

RFP: Datasets for Language in Sub-Saharan Africa

Questions and Answers

Note: This document was revised on 28 October to correct an error in Question 11.

0.Q	Who will be the reviewers? Are they drawn only from TAP, or do we need to nominate people in the proposal form?
0.A	The Technical Advisory Panel members, with support of the Steering Committee and other partners, will review proposals. You do not need to nominate reviewers in the proposal form.
1.Q	Can a proposal address multiple datasets, for example, if they are related?
1.A	<p>Yes. The scope of a proposal should be scientifically coherent and commensurate with the use cases, in NLP and for downstream applications, that motivate the creation of the dataset(s).</p> <p>For example, a proposal including datasets in multiple languages to improve the accuracy of a specific NLP task in a variety of contexts would be acceptable. Similarly, a collection of datasets for several NLP tasks in a particular language or context that would jointly enable a use case with real-world impact would be acceptable.</p> <p>Proponents may also submit more than one proposal in the case of multiple unrelated projects. Please also see Question 11 on additional considerations for submitting larger proposals. The Technical Advisory Panel may elect to fund any proposal in whole or in part.</p>
2.Q	Any ideas on size of datasets?
2.A	The proposal narrative (See Section 4: Proposal Requirements) asks proponents to discuss their proposed dataset, including dataset size, in the context of proposed use cases for the data. Proposals should ensure that the size of the proposed dataset enables valuable analysis of the phenomena captured in the dataset and/or sufficient training and evaluation for ML models.
3.Q	<ol style="list-style-type: none"> 1. What is the total pool of funding? 2. How many projects are you looking to fund?
3.A	The number and total cost of projects funded will depend on the quality of proposals. The Fund will aim to fund a range of projects across geographies and scales.
4.Q	Does a lead applicant need to be a data scientist? I am an AI ethicist and international

	project manager and can assemble a team that can fulfill the development of a data set. Can I still apply?
4.A	No, the lead applicant does not need to be a data scientist. The team should have the requisite expertise to complete the project. See the eligibility criteria in the RFP and “Quality” under the Evaluation Criteria for further details on how qualifications will be evaluated.
5.Q	Can you please further elaborate on the issue of ethical approval? Does it have to come with the proposal?
5.A	Ethical approval does not have to be completed at the time of proposal submission. Proposals should demonstrate an awareness of potential ethical issues they may encounter, how they plan to address them, and any institutional or national processes they will need to undergo (See Section 4: Proposal Requirements). As part of the review process, the Technical Advisory Panel may flag ethics concerns to be addressed if the grant is awarded.
6.Q	Can we collect data from one source? In one community, you can have more languages. In this situation what can one do?
6.A	Data collection sources and methods should be technically sound and align with the ML use cases that the proposed dataset will enable. Collection sources and methods should also ensure that the data is representative, and to the extent possible, unbiased. See Question 1 for additional details on proposals addressing multiple languages or other related datasets.
7.Q	Can multiple submissions (all not necessarily in NLP) be made from the same research institution?
7.A	Yes. However, please note that this RFP is solely for training and evaluation datasets for African languages. See Lacuna Fund’s website and subscribe to our newsletter to stay up to date on future funding opportunities in other domains.
8.Q	Can institutions submit a joint application?
8.A	Yes. We encourage collaboration between institutions. However, only one “lead” institution will receive funds. Other entities should be listed as partner institutions and structured administratively as subcontractors.
9.Q	Can the team be composed of members from different institutions?
9.A	Yes. We encourage collaboration between institutions.
10.Q	What are the evaluation criteria for deciding which applications you will prioritize for funding?

10.A	Please find a full list of evaluation criteria on page 3 of the RFP .
11.Q	1. How do you define "large transformational projects"? 2. Any examples of what small/medium projects are vs large/transformational projects?
11.A	<p>The Fund and the Technical Advisory Panel have no strict definition of a “large transformational project.” However, in a broad sense, transformational projects should make an impact in data availability or equity that enables significant progress in the field of African NLP.</p> <p>As an illustrative example, a small dataset might be a dataset oriented towards a single NLP task in a single language, whereas a medium dataset might be either more in-depth or comprehensive in a single application or language or a dataset for several tasks in a single language. A transformative dataset might address a common NLP problem or enable advances across multiple languages or contexts. A transformative project might also create new annotation tools/platforms or community processes that could accelerate future efforts in the field.</p> <p>Please note that “transformational” projects (in the budget range of 100,000-250,000 USD) should clearly demonstrate their potential impact on the field, feasibility, and sustainability beyond the grant period to the Technical Advisory Panel.</p>
12.Q	Can one team propose addressing multiple datasets?
12.A	Yes. See Question 1 for guidance on multiple datasets. In such a case, the team should demonstrate that they have the bandwidth to carry out multiple projects at the same time, if selected.
13.Q	Are there certain qualifications that the 'lead researcher' must have?
13.A	No. See Question 4.
14.Q	Does 'research institute' include universities?
14.A	Yes. Please find a full list of eligibility criteria on page 3 of the RFP .
15.Q	Could you please explain what national approvals mean?
15.A	Conducting research in certain countries may require the approval of relevant national entities (e.e. from an education ministry if research is conducted in school or from a science and technology ministry for human subjects research). This approval is not required upon submission of the proposal; however, proponents should demonstrate awareness of approval processes they will need to undergo.
16.Q	Is there any preference or weighting for lead organizations based in Africa as opposed to a lead organization outside of Africa with "substantial" partnerships in Africa? In other words, how interested are you in developing this competence and organizations on the African continent?

16.A	<p>There is no preference. However, Lacuna Fund has a strong emphasis on a participatory approach and proposals will be evaluated based on this criteria (See “Evaluation Criteria” in the RFP).</p> <p>While we welcome proposals from both lead organizations based in Africa and lead organizations outside of Africa with "substantial" partnerships in Africa, all institutions should be engaged in the technical aspects of proposed work and share the knowledge or benefits that accrue from it.</p>
17.Q	While no end-to-end applications are supported, does Lacuna support some prototype verification type work for these datasets? E.g. sample models that make it easy for other people to use the datasets.
17.A	As long as the main purpose of model development is to assess the quality/utility of the developed datasets (i.e. validation), or further encourage work on the datasets, some budgeting for model development is allowed within the scope of a supported project.
18.Q	How is Lacuna thinking about “maintenance” of the datasets after the 1 year of funding?
18.A	Lacuna Fund does not host datasets, because we want to give flexibility to teams to host their data in a way that makes sense for those that would use the data. Proposals may include an item in budget for hosting in perpetuity. In addition, any ideas proponents have for incentivizing a community to continue to contribute/maintain a dataset would strengthen a proposal. See pages 4 and 8 of the RFP for further information about sustainability.
19.Q	Do you encourage collaborations with EU/US institutes working sub-Saharan Africa languages?
19.A	See Question 16.
20.Q	<p>1. We are a startup from [A North African country] and willing to participate, are we eligible?</p> <p>2. Is this RFP for all African languages or for only sub-Saharan African ones? If it is only for sub-Saharan African languages, can we participate as [North Africans] by creating a sub-Saharan African languages?</p>
20.A	<p>Please find a full list of eligibility criteria on page 3 of the RFP.</p> <p>The geographic scope of this RFP is limited to sub-Saharan Africa as defined by the World Bank: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups</p> <p>However, organizations outside sub-Saharan Africa may partner with organizations in the region on proposals.</p> <p>The RFP is focused on languages in sub-Saharan Africa, so the content of your proposal should therefore be addressing a dataset relevant to those languages.</p>

21.Q	Any thoughts on the Hausa language in general? Has there been any applicants or existing available datasets for Hausa yet?
21.A	Dataset creation for any language from sub-Saharan Africa, including Hausa, is within the scope of the call for proposals. The responsibility for surveying and understanding the existing dataset landscape for any particular language falls upon the proposers. Please see the “Resources” page on the Fund’s website for a non-exhaustive list of existing efforts.
22.Q	Will a recording of the webinar from 2 October be made available?
22.A	Yes, a recording of the webinar is available here .
23.Q	Further to the funding posted for Datasets for Language in Sub Saharan Africa, could you please let us know if applicant can re-apply for the RFP (either due to partnership change or for better proposal)?
23.A	Institutions may apply to more than one RFP from Lacuna Fund. Please note that the focus of this recently released RFP is on datasets for language, rather than agriculture.
24.Q	<ol style="list-style-type: none"> 1. Can one proposal address multiple datasets? What about addressing multiple languages at a time for a given dataset, for example POS tag dataset for multiple languages? 2. Can one lead apply for two proposal, if in case one proposal should address one dataset? 3. What are the allowed materials to be included in the proposal, machines included?
24.A	<ol style="list-style-type: none"> 1. Yes. Please see Question 1. 2. Yes, a lead can apply for two proposals. In such a case, the team should explain if they have the bandwidth to carry out both projects at the same time, if selected. 3. Application requirements are included in the RFP under Section 4. Materials beyond the proposal elements specified in Section 4 will be considered at reviewers’ discretion.
25.Q	Is it possible to have a single proposal for the same dataset in multiple languages? For example, can I submit just 1 proposal if I am creating a sentiment analysis dataset for 3 African languages?
25.A	Yes. Please see Question 1.
26.Q	<p>[International organization], through its [branch in Africa] has been working on NLP tools for languages in Africa with Academic Institutions in the region during the last 6 years.</p> <p>The grants offered by Lacuna Fund will give us the opportunity to improve and open resources that are the result of years of work.</p>

	My inquire is the following: is [international organization] eligible for a grant under the “language” call?
26.A	Yes, international organizations are welcome to apply. If they are not headquartered in Africa, they should be in a substantial partnership with such an organization.
27.Q	<p>Would you accept one application from a consortium of organizations and researchers? Theoretically, there would be different dataset proposals within the consortium proposal, and while there would be one lead organization (primarily to leverage their administrative capacity), there then might be several sub-grantees to this lead institution which would ideally be affiliated with the dataset proposal's PI and would therefore be the intellectual lead organization of each dataset proposal.</p> <p>Each dataset proposal would be independent in terms of implementation and would have a clear separate budget for its activities, such that if one was rejected on its individual shortcomings, we hope that this would not affect the chances of the rest of the proposals in the consortium application.</p> <p>Through such a structure, an organization with grant administrative capabilities might then lend this capacity to other organizations that have the technical capacity. And the consortium might benefit from shared legal and ethical research partners.</p>
27.A	<p>Yes. Please note that only one “lead” organization will directly receive funds and that partners should be structured administratively as subcontractors. Please see Question 1 for information on proposals for multiple related datasets.</p> <p>Given that such a consortium might constitute a large/transformational proposal (>100,000 USD), see also Question 11.</p> <p>In particular, the Fund welcomes proposals that provide a compelling rationale for a consortium or partnership approach over proposals for each element of the partnership taken individually.</p>
28.Q	<p>1. Can we apply with more than one project? For instance, apply for different languages applying same methodologies in separate project applications?</p> <p>2. Can we apply with non-registered groups as partners?</p>
28.A	<p>1. Yes, see Question 24. See Question 1 for information on including related datasets in a single proposal.</p> <p>2. Lacuna Fund may only make grants to eligible lead organizations. Individuals or non-registered organizations should engage in the project through an institutional sponsor.</p>
29.Q	You mention eligible organizations to be non-profits, research institutions and for-profit social enterprises. My question is, what exactly qualifies an organization as a social enterprise and specifically does our organization qualify? We are a business but one of our primary goals is financial inclusion and towards that end we are developing models that

	<p>allow people with low English literacy to access digital financial technology in their own native languages. We want to collect and make available data that helps that purpose. Do we qualify to apply?</p>
29.A	<p>There is no formal designation for a social enterprise—all for-profit organizations with a mission related to social good are welcome to apply. Please review other eligibility criteria on page 3 of the RFP to confirm that you meet all other criteria.</p>
30.Q	<ol style="list-style-type: none"> 1. Can the proposal address more than one language? 2. Can we apply to address 2 languages in the response to the RFP? 3. For the parallel corpora, can we just use English as the mainstream language, hence the data will be eg English-[African language]? 4. Can we work on an unparalleled dataset, but the same domain for Unsupervised learning? 5. Can the staff salary be on a monthly basis instead of per hour /per day? 6. Should all the salaries be inclusive of taxes eg, a case where you will hire a consultant? 7. Could you provide an example of a filled Budget template? this will really be important to guide through the filling of the budget in the right manner 8. If the plan of collecting dataset includes incentivizing people to share the data and in return rewarding them with Phone credit etc., where should this cost fall? 9. Can office rent or communal space to execute the work from being indirect costs? 10. What are the different budgets (max and min) related to the different scope of the project? 11. Do we have at least a minimum size of the data required per language in case one is creating new data? E.g., number of sentences or words, etc. 12. Can we use GitHub for data storage after the project is over? Free repository?
30.A	<ol style="list-style-type: none"> 1. See Question 1. 2. See Question 1. 3. Proposed datasets should be suitable for the use cases included in the proposal. 4. Yes. Please see Section 3: Purpose and Need of the RFP, in particular: “Labeled and unlabeled text corpora for use as training data.” 5. Yes, it can. Please indicate the units in the budget.

	<p>6. Yes, they should.</p> <p>7. A sample budget will be made available on the Apply Page here.</p> <p>8. On the budget, this cost will fall under Other Costs.</p> <p>9. Yes, rent and communal space are considered indirect costs and would fall under the 12% indirect cost allotment.</p> <p>10. Please refer to the "Budget" section on page 9 of the RFP: "We are anticipating proposed budgets in the range of \$20k-100k for small to medium-sized projects and as much as \$250k 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."</p> <p>11. See Question 2.</p> <p>12. Proposals should include a sustainability plan, including a hosting solution that fulfills the "Accessibility" and "Sustainability" criteria included in the Evaluation Criteria section of the RFP.</p>
31.Q	<p>1. What is the position of the fund on outsourcing of certain aspects of the project activities to a competent service provider whose involvement will speed up the project?</p> <p>2. What is the budget limit for consultancy?</p>
31.A	<p>1. The Fund has no position. We recognize that proposals could be strengthened by the involvement of relevant consultants (e.g. linguists or legal specialists). However, proponents should review the principles of Lacuna Fund, including the focus on a participatory approach and ethics best practices, including fair labor standards. Proposal and partnership structures should reflect these principles.</p> <p>2. We do not have a budget limit for consultancies.</p>
32.Q	<p>1. How many projects are likely to be funded in the course of LACUNA FUND: Language RFP 2020?</p> <p>2. Is that ok if a datasets collection comprises training data for speech recognition and automation of morphological analysis; in other words, is it permitted to produce hybrid collection of datasets?</p> <p>3. Who is responsible for financial management of the grant? Is it the Principal Investigator or the financial department of the host institution?</p>
32.A	<p>1. See Question 3.</p> <p>2. Yes. See Question 1.</p>

	3. That determination is the responsibility of the proponent. The grantee is responsible for financial management and record retention (if funded).
33.Q	Does everyone on a joint application have to be affiliated with one of "non-profit entity, research institution or for-profit social enterprise" or just the lead applicant? Is there a fixed definition of what a "for-profit social enterprise" is?
33.A	See Questions 28 and 29.