

The Language Digitisation Debate: Big Tech vs Indigenous Communities

Advocates for indigenous language communities want to see their languages thrive online without relinquishing sovereignty of their language data to Big Tech.

London, United Kingdom Jul 2, 2024 ([IssueWire.com](https://www.issuewire.com)) - On May 21, Translation Commons hosted the [Indigenous Languages and Technology seminar](#), in partnership with UNESCO. The event was in honour of the International Decade for Indigenous Languages (from 2022-2032) and World Day for Cultural Diversity for Dialogue and Development. Experts in language technology, translation, and indigenous cultural preservation gathered to discuss challenges facing [88.1 million speakers of endangered languages](#) across the globe, many of whom belong to indigenous communities.

In a world ruled by digital communication, efforts to preserve, revitalise and promote linguistic diversity rely on collaboration between grassroots, indigenous organisations, and tech industry leaders. For such collaborations to work, the ethics and feasibility of projects must be carefully considered. As it stands, advocates for indigenous language preservation are using social media and creative software like Adobe Suite to make information about endangered languages more accessible. These endeavours are made more difficult due to the limited digital resources (such as keyboards and transcription tools) at their disposal.

Linguistic anthropologist Aiyana Twigg believes that “AI can and should be utilised to support indigenous languages”. She spoke at the event about her youth-centred approach to promoting the revitalisation of her native Blackfoot and Ktunaxa languages, noting that existing technologies like speech-to-text and auto-generated social media captions should be adapted to include indigenous languages. She acknowledged that these advancements can only take place if tech companies are willing to work closely with indigenous communities to ensure terms and conditions around data ownership, usage, and accessibility are “dictated by and aligned with the values of the communities [from which the languages originate].”

This requires a holistic approach wherein tech companies would have to invest in the communities themselves. Motorola is one of the few large tech companies to have adopted this approach. Their Edge Plus phones, released in 2022, include a Cherokee language interface, [developed with the support of Cherokee language preservationists and community leaders](#). On June 27, [Google Translate also updated their language database](#) to include 110 new languages including indigenous African languages like Fon and Kikongo and other languages spoken by “small communities of indigenous people.”

As well as direct consultation with Indigenous communities, Twigg advocated for youth scholarships and open-source software that allows indigenous community members to easily build their own solutions. She also sees a need for stricter privacy rules around how language data is used and by who. Her concerns about data security are shared by the staff at Te Hiku Media, a Māori language, a non-profit radio station that built pioneering language tech from a community-sourced database of te reo Māori audio. Since developing its speech recognition and speech-to-text tools in 2018, the company has been [fending off requests](#) from major tech corporations that want to use the data to bolster their own data sets.

CEO of Te Hiku, Peter-Lucas Jones, and his team are wary of offers from US-based firms such as Lion Bridge that seek to profit from languages they have no connection with. They have, however, formed partnerships with academic institutions issuing data licences along the very strict terms that any project

created with Māori data must benefit Māori people directly and ownership of said projects must remain within Māori communities. If Big Tech is to play a pivotal role in the kind of language revitalisation being spearheaded by companies like Te Hiku, the industry would have to be open to letting indigenous communities lead in ways that are as yet unprecedented. The future of [over three thousand endangered languages](#) will remain uncertain until key tech industry players seek the insight of indigenous organisers who are dedicated to language revitalisation.

Press release courtesy of Derivation, a language technology solution providing businesses with accessible global language data and analysis.

Media Contact

Derivation

mkonteh@derivation.co

Source : Derivation

[See on IssueWire](#)