

How to Use this Playbook

This playbook is for anyone who wants to collect voice data in an African language, whether the project is:

- Small — 50-100 speakers
- Medium — 500-1,000 speakers
- Large — thousands of contributors
- Academic
- Community-led
- Government-led
- NGO-led
- Commercial
- Open-source
- Designed for ASR/TTS
- Intended for language documentation
- Intended for AI research

It is designed to answer a simple question:

“If I want to collect voice data in my language, what exactly do I need to do?”

The playbook takes you through the entire lifecycle:

“

Plan ? Design ? Prepare ? Mobilize ? Recruit ? Consent ?
Train ? Record ? Validate ? Transcribe ? QA ? Pay ?
Document ? Release ? Sustain

Voice collection is not simply a recording exercise. It is an operational system involving people, language expertise, community relationships, technology, quality control, ethics and data management. The AfriVoices-KE project organized these functions through language leads, resource persons, mobilizers, contributors, validators, transcribers, super-reviewers and administrators.