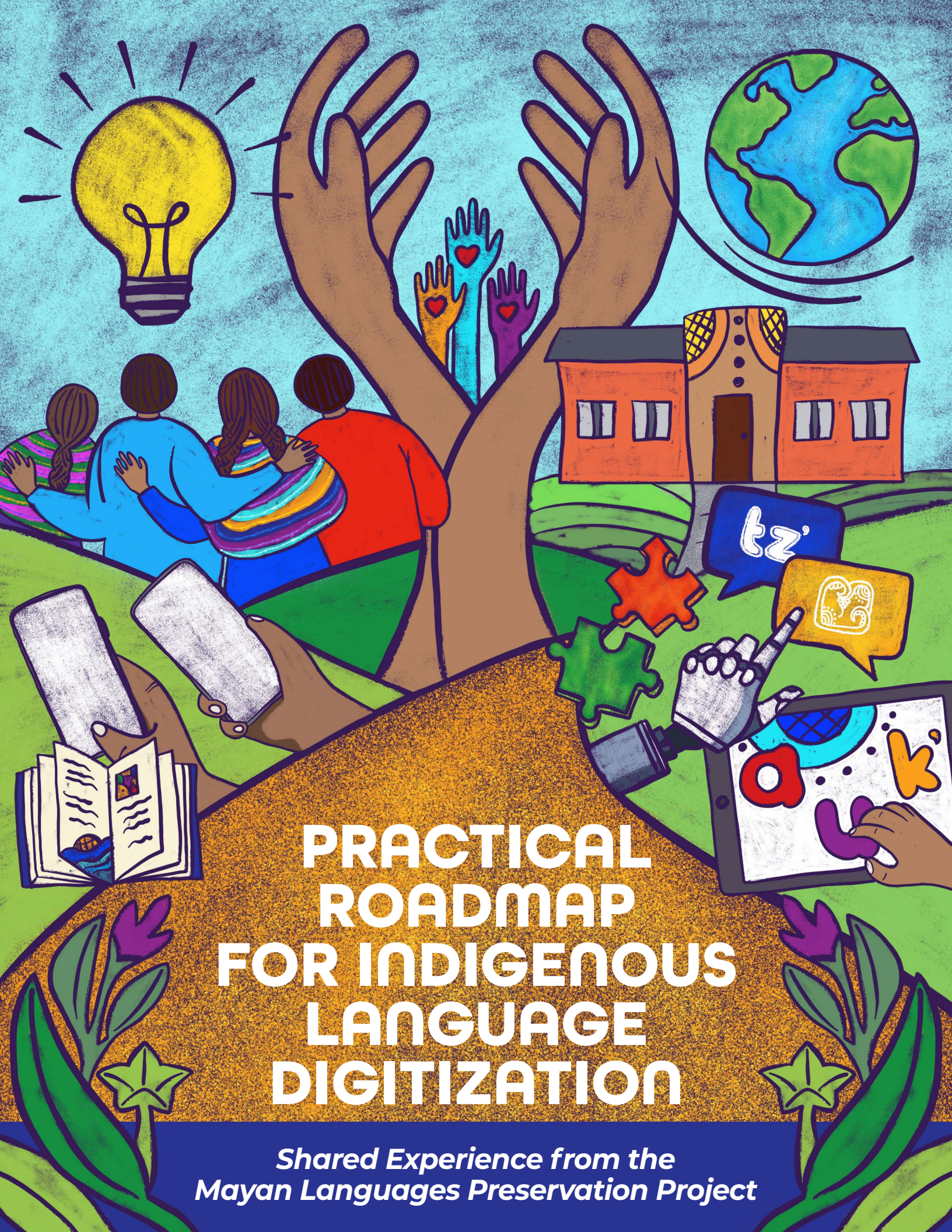


PRACTICAL ROADMAP FOR INDIGENOUS LANGUAGE DIGITIZATION

Shared Experience from the Mayan Languages Preservation Project

Acknowledgements

This Roadmap grows out of the preservation and digitization efforts of the Mayan Languages Preservation Project (MLPP). We are deeply grateful to the team at MasterWord Services and to the company's CEO and President, Ludmila Golovine, for sustained support and collaboration. Their staff, volunteers, and interns worked alongside our language professionals to develop a practical, real-world example of how Indigenous languages can be digitized in ways that honor community leadership.

We also thank the many departments at UNESCO and the International Decade of Indigenous Languages for their support and collaboration, and for inviting us to share the outcomes of our first two years of work. In February 2025, UNESCO and the Language Technology for All (LT4All) conference provided space for our team to present in Kaqchikel. This was the first time a presentation at a UNESCO event was delivered entirely in an Indigenous language, and it marked an important step toward greater visibility for Indigenous languages in global forums.

To our many team members, board of advisors, and institutional and technical partners, we offer sincere thanks. You helped us navigate shared challenges and kept us focused on a common goal: expanding safe, equitable digital spaces for Indigenous languages.

Most of all, we acknowledge our Mayan language community partners and language activists. Your time, your vision, your guidance, and your care for your linguistic communities shaped every part of this work. Your efforts in documenting, teaching, and revitalizing your languages are the foundation of this Roadmap and a guide for others who wish to protect their own languages and futures.

On behalf of the Mayan Languages Preservation Project,

Winston K. Scott, PhD
Director, Mayan Languages Preservation Project

Roadmap coordination: Ryan Foley
Design: Marla Garcia

In partnership with: MasterWord®

In support of: 2022-2032 | INTERNATIONAL DECADE OF Indigenous Languages

Who is this Roadmap for?

This Roadmap is for Indigenous communities who want to bring their languages online with little money, few formal contacts, and no promise of tech company or government backing. Maybe you are a language champion, teacher, interpreter, student, Elder, or parent who already works between languages every day. Maybe you are one person trying to figure out where to start.

We wrote this with you in mind. We share what worked for us, what did not, and how we kept going anyway. Our hope is that our experience helps you move faster, avoid some mistakes, and stay in control of your language and data.

WHY DID WE WRITE IT?

Across the world, Indigenous languages are losing reach in daily life, especially in digital spaces. Phones, computers, and online services often support only a few dominant languages. When your language is missing from those systems, people are pushed to switch languages to study, work, seek medical care, or access government services.

Digitization can help change this. When communities can type, read, listen, and create in their own languages online, they gain better access to information, services, and opportunities. We developed this Roadmap during the UNESCO International Decade of Indigenous Languages as one small contribution to a much larger movement for equal digital access for marginalized languages.

OUR STORY: STARTING FROM "ALMOST NOTHING"

In Guatemala today, there are 22 Mayan languages, and around 40 percent of the population identifies a Mayan language as the language they speak and understand best. Yet online, Mayan languages are almost invisible. There were almost no keyboards, no spell-checkers, very little news or public information, and few safe spaces to use these languages on the internet.

Faced with this gap, a small group of Mayan language professionals decided to start at the very beginning. We began our digitization efforts by using smartphones, a simple wiki, and community knowledge to build three talking glossaries. From there, the work grew into developing digital keyboards, a shared glossary platform, and early steps toward machine translation and text-to-speech in Mayan languages.

If your language also feels "almost invisible" online, you are not alone. Our situation may be different in the details, but the feeling of being left out of digital spaces is shared by many communities worldwide.

WHO ARE WE?

We are the Mayan Languages Preservation Project (MLPP). We started in May 2023 and joined in supporting the UNESCO International Decade of Indigenous Languages. Our work focuses on the more than seven million people who speak one of the 22 Mayan languages in Guatemala.

MLPP was created by and for Mayan language speakers in Guatemala and in the diaspora. Our team includes interpreters, translators, teachers, students, Elders, community leaders, and activists. Together we collect, review, and digitize glossaries, tools, and educational resources in languages such as Q'eqchi', K'iche', Kaqchikel, and others.

We are guided by a small steering group of Mayan language experts, community partners, and sponsors. Our project director is Dr. Winston K. Scott, PhD, a cultural anthropologist who has worked with Mayan communities for three decades. Advisors such as Kaqchikel author and teacher Baldomero Cuma and community Elder and language professional Misael Itzep help keep the work rooted in community needs. Our shared goal is to build safe digital spaces where Mayan languages can be used for learning, creativity, and everyday life, while communities stay in control of their knowledge and data.

"It gives my soul immense joy to see our linguistic communities inspired to preserve its language, culture, and traditions and to see institutions honor and recognize our work. This project is an example of how people working collectively toward a common cause may work seamlessly and naturally as we recognize and promote or languages and honored worldview."

– Kawoq Baldomero Cuma, Chair, MLPP Board of Advisors

Visit us at
www.mayanlanguagespreservation.org

How to use this Roadmap

WHAT YOU'LL FIND

This Roadmap gives you a clear, step-by-step framework with simple checklists and case examples from the Mayan Languages Preservation Project (MLPP). Each step points you to short, practical tools such as "how-to" guides, picture guides, checklists, and templates. You can find these on the Roadmap Resources page on our website using the QR code or link on this page.

You can read the Roadmap from start to finish or jump to the parts that match what you are working on right now.

WHAT'S DIFFERENT HERE

If you already know about the Zero to Digital resources developed by Translation Commons, you can think of this booklet as one concrete story of how a community put many of those ideas into practice with limited money, limited internet, and technology that is readily available in rural communities.

This Roadmap sits beside it as a "boots-on-the-ground" companion. It shows how one group of Mayan language communities used those principles to build talking glossaries, keyboards, a glossary platform, early machine translation tests, and text-to-speech, while protecting community ownership of the data.

WHAT WE MEAN BY "BOOTS ON THE GROUND"

When we say, "boots on the ground," we mean real work in real conditions that go beyond theoretical concepts to practical use. Our team:

- Started with grammar books, dictionaries, and personal wordlists already in the community
- Used phones, low-cost laptops, and shared spaces for recording and testing
- Built tools that people could use right away in schools, families, and community organizations

At every stage, Mayan language professionals and community leaders decided what to build, in what order, and why.

START HERE: DECISION GUIDE

Use this simple guide with the QR code or link to the **Digitization Roadmap Resources** page:

- **If you need ethics or consent forms or help with data agreements**
  Go to **"Ethics, Consent & Data Governance."**
  There you will find sample consent scripts, simple forms, and examples of community-led agreements.
- **If you are building or improving a keyboard so people can type in your language**
  Go to **"Digital Keyboards."**
  There you will find notes from the MLPP keyboard builds, plus tips you can share with developers.
- **If you are recording stories, songs, or other audio**
  Go to **"Recording & Asset Management."**
  There you will find guidance on planning recordings, naming and storing files, and keeping control of how your recordings are used.

You do not have to use every step or every tool. Take what fits your community, adapt it, translate it into your language, and add your own experience.

https://mayanlanguagespreservation.org/roadmap-resources

Contents

Guiding Principles: Community Leadership, Ethical Collaboration, Data Sovereignty

COMMUNITY LEADS

Speakers and community professionals decide what to build, in what order, and why. Interpreters, translators, teachers, Elders, and youth know the real situations where language access fails, such as hospitals, courts, and schools. They also know which terms are missing, which tools people can actually use, and which stories should never go online.

In this Roadmap, community leadership means forming small advisory groups, listening before acting, and checking in often. For our project, that means that Mayan language professionals advise on priorities, choose which tools to develop first, and define what "success" looks like for their communities, not for outside institutions.

Example: Interpreters report that certain legal or medical terms cause misunderstandings every week. The advisory group then decides that a talking glossary for those terms should come before more advanced tools like machine translation.

ETHICS UP FRONT

Ethics come before technology. People who share stories, words, or songs have the right to decide what is recorded, how it is used, and who can hear it. Many communities have specific rules about sacred stories, seasonal songs, or knowledge that should only circulate within families or certain groups. Past harms where outsiders took recordings or texts without consent make this even more important.

In this Roadmap, ethical collaboration means using clear, simple consent forms or scripts in the community's preferred language, explaining options, and making it easy to say "no" or to change one's mind. It also means being ready to stop, slow down, or change direction if participants or community leaders raise concerns.

Example: A storyteller agrees to record traditional stories so that local teachers can use them in class or to train a Machine Translation (MT) model but does not want to post them publicly. The group decides together that these recordings will be stored in a community-controlled space and will not be posted on public video platforms.

SOVEREIGNTY IN PRACTICE

Language materials are more than data. Recordings, texts, glossaries, and other resources inform aspects of history, cultural identity, environmental stewardship, and kinship relationships. Data sovereignty means that the community keeps control over these materials, even when partners help with funding, equipment, or technology. The community decides where files are stored, who can access them, and whether they may be used to train tools such as speech recognition or machine translation.

In this Roadmap, sovereignty in practice means putting agreements in writing, choosing storage options that match local comfort and capacity, and saying "no" to extractive projects that want language data without real benefit to the linguistic community or shared decision-making. When new tools are developed, communities have a voice in how models are trained, where they are hosted, and how others may connect to them.

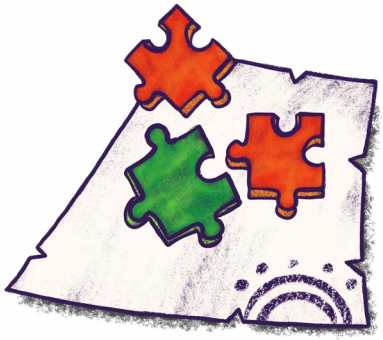
Example: A university offers to host copies of recordings on its servers. The community signs an agreement that keeps ownership with the community, limits how the university can use the files, and requires written permission before any new projects can use those recordings.

https://mayanlanguagespreservation.org/roadmap-resources

STEP 1

Identify Challenges and Listen

▲ **Indigenous teacher, Mérida Hún Caal**



START BY LISTENING

Before choosing tools, start with conversations. Sit with teachers, parents, youth, interpreters, Elders, and community organizers. Ask them what makes it hard to use the language in daily life, especially when they read, write, or go online.

In Guatemala, there are 22 Mayan languages, and a large portion of the population speaks a Mayan language as their primary and preferred language. Yet when people want to read, type, search the internet, or find entertainment, they are often required to use Spanish. Keyboards do not exist for most Indigenous languages. There is little or no news, subtitles, or spell-check capabilities in Mayan languages. Many everyday activities require proficiency in speaking and reading in a colonial language.

Listening to these experiences is the real starting point for any digital project. It helps you see the gap between how much your language is spoken and how little it appears in digital spaces.

ALLOW COMMUNITIES TO DEFINE THE CHALLENGES

Use your conversations to map the main barriers. Keep the language simple and concrete. For example:

- "We cannot type our language on our phones or computers."
- "There are no news stories or videos with subtitles in our language."
- "Children can only play games and watch shows in another language."
- "We cannot spell-check documents in our language."

Group these barriers into themes that make sense to your community. You might group them as "typing and keyboards," "media and content," "schools," or "public services." The key point is that community members decide what matters most.

This also sets up good governance. Governance means that the community leads decisions from the first step. Listening sessions are part of that. They show that people are owners of their knowledge and that they will guide what belongs online and how it should be used.

WHAT YOU DO

- Invite a small group that reflects your community's demographics: language professionals, teachers, youth, parents, and community leaders.
- Ask open questions such as "When do you wish you could use our language on phones or computers?" and "Where do people still have to switch to another language?"

TRY THIS

For your first meeting, you can ask three simple questions and write the answers on a large piece of paper where everyone can see them:

1. "Where do you wish you could use our language but cannot?"
2. "What makes it hard when you try to read or write our language with current tools?"
3. "If we could fix one thing in the next year, what should it be?"

The answers to these questions become your roadmap.

- Write down each challenge in simple words, on paper or a shared screen.
- Look for patterns: which problems appear repeatedly across groups and locations.
- Document every win that might be accomplished with low-cost tools, and bigger challenges that will take more time and the cooperation of partners.

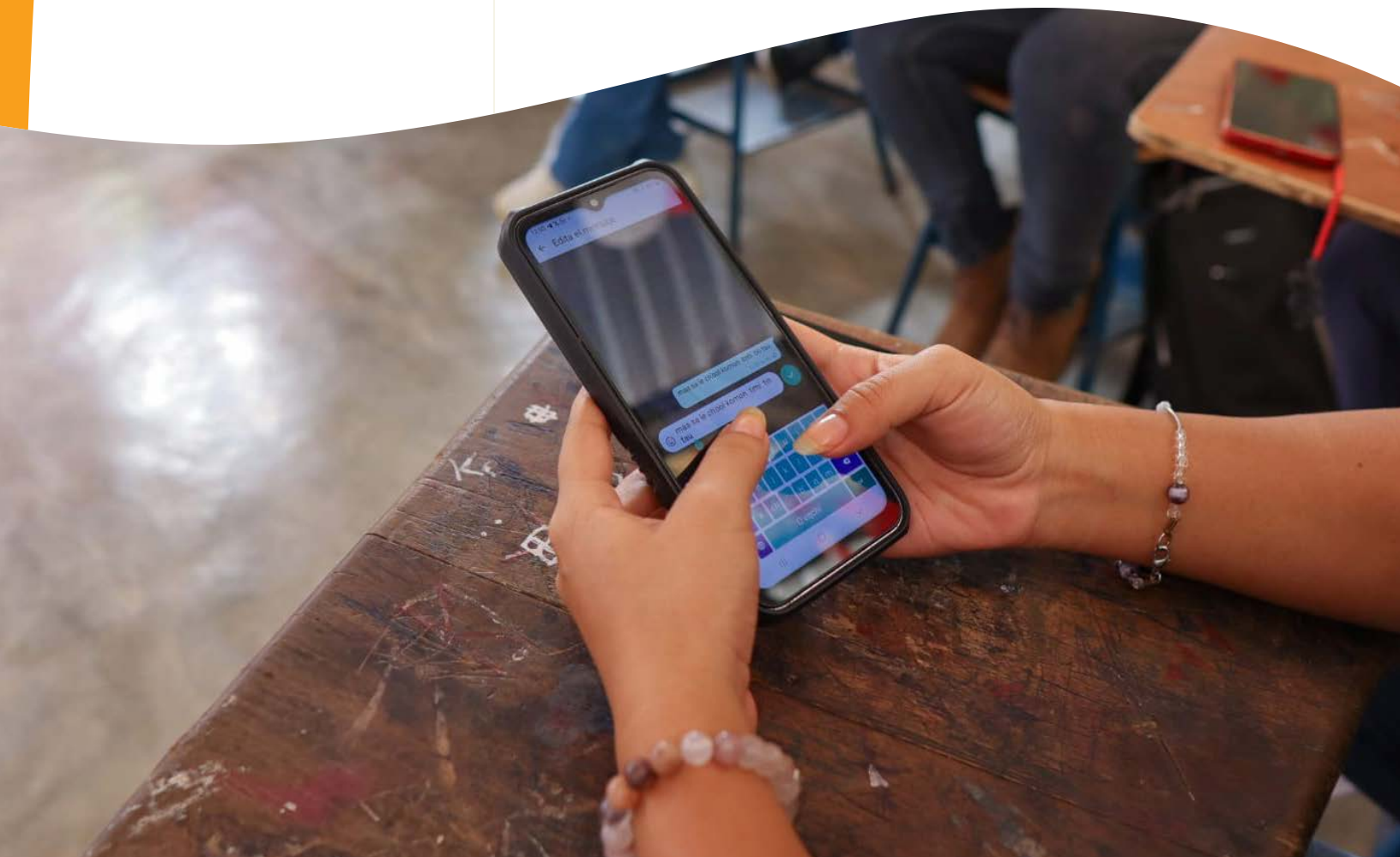
LESSON LEARNED

In our work with Mayan languages, listening sessions showed that low literacy and the cost of devices shape almost every design decision. Many people access the internet only through shared or low-cost Android phones. Many community members are stronger in speaking than reading.

Because of this, we made two early choices:

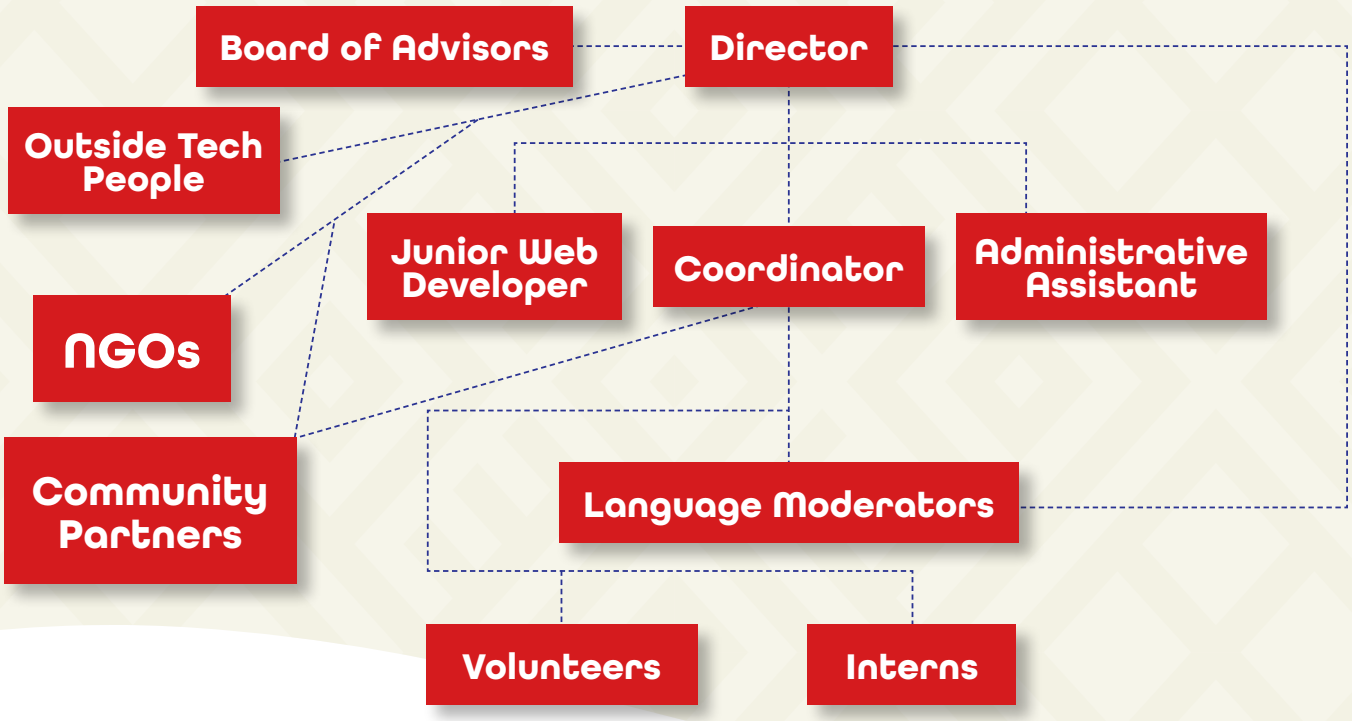
- We prioritized full audio for every glossary entry so that people could listen, not only read.
- We focused first on Android keyboards so that people could type in their language on the phones they already use.

Starting with listening made later community adoption of new tools like keyboards much easier. People saw their own needs and stories reflected in the tools. They already understood why certain features had been chosen and felt more comfortable suggesting changes.



STEP 2

Assemble a Team



Digitization is not a solo project. Even a small team can do a lot if people have clear roles, shared standards, and support to learn new tools. In MLPP's experience, a collective community-led effort was the only way to move from an idea in a meeting to real glossaries, keyboards, and a living website.

WHO'S AT THE TABLE

Start with a small core team and grow as the work grows. In our case, the first group was only a few people. Within two years, it became a network of employees, and more than a hundred regular volunteers.

For your own project, aim for roles like these:

- **Advisory board of native language professionals**
  Interpreters, translators, teachers, and community leaders who represent different regions and experiences. They help set priorities, decide which tools to build first, and keep the work aligned with community values.
- **Project director or coordinator**
  Someone who keeps track of tasks, meetings, and timelines, and who makes sure people have what they need to contribute. This person often acts as a bridge between technical partners and community members.

TRY THIS

To setup your own team, you can:

1. Invite 5–10 people who already support your language in different ways (interpreting, teaching, organizing, youth work).
2. In your first meeting, agree on your shared goals for the next 6–12 months and write them down in simple language.
3. Decide on one shared standard, such as which spelling system you will use, and add it to your written goals. Where regional variations occur, discuss adding notes to highlight and celebrate these variations.
4. Choose one person to act as coordinator for three months and support them with regular check-ins.

These early decisions will make every later step more clearly led by the community.

- **Language collaborators and team members**
  Speakers who suggest terms, review translations, record audio, take photos, and share feedback. Many may already work as interpreters or teachers; others may be students or community organizers.
- **External collaborators**
  People from outside the language community, such as technical developers, interns, or university partners. Invite them to collaborate when they are ready to follow community leadership and respect data sovereignty.

From the beginning, agree on shared language standards. MLPP used the orthographies of the Academia de Lenguas Mayas de Guatemala (ALMG) so that spellings would match across pages and tools.

TRAIN AND DOCUMENT

Once your team is in place, make it open and engaging for people to participate. We chose an open platform that community members could learn with practice. To help new contributors, the team:

- Created short video tutorials in multiple languages
- Showed how to suggest new terms, upload audio, and add images
- Held regular online meetings to answer questions and listen to ideas

These trainings did more than teach how to use software. They helped people feel confident, see that their contributions mattered, and understand how the pieces fit together. Over time, contributors moved from “I am afraid of breaking the website” to “I know exactly where to add this new term.”

Write down simple instructions for common tasks, such as:

- “How to add a new term”
- “How to record and upload audio”
- “How to correct a spelling”

Keep these guides short, with screenshots or pictures when possible. Link them on your project website so new team members can find them quickly.

LESSONS LEARNED

We learned three key lessons in this step:

1. **Start small, then grow.** A small, committed advisory group can set a clear direction. New contributors can join as specific needs arise.
2. **Agree on standards early.** Shared orthography and guiding principles help keep hundreds of pages consistent, even when many people are contributing.
3. **Invest in training.** Short videos, patient coaching, and open meetings build trust and confidence. This makes it easier to move later to more advanced platforms and tools.

https://mayanlanguagespreservation.org/roadmap-resources

STEP 2A

Get Started with Talking Glossaries

Talking glossaries are a simple way to turn your team's knowledge into a digital tool that people can actually use. A talking glossary is a set of terms where each entry has:

- the term in your language
- a clear explanation
- an example sentence
- audio so people can hear it
- and, when possible, an image

For MLPP, our first three talking glossaries became the base for many later categories and tools.

PUT COMMUNITY NEEDS FIRST

Start where people feel the most pressure to switch languages.

We began with terms from medical, legal, and social services, because interpreters and community members reported daily misunderstandings in clinics, courts, and public-service offices. The team gathered definitions from existing professional glossaries and then created written and oral definitions and example sentences in each Mayan language.

See our glossaries at https://mayanlanguages.wiki

TRY THIS

To start your own talking glossary:

1. Choose one high-need area. For example: hospital visits, school enrollment, or social-service appointments.
2. Ask interpreters, teachers, or Elders for a list of 20–30 terms they see or hear all the time in that area.
3. For each term, collect:
   - a simple definition in your main contact language and in your language
   - at least one example sentence
   - an audio recording of the term and example
   - an image, if it helps explain the idea
4. Invite others to add categories that feel important, such as foods, kinship, or local places.
5. Make it fun. You can hold a small "term day" or friendly contest where people suggest new terms for a chosen category, then vote on which ones to add first.

These first talking glossaries do not need to be perfect. They are a place to gather the words your community already uses and to practice adding audio and visuals in ways that make your language feel alive in digital spaces.

For each term, they aimed to include:

- a simple definition
- one or more example sentences
- full audio for the term, the definition, and the example
- an image where it helped explain the idea

Full audio and visuals were especially important because many speakers of Mayan languages did not have the chance to complete formal schooling. Listening and looking together made the glossaries usable for people of many ages and literacy levels.

GROW WITH COMMUNITY INPUT

Once our first glossaries were in place, ideas started to grow. Team members and volunteers suggested new categories that mattered to them and their families.

They moved from only "essential services" to include terms related to:

- favorite foods and local dishes
- animals and nature
- numbers
- kinship
- school and classroom
- civic participation
- and language where loan words were encroaching

These conversations did more than help us add terms. They opened space to talk about where loanwords from another language felt helpful and where they felt like a loss. As a team, they agreed that adding a category meant committing to preserving those terms.

Over less than two years, the number of categories grew from three to around thirty, each rooted in what community members felt was important.

LESSONS LEARNED

From this step, we learned:

- **Start with urgent domains.** Beginning with medical, legal, and social services built trust because people could see direct benefits for real-life situations.
- **Design for listening, not just reading.** Full audio and images made the glossaries useful even for people who are stronger speakers than readers.
- **Use fun to build unity.** Friendly contests and games, like asking people to suggest favorite foods or animals for new pages, sparked creativity and strengthened ties across regions and languages.

https://mayanlanguagespreservation.org/roadmap-resources

STEP 3

Digitize (Get the First Data In)

▲ **Henry Per and Elizabeth Choc Maas (recording audio files for Kaqchikel and Q'eqchi')**

To build any digital tool, you need data: words, sentences, stories, and recordings in your language. This step is about getting the first good data in, using what you already have and what you can record now.

WORK WITH WHAT EXISTS

Start with materials that already exist:

- dictionaries and grammar books
- school materials and community wordlists
- notes from interpreters and teachers
- open online repositories such as GitHub

Use these as a base for your first digital files. You can:

- type out key terms and example sentences
- take photos or scans of short sections you are allowed to reuse
- use printed content to guide new recordings in your language

If you use materials created by others, do a simple permissions check. When you are not sure, do not copy or post the material. Instead, use it as a reference to create new content with your own team and always provide citation where needed.

RECORD NEW AUDIO

Printed text does not capture how people actually speak. Short audio recordings bring your language to life and will be useful later for tools like talking glossaries and text-to-speech.

You can:

- record term explanations and example sentences
- record short stories, conversations, or dialogues
- use phones or simple recorders in a quiet space

Keep recordings short and focused. It is easier to save, label, and reuse ten short clips than one long session. Always follow your ethics and consent practices: explain the purpose, get verbal or written permission, and respect any decision to pause or stop.

"Safeguarding Your Language Through Documentation" is a very helpful UNESCO resource available at https://en.ccunesco.ca

▲ **Claudia Pacay Col**

SAVE IN THREE PLACES

Plan for safety from the beginning. Save each important file in at least three places. For example:

- your main laptop or office computer
- an external hard drive or USB stick stored somewhere safe
- a trusted online or archival option

Choose storage options that match your community's comfort level and capacity. If an archive, museum, or library offers to store copies, make sure you have a clear written agreement about ownership, access, and how files may be used in the future.

BE CAREFUL WHEN SCANNING PRINTED BOOKS

If you are thinking about scanning entire books, pause first. Even if a book is in your language, the author or publisher may still own the rights. When you are unsure, do not scan or upload the book. Instead, you can:

- create new recordings based on the ideas
- write your own explanations and examples
- clearly credit the source in your notes

This respects both legal rights and community protocols.

LOOKING AHEAD

The text and audio you gather in this step will feed later work on keyboards, talking glossaries, machine translation, and text-to-speech. You do not need to understand all those tools now. What matters is that your first data is:

- clearly labeled
- stored safely
- collected under community rules

If you do this well, you will be ready when you decide to work with technical partners or move into Steps 3A–3D.

https://mayanlanguagespreservation.org/roadmap-resources

STEP 3A

Working with AI

WHY AI GETS MORE WRONG IN INDIGENOUS LANGUAGES

Most AI models learn from very large text collections on the internet. Those collections are rich for a small group of big world languages, but much thinner for Indigenous languages.

Because there is far less real material to learn from, AI systems are more likely to invent or distort meaning in Indigenous language output, and when translating between languages, even when the sentences look smooth and correct.

If you are used to strong AI results in English or Spanish, it is easy to expect the same quality in your own language and to miss these hidden errors. This is why you should only use AI for low-risk drafts, never for health, legal, or other important content, and always have trusted speakers review and approve any text in your language before you share it.

AI tools are becoming more common, so it is natural to ask, "Can AI help our language?" This step helps you test AI models safely and measure what they can and cannot do in your language.

WHAT WE MEAN BY AI

In this section, we will reference large language models (LLMs) available through websites or apps (for example, chat-style tools). These systems:

- read prompts and text that you type in
- generate new text in many languages
- can sometimes attempt translation or explanation

For Mayan and other Indigenous languages, these AI models are still very limited. They often:

- know very little about the language
- "make things up" and sound confident when they are wrong
- mix in other languages or invent words
- use external linguistic terms because they overly rely on religious texts in their training

You should never rely on AI output without having a careful review and approval process in place.

ALSO WORTH READING

UNESCO's 2025 publication *Global Roadmap for Multilingualism in the Digital Era* explains how language technologies and AI can support all languages, not just a few big ones. It highlights community leadership, data rights, and fair access to tools, all key issues for any Indigenous language digitization project.

Global Roadmap for Multilingualism in the Digital Era

WHAT WE LEARNED WHEN WE EXPERIMENTED WITH AI

We decided to run small, careful tests so that we could speak from experience instead of guessing what these tools can and cannot do. In our experiments, we initially tested several models that were state-of-the art at the time, including GPT-4o and Claude 3.5 Sonnet.

We:

- saw that most models performed very poorly with Mayan languages and often hallucinated content
- identified one model (Claude 3.5 Sonnet) that showed basic ability in Kaqchikel

DO YOUR OWN EXPERIMENT

For a more structured experiment like MLPP's, visit our Roadmap Resources page and download our simple guide:

"Testing AI with Reference Materials in Your Language"

It walks you through:

- choosing a short, non-sensitive text
- building a small reference pack from your glossaries and grammar notes
- prompting the AI
- scoring the output together with skilled speakers

Our approach was:

1. We chose short, non-sensitive texts.
2. Using an approach called in-context-learning, we built a reference pack with dictionary entries and short grammar explanations.
3. We put the instructions, the reference pack, and the text together into one long prompt.
4. We asked the AI to produce output in Kaqchikel while following the reference pack.

What we saw in the results:

- When we added dictionaries, the output improved somewhat.
- When we added grammar explanations as well, it improved again.
- The process was still slow, costly, and every output needed careful human checking.

For us, this is promising research, but it is not ready for everyday translation or large projects. At this stage, we see AI as something to study and test as it improves, while using it only in a supporting role with human review for low-risk tasks.

https://mayanlanguagespreservation.org/roadmap-resources

STEP 3B

Digital Keyboards and Apps in Our Languages

BUILDING A DIGITAL FOUNDATION

If people cannot type in their language, they are pushed to switch to another language every time they send a message, search the internet, or write a document. In our conversations with Mayan communities, people told us they had to use Spanish for reading, typing, and online activities. There were no keyboards for Q'eqchi', no spell-check for Q'anjob'al, and no subtitles or interfaces in Kaqchikel.

Creating digital keyboards became one of our first practical tools. Keyboards are not just "nice to have." They are a foundation for education, work, and daily communication in our languages.

DIGITAL KEYBOARDS

When we looked at what devices people used most, community feedback was clear: Android phones were the most common and most affordable devices in many Indigenous communities in Guatemala. We focused first on Android so people could use our keyboards in the apps they already use every day: texting, search, Facebook, WhatsApp, email, and more.

Later, we added a desktop option so people could also type in their language on Windows computers.

Two tracks we used

We took two parallel paths:

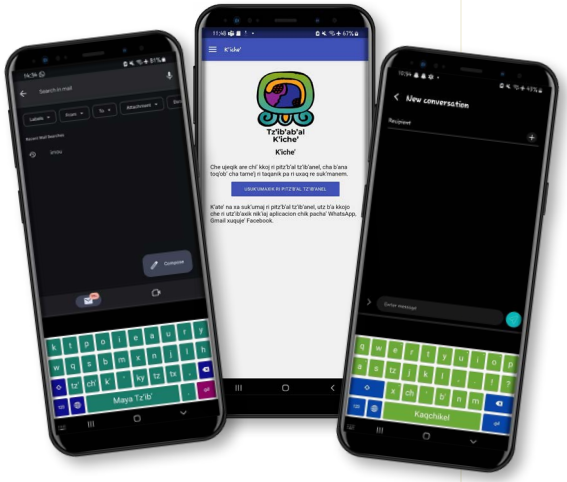
- **Android track (Google Play)**
  We built Maya Tz'íb', a "universal" keyboard that supports several Mayan languages, plus language-specific keyboards for Kaqchikel (Pitzb'äl Kaqchikel), K'iche' (Tz'ib'ab'al K'iche'), and Q'eqchi' (Tz'iib'leb'aal Q'eqchi'). As the application of the first Mayan language keyboards proved successful, we continued to create keyboards for Akateko, Mam, and Q'anjob'al.

  We worked with volunteer developers who used Android Studio and the Google Play Developer Console to build, test, sign, and publish these as free apps in the Play Store.

  The technical details (developer account, app bundles, screenshots, store listing, and review process) are covered step-by-step on the Roadmap Resources page.

- **Windows track (Microsoft Store)**
  For desktop users, we created a Windows keyboard using Microsoft Keyboard Layout Creator (MKLC). We compiled installers, code-signed them, created a Microsoft Store listing, and published the keyboard so people could install it like any other Windows keyboard.

  We are upfront about current limits: for example, our Windows keyboard does not yet include predictive text or spell-check.



TRY THIS

Even if you are not ready to build a keyboard yet, you can take first steps:

1. **Ask which devices people use most.** Are Android phones, iPhones, or shared computers more common?
2. **Make a "character list" for your language.** Invite speakers to write down all the letters and symbols they use in writing, and mark which ones are most important.
3. **Check what already exists.** Look for existing keyboards for your language or related languages. Note what works and what does not.
4. **Talk with a potential technical partner.** Share your device information and character list. This will help a developer understand what kind of keyboard or app would be most useful.
5. **Plan for testing.** When you do have a keyboard or app, set up a small test group of speakers to try it, report problems, and suggest improvements before a wide release.

These small steps will prepare you to work with developers, strengthen your case with organizations who may want to fund your work, and make sure that any future keyboards or apps are grounded in real community use.

DESIGN WITH SPEAKERS, NOT JUST DEVELOPERS

From the start, we asked speakers to guide the keyboard layouts. We invited K'iche', Q'eqchi', Mam, Kaqchikel, Chuj, Q'anjob'al, Poqomchi', and Akateko collaborators to list the characters they use most in their languages. We then prioritized those characters so they would be easy to type on a phone or computer.

When we did not yet have collaborators for a language, we used grammar books and dictionaries from Guatemalan publishers to guide the layout. In all cases, the goal was the same: make typing easier and faster for native speakers, not just copy a Spanish keyboard and add a few extra letters.

MAKING WEBSITES INCLUSIVE BY DESIGN

Keyboards are only part of the story. We also wanted the apps and websites to appear in our languages, not just in a bridge language.

MLPP Website:
www.mayanlanguagespreservation.org

MLPP Glossaries:
https://mayanlanguages.wiki

On our glossary platform, we worked to translate the full user interface into Spanish, K'iche', Q'eqchi', Kaqchikel, and Q'anjob'al, with more Mayan languages on the way. This includes everyday elements like account pages and menus. For example, a line like "Account name: @username" must be translated in a way that still works when the system inserts each user's name. We pay close attention to grammar so that the whole phrase sounds natural in the target language.

Our guiding question is simple:

"If we can provide the full user experience in our languages, why would we force people to use another language in the interface?"

We test these translations with contributors, update them when people find better wording, and keep adding new interface elements in multiple Mayan languages as the site grows.

LESSONS LEARNED

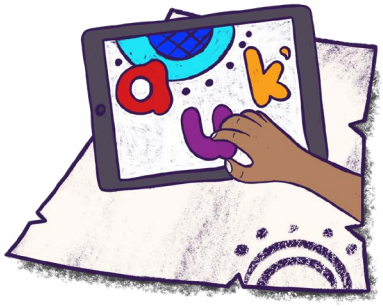
From our keyboard and website work, we learned that:

- **Device choice should follow community use, not trends.** We focused on Android first because community members told us it was the most widely used and affordable option.
- **Speakers must lead layout design.** Asking speakers which characters they actually use helped us build keyboards that feel natural in daily use.
- **Be honest about limits.** It is better to release a simple, stable keyboard without predictive text than to promise features we cannot yet provide.
- **Interfaces matter as much as text.** Translating the full user interface (menus, buttons, account pages) helps people feel that the app is truly "theirs," not a foreign tool with a few translated words.

https://mayanlanguagespreservation.org/roadmap-resources

STEP 3C

Text-to-Speech (TTS)



WHY PRIORITIZE TTS?

Text-to-Speech (TTS) lets a device read written text aloud in your language. For communities with low literacy or few written materials, this can be a powerful bridge. Grandparents can listen along with children. Learners can hear correct pronunciation as they read. People who are tired or busy can still hear messages and stories in their own language.

For us, TTS opens another way for all generations to use and enjoy their language with the tools they already have.

PARTNER CAREFULLY

TTS is not something most communities can build alone. It usually requires:

- a large amount of clean audio
- exact transcripts of that audio
- technical expertise to build and deploy the model

In our case, we reached out to **Microsoft** about adding **Q'eqchi'** TTS to their existing platform. Their team agreed to:

- use our recordings and transcripts only to build a Q'eqchi' voice in their products
- make this Q'eqchi' TTS available at **no extra cost** to users of their software
- work with us under written assurances that explained how the data would and would not be used

When working with partner organizations, written agreements are essential. They should name who owns the data, what products it may be used for, and who has the right to approve any future uses.

DATA SOVEREIGNTY AND WRITTEN ASSURANCES

From the beginning, we treated this as an ethical collaboration, not just a technical project. We entered the agreement only after receiving written assurances that:

- the recordings and transcripts would be used **only** to create Q'eqchi' TTS in Microsoft products
- users would not pay extra to access Q'eqchi' TTS once it was available
- our data would not be repurposed for other uses without our consent

For your community, the key lessons are:

- **Do not give away recordings without clear agreements.**
- **Make sure any TTS tool remains affordable and accessible to your people.**
- **Document who owns the data and who approves any future use.**

If you decide to work on TTS with a trusted partner, you can follow a simple five-step pipeline:

1. CHOOSE WHAT TO RECORD

- Pick non-sensitive texts from everyday life (announcements, stories, dialogues, glossary entries).
- Avoid sacred content or texts where rights are unclear.
- Make sure the language and style match how people really speak.

2. RECORD CLEAR AUDIO

- Invite 1–3 fluent readers (you may want both male and female voices).
- Use the quietest space you can find and good microphones if possible.
- Record short clips instead of long sessions (for example, 1–3 sentences per file).

3. CHECK QUALITY AS YOU GO

- Listen to each file and reject any with echo, noise, or unclear speech.
- Re-record files that are hard to understand.
- Keep a simple log of which files are "OK" and which need to be fixed.

4. TRANSCRIBE CAREFULLY

- Write out exactly what is said in each file, "word by word and letter by letter."
- Make sure spelling follows your agreed orthography.
- Give each file a clear name so you can match audio and text later.

5. SHARE WITH YOUR TTS PARTNER UNDER AN AGREEMENT

- Once recordings and transcripts are ready, provide them to your technical partner.
- Make sure you have a written agreement that explains:
  - who owns the recordings and transcripts
  - what tools or products they can be used for
  - what requires new permission in the future

https://mayanlanguagespreservation.org/roadmap-resources

STEP 3D

Using a Translation Management Tool

HOW A TRANSLATION MANAGEMENT TOOL CAN HELP YOUR COMMUNITY

If you work with a developer who uses Drupal or another content management system (CMS), ask if they can add a translation management tool. This kind of tool should help you:

Keep content safe
People can translate terms, example sentences, or interface text without touching the layout or technical setup.

Assign clear tasks
You can say, "Ana translates Q'eqchi' interface labels" or "Luis checks K'iche' example sentences," and the system will show each person only the items they are responsible for.

See progress by language
You can see which parts are finished in each language, where there are gaps, and which sections need attention next.

For your team, this reduces confusion and makes it easier to invite new contributors into the project.

As your project grows, you may need a website or app with many terms, possibly many languages, and many contributors. If everyone can touch everything, people get scared of "breaking the site," and work slows down.

A translation management tool helps you keep things organized. You do not have to understand all the technical details. The key idea is to use tools that make it easier for your community to work together, instead of making people afraid to touch the project.

OUR EXPERIENCE: MOVING FROM MEDIAWIKI TO DRUPAL

We first used MediaWiki (the software behind Wikipedia) for our glossaries. At first, it looked like a good fit: anyone with an account could add terms, images, and audio. In practice, we ran into problems:

- every contributor could change the page layout and special formatting
- small mistakes could break whole pages
- many people became afraid to edit at all

To solve this, we moved to Drupal, which is a content management system (CMS). Drupal let us:

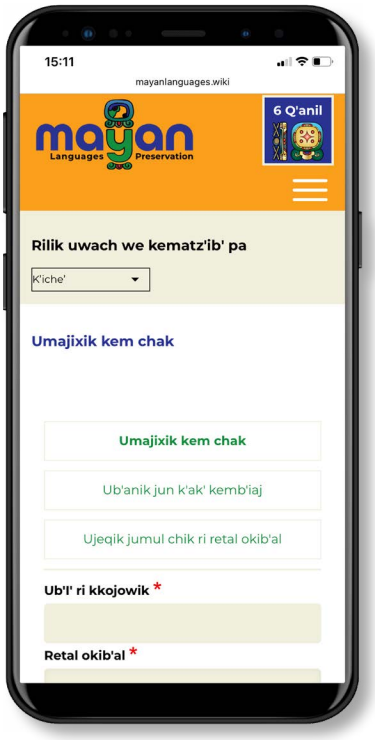
- separate terms, audio, and images into clear fields
- give each contributor a role, so they could only edit what they needed
- add a translation management tool on top of that

Inside Drupal, we used a specific module called the Translation Management Tool (TMGMT). It is a tool inside Drupal that helps manage translation work.

This tool allowed us to:

- group content by language
- assign translation tasks to specific users
- track which items were translated, and which were still pending

▲ **Dalila Choc**



Once we made this change, people were more willing to contribute. They could focus on language, not on website code.

LOCALIZE THE USER INTERFACE

We also used our translation management tool to translate all the interface text on our glossary site, including:

- login and logout messages
- password reset emails
- menu labels and buttons
- system messages and alerts

We treat this interface text as part of our language work, not as an afterthought.

We pay attention to grammar even in short phrases that use placeholders, such as:

- “Account name: @username”
- “You have 3 new messages”

Our goal is simple:

If someone chooses Q’eqchi’ or K’iche’ in the settings, they should be able to use the site fully in that language, without having to switch back to Spanish or English.

You can set the same goal for your own site or app: make the whole experience available in your language, not just the content pages.

TEST AND IMPROVE

After we moved to Drupal and added the translation management tool, we:

- asked contributors to use the site in their own languages
- watched where they got stuck or confused
- asked which labels and steps felt unclear

Based on this, we:

- changed button text and menu labels
- simplified some screens
- adjusted roles if people are still worried about breaking something

You can follow a similar process:

1. Ask a small group of contributors to use the site in your language.
2. Note where they hesitate or ask for help.
3. Change labels, instructions, or roles to make things easier.
4. Repeat later as you add more content and contributors.

https://mayanlanguagespreservation.org/roadmap-resources

STEP 4

Community Use in Schools – A Case Study In Tool Adoption

HOW YOU CAN TRY SOMETHING SIMILAR

You do not need a large program to begin. You can:

- invite one school, class, or after-school group to pilot your tools
- ask students to write short stories, poems, or dialogues using your keyboards and glossaries
- collect simple feedback from teachers and students: What was fun? What was hard? What should we change?

Every new word typed, spoken, or read in your language at school is a small step toward long-term adoption of your digital tools and toward a future where children see their language as part of education, not separate from it.

Digital tools only become real when people use them in daily life. Schools are a powerful place to start. When children see, hear, and type their language at school, they begin to connect it with learning, creativity, and their future.

A CLASSROOM CASE STUDY: CENTRO TECNOLÓGICO BILINGÜE (CTB)

At the Centro Tecnológico Bilingüe (CTB) in Tucurú, Guatemala, Q'eqchi' students and teachers worked with us to try the new Q'eqchi' keyboards and interactive story workbooks. Students wrote poems, fables, and short stories in Q'eqchi', added drawings, and used the keyboards to type the full orthography, including characters that standard keyboards ignore.

Teachers told us that this:

- increased students' motivation and imagination
- helped children value Q'eqchi' and improve spelling in both Q'eqchi' and Spanish
- It is important because we are able to support our community families in school to foster the value of our students' native Q'eqchi' language.

Students said the project helped them "find their voice" and see Q'eqchi' as a language for books, to learn about animals and science, and to tell stories, not only for use at home with family.

You will be able to read more about this collaboration in the **"Our Languages, Our Voices, Our Future"** case study on the MLPP website at: https://mayanlanguagespreservation.org/case-study-ctb

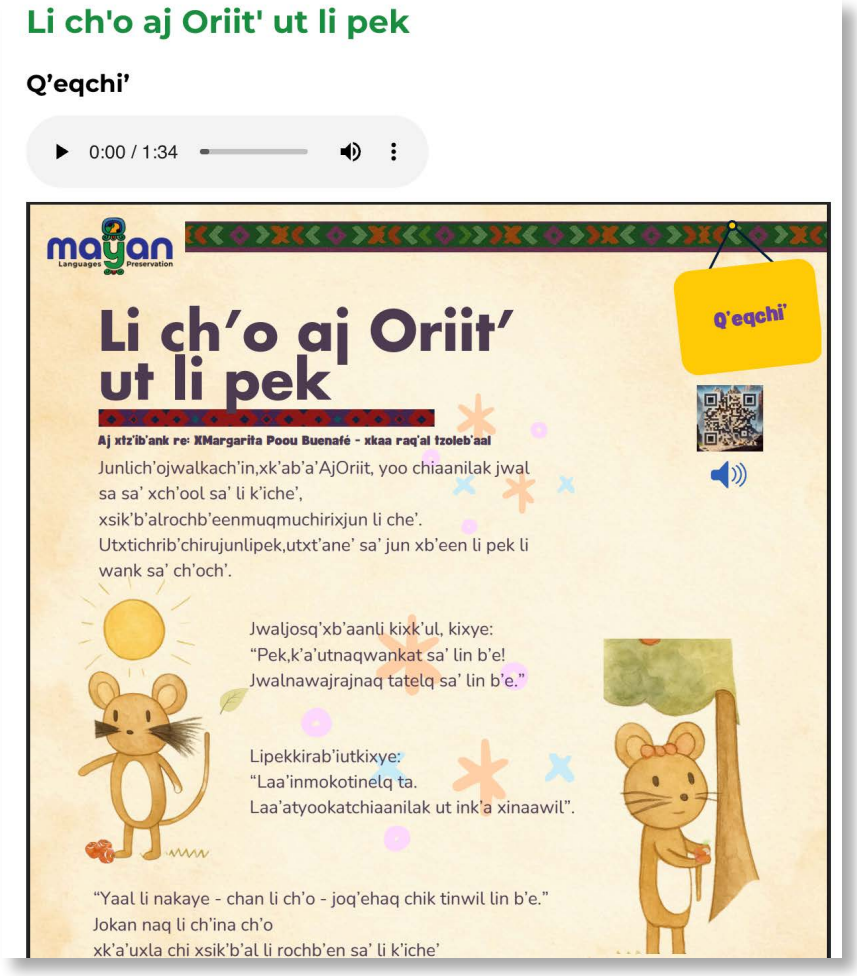
▲ Centro Tecnologico Bilingüe teachers testing Q'eqchi' keyboard for Android

A VIRTUOUS LOOP: SCHOOLS AND GLOSSARIES HELPING EACH OTHER

Working with CTB was not a one-way “pilot.” It was a mutual exchange.

- **From the glossary to the classroom:**
  Teachers used glossary terms, example sentences, and images to build lessons, writing prompts, and activities. Digital tools made it easier to plan Q’eqchi’ lessons and to show correct spelling and punctuation.
- **From the classroom back to the glossary:**
  Students and teachers created new content including stories, poems, local place names, plants, foods, and songs. They recorded new audio, suggested better example sentences, and drew images. These materials then fed back into the glossary and other tools.

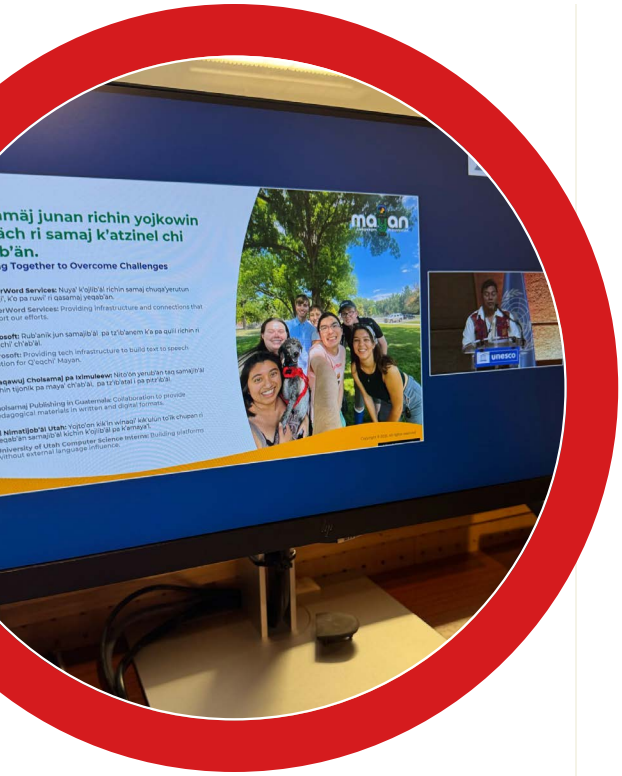
This “virtuous loop” means the tools and the school strengthen each other. The glossary supports teaching. The school returns real language, voices, and images that keep the tools alive and growing.



https://mayanlanguagespreservation.org/roadmap-resources

STEP 5

Share with the World



When your tools are working, you can choose how and where to share your story. Sharing shows that your language belongs in public and digital spaces, not just at home.

Telling your story can:

- inspire other communities to start their own projects
- show governments, schools, and funders what is possible
- challenge the idea that only big world languages “fit” in technology

You decide what to share and what to keep private. Sharing should always happen on your terms.

AN EXAMPLE: SPEAKING AT UNESCO

As part of the UNESCO International Decade of Indigenous Languages, Kawoq Baldomero Cuma, board of advisors for MLPP shared MLPP’s work at an UNESCO event in Paris, in Kaqchikel. This was the first presentation at an international UNESCO event delivered fully in an Indigenous language.

In that presentation, we:

- showed how Mayan communities are building their own tools, such as glossaries, keyboards, and learning resources
- explained why equal digital access matters for Indigenous and marginalized language speakers
- spoke as authors of our own story, using our own language

▲ **Kawoq Baldomero Cuma speaks about the importance of Indigenous language inclusion in technology following his historic speech at the LT4all conference at UNESCO.**

You can read more in the press release:

"Historic First: Baldomero Cuma Chavez Delivers UNESCO Presentation in Kaqchikel, an Indigenous Mayan Language of Guatemala"

This was not just a showcase. It was a way to say in a global space that our languages deserve the same respect and technical support as any dominant language.

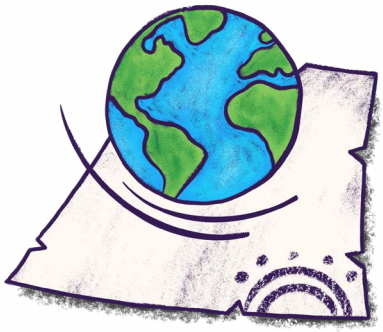
SHARE ON YOUR OWN TERMS

When you start to share your work more widely, you can ask three simple questions:

1. **What do we want to show?**
   For example, youth writing in the language, school projects, or ethical tech partnerships.
2. **Who is this for?**
   Local schools, other Indigenous communities, potential funders, or international bodies.
3. **What protections do we need?**
   Decide in advance what can be recorded, posted, or quoted, and what should stay within the community.

You can share through:

- local events such as school fairs, community meetings, or radio programs
- national or regional gatherings and conferences
- online spaces you control, like your website, social media, or community pages
- invitations to speak at international events, when that feels right for you



https://mayanlanguagespreservation.org/roadmap-resources

Getting Started Checklist

You do not have to do everything at once. This checklist gives you some first steps you can take over the next few weeks and months.

1. Listen and set direction

- ❑ Hold at least one listening session with language champions, teachers, youth, Elders, and community leaders.
- ❑ Write down the main challenges people name (for example, keyboards, school materials, recordings, or public services).
- ❑ Choose 1–3 challenges to focus on first.
- ❑ Form a small advisory group of trusted language professionals to guide decisions.

2. Build your team

- ❑ Invite 5–10 people who already support your language (interpreters, teachers, organizers, students).
- ❑ Agree on a few shared goals for the next 6–12 months and write them in simple language.
- ❑ Decide which spelling system or orthography you will use in your project.
- ❑ Choose one person to act as coordinator for the next few months.

3. Start gathering and creating data

- ❑ Make a list of existing materials you can draw on (dictionaries, grammar books, school materials, wordlists).
- ❑ Choose one urgent area (for example, hospital visits, school enrollment, or social services) and start a small talking glossary.
- ❑ Record 10–20 short audio clips (terms or example sentences) in a quiet space, following your consent practices.
- ❑ Save your files in at least three places (for example, computer, external drive, and a trusted online or archival option).

4. Plan for tools that match real use

- ❑ Ask which devices people use most (Android phones, iPhones, shared computers).
- ❑ Create a "character list" for your language with the letters and symbols speakers actually use.
- ❑ Look for any existing keyboards or tools in your language or related languages and note what works and what does not.
- ❑ Talk with at least one potential technical partner about your priorities and your community's device situation.

5. Try your tools in one real setting

- ❑ Invite one school, class, or community group to try your first tools (for example, a talking glossary or a simple keyboard).
- ❑ Ask students or participants to create short texts, drawings, or recordings using your tools.
- ❑ Collect simple feedback: What was helpful? What was difficult? What should we improve?
- ❑ With consent, add some of the new words, sentences, audio, or images back into your glossary or resource page.

6. Protect ethics and data

- ❑ Draft a simple consent script or form in your language and a contact language.
- ❑ Agree where recordings and texts will be stored and who has access.
- ❑ Write down basic rules for what can be shared publicly and what must stay in community-controlled spaces.

7. Decide if and when to explore AI and advanced tools (optional)

- ❑ Read the AI and TTS sections of this Roadmap and talk with your advisory group about risks and benefits.
- ❑ If you decide to experiment with AI, use only non-sensitive texts and follow the MLPP testing worksheet from the Resources page.
- ❑ Keep notes on what the tools get right, what they get wrong, and where you will and will not use them.

You can print this page and use it as a working checklist. Each item you check off is a real step toward bringing your language into digital spaces on your own terms.

mayan
Languages Preservation

Mayan Languages Preservation and Digitization Project

contact@mayanlanguages.org

instagram.com/mayanlanguages

mayanlanguagespreservation.org

MasterWord®

2022-2032 | INTERNATIONAL DECADE OF
Indigenous Languages