

Lessons Learned – 2020 Grantees

August 2022

Grantees from the initial rounds of Lacuna funding gathered in August 2022 to share lessons learned and discuss the future of their projects. Grantees' projects focused on the Agriculture and Natural Language Processing (NLP) domains. We share their lessons learned to help future grantees and others working in these domains.

Natural Language Processing

Participants in the Natural Language Processing domain gathered to share successes, challenges, and lessons learned from their projects. Lessons will be shared with future grantees. Project representatives are encouraged to add to this list.

SUCSESSES

- Projects increased access to Natural Language Processing technologies and African language datasets across the continent!
- Project teams learned a new process that can be shared with other NLP experts in the future. African NLP grew a lot in the last three years – these projects created new processes and systems that can serve as the starting point for future projects and researchers.
 - The field will be ready for bigger projects now that these foundational projects are complete.
 - These projects built a portfolio for African NLP!
- Interdisciplinary collaborations with linguists were successful and an important opportunity to take advantage of. Linguists have a lot of data available but lack the data science/ML expertise to use the data.
- Teams appreciated Lacuna Fund's flexibility with revised timelines and project scopes as they learned what was possible in this first round.
- The NLP community was important for sharing knowledge and helping each other learn. The African NLP community has moved to self-assessment and producing impact.

LESSONS LEARNED

- Disbursing funds to translators, annotators, and other partners across the continent was challenging. Transparency, communication, and clear contracts with those receiving funds were key to navigating this challenge.
- A small budget for legal support was helpful in building contracts and answering copyright questions.

- Assigning tasks in smaller batches to more people improved the quality and reliability of the data.
- Some projects found it useful to have different people trained on different parts of the annotation process, so the process is done by a group of people.
- Hire professional translators to do any translation.
- Teams should be prepared to audit funds that were passed to other countries, but should not assume that they will all need auditing.
- Teams should include a research coordinator, business coordinator (to manage partnerships), data quality expert, and legal expert.
- Teams need strong monitoring for quality in data collection. For example, some translators may introduce bias into the dataset by using Google translate as a starting point for translation rather than translating from scratch.
- Teams learned a lot about a realistic quantity of data and timeline!
- Research teams should share the credit with the annotators and translators doing much of the work. The research community should support them with training, community building, and fair payment.
- Include time and budget for institutional approvals in the project plan. University processes in particular are complicated and time-consuming. Third-party fiscal sponsors may help streamline some of the approvals, but researchers employed at universities should consider whether that is the best solution in the long-term.

CHALLENGES

- Some teams experienced confusion with unpaid volunteers thinking they were being compensated when they were receiving something in-kind.
- Social media data is difficult to obtain. Lacuna Fund could consider building partnerships with key companies (e.g., Twitter) and coordinating with them to work with grantees on providing data and other support. In the long term, these relationships could lead to a bigger conversation, led by African researchers, with big tech about how African data is used and owned, and encouraging R&D to happen on the continent.
 - The African NLP community could draft licensing requirements for African researchers and institutions to maintain ownership of African language datasets, while also being represented and involved in development by big tech companies.
- African researchers need infrastructure (e.g., compute power) and capacity (e.g., more university programs and fellowships) to compete with big tech companies. Lacuna Fund partners could donate extra compute to support this goal.
- Projects needed speakers of each variety of a language in order to properly annotate, transcribe, or translate it.
- Tonal languages needed translators with that specific competency and a better transcription method that supports tonality.
- Translators based in areas with conflict (e.g., Tigray) became unavailable/offline with no notice. To mitigate this, the project only gave small quantities of tasks that could be completed quickly, then would assign more as each batch was completed.

Agriculture

Participants in the Agriculture domain gathered to share successes, challenges, and lessons learned from their projects. Lessons will be shared with future grantees. Project representatives are encouraged to add to this list.

SUCSESSES

- Community engagement was efficient, and community ownership made datasets more sustainable. For example, one project provided fish farmers with ponds and fish instead of financial compensation. Then, farmers learned when to change the water, cover the pond, etc.
- Trainings and workshops improved community engagement and ownership.
- Local partnerships and capacity building improved sustainability for datasets.
- Cameras mounted on cars/bikes (helmets) allowed for safe data collection at a large scale and the ability to expand to other regions.
- Private sector participation and ownership incentivized the open release of data.
- New partnerships created the capacity to process images.
- Community-based methods for community land use plans were successful.
- Starting with a pilot worked well.
- Generating data with sensors worked well.
- One partner went beyond the scope of their task, which improved accuracy.
- Historic data became valuable to one project.
- Including an agriculture expert/extension agent on the team was helpful. They could leverage agricultural institutions to create a two-way benefit for agriculture experts and farmers. The experts learned from the farmers involved in the project, and the farmers received more ag support services.
- Ability to monitor water quality.
- One project was for four local governments, and now the other 280 want the service too!

CHALLENGES AND LESSONS LEARNED

- Several teams faced challenges with cloud hosting. Images are large, the channels received in-kind were not enough to cover the project, and the teams will continue needing cloud hosting after the Lacuna Fund grant ends.
- Project management across multiple partners and countries was challenging. The teams improved their team management skills over the course of the project.
- Labeling takes time!
- Satellite imagery was not at high enough resolution. Planet only works in RGB and is not available everywhere on the continent. Some teams did not have the budget to purchase additional imagery.

- Ensuring the quality of sensors was challenging. Teams had to calibrate the sensors and confirm that they were in a standard range. It was suggested that the Wazi hub gives IoT devices, which is an avenue to be explored.
- Delays in the release of funds from banks due to COVID created time constraints on projects at the initial phase of the project.
- Power outages could be mitigated by investing in solar energy in advance.
- Using different models resulted in teams being uncertain about accuracy.
- Requests to scale up the projects could be challenging.
- Google Earth Engine was useful for verification.
- Data was sensitive to changes in land use. With the changes, the project will need to get government approval again.
- Photos were too granular to see the whole landscape. The team used a 5x5 box.
- One project used modelling to predict drought, but there were insufficient images with labels to monitor other risks (e.g., flooding).
- Training is key to ensuring the quality of crowd-sourced data.
- Administrative challenges included delays due to contracts and the expense of shipping equipment. The equipment expenses can be improved if equipment is distributed when people attend workshops together.
- Geopolitical and security concerns were mitigated with post-season instead of in-season labeling.
- Getting historic data to feed into the models would create more accurate models, but obtaining historic data is challenging.