a compelling theory of change demonstrating how the dataset will improve machine learning and be applied to help the poor and vulnerable.
- **Participatory Approach** - If the dataset has a geographical scope (e.g. language or geospatial datasets), 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 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 We 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

RFP Posted Publicly on Lacuna Fund Website	13 July 2020
Question and Answer Deadline Please submit questions to secretariat@lacunafund.org	27 July 2020
Answers Posted	3 August 2020
Applications Due	3 September 2020

Question and Answer Period: All questions related to the RFP should be submitted to secretariat@lacunafund.org with “Agriculture RFP 2020 Question” in the subject line. Questions submitted by 27 July will be de-identified and answered publicly by 3 August on the Lacuna Fund website in a document posted on the [“Apply” page](#).

3 - Purpose and Need

This request for proposals will fund labeled agricultural datasets for machine learning in sub-Saharan Africa, whether referenced to earth observation (EO) data or related to other aspects of the crop and animal agricultural system. Lacuna Fund’s mission is to support labeling and creation, expansion, and maintenance of transformative labeled datasets to fill gaps and improve accuracy in labeled data used for machine learning. Projects focused on building machine learning models or other applications will not be funded.

Purpose

In the agricultural AI for social good domain, recent advances in the analysis of remote sensing data have enabled improved accuracy of a variety of tasks, from prediction of crop types and yields to field boundary estimation worldwide.¹ Both startups and large development programs have built on this data availability to develop machine learning applications that allow for both more personalized services and better data for decision-making.

However, a lack of ground truth labels, as well as a lack of datasets to address unique challenges in mapping smallholder farms, hinders further progress towards building beneficial ML applications that are effective for under-served populations worldwide.² Alongside a need for greater open source, quality data, an African ecosystem capable of building on the data in real-world applications must be strengthened to ensure equitable benefit from any funding.

An explosion of EO providers and services, nascent standards, such as the SpatioTemporal Asset Catalog (STAC)³ and a preliminary version⁴ of best practices for ground reference data collection and cataloguing are creating greater coherency for ML-ready labeled data across the AI for agriculture landscape.

Need

This RFP prioritizes proposals for datasets in crop and animal agriculture. However, the TAP recognizes the importance of other forms of agriculture such as fisheries and apiary to gender equity and poverty

¹ Lei Ma et al., “Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review,” *ISPRS Journal of Photogrammetry and Remote Sensing* 152 (June 1, 2019): 166–77, <https://doi.org/10.1016/j.isprsjprs.2019.04.015>.

² Jadunandan Dash, “Satellites and Crop Interventions,” *Nature Sustainability* 2, no. 10 (October 2019): 903–4, <https://doi.org/10.1038/s41893-019-0402-3>.

³ <https://github.com/radiantearth/stac-spec>

⁴ <https://github.com/radiantearth/ground-referencing-guide>

alleviation and will consider proposals in these areas on a case-by-case basis. Given existing funding opportunities, the Fund will not make awards in areas related to forestry.

Building on the recent release of a challenge for crop type classification and field boundary mapping and associated datasets for Kenya, Tanzania, and Uganda as well as for wheat rust,⁵ the Technical Advisory Panel has identified a need for labeled datasets in the following areas. However, Lacuna Fund RFPs are intentionally open, to encourage new and innovative ideas that we may not have identified.

- Field boundary identification, particularly for smallholder farms.
- Crop type classification, including issues related to crop rotation or the combination of crops in a single plot.
- Yield estimation, including particular issues related to estimation for subsistence farmers.
- Pasture carrying capacity and/or early warning systems for forage availability and depletion in pastoral lands.
- Issues related to pastoral migratory patterns and anticipated conflicts with crop farmers.
- Prediction of the best locations to establish forage banks and support infrastructure along livestock migratory routes.
- Forage and livestock resource management in managed ecosystems such as community grazing lands and ranches.

The TAP also sees a need for labeled data that may be applicable to a range of needs or use cases, including, but not limited to:

- Agricultural practices (e.g. ploughing, harvesting, grazing, burning, irrigation, etc.).
- Soil health data and related labels.
- Pest and disease labeled datasets for all crop types (while widespread for certain staple crops, could be further broadened or made more context aware).

Project Timing – The TAP recognizes that the seasonality of agriculture affects when data can be collected. As noted above, the Lacuna Fund will consider projects that have both data labeling and/or collection aspects, recognizing that labeling of existing data could be completed on a shorter timeline.

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.

⁵ <https://www.cv4gc.org/cv4a2020/#ses11>

Proposal submissions will only be accepted through the 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

Your 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, the affiliated institution(s) ethical review processes, and the team's ability to gain national approvals.
- CVs for key team members.

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 may 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 and Proposed Solution** - Describe the problem or gap in labeled data and the proposed solution (dataset labeling, creation, augmentation, or maintenance).
- **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.
- **Theory of Change** - 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, ensuring that cellphone-collected data of pest imagery is georeferenced and timestamped could

increase its potential for future use in concert with remote sensing data. Note any practical constraints you may face (e.g. internet penetration).

- **Specifications and Deliverables for Proposed Data and Documentation** – Include the following:
 - Types and format of data and/or labels (must be time and location-stamped), as well as sample frame and size.
 - Proposed data collection and labeling techniques and information on planned sensor use or data interoperability with future sensor-collected data.
 - A plan to assess and mitigate error.⁶
 - 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.
 - Metrics to be used to assess desired outcomes of data creation. (i.e. fairness metrics in the dataset, QA/QC against a benchmark, etc.)
 - Any anticipated challenges or uncertainties in data collection and proposed countermeasures.
- **Data Management and Licensing** – Please describe:
 - Which metadata records will be captured and how they will support FAIR data principles. For EO data, metadata should be listed in the STAC standard.
 - Plans for licensing to maximize responsible downstream use. Per Lacuna Fund principles, 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.
- **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:
 - 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 and gender parity for field agents.
 - 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.
- **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

⁶ See Arthur Elmes et al., “Accounting for Training Data Error in Machine Learning Applied to Earth Observations,” *Remote Sensing* 12, no. 6 (March 23, 2020): 1034, <https://doi.org/10.3390/rs12061034>.

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.

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.

Proposed projects must be completed within 6-15 months from notification of award (notifications are planned for approximately October 2020).

Budget and Budget Narrative

The online application will require the applicant to upload a budget formatted in the Lacuna Fund budget template, available on the [application page](#) under “Open RFPs” and in the applicant portal. Budgets should be submitted in US Dollars. We are anticipating proposed budgets in the range of \$20k – 100k for small to medium-sized projects and as much as \$500k for large, complex projects. Note that 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;
- 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 costs must be limited to 13% of the proposal budget.*

See the instructions sheet in the budget template for further information on budget guidelines.

Budget Narrative: The budget template will also give space to state any budget assumptions or provide additional context.

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 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 organisation started operating as this entity.
2. Does your organisation 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:
 - (a) To the best of my knowledge, all information set forth in this anticorruption due diligence form is truthful, correct and complete;

- (b) 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;
- (c) 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;
- (d) 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;
- (e) 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;
- (f) 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.