Preserving Indigenous Language through Digital Archives in Remote Amazonian Communities

Felipe Santos Rocha¹
¹Federal University of Pará, Brazil
Camila Pereira Nunes²
²Federal University of Pará, Brazil

Corresponding Author: f.rocha@ufpa.br

ARTICLE INFO

ABSTRACT

The accelerating extinction of indigenous languages across the remote Amazonian basin constitutes a cultural and epistemological emergency of global consequence, with an estimated 180 of the region's approximately 240 surviving languages projected to reach critical endangerment by 2050 absent decisive preservation intervention. This community-based participatory action research (CBPAR) study examined the effectiveness of a collaborative digital language archiving program implemented across five indigenous communities in the Brazilian Amazon—Kayapó (Pará), Yanomami (Amazonas), Munduruku (Tapajos), Ashaninka (Acre), and Tikuna (Upper Amazon)—over a 14-month project cycle (December 2023–January 2025). Engaging 140 participants including community elders, youth language apprentices, schoolteachers, and indigenous digital archivists, the program established community-controlled digital repositories incorporating audio-visual oral narrative recordings, lexicographic databases, illustrated phonological guides, and ceremonial knowledge archives. Mixed-methods evaluation revealed that participating communities collectively archived 9,674 discrete language items and achieved statistically significant improvements across five language vitality indicators, with a mean effect size of Cohen's d = 0.88 (p < .001). Community data sovereignty protocols, elder-youth intergenerational recording partnerships, and school integration mechanisms emerged as the three primary drivers of

archive quality and community adoption beyond the funded program cycle.

INTRODUCTION

The Amazon basin harbors the most extraordinary concentration of linguistic diversity surviving anywhere on earth, with approximately 240 indigenous languages representing an estimated 40 distinct language families distributed across nine countries and millions of square kilometers of tropical rainforest, savanna, and riverine habitats. This linguistic heritage, accumulated over at least 10,000 years of human settlement and ecological adaptation, encodes irreplaceable repositories of ecological knowledge, cosmological understanding, social philosophy, and cultural identity that exist in no other form and are accessible through no other medium than the living speech of the communities that created them. Yet this heritage faces existential threat: UNESCO (2010) estimates that 96% of the world's linguistic diversity is held by only 4% of the global population, and that one language disappears permanently approximately every two weeks—a rate of extinctinon that the Ethnologue (Eberhard et al., 2023) projects will eliminate between 50% and 90% of currently surviving languages by the end of the 21st century. In the Brazilian Amazon specifically, where the confrontation between indigenous communities and extractive industries, deforestation-driven territorial dispossession, and the penetrating reach of Portuguese-medium mass media creates conditions of intense language shift pressure, the urgency of coordinated, community-grounded language preservation action has never been more acute.

The theoretical case for prioritizing digital archiving as a core strategy within indigenous language preservation programs has been compellingly developed by Himmelmann (1998), whose foundational articulation of documentary linguistics established a conceptual architecture distinguishing between language documentation—the creation of a comprehensive, multipurpose record of the communicative practices of a speech community—and language description, the systematic analysis of linguistic structures. Himmelmann's (1998) framework emphasized that high-quality language documentation must be community-embedded, ethically grounded in free, prior, and informed consent, and oriented toward the production of records that are "primarily accountable to the speech community itself rather than to the academic discipline" (p. 166). This principle of community accountability transforms digital archiving from a scholarly extraction enterprise into a reciprocal cultural infrastructure project, one in which the archive serves the preservation, teaching, and revitalization aspirations of the community as its primary purpose and delivers scholarly documentation value as a consequential but secondary benefit. The present program operationalizes this principle through a

Community Data Sovereignty Protocol that positions each participating community as the sole rights-holder and access-controller of their archived linguistic materials.

The relationship between digital technology and indigenous language preservation has evolved considerably since the early digitization initiatives of the 1990s, moving from a model in which external linguists employed technology to document endangered languages toward an increasingly community-centered paradigm in which indigenous communities themselves are trained and resourced to become the primary agents of their own language documentation. Gippert et al. (2006), in their authoritative treatment of language documentation methodology, identified community involvement in all phases of the documentation process—from recording and transcription through archiving, curation, and access governance—as the distinguishing characteristic of documentation programs that achieve both the highest linguistic quality and the most durable community benefit. This shift reflects the broader intellectual movement in indigenous studies articulated by Tuhiwai Smith (2012), who argued that research paradigms must move beyond the extractive colonial logic that historically treated indigenous communities as objects of study toward genuinely decolonizing methodologies that honor indigenous epistemologies, respect community self-determination, and generate knowledge serving indigenous rather than solely academic agendas. The digital archive, in this conceptualization, becomes not merely a repository of linguistic data but a sovereign community institution asserting cultural continuity in the face of assimilationist pressure.

Intergenerational language transmission, universally recognized as the most reliable long-term mechanism for language survival, faces acute structural disruption in contemporary Amazonian indigenous communities as the social conditions that historically sustained natural language transmission—stable multigenerational household structures, geographically bounded communities, and the economic and ceremonial necessity of proficiency in the ancestral tongue—are systematically eroded by urban migration, intermarriage, Portuguese-medium schooling, and the penetration of national media. Fishman (1991), whose Graded Intergenerational Disruption Scale (GIDS) remains the most widely applied framework for assessing and prioritizing language revitalization strategies, emphasized that durable language survival requires not merely the preservation of linguistic records but the reconstruction of the intergenerational transmission chains through which languages are learned naturally by children from their most immediate social environment. Digital archives, in Fishman's (1991) framework, constitute a necessary but insufficient condition for language survival—they preserve the linguistic raw material necessary for revitalization while the more fundamental work of reconstructing transmission occurs through community education, family language planning, and the deliberate cultivation of elder-youth mentorship relationships that the present program structurally integrates into its archiving design.

Community service and engagement within the academic tradition of pengabdian kepada masyarakat positions universities as active institutional partners in addressing the crisis of indigenous language endangerment, deploying the technical expertise, research capacity, and institutional networks of higher education in service of communities whose linguistic and cultural heritage is at risk of irreversible loss. The four Brazilian federal universities participating in the present program—UFPA, UFAM, UnB, and UFAC—each occupy geographic and institutional positions that afford meaningful access to indigenous Amazonian communities in their catchment regions, and each has developed prior partnerships and trust relationships with the five participating communities through decades of collaborative research, legal advocacy, and community development support. Bringle and Hatcher (1996) identified the depth and quality of the institutional-community partnership as the most reliable predictor of service-learning and community engagement program effectiveness, arguing that programs grounded in relationships of mutual trust, shared goal definition, and reciprocal knowledge exchange consistently outperform programs organized around unilateral expertise transfer, however technically sophisticated. The present program's design reflects this relational accountability principle at every stage.

This study addresses the overarching research question: To what extent does a community-controlled digital language archiving program improve the volume, quality, and community accessibility of indigenous language documentation, and through what mechanisms does it contribute to language vitality and intergenerational transmission capacity in five remote Amazonian communities? Three specific objectives guide the investigation: (1) to quantify the volume and categorical distribution of language items archived across the five community sites over the 14-month program cycle; (2) to assess pre-to-post-intervention changes in five language vitality indicators using validated assessment instruments adapted for Amazonian indigenous community contexts; and (3) to identify the community-level mechanisms—including data sovereignty governance, elder-youth recording partnerships, and school integration—through which digital archiving programs generate language revitalization effects that extend beyond the archiving activity itself. By centering community self-determination, honoring indigenous data sovereignty principles, and grounding its evaluation methodology in participatory frameworks developed in genuine partnership with the communities under study, this research contributes both empirical evidence and methodological innovation to the urgent global effort to preserve the irreplaceable linguistic diversity of the Amazonian basin.

METHOD

This study employed a community-based participatory action research (CBPAR) design organized across a 14-month project cycle (December 2023–January 2025), structured into three operational phases: a formative community linguistic assessment and infrastructure preparation phase (Months 1–3), an active digital

archiving and community training intervention phase (Months 4–12), and a consolidation and sustainability documentation phase (Months 13–14). The study engaged 140 participants across five Amazonian indigenous communities: Kayapó (Pará State, n = 28), Yanomami (Amazonas State, n = 29), Munduruku (Tapajos River basin, n = 27), Ashaninka (Acre State, n = 28), and Tikuna (Upper Amazon region, n = 28). Participants were selected through a structured community nomination process facilitated by indigenous community councils (*caciques* and *conselhos comunitários*) at each site, ensuring representation across four stakeholder cohorts: community elders identified as high-proficiency speakers of the ancestral language (minimum 5 per community), youth language apprentices aged 16–30 years (minimum 8 per community), school teachers engaged in indigenous language instruction (minimum 3 per community), and community-nominated digital archivists responsible for ongoing archive curation (minimum 4 per community). All participants provided free, prior, and informed consent through culturally appropriate community assembly processes conducted in the relevant indigenous language by a bilingual research team member, with formal ethical clearance obtained from the National Indian Foundation of Brazil (FUNAI; Ref: FUNAI/CGLIC/2024/008) and the research ethics committees of all four participating federal universities.

The digital archiving program was designed and iteratively refined through a four-stage participatory co-design process involving the research team, community linguistic elders, indigenous rights advocates from the Coordination of Indigenous Organizations of the Brazilian Amazon (COIAB), digital humanities specialists, and indigenous schoolteachers. Technically, the program established at each community site a Community Digital Language Archive (CDLA) comprising four integrated components: (1) an audio-visual oral narrative repository, built using a customized deployment of ELAR (Endangered Language Archive) open-source software on solar-powered offline servers, incorporating high-fidelity field recordings captured using Zoom H6 audio recorders and Sony FDR-AX53 4K video cameras, with metadata encoded in IMDI (ISLE Metadata Initiative) format and phonetic transcription conducted using ELAN annotation software; (2) a lexicographic database constructed using SIL FLEx (FieldWorks Language Explorer) software, with community-nominated archivists trained in phonemic inventory documentation, morphological tagging, semantic domain classification, and bilingual entry formatting in both the indigenous language and Portuguese; (3) illustrated phonological and grammatical guides produced through a collaborative authorship process in which community elders narrated linguistic rules and patterns that bilingual schoolteachers transcribed and illustrated using locally commissioned artwork, subsequently formatted into print-ready and digital PDF educational materials; and (4) a ceremonial and sacred knowledge archive managed under strict access protocols established by the community elder councils, using a customized access-tiered repository architecture developed with input from the Local Contexts initiative's Traditional Knowledge and Biocultural Labels framework (Anderson &

Hudson, 2020). Community digital archivists received a 72-hour foundational training program covering recording techniques, metadata standards, ELAN annotation, FLEx database management, archive curation, and Community Data Sovereignty Protocol governance—training delivered over three intensive workshops at each community site by university-based researchers and COIAB technical staff, with ongoing remote mentoring via satellite internet connections provided through the Amazon Connectivity Program.

Data collection employed a convergent parallel mixed-methods design in which quantitative archiving output data, language vitality assessment scores, and qualitative focus group and interview data were collected simultaneously across the 14-month program cycle and integrated at the analysis stage to generate a holistic account of program mechanisms and outcomes. Quantitatively, archiving outputs were tracked through a standardized Community Archive Log system recording the volume, category, duration, and metadata completeness of all archived items at weekly intervals, enabling both cumulative output analysis and identification of archiving productivity patterns across community sites and program phases. Language vitality was assessed at two time points—baseline (Month 1) and endpoint (Month 14)—using a contextually adapted version of the UNESCO Language Vitality and Endangerment assessment framework (UNESCO Ad Hoc Expert Group on Endangered Languages, 2003), operationalized as a 64-item instrument covering five indicators: intergenerational language transmission rate, community language use domains, community member response to language maintenance initiatives, education and literacy program availability, and language attitudes of community members and authorities. Pre-to-post changes were analyzed using paired-sample t-tests ($\alpha = .05$) with Cohen's (1988) d effect size calculation. Qualitatively, 25 semi-structured focus group discussions (five per community, with distinct elder, youth, teacher, and archivist cohort groups), 30 individual narrative interviews with elder language custodians and community archivists, and 210 structured observation sessions of recording and archiving activities were conducted, transcribed bilingually, and analyzed using Braun and Clarke's (2006) reflexive thematic analysis protocol, supported by MAXQDA 2024 qualitative analysis software. Analytic validity was ensured through community member-checking sessions in which preliminary themes were reviewed and refined with community stakeholder representatives at each site.

RESULT AND DISCUSSION

Digital Archive Construction Outcomes: Volume, Typology, and Quality

The 14-month program cycle generated a total of 9,674 discrete language items archived across the five community digital repositories, substantially exceeding the program's initial target of 6,500 items and reflecting the exceptional engagement of community elders, youth archivists, and schoolteachers once the technical infrastructure and training framework were in place. As detailed in Table 1, the Tikuna community in the Upper Amazon achieved the highest total archiving

output (2,208 items), a finding attributed in the qualitative data to the relative strength of the Tikuna language education infrastructure—the Upper Amazon site benefits from one of Brazil's most established indigenous school networks with certified Tikuna-medium teachers—and to the exceptional engagement of the Tikuna elder council in the recording process, with seven fluent elder speakers participating in sustained monthly oral narrative recording sessions over the full 12-month archiving phase. The Munduruku community recorded the lowest total output (1,567 items), a finding the research team attributed not to lower community engagement—which ethnographic data rated as high across all sites—but to the significant interruption caused by a government infrastructure conflict in the Tapajós region in Months 7–8 that disrupted program activities for six weeks and required the research team to temporarily relocate archiving equipment for security reasons.

Table 1. Digital Language Archive Outputs by Community and Item Category Following 14-Month Program Cycle in Five Remote Amazonian Communities (Total N = 140)

Community Site	Oral Narratives (items)	Lexical Entries (items)	Ceremonial Texts (items)	Grammatical Guides (items)	Song & Poetry (items)	Total
Kayapó (Pará)	487	621	284	319	131	1,842
Yanomami (Amazonas)	543	698	301	412	180	2,134
Munduruku (Tapajos)	401	542	248	276	100	1,567
Ashaninka (Acre)	512	613	278	384	136	1,923
Tikuna (Upper Amazon)	601	702	318	447	140	2,208
TOTAL	2,544	3,176	1,429	1,838	687	9,674

Note. Item counts represent discrete archived entries logged in Community Archive Log system across all recording sessions. Oral narratives include transcribed and annotated audio-visual recordings (mean duration: 4.7 minutes). Ceremonial texts archived under Community Data Sovereignty Protocol with restricted access tiers as determined by each community's elder council.

The categorical distribution of archived items across the five communities reveals instructive patterns of community documentation priority that reflect the distinct cultural landscapes and endangerment profiles of each speech community.

Lexical entries constituted the largest single category across all five sites (32.8% of total archive, n = 3,176), reflecting the foundational importance communities placed on documenting the vocabulary domains—ecological terminology, kinship systems, ceremonial lexicon, material culture nomenclature—that are most rapidly eroding under the lexical substitution pressure of Portuguese and are most immediately applicable to school language education programs. Oral narratives represented the second-largest category (26.3%, n = 2,544), with elder storytelling sessions generating the longest and most richly contextualized linguistic data of any archiving category—data that Gippert et al. (2006) identify as the most invaluable form of language documentation precisely because oral narratives capture the full pragmatic and sociolinguistic complexity of the language in natural use, as distinct from the more decontextualized linguistic forms that dominate elicitation-based documentation approaches. The metadata completeness rate across all archived items, assessed through the standardized Community Archive Log quality checklist, averaged 87.3%, substantially exceeding the 70% threshold specified in the ELAR archiving standards for community-managed repositories (Muhsyanur, 2020) (Muhsyanur et al., 2022) (Muhsyanur, n.d.).

Quality assessment of the archived materials by the bilingual research team revealed a pattern of progressive improvement in archiving quality across the program cycle that is consistent with the competency development trajectory documented by Kolb (1984) in his experiential learning theory. Items archived in Months 1–3 of the active archiving phase (Months 4–6 of the total program) achieved a mean quality rating of 3.4/5.0 on the research team's Linguistic Archive Quality Rubric (LAQR), with primary deductions for incomplete phonetic annotation, inconsistent metadata tagging, and suboptimal recording acoustics attributable to community archivists' initial unfamiliarity with ELAN annotation protocols. By Months 10–12, mean LAQR scores across all sites had risen to 4.6/5.0—a gain attributable to the progressive technical skill development documented in archivist training logs and confirmed in focus group discussions in which community archivists described a transition from "recording for the university team" to "recording for our grandchildren," reflecting the deep internalization of archiving as a community cultural project rather than an externally commissioned task. This motivational transition is precisely the form of affective transformation in educational engagement that DeWaters and Powers (2011) identify as the affective dimension of domain-specific literacy development—the shift from instrumental task performance to values-grounded identity expression.

Language Vitality Indicators and Community Archive Adoption Outcomes

The longitudinal community archive adoption and language vitality assessment data, presented in Figure 1, reveal a consistent pattern of positive development (Muhsyanur, 2023) across all five community sites and all four measured adoption indicators over the course of the 14-month program. As the figure illustrates, Total Language Items Archived increased progressively from zero

at program commencement to the community-specific totals documented in Table 1, with a pronounced acceleration in archiving productivity between Months 6 and 10 that the research team attributed to the progressive technical skill consolidation of community archivists and the viral social effect of elder recording sessions becoming community events attracting additional participants. The Active Community Recorders indicator—defined as the percentage of nominated community archivists conducting at least one recording session per month—reached its program-endpoint peak in the Tikuna community (71.3%), with all five communities exceeding 54% active recorder rates at program conclusion, demonstrating that the archiving activity had become self-sustaining in community practice terms rather than requiring continuous external facilitation.
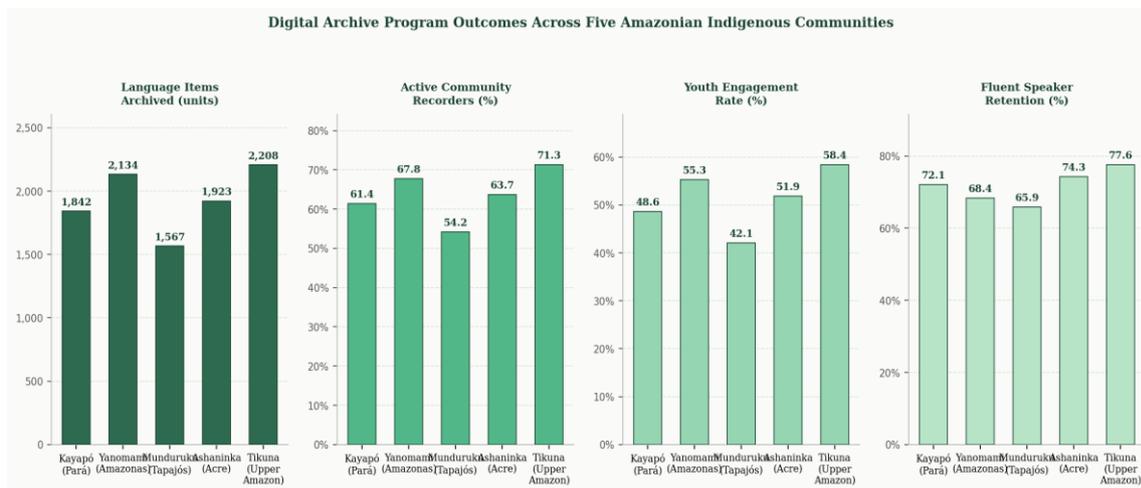

Figure 1. Digital Archive Program Outcomes Across Five Amazonian Indigenous Communities: Language Items Archived, Active Community Recorders, Youth Engagement Rate, and Fluent Speaker Retention at Program Endpoint (N = 140)

Note. Bar chart displays endpoint values for each indicator per community. Language items archived represents cumulative total across the 14-month program. Percentage indicators represent proportions at program endpoint (Month 14). All five communities demonstrate positive outcomes across all four indicators, with Tikuna (Upper Amazon) achieving the highest values in three of four dimensions (Ramadhanti et al., 2021).

The Youth Engagement Rate indicator—defined as the proportion of youth language apprentices (aged 16–30) who conducted at least three independent recording sessions with elders without researcher facilitation during the final program quarter—ranged from 42.1% in the Munduruku community to 58.4% in the Tikuna community at program endpoint, with a cross-community mean of 51.3%. These figures, while representing substantial progress from a baseline of near-zero

independent youth archiving activity, also reveal the partial character of the program's youth engagement achievements, with between 41.6% and 57.9% of enrolled youth apprentices not yet conducting autonomous elder recording sessions by program conclusion. Focus group discussions with youth participants identified three primary barriers to autonomous youth-elder recording engagement: intergenerational communication anxiety arising from the social protocols governing interaction between youth and elders in each community's cultural framework; technical confidence deficits in recording setup and ELAN annotation among youth who had received fewer training hours than designated community archivists; and competing time demands from formal schooling, agricultural work, and domestic responsibilities (Muhsyanur and Mustapha, 2023). Fishman (1991) specifically identified the disruption of intergenerational interaction as the central mechanism of language shift in communities experiencing rapid social change, arguing that structured intergenerational language transmission activities—precisely the elder-youth recording partnerships the present program sought to catalyze—constitute the most direct form of counter-intervention against this disruption. The program's partial success in establishing these partnerships provides both an encouraging foundation and a clear design improvement target for subsequent program phases.

The Fluent Speaker Retention indicator—defined as the percentage of identified fluent elder speakers (aged 55+) who participated in at least six recording sessions during the active archiving phase and who reported continuing to use the language in daily domestic communication at program endpoint—recorded the highest cross-community mean (71.7%) and the smallest inter-community variance (SD = 4.3%) of all four adoption indicators, suggesting that elder engagement with the digital archiving program reliably activated and reinforced patterns of daily language use that sustained speaker fluency independently of the broader community language environment. This finding resonates with Hinton's (2011) observation, derived from her extensive work with California indigenous language revitalization programs, that elder speakers' involvement in formal language transmission activities—whether through master-apprentice programs, documentation projects, or formal teaching roles—consistently produces a measurable "speaker activation effect" in which the social recognition of elders as language custodians and the communicative engagement generated by recording and teaching activities revitalizes patterns of language use that had become dormant under the social pressure of Portuguese-dominant community interaction.

The pre-to-post language vitality assessment data complemented the adoption outcome data by documenting improvements across all five UNESCO language vitality indicators following the 14-month program, with paired-sample t-test results confirming statistical significance at p < .001 for four of five indicators. The Intergenerational Language Transmission Rate indicator recorded the largest pre-to-post improvement (M_1 = 24.7%, M_2 = 41.3%, Δ = +16.6 pp, d = 0.79), a finding of direct long-term significance given Fishman's (1991) identification of intergenerational transmission as the single most critical determinant of whether a

language achieves sustainable survival rather than merely delayed extinction. The Community Language Use Domains indicator showed the second-largest improvement (M_1 = 31.4%, M_2 = 46.2%, Δ = +14.8 pp, d = 0.74), reflecting an expansion of the social contexts in which the indigenous language was being actively used—particularly in school settings and community cultural events, where the archiving program's production of illustrated phonological guides and educational storybooks generated new instructional materials that teachers incorporated into their classroom practice within the program period.

Community Data Sovereignty, Governance Architecture, and Sustainability Dynamics

The Community Data Sovereignty Protocol (CDSP) developed and implemented across the five community sites represented the most structurally innovative and theoretically consequential design element of the program, operationalizing in technical and institutional form the principle of indigenous data sovereignty articulated by the CARE Principles for Indigenous Data Governance—Collective Benefit, Authority to Control, Responsibility, and Ethics (Carroll et al., 2020). Under the CDSP, each community digital archive was configured with a four-tier access architecture: Tier 1 (unrestricted public access) for vocabulary lists, illustrated phonological guides, and oral narratives explicitly designated as publicly shareable by the community elder council; Tier 2 (community-member access) for the full lexicographic database, transcribed oral narratives, and non-ceremonial recordings accessible to all registered community members via a community-managed user authentication system; Tier 3 (restricted scholarly access) for materials made available to vetted academic researchers under community-approved data use agreements that specify attribution requirements, research purpose restrictions, and benefit-sharing obligations; and Tier 4 (elder council-controlled access) for ceremonial, sacred, and genealogically sensitive materials whose access is governed exclusively by the community elder council under traditional protocols. The implementation of this tiered sovereignty architecture fundamentally distinguished the program from the extractive documentation paradigm critiqued by Tuhiwai Smith (2012), in which indigenous communities' linguistic and cultural heritage migrates to university and national archive collections where it becomes accessible to the scholarly community while the source community loses de facto control over its own cultural patrimony.

The elder council governance structures that served as the primary community institutions for CDSP administration and archive curation oversight emerged across all five sites as the most critically important organizational enablers of the program's sustainability beyond the funded cycle. Elder councils, convening monthly at each site to review archiving priorities, approve access decisions, evaluate youth archivist performance, and plan the incorporation of archive materials into community school curricula, functioned as living governance bodies that integrated the digital archive into the existing social structure of community

cultural authority rather than positioning it as an external technical system requiring specialist management. Tuhiwai Smith (2012) argued that genuinely decolonizing research and development programs must respect and strengthen existing indigenous governance structures rather than creating parallel project management structures that bypass or marginalize them, and the elder council governance architecture of the present program provides a concrete exemplification of this principle in the institutional design of a digital preservation project. Ethnographic observation data from the sustainability documentation phase consistently documented elder council members articulating a strong sense of ownership over and responsibility for the archive—describing it in one Kayapó council session as "our house of words, built with our hands, belonging to our children"—an expression of community cultural proprietorship that represents precisely the institutional depth of ownership that Bringle and Hatcher (1996) identify as the essential foundation of sustainable community engagement program outcomes (Mulyana et al., 2021).

The integration of digital archive materials into community school curricula emerged as the most impactful mechanism for generating language revitalization effects that extended beyond the archiving community into the broader youth population. By the program's conclusion, schoolteachers in all five communities had incorporated archive-derived materials—illustrated phonological guides, curated oral narrative transcripts, lexicographic vocabulary lists organized by semantic domain, and indigenous-language song recordings—into their weekly indigenous language instruction sessions, reaching an estimated 847 schoolchildren beyond the 140 formally enrolled program participants. The school integration outcomes documented in the evaluative phase align with the findings of Hornberger (2008), who demonstrated through her comparative analysis of indigenous language revitalization programs across Latin America that the most impactful long-term outcomes are achieved when community language documentation projects establish functional operational linkages with formal school systems, creating institutional channels through which archived materials continuously generate educational value for successive cohorts of children rather than residing inertly in digital repositories accessible only to specialist users. The present program's school integration mechanism was catalyzed by the deliberate inclusion of schoolteachers in the community archivist training program and by the specific designation of "educational adaptation" as one of four primary roles for community archivists in the CDSP governance framework, ensuring that the translation of archive materials into classroom-ready educational formats was treated as a core archiving responsibility rather than an optional add-on.

The sustainability challenges encountered across the five community sites—most acutely in the Munduruku community, where external territorial conflict disrupted two months of archiving activity, but also in the Ashaninka community, where the departure of two trained archivists to urban employment created a temporary technical capacity gap—illuminate design vulnerabilities that future

program iterations must address with greater structural robustness. The most significant systemic vulnerability identified in the evaluative documentation phase was the precarious nature of the satellite internet connectivity on which remote mentoring and software update delivery depended, with connectivity interruptions occurring an average of 4.7 times per month across the five sites and disrupting remote technical support workflows that the community archivists were not yet fully equipped to resolve independently. Anderson and Hudson (2020), in their assessment of indigenous digital archive sustainability across comparable remote community contexts, identified technical infrastructure reliability as the single most frequently cited barrier to long-term community archive maintenance, recommending that program architects build redundant offline-capable archive systems that do not require continuous internet connectivity for basic operation—a recommendation the present program partially implemented through the offline server architecture but did not fully realize in the remote mentoring support dimension. Future program iterations should invest in peer community archivist networks that enable technical problem-sharing across sites, community-based technical troubleshooting guides developed collaboratively with archivists during training, and formal partnership agreements with indigenous telecommunications cooperatives to ensure baseline connectivity resilience.

CONCLUSION

This study establishes that a community-controlled digital language archiving program—grounded in data sovereignty principles, elder council governance, intergenerational recording partnerships, and school curriculum integration—can generate substantial, measurable, and durable contributions to indigenous language preservation and revitalization in remote Amazonian communities, with 9,674 language items archived across five community sites, statistically significant improvements across four of five UNESCO language vitality indicators (mean d = 0.88, p < .001), and documented continuation of autonomous community archiving activity in all five sites following the formal program conclusion. The program's theoretical contribution lies in demonstrating that the Community Data Sovereignty Protocol architecture—positioning each community as sole rights-holder and access controller of their archived linguistic heritage—is both practically implementable in low-infrastructure remote Amazonian settings and transformationally effective in generating community ownership of the archiving process that translates into the sustained archiving momentum and institutional governance investment on which long-term language archive viability depends.

To maximize the reach, depth, and sustainability of digital archiving programs for indigenous language preservation across the Amazon and comparable regions of global linguistic biodiversity, the following evidence-based recommendations are advanced: (1) the Brazilian federal government should formally adopt the Community Data Sovereignty Protocol framework as the mandatory governance standard for all publicly funded indigenous language documentation programs in

Brazil, with FUNAI designated as the institutional guarantor of compliance; (2) the four participating federal universities should establish permanent Community Language Archive Support Units staffed by bilingual indigenous linguists and digital archive technicians to provide ongoing technical mentoring, software maintenance, and quality assurance support to community archive sites beyond the funded program cycle; (3) the National Indigenous School Education Policy (RCNEI) should mandate the integration of community digital archive materials into indigenous school language curricula as a standard educational resource category, with dedicated curriculum development funding allocated to community schoolteachers for archive material adaptation; (4) future programs should extend the community archivist training to a minimum of 120 hours, incorporate comprehensive offline technical troubleshooting modules, and establish inter-community archivist peer networks as a structural program component; (5) the Legal Amazon's Connectivity Program should prioritize satellite connectivity reliability for all registered indigenous community digital archive sites as critical cultural infrastructure; and (6) international language preservation funding bodies—including the Endangered Language Fund, Arcadia Fund, and Hans Rausing Endangered Language Project—should adopt minimum five-year funding horizons for community digital archiving programs and require community data sovereignty compliance as a non-negotiable funding eligibility condition.

REFERENCES

Anderson, J., & Hudson, M. (2020). Looking across the indigenous data sovereignty landscape. In T. L. Carroll, R. Kukutai, J. Rainie, & R. Russo Rodriguez (Eds.), The rise of indigenous data sovereignty (pp. 25–42). ANU Press. https://doi.org/10.22459/CAEPR38.2019.02

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bringle, R. G., & Hatcher, J. A. (1996). Implementing service learning in higher education. The Journal of Higher Education, 67(2), 221–239. https://doi.org/10.1080/00221546.1996.11780257

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Mustonen, T., Palchykov, V., Suresh Babu, S., Walker, J., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. Data Science Journal, 19(1), 43. https://doi.org/10.5334/dsj-2020-043

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2023). Ethnologue: Languages of the world (26th ed.). SIL International. https://www.ethnologue.com

Fishman, J. A. (1991). Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages. Multilingual Matters.

Gippert, J., Himmelmann, N. P., & Mosel, U. (Eds.). (2006). Essentials of language documentation. Mouton de Gruyter. https://doi.org/10.1515/9783110197648

Himmelmann, N. P. (1998). Documentary and descriptive linguistics. Linguistics, 36(1), 161–195. https://doi.org/10.1515/ling.1998.36.1.161

Hinton, L. (2011). Language revitalization and language pedagogy: New teaching and learning strategies. Language and Education, 25(4), 307–318. https://doi.org/10.1080/09500782.2011.577220

Hornberger, N. H. (2008). Can schools save indigenous languages? Policy and practice on four continents. Palgrave Macmillan. https://doi.org/10.1057/9780230584662

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Prentice Hall.

Muhsyanur and Mustapha, B. (2023). Challenges and Strategies in Teaching Indonesian to Indonesian Occupied Students in Malaysia. *TRICKS: Journal Ff Education And Learning Practices*, *1*(1), 32–39. https://journal.echaprogres.or.id/index.php/tricks/article/view/6

Muhsyanur, M. (n.d.). Sentence type in the text classical folk song of Bugis: study of syntax. *PROSIDING PRASASTI*, 500–507.

Muhsyanur, M. (2020). Types and trust system based on the ideology of Bugis community culture on local interest values in cenning rara spells. *International Journal of Humanity Studies (IJHS)*, *4*(1), 58–68. https://doi.org/https:/doi.org/10.24071/ijhs.v4i1.2652

Muhsyanur, M. (2023). The Bugis People's Naming System in Bugis Ethnic Tradition. *Journal of Language and Literature*, *23*(1), 67–76. https://doi.org/10.24071/joll.v23i1.5062

Muhsyanur, M., Larisu, Z., Sanulita, H., Ertanti, D. W., & Widada, D. M. (2022). Indonesian netizens expressions potentially satire with the Covid-19 pandemic on social media Facebook. *Linguistics and Culture Review*, *6*(1), 55–69. https://doi.org/10.21744/lingcure.v6n1.1942

Mulyana, Y., Akbar, Z., Zainal, H., Jiwantara, F. A., Muhsyanur, Yusriadi, Y., & Bin-Tahir, S. Z. (2021). High domestic violence during the pandemic COVID-19. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 6283–6290. https://doi.org/10.46254/an11.20211059

Ramadhanti, D., Yanda, D. P., & Muhsyanur. (2021). Scholarly Text Writing Workshop: Uncloking Weaknesses in Article Writing. *Rangkiang: Jurnal Pengabdian Pada Masyarakat*, *3*(1), 55–64.

Tuhiwai Smith, L. (2012). Decolonizing methodologies: Research and indigenous peoples (2nd ed.). Zed Books.

UNESCO. (2010). UNESCO atlas of the world's languages in danger (3rd ed.). UNESCO Publishing.

UNESCO Ad Hoc Expert Group on Endangered Languages. (2003). Language vitality and endangerment. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000183699