



Language Documentation through Data Collection and Audio Recording

Kambai Akau

2 October | Zwat Swak | Antɔŋ | Oktooba | Kele | Ahinime 2026

13:30–13:50 (UTC)

Plenary room (Afzelia 2)

Content

- Why documentation of languages and their variants?
- What tools may be needed?
- Where may one host the data to be collected?
 - Wikidata
 - Wikimedia Commons
- How to get started
- The Wiki Audio Walk (WAW) Campaign Case Study
- Want to see more about the WAW campaign?



Why document languages and their variants?



Key Advantages

- Safeguarding knowledge.
- Uncovering broken links between dialects.
- Retrieving lost words.
- Assisting even growth of language variants and retention of identity.
- Reduce low self-esteem among speakers of undocumented variants.
- Providing resources for researchers.



What tools may be needed?

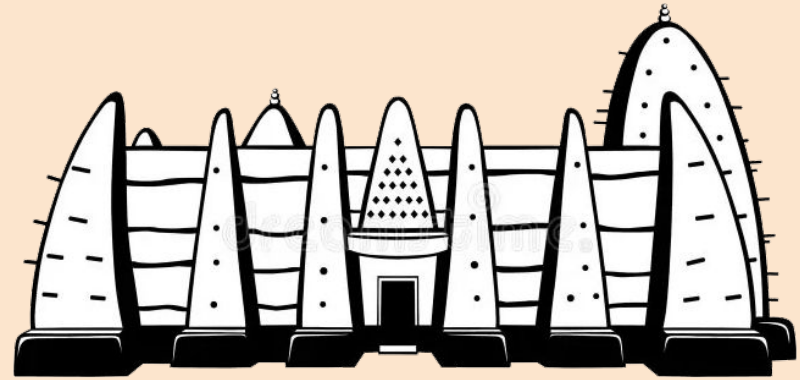


Basic Requirements for Data collection

- Recorder/mobile device
- Mic
- Writing materials
- Knowledgeable speakers
- Hosting website

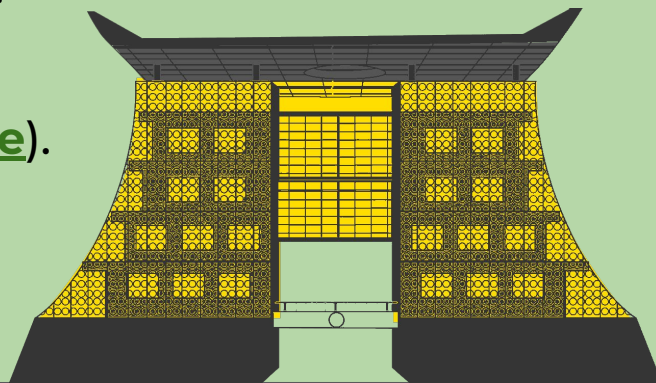


**Where may
one host the
data to be
collected?**



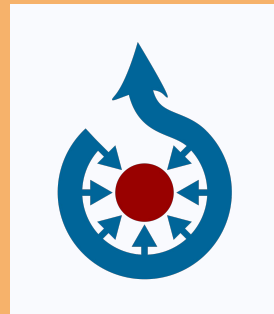
Wikidata

- Multilingual human and machine-readable [open-knowledge base](#).
- Structured data central storage for Wikimedia projects like Wikimedia Commons, Wikipedia, Wiktionary, etc.
- Can store lexemes of [enabled languages](#).
 - Search if the language is supported by language family [here](#) on Translatewiki.
 - Request for language to be enabled [here](#).
- Create Wikidata lexeme [here](#).
 - See examples [here](#) (more specifically, [here](#)).

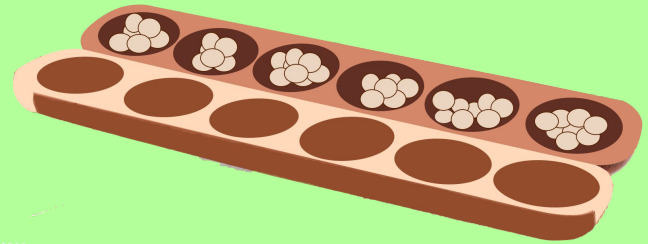


Wikimedia Commons

- Multilingual open-source media file repository.
 - Supports audio files, videos, images, textual files, and other media files of certain types only.
- Uploaded files are free licensed.
- Upload files directly via:
 - Wikimedia Commons' Upload Wizard, or
 - Lingua Libre
- Welcome to journey with us under the Wiki Audio Walk (WAW) campaign banner.



**Want to See More
About the Wiki
Audio Walk
Campaign?**



Wiki Audio Walk on Wikimedia Commons



Image by Friday musa. Source:
[Wikimedia Commons](#). Licensed under
CC BY-SA 4.0.



See What We do **Here!**

- View our 2025 Campaign **here.**
- View our 2026 Campaign **here.**

Contact: **tyapwikimedia@gmail.com**

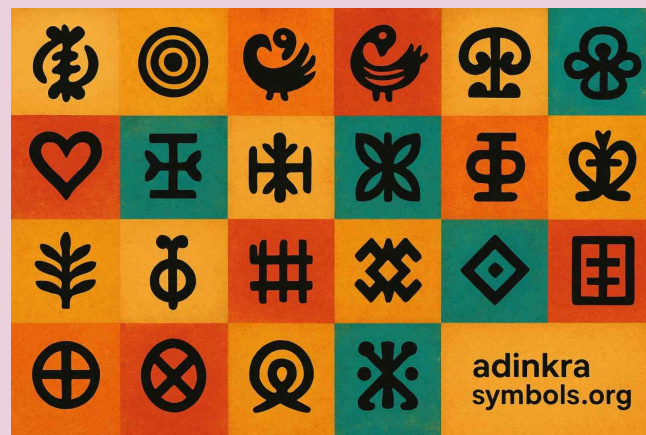


Take Home

- Every language is a library.
- Every dialect is a bookshelf.
- Every data documented is a memory saved.
- Every voice recorded today becomes knowledge that future generations can still hear.



Q&A



N gwai!
N payiya!
Akpe na wo!
Oyiwala donn!
Meda wo ase!
Thank you!

Kambai Akau
<https://linktr.ee/tlkambai>
camylevsky@gmail.com

