Inclusive Climate Datasets for Artificial Intelligence

Lacuna Fund RFP Webinar

17 May 2022

Lacuna Fund

A lacuna is a gap, or a missing portion, of a manuscript or piece of music.

Lacuna Fund was formed to provide resources to data scientists in low- and middle-income contexts to fill the gaps in ML datasets needed to solve urgent problems in their communities.

Climate Funding Partners

Google.org

Federal Ministry
for Economic Cooperation
and Development

FAIR FORWARD

Artificial Intelligence for all.

The ROCKEFELLER FOUNDATION

Structure

- **Steering Committee**: Strategic direction
- **Technical Advisory Panels**: RFP design and proposal selection
- **Secretariat**: collaboration, convening, facilitation, communication, and logistics

Principles

- Accessibility
- Equity
- Ethics
- Participatory Approach
- Quality
- Transformational Impact

Funding

Open solicitation process (Requests for Proposals) in different domains:

- Agriculture
- Language
- Health
- **Climate**
- Other areas where the Steering Committee identifies a need

Lacuna Fund provides resources for dataset creation, expansion, or maintenance.

Requirements for Climate RFPs

Climate & Energy and Climate & Health

What We Are Looking For

- ML Datasets!
  - Collecting and annotating new data
  - Annotating and releasing existing data
  - Expanding existing datasets and increasing usability
  - Linking and harmonizing existing datasets

Proposal Components

- Applicant information
- Proposal narrative
- Project timeline
- Budget and budget narrative
- Due diligence self-reporting form

Key Dates

20 May: Question and Answer deadline

3 June: Answers posted

17 July: Full Proposals due

September: Technical Advisory Panel makes decisions on grant awards

October: Grant Awardees are notified

Elements of Strong Proposals

Elements of Strong Proposals

- **Multi-disciplinary team** - experience in data science, climate, and health or energy
- **Use cases and community engagement**
- Clear **problem statement** and how the proposed dataset or aggregation will help solve the problem
- Specificity about dataset size – should be of **sufficient size and quality** to be useful in future applications
- **Budget** is appropriate for the size of the dataset produced

Elements of Strong Proposals

- **Equity** considerations (gender, socio-economic status, ethnicity, etc.)
- Consideration and plan for potential **privacy and ethics** issues
- Plan for **data management and licensing**
- Working across areas where possible and relevant: time, diseases, regions, climate events
- **Budget** is appropriate for the size of the dataset produced

Community Engagement

- Describe previous consultation and/or proposed collaboration with intended beneficiaries.
  - When and where you have met/ will meet with partners
  - How partners will be involved in:
    - Identifying data needs
    - Data collection, labeling
    - Data governance
    - Dataset use
  - How partners will benefit from the new/expanded dataset

Privacy and Ethics

Explain how your team will address:

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.

Describe the process you will use to screen for potential ethical issues (e.g., an institutional review board, etc.).

Data Standards and Sharing

- **F**indable - easy to find on a public, widely used platform
- **A**ccessible - open format (CCBY 4.0)
- **I**nteroperable - data format
- **R**eusable - metadata

- Sustainable - plan for maintenance
- Shared – engaged community of users & plan for sharing dataset once complete

https://www.go-fair.org/fair-principles/

What relevant platforms are you seeing for storing and sharing data in your domain?

Please answer in the chat.

Budget and Allowed Costs

- Must be submitted in the Excel template.
- When in doubt, explain!
- Allowed costs include:
  - Training/capacity building;
  - Data collection; labeling; QA/QC or verification; and post-processing;
  - Licensing
  - Dataset hosting;
  - Open access publication of results;
  - Metadata or data statement;
  - Crowd-sourcing efforts, such as label-a-thons.

Climate & Energy Request for Proposals

Why Climate & Energy Datasets?

Goal: 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 & Energy Dataset Needs

- **Energy demand data**, at the building, community, or grid level.
- **Energy supply data**
- **Grid infrastructure location data**
- Electricity reliability, hazards, exposure and vulnerability, as well as fault detection data
- **Household-level survey data**, including socioeconomic information for consumer demand estimation.

Climate & Energy Dataset Needs

- **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.

Climate & Energy Dataset Needs

This is just a start!

The Technical Advisory Panel and Lacuna Fund welcome dataset ideas outside of the scope above.

Other Considerations

- Energy solutions to address locally-identified needs & appropriate to local environments and economic contexts
- Existing hardware/infrastructure preferred for cost efficiency
- Evidence of contact or existing partnership with utilities, operators, government agencies, or other partners where relevant
- Working with existing data encouraged – aggregating or making public

Climate & Health Request for Proposals

Why Climate & Health Datasets?

Goal: Open and accessible datasets for machine learning applications to understand and mitigate climate harms on health in Africa, South and Southeast Asia, and Latin America.

Climate & Health Dataset Needs

- Linking **vital registration data** to climate data
- Collecting and linking **disease vector and outbreak data** with climate data
- **Non-communicable diseases** and climate change data
- **Extreme weather events** data for early warning systems
- Aggregating Water, Sanitation, and Hygiene (**WASH**) system data with climate datasets

Climate & Health Dataset Needs

- Climate **impacts on health**
- **Mental health effects** of climate change, and data on potential interventions
- **Health co-benefits** of climate action data, linking mitigation measures and health gains
- **Gender-disaggregated data** of climate harms on health
- **Indirect impacts** of climate change on health (e.g., related to food security, migration, mobility)

Climate & Health Dataset Needs

… and many more!

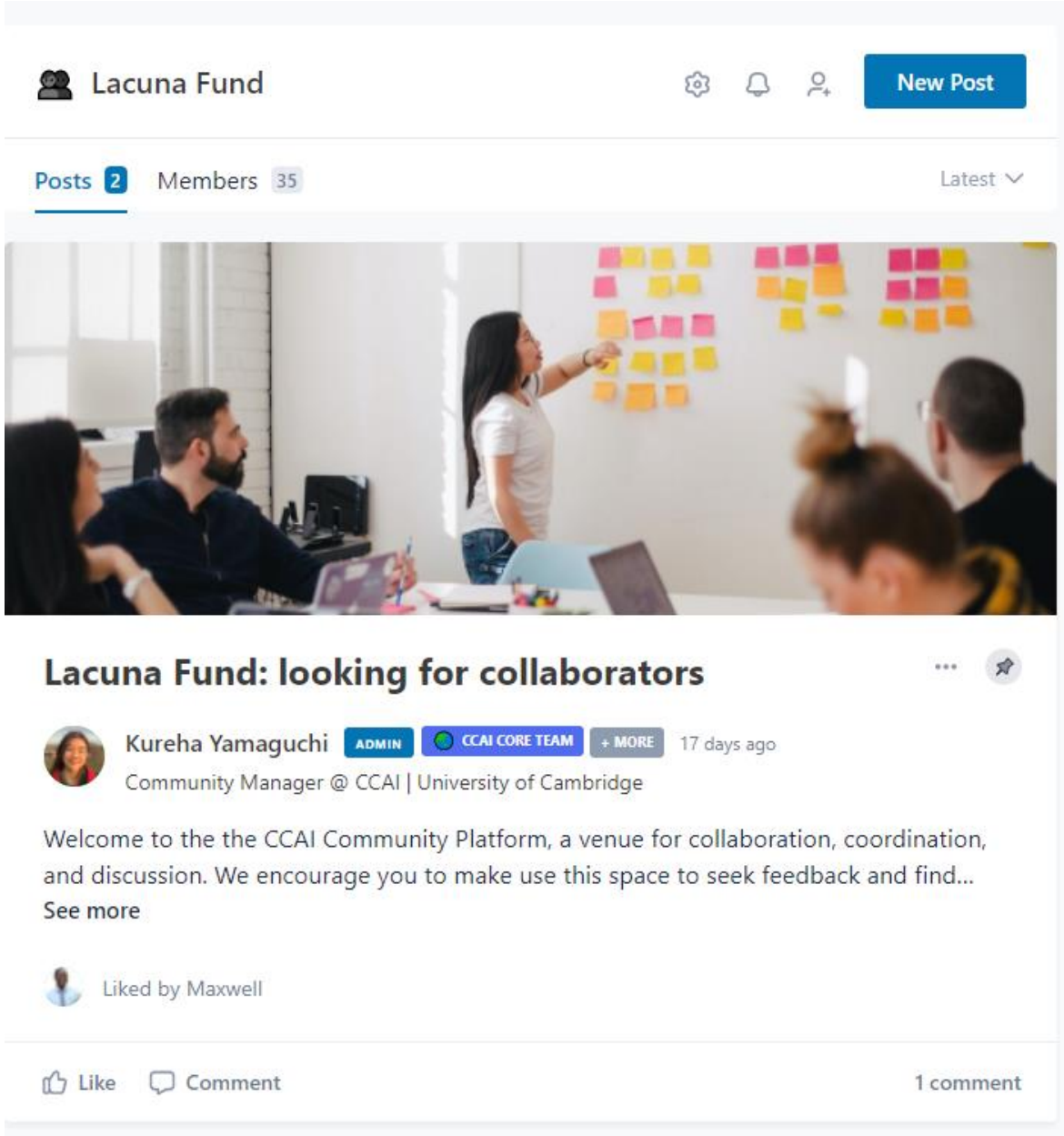
These are just a few ideas and key points. The TAP and Lacuna Fund welcome proposals outside these topics.

Other Considerations

- Ethical, privacy and access considerations on health data
- Leverage existing resources, link available datasets on climate factors and health outcomes
- Strong collaboration and contribution from data scientists, health professionals, government agencies, and/or the meteorological field on project teams is encouraged

Online Community Platform for Collaboration

Presented by Climate Change AI



Please find more information at

https://www.climatechange.ai/calls/lacuna

Questions?

Next Steps

- **Submit additional questions to secretariat@lacunafund.org by 20 May**
- **Answers posted** publicly on Lacuna Fund Apply page on **3 June**
- **Proposals due on 17 July**