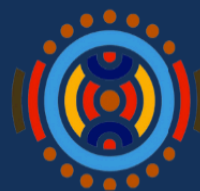


Introduction to the Zero to Digital Guidelines

TRANSLATION
COMMONS



2022-2032 | INTERNATIONAL DECADE OF

Indigenous Languages

Table of Contents

Language Digitization	2
The Zero to Digital Series	3
The Abridged Zero to Digital Series	3
Who Should Use These Guidelines?	4
Translation Commons	4
UNESCO	4

Indigenous languages worldwide are facing an unprecedented crisis. With each passing generation, many languages are disappearing, resulting in the loss of invaluable cultural knowledge and identity. When a language goes silent, the world may retain only a dimmed recollection, if any, of the lore, ecological practices, oral histories, art, music, knowledge of medicinal plants, technologies, spiritual beliefs, values, and food practices of that community.

A growing factor in the marginalization of Indigenous communities in the modern world is the inability to use their native language on digital systems. This can have profound consequences, hindering efforts to address their unique needs and perspectives. For example, this hampers participation with organizations that shape policies and economies that affect their lives. It is an obstacle to contributing their knowledge to humanity or to expressing their opinions fully, precisely, and when needed, forcefully to people outside their community. Further, it hinders access to the vast information resources and tools available online. Gaining consensus among its members may be impacted, especially across a diaspora. And finally, the isolation adds to the risk of outsiders misrepresenting their values and culture.

Language Digitization

Digitizing their native language is but one impactful step for an Indigenous community to counter this marginalization. It allows them to participate more equally on the world stage while preserving the uniqueness of their language. It enables mobile and desktop software to support their language. With digitization, native language speakers can communicate online, share knowledge, and utilize popular software and devices.

It is also important that each community initiate and manage the digitization process of its language. This allows communities to maintain ownership and governance of their language data and limit cultural misappropriation.

A community seeking to digitize their language will need to understand the best practices for:

- collecting, annotating, and storing materials about their language;
- sharing their language data while maintaining their data sovereignty;

- inclusion in language and technology standards organizations;
- achieving implementation of their language on digital systems enabling communication, publication, and native language applications;
- developing advanced natural language processing tools in support of their grammar, orthography, as well as machine translation, AI, etc.

The *Zero to Digital* Series

The *Zero to Digital* series is an ongoing project developed by Translation Commons in support of the United Nations International Decade of Indigenous Languages (IDIL). It addresses this information gap by providing a comprehensive step-by-step framework for collecting and organizing Indigenous language data and bringing the language online.

The already published guidelines, available in several languages, include:

- *Introduction to the Zero to Digital Guidelines*
- *Zero to Digital: A Guide to Bring Your Language Online*
- *Language Data Gathering Guideline*
- *Terminology Guideline*
- *Benefits of Digitization of a Language*
- *Font Creation Guidelines*

The Abridged *Zero to Digital* Series

Translation Commons is creating abridged versions of these guidelines to be shared in the UNESCO library. To date, the following abridged versions have been published:

- *Introduction to the Zero To Digital Guidelines* - this document
- *Language Data Gathering Guidelines Abridged Version*

Please visit [our resource page](#) for the original English version of this document, as well as its translations and other guidelines from Translation Commons.

Who Should Use These Guidelines?

These guidelines are intended for a diverse range of individuals and organizations involved in Indigenous language revitalization:

- Indigenous language advocates and communities
- Researchers and academics
- Government agencies and NGOs
- Technology developers

Translation Commons

Translation Commons is a nonprofit volunteer community that supports the digitization of languages, mentors language professionals, and provides courses and resources for the language industries. One of our principal programs is the **Language Digitization Initiative**, which provides a roadmap for Indigenous and minority language communities who want to digitize their language. We provide tutorials and workshops, create tools and applications, and bring language advocates together with industry experts. By supporting equitable digital access, we further our mission to enable interested language communities to participate in global online activities and pursue the benefits of computer applications in their native language.

UNESCO

Since adopting the 2003 Recommendation on the Promotion and Use of Multilingualism and Universal Access to Cyberspace, UNESCO has been actively supporting language digitization practices. This aligns with the development of multilingual content and systems in digital form, ensuring that all cultures, including indigenous ones, can express themselves and access cyberspace in their languages.

Additionally, the Recommendation urges Members to encourage and support capacity-building for producing local and indigenous content on the Internet. Consequently, UNESCO supports the digital empowerment of Indigenous Language Communities within the framework of the International Decade of Indigenous Languages (IDIL2022-2032).